

석 사 학 위 논 문

지도교수 고 창 수

주어의 의미역 연구

A Study on the Thematic role of the Subject

2006년 6월 일

한성대학교 대학원

한국어문학과

한국어정보처리 전공

권 정 현

목 차

1. 서론	1
1.1. 연구의 목적	1
1.2. 연구의 방법	6
1.3. 선행연구	8
2. 주어의 개념과 범위	14
2.1. 주어의 개념	15
2.2. 주격표지와 주어	18
2.3. 주격 중출 현상	23
3. 주어의 의미역 설정	27
3.1. 의미역 체계의 기준 및 점검	28
3.2. 자질연산문법에서의 의미역 체계	35
3.3. 주어의 의미역 설정	43
4. 주어의 의미역 부여 방법	46
4.1. 의미역 데이터베이스 구축 방법	47
4.2. 논항과 동사의 의미자질	57
4.3. 주어 의미역 연산 절차	61
5. 결론	64
참고문헌	67

[부록]	76
------	----

ABSTRACT	91
----------	----

〈표 차례〉

<표 1> Fillmore의 내면격 분류의 변천	10
<표 2> Chafe(1970)에 설정된 동사 자질과 관련 의미역	14
<표 3> 김민수(1981: 171)의 11격	30
<표 4> 최호철 외 (1998)에 제안된 39개의 의미역	33
<표 5> 21세기 세종계획(1998~2002)의 의미역 분류	34
<표 6> 김원경(2000)의 내면격(의미역) 명칭과 정의	37
<표 7> Longacre(1976)의 의미역 기준	38
<표 8> 한국어의 의미역 체계 및 배당 규칙(고창수 외 2006)	40
<표 9> 오충연(2001)에 제안된 주어의 의미역	44
<표 10> 김원경(2000)에 제안된 주격 명사구의 내면격 부여 알고리즘	44
<표 11> 주어 의미역의 자동 배당 알고리즘	46
<표 12> 남경완(2005)에서 제안된 용언의 의미 분석 기준과 의미적 속성	59

〈그림 차례〉

<그림 1> 형태소 통계	48
<그림 2> 최빈도 동사 통계창	49
<그림 3> 불린 검색식	50
<그림 4> 「말샘41」의 어휘 빈도 처리 결과	51
<그림 5> 단문분할 작업과 병행하는 구문분석기 워크벤치(workbench)	52
<그림 6> 주어와 서술어의 「색표시 처리」 최초과정	53
<그림 7> 용례의 단문분할을 위한 「색표시 처리」	54
<그림 8> 엑셀을 이용한 의미역 자동배당 규칙 「데이터베이스」	55
<그림 9> 말뭉치 가공 절차 순서도	56

1. 서론

1.1. 연구의 목적

이 연구는 효율적인 한국어정보처리, 특히 구문분석을 위한 한국어 주어의 의미역 자동 배당 알고리즘¹⁾을 제안한 것이다. 이를 위해서 먼저 한국어의 격 체계에 따른 주어의 개념을 살펴보고, 의미역 체계의 전반적인 모습에 대해 논의하고자 한다.

한국어정보처리는 자연언어를 컴퓨터가 인식하여 처리할 수 있도록 언어학적인 모형을 마련해 주는 응용언어학의 한 분과이다. 이 분야는 자연언어처리²⁾에서 목표로 하는 정보검색·질의응답·기계자동번역 등의 실제 프로그램 구현을 위해 언어적 절차를 제공하는 데 목적이 있다. 따라서 한국어정보처리에서 다루는 문법은 실제 프로그램에 적용 가능한 알고리즘이어야 한다. 이를 위해 이 연구에서는 언어학 연구의 문법적 절차가 이론적 정합성을 가지는 것뿐만 아니라 그 실용적 가치에도 적용되는가를 실제 말뭉치 분석을 통해 점검해 보고자 한다.

이 연구의 궁극적인 목적은 지금까지 시스템 개발에서 문제가 되었던 의미역 배당을 정밀하게 하는 문제에 천착하여 효율적인 의미역 자동 배당 방법을 제안하는 것이다. 이를 위해서 세종 말뭉치에서 추출하여 가공한 표준 100문장을 통해 의미역 배당의 적절성을 점검하였다. 이러한 연구는 향후 시스템 개발을 위한 기반 자료로 이용될 수 있으며, 먼저 이를 통해

1) 이 연구에서 제시하고 있는 알고리즘(algorithm)은 동사가 주어에 의미역을 배당하기 위한 자질연산 절차를 말한다.

2) 컴퓨터 공학에서 자연언어를 인식하여 처리할 수 있는 컴퓨터 프로그램 개발 분야를 일컫는다.

서 이 연구의 이론적 기반인 자질연산문법(Feature Computation Grammar, 이하 FCG)의 이론적 우수성을 확인할 수 있을 것이다.³⁾

이를 위해서 각 장에서는 다음과 같은 연구를 진행하였다. 2장에서는 전통적인 주어 개념 확립에서 제기되어 왔던 제반 문제들을 점검하였다. 그리고 3장에서는 지금까지 논의된 의미역 체계의 문제점을 살펴보고 정보처리를 위한 대안을 제안하였다. 4장에서는 동사 및 명사의 고유 어휘 자질을 이용하여 한국어 정보 처리 시스템에 적용 가능한 효율적인 주어 의미역 연산 알고리즘을 제안하였다.

격은 서술어에 대한 명사의 관계 개념이다. FCG에서는 이러한 관계가 각 어휘 자질의 연산에서 도출된 결과라는 입장을 취하고 있다. 이 연구에서는 서술어와 명사구와의 관계개념을 결정하는 직접적인 정보 요소가 무엇인가를 살펴보았다. 이 연구에서는 동사의 문법자질인 [수동성(passive)], [사역성(causer)] 등과 의미자질인 [인지성(cognitive)] 그리고 명사구(주어)의 대표적 의미자질인 [유정성/animate)] 등이 주어의 의미역을 담당하는 결정자질이라고 보았다.

또한 이 연구에서는 Chomsky(1986a:59-60)에서 제안하였듯이 동사가 논항에 의미역을 담당할 때 목적어에는 직접 의미역 표시를 하지만, 주어에는 간접 의미역 표시를 한다는 논의를 실제 자질연산절차를 통하여 확인하였다. 그리고 이러한 현상이 전접 명사구와 동사를 ‘구 묶음(chunking)’하여 자질 승계에 나타난 현상이라는 것을 논의하였다.⁴⁾

3) 이 이론은 고창수(2002)에 정리되어 있으며, 김원경(2000)에서는 이 연구 방법론을 이용하여 구조적 정보의 연산에 적용한 바 있다.

4) 최운호(2005)에서 논의된 자연언어처리 분야의 용어로 용언의 경우 보조용언 또는 명사구 전접 관형어 등의 처리를 위해 제시되었다.

(1) 현주는 [욕심이 생겼다.]

animate cognitive ↓

구뭉음 후 자질 승계

(1)에서 주어인 ‘현주’는 [유정성]을 가지는 명사구이고 ‘생기-’는 동사 자체가 [동작성(action)]을 가지고 있다. 그런데 구뭉음을 해 보면 전 접 명사구의 명사자질이 [인지성]이 있기 때문에 구뭉음을 하게 되면 동사가 자질을 승계 받게 된다. 그래서 구뭉음 된 동사구가 가지는 [인지성]이 주어인 ‘현주’에게 <경험자역(Experiencer)>을 배당하는 결정자질이 된다. 이러한 과정에 대한 보다 구체적인 논의는 4장에서 다룬다.

이 연구가 기존의 의미역 체계에 대한 연구와 다른 점은 실재하는 말뭉치를 자료의 성격에 맞게 가공하여 데이터베이스화하였다는 것이다. 이 연구는 1000만 어절 규모의 ‘21세기 세종계획 균형말뭉치(국립국어원 구축자료, 이하 세종 말뭉치)’에서 선별한 자료를 대상으로 하였다. 이 자료를 기반으로 해서 기존의 논의와 다른 방식으로 의미역을 분류함으로써 연구자의 직관에 의존한 모호한 의미 기준에서 벗어나 한국어 정보처리에 적용 가능한 객관적 알고리즘을 제시하였다고 할 수 있다.

이 연구에서 다루는 의미역과 관련된 개념들은 통사적·의미적인 두 층위에서 다루어진다. 문장의 기본 구성은 명사구와 서술어이다. 명사구는 문장의 핵이 되는 서술어에 대해 일정한 역할을 가지는데 명사구가 갖는 역할은 통사적 기능과 의미적 기능 두 가지다. 통사적 기능은 주어, 목적어, 보어 등의 문장 성분을 뜻하며, 의미적 기능은 행위자, 대상, 처소 등 의미적 역할 즉 의미역을 뜻한다.

이 연구에서 상정하는 주어는 생성문법의 구조격 개념에 기반을 두고 있다(고창수 1992). 구조격이란 어떤 명사구가 통사 구조상 보이는 관계 개념으로 볼 수 있는데, 주격은 어떤 논항이 주어임을 나타내는 문법적 절

차라고 할 수 있으며 주격표지나 통사구조 상의 위치로 실현된다. 한 문장의 주어는 서술어의 의미자질과 연산되어 특정한 의미역을 배당하게 된다. 따라서 한 논항의 의미역을 산출하기 위해서는 통사 구조상의 관계 개념과 의미상의 관계 개념 양 측면에서의 접목이 필요하다(김원경 2000).

지금까지 의미역과 격에 관한 연구를 간단히 살펴보면 크게 ‘의미적인 격’과 ‘문법적인 격’ 두 가지 흐름으로 연구되어 왔음을 알 수 있다. 그 중 Fillmore(1968)에서는 격 관계를 내면구조에서 명사구와 서술어가 지닌 의미론적으로 적합한 구문관계를 의미한다고 보고 보편적이고 근본적인 개념의 집합으로 내면구조상의 격을 설정하였다. 즉 표면상의 격 관계를 부정하고 심층의 격 관계와 보편성을 지닌 격 체계의 수립을 시도하였다. 이러한 시도에서 나온 의미역은 명사항의 술어에 대한 논리적 역할의 개념인 Fillmore(1968)의 ‘의미격(Semantic case)’에서 발전한 것이라 볼 수 있다.

이에 반해 변형문법에서는 의미격이 통사적인 격 관계와 전적으로 일치하지 않는 점을 파악하고 구문 구성단위들의 통사적인 관계를 중시하여, 의미격 보다는 구문구조상의 격 관계를 염두에 둔 통사적 기제로서 의미역을 설정하게 되었다. Chomsky(1981)에서는 이러한 의미에서 구조격과 본유격, 내면격을 나누어서 제시하고 있는데 여기서 새로이 도입된 본유격(Inherent Case)은 격 형태로서 의미적인 관계를 그대로 표시하는 것을 말한다.⁵⁾ 이러한 연구의 흐름을 볼 때, 이 연구에서 제시하는 의미역의 개념은 구조격과 의미역의 통합개념이다.⁶⁾ 이는 연구의 대상인 말뭉치를 분

5) 구조격으로 일컬어지는 주격 및 목적격은 해당 의미역이 고정적이지 않고 통사적 관계에 의해서 설정되는 반면 속격을 제외한 기타의 격들은 의미역이 대부분 그대로 형태적으로 표현된다는 특성이 있다(Chomsky 1981).

6) 의미격(Semantic Case) 혹은 심층격(Deep Case)에 대한 이론은 Fillmore(1968)로부터 시작되었다. 그러나 Chomsky(1981)에서 제안된 이른바 의미역 이론(θ -theory)에서는 격 이론과 별도로 논항이 가지는 의미적 역할이 문법의 중요한 구성 요소로 다루어지고 있다. 이 연구에서는 의미격이라는 용어가 표

석할 때, 구조격 표지로 주어를 찾고 다시 주어에 의미역을 부여하는 방법적 절차로 확인할 수 있다.

이 연구에서 제안하는 의미역 체계는 성광수(1974)를 이어받은 김원경(2000)의 체계를 의미적 관점에서 보충하여 설정하였다. 이를 위해서 이 논의에서는 여기에 전산언어학 분야에서 일반적으로 논의된 Longacre(1976)의 의미역 체계를 수용하였다.

주어는 주로 <행위자역(Agent)>과 <주제역(Theme)>을 배당받는다고 알려져 있다. 주어의 의미역을 설정하는 데에 있어서 <행위자역>과 <주제역>의 세분화가 본 연구의 관건이 된다. 의미역 설정의 목적이 언어의 효율적이고 객관적인 정보 처리에 있는데, 실제 말뭉치를 살펴보면 <행위자역>과 <주제역> 두 가지로 모든 주어의 의미역을 설정하기에는 무리가 있기 때문이다.

이제까지 의미역은 몇몇 특정의 서술어와 논항 표지를 중심으로 논의되어 왔다. 그러나 의미역은 앞에서 말했듯이 서술어와 논항의 단선 관계는 아니다. 각각의 서술어가 지닌 개별 의미자질과 문장 내에서 통사적 기능을 하는 명사구 사이의 함수관계를 따져 보면 복잡하고 가변적인 의미역 배당의 양상을 알고리즘화해서 볼 수 있다. 따라서 기존의 의미역 체계 설정 방법에서 벗어나 새로운 설정 방법을 모색해 보았다는 데 이 연구의 의의가 있다.

종합해 보면, 이 연구에서는 모호한 의미역의 정의나 연구자의 직관에 의존하여 일부 문장을 설명하는 방식에서 벗어나 명사구나 동사구가 지닌 고유 자질 및 외면적인 격 형태나 구조격 자질 등을 기계적으로 검색하여 의미역을 배당하는 알고리즘 절차를 제안하였다. 이러한 연구를 통하여 논항과 술어 사이에 관계된 의미적 문제들이 실제 말뭉치에서 어떻게 발생하는

면적이고 형태적인 격 개념에 대해 혼란을 야기할 수 있으므로, 의미론적 격을 의미역(θ -role)으로 지칭하도록 하겠다.

지를 쉽게 알 수 있었다. 그리고 이 연구에서는 주어의 의미역 배당방법을 제안하고 100문장의 의미역 배당 데이터베이스를 구축하여 의미역과 관련된 다른 언어 현상을 손쉽게 연구할 수 있게 하였다.

1.2. 연구의 방법

이 논의의 연구 방법론은 한국어정보처리에 효율적인 자질연산문법 (Feature Computation Grammar)이다. FCG는 ‘언어 표현의 모든 요소를 자질 정보로 환원하며, 복잡한 연산 절차를 통해 자율적으로 표현’ 하는 문법 이론이다(고창수 2002:4). 이 방법은 문장을 분석한 후, 각각의 어휘자질을 명세화하고 이것을 연산해서 일정한 값을 낸다. FCG는 이러한 값이 의미 분석에 효율적으로 작용하는 것을 목표로 하고 있다.

이 연구는 자질연산문법에 의한 의미역 연산 절차를 이용하였다. 이러한 연구방법을 통해 이에 대한 의미역 설정의 모호성을 해결하여 신뢰성 있는 자료체를 구축한 후, 검증을 시도하였다. 이 연구를 위한 기본 자료는 세종 말뭉치를 가공한 30만 문장 집합에서 주어 의미역을 보여주는 100개의 표준 문장이다. 이러한 100개의 표준 문장을 가지고 FCG를 통해 자질별로 귀납하여 말뭉치에서 반복적으로 나타나는 규칙들을 포착하였다.

FCG는 고창수(1994)에서 처음 제안되었는데 이 논의에서는 고유의 어휘 자질에 의해 시제 정보를 연산하는 형태론 절차를 다루고 있다. 이후 고창수·김원경(1998)에서는 이 방법을 통해 선어말어미로부터 추출할 수 있는 정보 유형과 추출 방법을 논의하였다.

또한 김원경(2000)은 이제까지의 이론을 명세화하고 모형화한 구체적인 시도로, FCG를 이용하여 각각의 격조사에 적절한 구조격(Structural

Case)과 내면격(Semantic Case) 정보를 부여하는 방식을 제시하였다. 여기서 기계 번역 시스템과 대화형 시스템 등에 실제 활용될 수 있는 격 정보 처리 방법을 보여 주고 있다.

이밖에도 김유정(1996)에서는 영어를 국어로 번역하는 기계적 처리에서 나타나는 시제 형태소들 간의 대응관계를 자질 연산 알고리즘으로 기술하였고, 황화상(1998)에서는 자연 언어 처리를 위한 형태소 분석 방법을 논의하였다. 또한 김원경(1999)에서는 처격 조사의 자질 연산을, 이동혁(2000)에서는 한국어 부사의 자질 연산을 다루었으며, 이동혁(2004)에서는 한영기계번역기를 개발하기 위하여 채택하였던 FCG를 기본 문법으로 받아들여 언어 관계를 처리하였다. 한편, 유혜원(2002)에서는 중의성 문제를 중심으로 격 교체 구문을 논의하였다.

이상의 연구에서 볼 수 있는 것과 같이 FCG는 주로 형태소 분석 시스템이나 구문 분석 시스템 혹은 기계 번역 시스템과 같은 시스템 개발 과정에서 자연스럽게 도출된 연산 규칙을 정형화하면서 발전되어 왔다. 이런 점에서 FCG는 실제 작동하는 시스템에 기반을 둔 충분히 의미있는 이론이라고 할 수 있다.⁷⁾

이 연구에서는 주어의 의미역 산출을 위하여 FCG에서 제안하는 두 가지 연산 원리를 적용하였다. 하나는 자질 포화 원리이고 다른 하나는 동일 자질 결합 원리이다. 이에 대한 구체적 절차는 3.2 장에서 논의할 것이다.⁸⁾

FCG의 자질은 김원경(2000:47)에서 제시하는 바와 같이 음운적 자질과 문법적 자질, 의미적 자질, 시스템 관련 자질 등으로 나눌 수 있다. 본 논

7) 실제 이 연구에서 이용된 형태소 분석 및 어휘 통계 시스템 ‘말샘41’과 ‘형태소 분석, 구문 분석 워크벤치’는 이 이론에 기반한 고창수·김원경의 특허 기술로 개발된 것이다.

8) FCG는 원리 중심의 문법 이론으로, 복잡한 연산절차를 세부적으로 규정하는 이론이 아니다. 다시 말해서 FCG는 대규모 말뭉치를 대상으로 복잡한 연산규칙을 구성할 수 있는 기본적 원리 및 제약 조건들에 대한 이론이며, 각 세부절차들은 요소들의 자질연산이라는 단순한 방법으로 개별적인 문법 절차로 기술된다.

의에서는 논항과 동사의 의미적 자질에 중점을 두고 연산을 진행해 나갈 것이다. 의미 자질의 상정은 말뭉치를 대상으로 이루어지며 의미자질은 의미역을 산출하는 전 단계로써 논항과 동사 부류를 나눌 때 배당할 것이다.⁹⁾

1.3. 선행연구

이제까지 의미역에 관한 국내의 연구들은 국어학과 한국어정보처리의 두 분야에서 이루어져 왔다. 국어학에서는 의미론과 통사론의 입장에서 문장 단위의 관계개념 규명에 목적을 두고 연구가 이루어져 왔는데, 명사구가 차지하는 문법적 위치와 의미기능에 관한 연구로 나누어 논의되었다. 한국어정보처리 분야에서는 언어의 전산처리를 보다 쉽고 효율적으로 하기 위해 의미역 분류와 의미역 배당 방법에서 중점적으로 연구가 이루어져 왔다.

먼저 국어학 분야에서 진행된 의미역 연구를 살펴보면, 1970년대에 외국의 언어이론이 본격적으로 도입되면서 의미역 분류에 초점을 맞추고 활발히 진행되었다.¹⁰⁾ 국어의 연구에 주로 영향을 준 외국의 이론은

9) 각 해당 의미자질에 대한 구체적인 설명과 말뭉치 가공 절차는 본 4장에서 논의한다.

10) Chomsky(1981)의 의미역 이론(Theta Theory)과 격 이론(Case Theory)은 J.Katz의 의미관계(Semantic Relation), R. Jackendoff의 의미역 관계(Thematic Relation), C. Fillmore의 격 관계(Case Relation)의 영향 하에 발전된 것으로서, 문장 내 주어나 목적어 등의 명사어가 그 문장에서 갖는 의미적 관계를 다루는 이론이다. 이 논리적 의미 관계를 나타내는 명사 자격의 논항(argument)은 몇 가지 역할(Agent, Patient, Theme 등)로 나눌 수 있는데, 이러한 문장 내 모든 명사어는 의미역 위치에 있으며 반드시 격 할당을 받아서 실현된다는 것이다.

Fillmore(1968)의 격 문법과, Chomsky(1980, 1981) 이래로 활발히 논의된 의미역 이론, Jackendoff(1990)의 어휘 개념 구조와 Grimshaw(1990)의 논항 구조에 관한 연구 등이 있다.

격 문법과 후기 생성 문법 등에서는 명사구와 관련된 의미역을 중요한 연구의 대상으로 삼았다. 이러한 이론에 기반을 두고 명사구의 의미 기능이 활발히 연구되고, 개별 형태들이 나타내는 의미의 체계가 정리되었다. 여기에 어휘부 중심의 언어 이론이 가세하여, 각각의 동사들이 가지는 논항 관계, 결합가 등이 상술되었다. 이후에는 문법 기능과 의미 기능이 통합된 분류 체계가 나타나게 된다.

‘의미역(Thematic) 이론’은 Gruber(1965)와 Fillmore(1968)에서부터 본격적으로 논의되었다. Fillmore(1968)는 격 문법에서 술어동사와 해당 명사구 사이의 의미관계를 표시하는 격(Case)이란 개념을 논의하였다. 여기에서 격은 내면격(Semantic Case)이라고도 한다. 격 문법은 그전의 변형생성문법이 해결하지 못한 중요한 문제를 명확히 설명하였다. 예를 들어 <행위자역>이나 <경험자역> 등의 내면격과 주어와 서술어라는 문법관계의 차이 및 영역간의 층위를 제시하였는데 이는 아래의 예문에서 살펴볼 수 있다.

(2) 가. 경수가 열쇠로 문을 열었다.

나. 열쇠로 문을 열었다.

다. 문이 열렸다.

라. 경수는 문이 열리게 했다.

(2)에서 밑줄 그은 주어가 매개 문장에서 의미적으로 담당하는 역할(role)들의 차이가 인식된다. 즉 위 예문은 각각 <행위자역>, <도구역(Instrument)>, <피행위자역(Patient)>, <사역자역>이 된다.

이러한 격의 양면적인 특성으로 인하여 전통적으로 국어 문법에서 격 연구는 크게 두 방향으로 진행되었다. 하나는 논항 간의 의미역 관계를 중심으로 한 의미적 격 연구이고, 다른 하나는 논항의 서술어에 대한 문법적인 관계 즉 문장 성분 표시를 중심으로 한 통사적 격 연구이다. 이 흐름은 Chomsky(1968)에 영향을 받은 표면 구조상의 구조격에 대한 연구와 Fillmore(1968)에 영향을 받은 심층격으로서의 의미격에 대한 연구로 이어진다.

Fillmore가 제안한 격 유형과 격 명칭은 다음과 같다.

연대	격 명 칭
1966	Agentive, Instrumental, Objective Location, Time, Dative, Benefactive, Comitative
1968	Agentive, Instrumental, Objective Factitive, Location, Time, Dative Benefactive, comitative, Essive Translative
1971	Agent, Experiencer, Instrument Object, Source, Goal, Place Time, Path
1977	Agent, Patient, Source, Goal Range

<표 1> Fillmore의 내면격 분류의 변천

의미격의 분류방식은 연구자들마다 다양한 견해를 보인다. Fillmore 본인도 의미격의 정확한 수를 결정짓지 못하였고 지금도 의미격의 분류 방법에 대한 논의는 지속되고 있다.

1980년대 Chomsky는 C.J. Fillmore, R. Jackendoff, J.J. Katz 등에서

이루어진 이론의 발전에 따라 이 개념의 명칭을 의미역(Thematic Role, 혹은 Theta-role, 약칭으로 θ -role)으로 고쳤을 뿐만 아니라 이것을 원리와 매개변항 이론에 받아들여 의미역 이론(Theta Theory)으로 발전시켰다.

Chomsky(1981)의 ‘지배 결속 이론’의 등장 이후 국어의 격 연구는 문법적인 관계를 표시하는 구조격을 중심으로 격 형태인 주격 표지나 목적격 표지가 격을 할당할 수 있는 조건과 그 구조를 중심으로 논의가 이루어져 왔다. 하지만 최근에는 동사의 ‘논항구조’ 및 ‘어휘개념구조’를 중심으로 한 연구가 부각되면서 의미 관계를 중심으로 ‘논항’과 ‘의미역’에 대한 연구가 확대되기에 이른다.

논항의 의미 관계를 중심으로 격을 논한 연구로는 이선희(2004)를 들 수 있다. 이 연구는 {-를}의 문법적 실현과 관련된 다양한 의미·통사 제약들을 형식화하였다. {를}논항에 대한 연구는 이외에도 주격 표지에 비해 다양하게 이루어졌는데 홍종선·고광주(1999)를 대표적인 것으로 꼽을 수 있다. 이 연구는 {를}논항이 실현되는 양상과 분포를 분석하고, 이에 따른 유형을 의미적인 자질에 따라 체계화하였다. 그러나 이들의 의미역 연구에서 아쉬운 것은 {를}논항에만 초점을 맞춘 의미적 기능 연구는 비교적 많았으나 의미역 연구에서 주어의 의미역을 다룬 논의들은 거의 찾아볼 수 없었다는 점이다.

최근 들어서는 의미역을 정보처리 관점에서 기술한 국어학 분야 연구들이 나오고 있다. 대표적인 연구로는 김원경(2000)을 들 수 있는데 여기서는 자질 연산 문법이라는 방법론을 제시하면서 의미역의 배당절차를 제시하고 있다. 또한 유혜원(2002)에서는 격 교체 현상의 예들을 살펴보고 기계번역에서 격 교체 정보를 주어서 보다 안정된 정보처리가 가능하도록 하였다.

정보 처리 분야에서도 명사구의 문법적인 기능과 함께 의미 기능을 연산

해 내는 것은 매우 중요한 과제이다. 의미역은 언어 전산 처리의 성능을 향상시키는 데에 큰 역할을 한다. 그러나 의미역 분류 체계의 수립은 어려운 문제이므로 한국어정보처리 분야에서는 국어학에서의 의미역 분류방식을 그대로 이용하거나 전산처리에 용이하게 변형하여 사용하는 경향을 보이고 있다.

전산 언어학에서 조정미(1996)에서는 한국어의 의미역을 30가지로 구분한 후, 23가지의 대표 조사를 그 의미격에 따라 분류하였으며, 명사와 동사의 의미부류와 조사만을 이용해 의미역을 결정하는 신경망 기반 방법을 제안했다. 실험 결과는 제시하지 않았으나, 세 개의 자질(feature)¹¹⁾만으로는 의미역을 결정하기에 부족하다는 사실을 이 연구결과를 통해 짐작할 수 있다.

양단희(1998)에서는 격 원형성(case prototypically)이라는 개념을 도입하였는데¹²⁾, 이는 모든 격에 대해 명사와 동사가 갖고 있는 의미의 정도를 말한다. 각 용언과 명사가 지닌 격 원형성을 말뭉치로부터 미리 계산해 둔 후, 논항의 격을 이로부터 결정하는 방법을 제시하였다. 이 방법은 기계학습을 통해 말뭉치로부터 지식을 구축했기 때문에 은유나 환유현상을 다룰 수 있는 장점이 있으나, 대량의 학습데이터가 필요하며, 격조사가 표현할 수 있는 격 종류를 3가지로 제한한 점이 문제점으로 나타난다.

박성배(1997)에서는 한영 기계번역을 위해 구축한 동사 패턴에 수작업으로 의미정보를 추가하여 규칙을 만들고, 이러한 규칙을 이용해 통계 정

11) 여기에서 자질은 자질연산문법을 염두해 두고 논의된 것은 아니나, 의미역 결정에 이용되는 정보를 뜻한다.

12) Dowty(1991)의 ‘원형 의미역 이론’은 모형 이론적 의미론(model-theoretic semantics)이라는 모형 이론(model Theory)에 바탕을 둔 진리조건적 의미론으로서 국어 의미역 연구와 관련하여 유현경·이선희(1996)에서 언급된 이후 최근 이선희(2004)에서 본격적으로 소개되었다. 이 이론의 특징은 ‘행동주’나 ‘대상’과 같은 의미역이 ‘개별적인 유형’으로 존재하는 것이 아니라 좀더 일반적인 일군의 의미적 함의들의 집합에 해당하는 의미역이 ‘원형(prototype)’으로 존재한다는 것이다.

보를 추출하여 고빈도의 의미역을 할당하는 규칙을 자동으로 생성하였다. 이 연구는 한영 기계번역에서 관계절의 의미역을 배당하기 위한 목적으로 이루어졌다. 그러나 수작업에 따른 비용 증가와 어휘를 이용한 학습으로 인한 규칙의 적용률 저하가 문제점으로 지적되었다. 박성배(2000)에서는 결정트리(decision tree)를 사용하여 한영 기계번역에서 나타나는 부사격 조사의 의미 중의성을 해소하기 위한 연구를 하였다. 여기서 사용한 자질은 200개의 클래스로 클러스터링된 명사와 동사의 의미부류, 보조사 유무, 명사와 동사가 떨어진 거리, 전체 문장에 대한 거리들의 상대 거리인데, 실험 결과를 통해 각각의 거리와 상대거리는 의미역 결정 능력이 약하다고 밝히고 있다.

Fillmore(1968)의 영향을 받은 외국의 전산 언어학자들로는 Somers(1987)에서 제시한 Chafe, Cook, Longacre 등을 들 수 있는데 이들은 Fillmore(1968a:27)에서 문장의 격틀을 결정하는 것이 동사 자체라는 관점을 채택하여 발전시켰다. 따라서 이들이 내세우는 ‘격틀 자질(case frame feature)’은 명시적으로 주어지는 동사의 어휘내향을 포함하는 어휘부의 개념을 아주 강하게 암시하는 것이다. 그래서 격은 동사에 대한 어떤 종류의 ‘결합가(valency)’ 표지가 된다. 이들의 연구에서도 동사의 자질 설정이 결합가를 결정하는 핵심인자로 자리잡고 있다. Chafe(1970)가 제시한 동사 자질과 이에 관련된 의미 관계(의미역)는 다음의 <표 2>와 같이 제시되어 있다.

주어의 의미역 연구
2. 주어의 개념과 범위

동사 자질	관계
state[상태성]	Patient<수동자역>
process[과정성]	Patient<수동자역>
action[동작성]	Agent<행위자역>
ambient[무주체성]	—
experiential[경험성]	Experiencer<경험자역>
benefactive[수혜성]	Beneficiary<수혜자역>
completable[완전성]	Complement<보충역>
locative[처소성]	Locative<처소역>
—	Instrument<도구역>

< 2> Chafe(1970) | 설정된 동사 자질과 관련 의미역

여기까지 국어학 분야와 한국어 정보 처리 분야에서 다각도로 이루어지고 있는 의미역에 관련된 논의들을 살펴보았다.

이 연구에서는 이제까지의 의미역 분류와 배당 방법에 관한 논의들을 이어받아 지금껏 주목 받지 못했던 주어의 의미역 개념규정과 의미역 설정 방법에 관한 논의들을 진행해 나갈 것이다.

2. 주어의 개념과 범위

2장에서는 주어가 가지는 범주적 위치를 점검해 보고 주어를 설정하는 조건을 점검하였다. 주어 설정에 따르는 제 문제들을 점검해 보면 우선 주격표지 범주 설정의 문제가 있다. 주어를 설정할 때 주격표지가 있다고 해서 반드시 주어가 되는 것은 아니기 때문이다. 또한 주격표지의 중복 출현

은 문장안에서 주격 중출 현상이라는 문제를 야기시키는데, 이는 주어의 의미역 설정을 위해서 반드시 점검할 필요가 있다. 이 장에서는 주어를 설정하기 위한 객관적인 기준을 점검하고 주어표지 설정문제와 주격 중출 현상을 살펴봄으로써 주어의 의미역을 실현하기 위한 제반 조건을 알아보고자 한다.

2.1. 주어의 개념

전통적으로 주어는 동사에 대한 행동의 주체로 본다. 주어는 문장의 출발점이며, 문장의 중요한 구성요소이다. Jespersen(1987)에서는 ‘외연(denotation)’이라는 개념을 적용하여 주어를 술어에 대한 외연적 집합이라고 정의하면서 주어는 술어에 비해 비교적 한정적이고 특수하며 고정된 요소라고 설명하고 있다. 이 정의는 비교적 주관적인 주어의 정의에서 벗어나 보다 객관적으로 주어를 명시하려는 입장을 보이고 있다. 그러나 이러한 정의만으로는 주어를 기계적으로 설정할 수 없다. 그러므로 여러 학자들의 주어에 대한 주장을 점검해 보겠다.

생성문법에서는 주어를 통사론적인 관점에서 S의 직접관할(dominate)을 받는 명사구, 즉 명사구+동사구의 연속체인 기저문장에서의 명사구가 갖는 문법적 기능이라고 한다(Chomsky 1965).

학교문법에서는 남기심·고영근(1987:234)에서 주어[임자말]를 일반적인 문장의 형식인 ‘무엇이 어찌한다, 무엇이 어떠하다, 무엇이 무엇이다’의 ‘무엇이(또는 ‘누가’)’에 해당하는 것으로 ‘어찌한다, 어떠하다, 무엇이다’의 주체를 표현하는 말이라고 하였다. 또한 주어의 성립조건으로는 체언이나 체언의 기능을 하는 말에 주격조사가 붙은 것으로 제시하고 있

다.

왕문용·민현식(1993:198)에서는 주어는 서술의 주체가 되는 성분으로 통사적으로는 그것이 단어이든, 구, 절이든 주격조사 ‘이/가’가 통합 가능한 성분이라고 하였다. 또한 서술어의 주체는 ‘이/가’에 의해서 기본적으로 실현되고 ‘은/는’은 대조적 화제 기능의 주어 표현에서 쓰일 뿐이며 주어로는 대개 체언이나 체언 상당어가 쓰인다고 하였다.

이러한 주어에 관한 정의들을 살펴볼 때 주어는 외현적으로 주격조사가 형태적으로 실현될 수 있는 것들로 볼 수 있다. 흔히 주격조사로 처리하고 있는 형태로는 ‘이/가’, ‘께서/에서’, ‘은/는’, ‘ \emptyset ’이 있다. 변형 생성문법이 지배·결속이론으로 발전하면서 양동휘(1998)에서는 한국어의 주격이 주격조사 그 자체에 의해 부여된다고 보았다.¹³⁾

양동휘(1998)의 견해를 받아들이면 주격표지 ‘이/가’는 주격 부여 절차에 기본적으로 적용될 수 있다. 그러나 같은 형태의 격이라 하더라도, 그 기능(목적어 또는 보어)이나 의미가 다른데도 모두 주격으로 처리할 수 있는가의 문제가 발생한다. 이러한 논점에서 ‘되다’ 구문의 ‘이/가’ 주격 중출 현상과 ‘은/는’의 경우 주체격 처리 문제, ‘ \emptyset ’의 주어 설정 문제는 논란의 여지가 많다. 특히 한국어의 문장은 성분어 위치에 따르는 주어 중심 언어라기보다는 상황 중심 언어이기 때문에 이러한 문제가 더욱 증폭된다. 이러한 견해에서 나오는 주장으로 ‘은/는’을 ‘주어’와 ‘주체’로 분리해서 보는 견해가 있다. 이러한 논의에 대해서 임홍빈(1987:21)에서는 “국어 문장의 기본 구조는 ‘주제-설명(topic-comment)’ 구조를 그 근간으로 한다. 주어-서술어 관계는 2차적인 것이다.”라고 언급하였다. 예를 들어 ‘논문이 이러니 참 큰일이다’ 같은 경우 주어를 지칭하는 명시적인 주격조사가 붙은 말이 없으며 문장의 ‘주체’도 명확하지 않기에 이러한 격 표지만으로는 주어를 설정할 수 없다고 본다. 이러한 측면에서 볼

13) 주격조사 그 자체가 명사구에 주격을 부여한다(양동휘 1998:14).

때 형태, 의미, 기능의 여러 관점에서 그간 논의되었던 정의들이 재정립될 필요가 있는 것이다.

이러한 문제점을 해결하기 위해서 먼저 Chomsky(1965:68-74)와 Fillmore(1972:1)의 견해를 참고해 볼 수 있다.

Chomsky(1965:68-74)에서는 “문장의 직접 지배를 받는 명사가 주어이다” 라고 하였는데, 이는 통사론적인 정의로 ‘문장의 주체’를 이해하는데는 큰 도움이 되지 못한다. Fillmore(1972:1)에서는 “주어는 서술어와 대립하여 인식되는 것으로서 서술어에 대하여 진술이나 질문이 이루어지는 사람 또는 사물을 가리킨다” 고 하였다. 남기심(1985)에서도 Fillmore(1972:1)의 견해를 지지하며 ‘주어란 서술어에 대립하여 진술이 이루어지는 것’이라는 점에 주목한다. 따라서 능동사일 경우는 능동주가 주어가 되고 피동사일 경우는 피동주가 주어가 된다는 견해가 논리적 타당성을 얻는다. 논항이 술어의 의미에 따라 실현되고 배열된다는 생각이 타당성을 획득하며 결국 이것은 주어의 의미역 설정 개념으로 귀결될 수 있다.

이밖에도 Jackendoff(1972), Bresan & Kanerva(1989)에서는 논항의 주어로서의 문법적 실현과 의미역 사이의 연관성을 포착하기 위하여 의미역 계층 가설을 제시한 바 있다. 하지만 의미역 계층은 학자마다 의견이 다양하며 많은 예외 현상이 존재하는 문제점을 가지고 있다.¹⁴⁾ 주어를 정의할 때 의미역 설정 개념을 제안한 Chafe(1970:95-104)에서는 다음과 같이 정의내리고 있다.

14) Bresan & Kanerva(1989)에서 다음과 같이 의미역 계층을 제시하였다.

행동주(agent)> 수혜자(Beneficiary)> 경험주/목표점(Experiencer /Goal)> 도구(Instrument)> 대상/피행동주(Theme/Patient)> 목표점(Goal)

의미역 계층의 우위를 차지하는 논항일수록 통사 구조에서 상위의 요소, 즉 주어 자리를 차지할 가능성이 높은 것으로 간주된다.

주어의 의미역 연구 2. 주어의 개념과 범위

주어를 서술어의 의미적인 측면에서 규정할 때, 동작을 표현하는 서술어에 대해서는 <동작주(agent)>, 인식이나 감각을 표현하는 서술어에 대해서는 <경험주(experiencer)>, 상태나 성질을 표현하는 서술어에 대해서는 그 <대상물(object)>이 각각 주어이다.

정리하면, 이 논의에서는 일차적으로 주어표지는 구조격 ‘이/가’로 설정하는 것이 바람직하고 본다. 그 다음으로 ‘은/는’과 무표지 문제를 해결하는 데 있어서 Fillmore(1972)의 정의를 수용하며, Chafe(1970)와 Longacre(1976)의 주어 의미역 분류를 일부 발전 적용시켰다. 또한 이 연구에서는 구체적인 방법론으로 김원경(2000)의 입장을 수용하여 의미역을 설정하는 전단계로써 구조격을 먼저 부여하여 격 교체를 시도하고, 의미역을 부여하는 이중 절차를 줌으로써 앞서 논의하였던 주어 설정의 문제점을 보완하였다.

이 장에서는 주어 설정에 따른 논의들을 정리하면서 주어에 관한 개념을 다시금 점검해 보았다. 하지만 본 연구에서 다루고자 하는 것이 주어 설정 문제의 해결만은 아니다. 앞서 제시하였듯이 주어의 의미역을 상정함으로써 주어 설정 문제를 넘어서 외재논항임에도 불구하고 주어가 동사와의 관계에서 보이는 관계규칙을 밝혀보고자 한다. 이를 위하여 이어지는 절에서는 주어를 결정짓는 주격표지와 주격중출에 관한 논의들을 살펴보고 주어 설정의 기준을 점검해 보려 한다. 그리고 주어의 의미역을 설정할 때 살펴보아야 하는 문제점들을 짚어볼 것이다.

2.2. 주격표지와 주어

앞에서 논의했듯이 명사구는 주격표지가 붙어 있다고 해서 모두 주어가 되는 것이 아니다. 하지만 주어를 설정하는 결정적인 잣대가 되는 것 또한 주격표지임을 부정할 수는 없다. 이 연구에서는 주격표지에 대한 기존의 논의들을 살펴보고 ‘이/가’, ‘은/는’, ‘ \emptyset ’가 가지는 성분적인 위치를 점검해 보고자 한다.

지금까지 주격표지에 대한 연구는 표지가 실현되는 환경 및 분포에 관한 연구와 격표지 자체의 의미 기능에 관한 연구로 나뉘어 논의되어 왔다. 이러한 연구들을 살펴보면 주격표지가 지닌 기능을 알게 되고 주격표지의 범위를 설정할 수 있다. 따라서 이 장에서는 각 표지가 지닌 환경과 분포, 기능을 연구자별로 알아보고 주어 설정의 절차를 제안하겠다.

주격표지에 대한 연구는 이기용(1969), 이기동(1981), 류동석(1984), 최수영(1984), 최규수(1991) 등을 대표적으로 들 수 있다. 이기용(1969)은 ‘가’와 ‘는’이 쓰이는 여러 환경, 특히 통사구조상의 환경을 분석함으로써 이들의 차이점을 지적하고 있는데, ‘가’는 주격조사로서 주어-동사의 관계를 강조 또는 명확히 하는 반면, ‘는’은 중성 독립격조사로서 어떤 주어가 통사구조상 비강조될 필요성이 있거나 또는 어휘 배열 상 그 중점인 동사와 분리될 필요성이 있을 때 쓰인다고 설명하고 있다.

이기동(1981)에서는 ‘는’의 의미를 살펴보기 위해서는 체계 속에서 다른 기호와 견주어 보아야 한다고 하면서 ‘는’의 의미를 ‘ \emptyset ’와 ‘가’와의 비교를 통해 살피고 있는데 ‘가’나 ‘는’의 선택이 기계적으로 처리되는 것이 아니라 화자·청자의 의식의 문제와 청자의 생각에 대한 화자의 여러 가지의 평가가 수반되는 문제임을 논하고 있다. 곧 ‘는’이나 ‘가’와 같은 문법표지의 기능을 이해하기 위해서는 인간의 인지문제를 떼어놓을 수 없다는 것이다. 예를 들어,

(3) 네가 먼저 가라.

(3)과 같은 말을 화자가 할 때, 주어에 ‘가’가 선택되어 쓰이는 것은 ‘네가’의 대상인 청자에게 초점을 맞추어 화자가 지시하는 것으로 보고 있다. 그리하여 ‘가’의 쓰임을 화자가 청자의 주의를 특별히 끌려고 할 때 쓰이는 것으로 일반화하고 있다.

류동석(1984)은 ‘이’뿐만 아니라 ‘을’, ‘은’과 ‘ ϕ ’를 언어의 통보기능¹⁵⁾과 관련하여 파악하고 있는데, 문 구성체 ‘XaY’ (서술의 핵은 Y)에서 ‘X’가 ‘Y’보다 통보기능량이 클 때 a는 ‘이/을’이 선택되고 작을 때 a는 ‘은’이 선택되며 ‘X’와 ‘Y’의 통보기능량이 같을 때 a는 ‘ ϕ ’가 선택된다고 설명하고 있다. 그리고는 이전의 논의들에서 ‘ ϕ ’를 ‘에’, ‘로’, ‘와’ 등과도 관련지어 온 데 대해 이들이 ‘ ϕ ’로 되었다고 하는 곳에서는 ‘이’나 ‘을’이 실현될 수 있음을 들어, ‘ ϕ ’는 ‘이’, ‘을’, ‘은’과만 관련지어야 한다고 주장하고 있다.

최수영(1984)에서는 국어는 주제가 현저한 언어, 곧 주제-평언(topic-comment) 구조로 되어 있다고 하면서 ‘가’가 주제를 표시할 때도 있음을 논하고, ‘는’은 대조의 의미를, ‘가’는 지정 혹은 배타적 독립의 의미를 갖는다고 하고 있다.

또한 최규수(1991)에서는 문두에 놓이는 ‘는’요소를 주제임자말, ‘가’요소를 제시임자말이라 규정지으면서, 주제임자말은 앞뒤 문장과의 관계로 볼 때는 ‘연음’의 기능과 ‘담화 전개’의 기능을, 또 그 주제어를 제외한 나머지 요소인 설명어와의 관계로 볼 때는 ‘주의 환기’의 기능을 가지며, 제시임자말은 앞선 담화 상황에 없던 요소를 새로 도입하는 기능

15) 화자는 발화조직 층위에서 통보기능량(communivative dynamism)이 적은 요소를 앞세워 언술의 출발점으로 삼고 새로운 정보를 뒤따르게 하여 설명하는 구조로 문장을 형성하는데, 통보기능량이 적은 요소가 ‘언급 대상성(aboutness)’의 의미를 가진 주제가 된다. 한국어에서는 문두에 실현되는 ‘은/는’이 통합된 성분을 주제로 보며 이런 주제는 옛 정보이거나 주어진 것, 한정성, 총칭성 따위의 의미론적 특성을 가진다고 한다(채완 1979; 유동석 1998).

을 담당한다고 설명하고 있다.

이러한 주격표지에 대한 연구들을 보면 주격표지의 어휘적인 의미나 격적인 관계의미에 대해 연구가 주로 이루어졌으며, 담화·화용적인 의미기능을 다루었더라도 그것이 ‘는’ 과의 비교만을 통해 이루어졌고 격 배당의 관점에서 주격표지 ‘은/는’ 과 ‘ \emptyset ’ 를 어떻게 볼 것인가에 대한 깊은 논의가 부족하다는 것을 알 수 있다. 또한 ‘ \emptyset ’, ‘가’, ‘는’ 을 한 범주로 놓고 볼 때 각각의 주격표지가 어떠한 다른 의미기능으로 인하여 주격표지 설정 논란을 기본적으로 가지는가, 또한 문장 속에서 어떠한 공통적인 표현기능을 가지는가를 분명하게 제시한 연구도 부족하다는 것을 알 수 있다. 이것은 결국 아직 구조적 격 표시 ‘이/가’ 에 중점을 두고 연구가 진행된 것이라고 볼 수 있다.

최근의 연구 경향은 이러한 주격표지의 기능과 분포를 보다 명확히 하여 한국어 정보 처리에 접목한 연구로 격 교체 현상에 관한 연구가 활발히 진행되고 있다. 유혜원(2002)과 김미령(2006) 등의 논문에서는 자질연산문법의 방법론을 적용한 격 교체 현상을 다루어 기계번역에서 많은 해결점을 제시하였다.

이 연구에서는 이러한 논의들을 종합해 볼 때 ‘이/가’ 라는 구조적 표지는 주어 설정에 1차적인 전제라 할 수 있다. 다만 여러 논의에서 문제되어 왔던 ‘은/는’ 과 ‘ \emptyset ’ 의 주격표지 설정 여부는 ‘이/가’ 의 격 교체 작업을 통해 완성할 수 있을 것이다.

예를 들어 ‘현주는 얼굴이 예쁘다’ 의 경우, ‘현주가 얼굴이 예쁘다’ 로의 격 교체가 가능한 것으로 보아 주어 설정에 이상이 없다. 하지만 의미·화용상의 기준으로 확대하면 주어의 설정은 아직 객관적인 기준이 세워지지 않았다. 다만 이 연구의 목적이 구문분석 작업에 있기 때문에 주어 의미역 설정에 앞서 주격표지 ‘이/가’ 의 격 형태를 기준으로 삼을 뿐이다. 이러한 논의의 배경은 김원경(2000:138)에서 제시하는 주격 부여 절

차와 주격 부여 원리에 있으며, 본 논의에서는 이를 적용하여 따르기로 한다.

그 구체적인 설명은 아래와 같다.

① 주격 부여 절차

1. 문자열의 동사로부터 우-좌 방향으로 주격 조사 ‘가’가 결합된 명사구가 있는지 확인하라.
2. 이 명사구에 Sub를 부여하라.
3. 문장의 왼쪽 끝까지 격이 확인되지 않으면, 가장 좌측 무표 명사구에 Sub를 부여하라.
4. 주격과 목적격이 모두 무표일 경우, 주격 명사구와 서술어의 선택 제약을 점검하라.
5. 만약 선택 제약을 어겼으면, 목적격 명사구와 주격 명사구를 바꾸어라.

② 주격 부여 원리

: 서술어가 [act]이면, 주격은 오직 하나만 부여한다.

(김원경 2000:138)

이러한 주어설정에 기준이 되는 주격표지 범주에 대한 문제와 주격표지 중출 현상에 대한 문제는 주어의 의미역 설정을 위한 중요한 위치에 자리하고 있다. 이 연구에서는 대용량 말뭉치에 나타난 주격표지가 붙은 명사구들을 서술어와의 호응을 통해 단문분할 작업을 하게 된다. 따라서 앞서 제시한 격표지 대체 작업인 ‘은/는’과 ‘ \emptyset ’의 ‘이/가’ 치환 방법을 적용하여 주어를 상정할 것이다. 그리고 다음 절에서는 단문내에서 이중으로 나타나는 주격표지 중출 현상을 조망해 보고 본 연구의 관점을 제시할 것이다.

2.3. 주격 중출 현상

국어에서 주어를 설정하는 데에 있어서 가장 문제가 되는 것은 바로 주격 중출 현상이다. 주격 중출 현상은 이미 오랫동안 논의되어 온 문제이다. 특히 이중주어에 대한 문제는 초기 문법학자들에서부터 제시되었으며¹⁶⁾ 한국어에서 주어에 관한 연구는 다중 주격 또는 이중주어 구문에 관련된 연구가 전부가 할 수 있을 정도로 이중 주어 문제가 최대 논의점이었다.

이중 주어(중주어)는 단문내에서 공기할 수 있는 주어들을 지칭하는 것이다. 이러한 이중주어에 대한 논의는 결국 범주 설정의 문제로 돌아가서 단문내에서 이중주어를 인정하는가, 또는 인정하지 않는가의 견해로 갈리게 되며 이중주어를 인정하는 논의들에서는 이들의 구분을 어떻게 할 것인가에 대한 다양한 대안들이 논의되어 왔다. 따라서 이 절에서는 기존의 오랫동안 논의되었던 주격 중출 현상들을 각각의 관점에서 점검해 보고 이를 통해 주어의 설정과 이중 주어문의 의미역 처리 해결점을 모색해 보았다.

주격 중출 현상에 관한 대표적 연구로는 남기심(1968), 서정수(1968:12)¹⁷⁾, 박순함(1970:12), 양인석(1972), 김영희(1973)¹⁸⁾, 김민수(1971)¹⁹⁾ 등의 제 연구가 있다.

16) 이중주어에 관한 최초의 연구는 유길준(1909)의 대소주어설로 시작된다.

* 아이가 얼굴이 예쁘다.

[대주어] [소주어]

박승빈(1931, 1935): * 코끼리는 코가 길다. * 장사는 머리털이 관을 찌른다.
(문주) (주어) (문주) (주어)

김윤경(1946): * 그 사람이 힘이 세다.
(큰임자) (작은임자)

홍기문(1947): * 사내가 하나가 간다. * 네가 바둑이 선수다.
(대주어) (소주어) (대주어) (소주어)

17) 변형생성문법의 표준이론에 의한 분석을 시도한 연구.

18) 박순함(1970), 양인석(1972), 김영희(1973)는 격 문법에 의한 변형론적 연구 제시.

이중주어를 인정하지 않는 대표적 논의로는 최현배(1937)에서 제시한 포유문설(또는 서술절설)이 있다. 서술절이라는 것은 주격 중출문을 복합문으로 보았다는 것이다. 그러나 이 논의는 명사구와 문장형식을 갖춘 절이 복합된 형태이기 때문에 (5)와 같은 내포문 형식의 절표지가 있는 일반적인 복합문과 하나의 논리로 설명할 수 없는 문제점이 있다. 또한 기존의 명사절이나 부사절 등의 일관성 있는 명칭부여에도 ‘서술절’의 하위범주를 물을 때 문제가 된다.

또한 성기철(1985, 1987)에서는 이중주어문에서 존대 받는 대상에 주목하면서 이중주어문이 서술절을 이루는 구조라 하였다. 그러나 남기심(1986)에서는 보문소의 개념을 들어 이 논의를 반박하였고 내포문의 설정에 문제를 안겨준다고 하였다.

(4) [현주가] [얼굴이 예쁘다]

주어 서술절

(5) 현주는 늘 [얼굴이 예쁘다]고 생각했다.

한편 이중주어를 인정하는 논의로는 홍기문(1947)의 연구가 있는데 겹주격 문장의 주어관계를 두가지 면으로 나누어 설명하였다. 하나는 문장표면에 나타나는 주격의 위치에 따라 대주어와 소주어 관계로 보고, 또 하나는 그 실질적 기능에 따라 본주격과 부주격으로 분석하였다.

(6) 승수가 재주가 많다.

(대주어) (소주어)

현수가 몸매가 최고다.

(대주어) (소주어)

19) 보어와 중주어를 동심, 이심적 구성에 의한 구분.

부주격 본주격

승수에게 채주가 많다.

(대격)

현주의 몸매가 최고다.

(속격)

김민수(1971)에서는 다음 문장을 ‘중복주어구문’으로서 단일문이라 하였다. 이러한 다주어설은 주어를 대소주어로 구분하지 않았기 때문에 대소주어설과는 또 다른 것이다.²⁰⁾

(7) 가. 연탄이 금이 간다.

나. 코끼리가 코가 길다.

박순함(1970)에서는 격 문법을 따르면서, 심층구조의 격 범주가 주제화를 일으킨다고 하였다.

성광수(1974)에서는 격 문법을 바탕으로 이중주어문제에 접근하였다. 여기에서 주어화와 주제화를 구분하였다.

(8) 가. 서울이 사람이 많다.(주어화)

나. 서울은 사람이 많다.(주제화)

다. 서울에는 사람이 많다.(주제화)

(8가)는 처격이 주어로, (8나)는 처격에 후치사 「-는(은)」이 연결된

20) 이는 주어와 보어의 분간 문제에서 출발하는데 이를 구분하는 것은 쉽지 않다. 주어를 세우지 않는다면 주어 하나로 일관돼서 간단하나, 또한 주어의 하위에서 보어에 해당하는 것들을 어떻게든 분간해야 할 것이다. 그 일례로 은주어(대주어·문주어)와 분주어(소주어·보주어)로 보는 견해를 들 수 있다(김민수 1971:92-93).

채 주제화된 것이다. 이 경우, (8나)의 「서울은」은 주제어인 동시에 주어의 기능을 내포할 수 있다. 그러나 일반적인 해석은 (8다)의 뜻으로 가해져야 할 것이다.

(9) 가. *내가 이 문제가 쉽다.

나. 나는 이 문제가 쉽다.

다. 나에게서는 이 문제가 쉽다.

주어화와 주제화의 구분이 불가피한 것은 위의 예문처럼 주제화는 될 수 있으나 주어화는 될 수 없기 때문이라고 하였다. 원래의 주어와 변형으로 도출된 주어가 있으므로 다주어설을 지지하는 논의이다.

서정수(1971), 임홍빈(1972)에서는 주제화 변형을 일으키는 술어를 구분하였다. (10가)처럼 주제화 변형을 일으킬 수 있는 술어는 정태성 술어이고 (10나)처럼 주제화 변형을 일으킬 수 없는 술어는 비정태성 술어라고 하였다. 임홍빈(1972)에서는 국어에 있어 주제화가 언어적인 특수한 사건으로 존재하는 것이 아니라, 보편적인 언어현상으로 존재하는 것이라 하였다.

(10) 가. 그가 마음이 약하다. → 그의 마음이 약하다.

나. *중소리가 아이가 놀랐다. → 중소리에 아이가 놀랐다.

변형이론에 충실하게 기반을 둔 논의는 서정수(1971)와 김영희(1978)이 있었다. 이들은 이중주어는 (11)처럼 다른 기능의 명사구가 변형되어 주격형태를 드러낸 것이라고 하였다.

(11) 가. 코끼리가 코가 길다.

나. 코끼리의 코가 길다.

여기까지 국어의 이중주어에 대한 연구를 검토하였다. 위 논의의 핵심은 이중주어를 인정하는가와 그렇지 않은가로 나뉠 수 있다. 이중주어를 인정하는 논의는 대주어 소주어설과 주제(주어이면서 동시에 주제어)주어설이고 인정하지 않는 견해는 서술절의 설정과 주제(주어가 아니라 독립된 주제)주어설이라고 하겠다.

이 연구에서는 성광수(1974)에서 제시하였던 주어화와 주제화를 구분한 논의를 그대로 수용하였다. 그 이유는 주어의 생략이 없는 영어와 달리 한국어는 상황중심구문이기 때문이다. 그래서 주어의 생략현상도 빈번하고 주어표지도 생략되는 경우가 많다. 따라서 의미역을 설정하기에 앞서 이중주어의 구분 문제를 이러한 한국어의 특성에 맞게 적용하여야 한다.

이 논문에서는 이중주어를 인정하는 방향에서 논의가 출발한다. 주어 설정 후에 주어 의미역에 대한 논의와 절차는 3장과 4장에서 진행하도록 하겠다.

3. 주어의 의미역 설정

본 장에서는 국어학에서 논의되었던 기존의 한국어 의미역 체계를 살펴보고 자질 연산 문법에 기대어 주어의 의미역을 재설정하려는 것을 시도하려 한다. 이제까지 연구자들에 의해 세워졌던 의미역 체계 기준은 서론에서 밝힌 바와 같이 이렇다 할 기준이 없었다. 다만 연구자의 직관과 모호한 의미적인 정의에만 의존했을 뿐이다. 한국어 의미역 체계에서는 격 표

지에 의한 분류가 그 대부분을 차지하고 있었으며 최근에 들어서야 말뭉치 기반의 체계 설정이 시도되고 있다.²¹⁾ 3장에서는 기본격(기본적 의미역)의 기준과 원리로 성광수(1972, 1974)를 전적으로 수용하여 고찰을 시도하겠다.

3.1. 의미역 체계의 기준 및 점검

한국어 의미역 체계의 기준은 Fillmore(1971)의 격 분류 기준을 그대로 수용하였다고 볼 수 있다. 격 범주에 대한 Fillmore(1971)의 원칙을 제시하면 다음과 같다(성광수 1999:15).

1) 일문일격 원리(one-instance-per-clause principle)

이 원칙은 단문내 어떤 격 역할을 하는 명사구는 하나라는 뜻으로 동일한 격이 단문내에 여럿 존재할 수 없다는 것이다.

(12) 물이 얼음이 되었다.

이 예문은 “물이”와 “얼음이”를 명사구의 의미상 같은 <대상역(Object)>으로 볼 수도 있다. 하지만 그렇게 될 경우 주격 중출이 일어나 일문일격 원리에 위배되고 만다.

21) 세종계획(전자사전 개발 2000)의 용언사전에서는 말뭉치에 기반하여 술어가 요구하는 통사적인 논항뿐만 아니라 의미적인 논항에 대해서도 의미역을 정의하여 격 틀정보를 구축했다. 또한 이희자(1998)는 말뭉치 분석을 통해서 국어 조사에 대한 특성을 밝혔으며 약 30개 정도의 조사를 가지고 기존의 연구와 비교, 재해석하였다.

2) 대조 원리(principle of contrast)

하나의 서술어에 대해 단문내 각각의 구문 위치에 있는 명사구들이 서로 상이한 의미기능 및 대조되는 격 범주에 속하는가 그렇지 않은가를 판단하는 원칙이다.

(13) 가. 현주와 그 딸이 아름답다.

나. *그 영화와 나는 슬프다.

예문 (13가)에서는 ‘현주’와 ‘딸’은 동일한 의미자질을 가진 <공동역(Comitative)>이기 때문에 문법적으로 성립이 되지만 (13나)의 경우는 ‘영화’와 ‘나’가 동일격을 받을 수 없기 때문에 비문법적인 문장이 된다.

3) 상보성 원리 (principle of complementarity)

상보성의 원리는 동일격소의 변이격(allo-cases of same caseme)을 판정하기 위한 원칙이라 하겠다.

(14) 가. 광호가 서울에서 부산까지 걸었다.

나. 광호가 7시에서 12까지 걸었다.

다. *광호가 서울에서부터 12시까지 걸었다.

(14가)는 공간을 의미하는 명사구에, (14나)는 시간을 의미하는 명사구에 원격(Source)과 달격(Goal)이 각각 배당된다. 그러나 (14다)의 경우 동일 단문내에 공간을 의미하는 명사구와 시간을 의미하는 명사구가 공기(co-occurrence) 관계에 놓였기 때문에 비문이 된다.

이러한 격 설정 원리와 격 문법적 이론에 근거한 국어의 격 연구는 학자

주어의 의미역 연구
3. 주어의 의미역 설정

마다 기본격 설정 수의 차이를 두고 진행되었다. 이를 살펴보면 다음과 같다.

먼저 김영희(1973)에서는 한국어의 격 범주를 행위격, 경험격, 도구격, 수여격, 대상격, 시원격, 도달격, 처소격, 시간격의 9개의 기본격과, 거리격, 기간격, 경로격의 3개의 상위격으로 분류했다.

성광수(1977)에서는 언어 일반의 논리적인 내면격의 범주로 위격, 여격, 구격, 객격, 공격, 원격, 달격, 처격, 시격, 노격 등이 설정되었으며²²⁾ 김승곤(1980)에는 한국어의 격 범주로 행위자격, 여격, 연루격(instrument), 시발격, 도달격, 처소격, 시간격, 경로격, 자질격, 목적격 등이 설정되었다.²³⁾

김민수(1981:171)에서는 내면격으로 동주격, 기점격, 경험적, 목표격, 대상격, 원인격, 도구격, 위치격, 시간격, 공동격, 상대격 등 11격이 제시되어 있고,²⁴⁾ 이중격 설정목록을 아래와 같이 함께 제시하였다.

내면격	동주격(A) 원인격(C) 기점격(S) 도구격(I) 경험격(E) 위치격(L) 목표격(G) 시간격(T) 대상격(O)
이중격	동주기점격(AS) 허용동주격(PA) 기점대상격(SO) 동주대상격(AO) 사동동주격(CA) 동주목표격(AG)

〈표 3〉 김민수(1981: 171)의 11격

22) 이것은 서술어와 명사어에 관계되는 보편적 의미 자질 [의지], [원인], [수단/방법], [지배], [유정], [구체], [대칭], [가산], [시간], [장소], [방향/과정]을 근거로 설정된 것이다.

23) 이는 조사의 어원에 따라, 문장 중의 서술어 자질과의 관계에 의하여 설정된 것이다.

24) 격의 설정은 그 체계를 선명하게 하기 위해서 다단계의 분류가 불가피하다. 첫째는 구문기능인 구문요소에 따라 나눌 일이다. 격은 본질적으로 구문상 위치의 지표요, 대상물에 대한 위치의 표시는 아니기 때문이다. 둘째는 격형에 따른 뜻으로 구분할 일이다. 이 뜻이란 지표로서 가지는 지향성이요, 기능어의 가의성(加意性) 같은 것이 아니다. 셋째는 휘뚜루 쓰이는 통격을 가려놓는 일이다. 이는 다른 지향성의 같은 격형을 해명해야 하겠기 때문이다(김민수 1982:163).

주어의 의미역 연구
3. 주어의 의미역 설정

김일웅(1984)에는 주체격(존재/속성/상태, 변환/동작/경험/말바꾸기의 주체), 대상격, 상대격, 위치격, 장소격(통과/출발), 비교격, 원인격, 시간격, 공동격, 자격격, 도구격, 수단격, 재료격, 이유격 등이 제시되어 있다.

배광희(1991)의 모형 문법에서는 동사를 중심으로 이에 관여하는 명사의 격 의미와 어휘적 연관을 다루었다. 그의 모형 문법은 구체적 의미 관계로 35개의 내면격(동작 주격, 대상격, 영향격, 시간격 등)과 12개의 명사 의미군(활동 물질, 시간, 작용 등)을 이용한 것이다. 배광희(1991)에서는 35개의 내면격과 12개의 명사 의미군의 전모가 드러나지 않아 그 구체적인 내용은 알 수 없으나, 이와 같은 언급으로 보아 북한에서도 기계 번역을 위한 격의 의미적 고찰이 상당히 진행된 것으로 보인다.

논항의 의미역에 대한 연구는 격 문법의 쇠퇴와 더불어 논의가 침체되었다. 그러나 후기생성문법에 도입된 의미역 이론과 생성형태론에서 출발한 논항 구조 이론에 힘입어 논항의 의미역에 대한 부분적 서술이 명사구 관련 연구의 주류가 되었다. 특히 Somers(1987)에 지적된 대로 격 문법의 이론은 자연언어처리를 위한 분해 구조 생성에서 통사구조를 대신할 수 있기 때문에 새로운 조명을 받게 되었다. 즉 논항에 부여된 특정한 의미역 표지는 통사 구조가 보이는 명사구의 위치 정보를 대치할 수 있을 뿐 아니라 보다 풍부한 정보를 함축한다.

논항의 의미역을 부여함에 있어 가장 문제가 되는 것은 의미역 부여를 위한 일률적인 원리가 부재하다는 데에 있다. 이를 위하여 명사구를 구성하는 격 조사의 형태적 분류가 정형화될 필요가 있다. 최호철(1993)에서는 격을 그 형태를 중심으로 주체격, 객체격, 방편격, 기준격, 위치격으로 나누었다. 각 격은 그 의미역에 따라 세분되는데, 이 때 세분되는 의미역은 서술어와 논항의 상호 작용에 의한 것이다. 격 형태를 의미역 설정의 기준으로 삼는 것은 한국어처럼 의미역 부여가 체계적으로 이루어지기 힘들다.

각 언어에서 실현되는 문장의 형식은 보편적 개념이 일정한 표현 수법에 의해 실현된 것이라 할 수 있다. 이러한 측면에서 볼 때 문장에서 기능하는 주어, 객어 등은 보편적 개념으로 이미 정해진 주체, 객체 등의 실현이라 할 수 있다. 따라서 여기에서 주체, 객체 등은 바로 서술어의 어휘적 특성에 의하여 결정된 것이다.

영어에서의 위치와 전치사, 한국어에서의 조사는 그 명사가 주체, 객체, 방편 등임을 외현적으로 보여 주는 위치/표지에 불과하므로, 주체, 객체 등은 실현된 문장에서만 드러나는 것이 아니라, 문장 생성 이전에 특정 서술어와의 관계에서 결정된 개념이라 할 수 있다. 그러므로 이 연구에서 주체, 객체 등은 주어, 객어 등의 의미적 범주로 사용된다.²⁵⁾

이상에서 보듯이 주체, 객체 등의 실현이 명사의 위치 또는 그에 붙은 표지에 의해 이루어진 것이라면 명사와 서술어와의 관계인 의미적 범주는 일차적으로 위치/표지를 근거로 설정할 수 있다. 영어에서 주체는 동사의 앞에서 실현되고 객체는 타동사의 뒤에서 실현되며 나머지는 전치사에 의해 실현되는데 한국어에서는 모두 조사에 의해서 실현된다.

국어의 의미격이 실현된 표지로 성광수(1977)에는 ‘에게, 에(게)(서)(부터), 로, 에(게)로(까지), 와, 하고, 한테, 이, 을’ 이, 김승곤(1980)에는 ‘에, 에서, 한테, 로부터, 에게, 로써, 로서, 까지’ 가 제시되어 있고 김민수(1982)에는 ‘이, 의, 을, 로, 에, 와, 야, 이(술격)’ 가 김일웅(1984)에는 ‘이, 에게, 와, 을, 에, 에서, 로’ 가 제시되어 있다.

이 가운데에서 ‘까지, 부터, 야, 의, 이(술격)’ 는 서술어와 관련을 맺지 않는 것이므로 제외된다. 그리고 격 구조의 차원에서 ‘로써, 로서, 에로, 에게로’ 는 ‘로’ 에 통합되고 ‘하고’ 는 ‘와’ 에 통합되며, ‘한테’ 는 ‘에

25) 주어, 객어, 보어는 통사적 범주로, 주체, 객체, 위치, 방편 등은 의미적 범주로 구분되는데 주어는 주체와 대응되고 객어는 객체와 대응되며 보어는 위치, 방편 등과 대응된다.

주어의 의미역 연구
3. 주어의 의미역 설정

게’에 통합되므로 의미격 실현의 표지는 ‘이, 을, 로, 와, 에’ 형으로 조정된다(성광수 1977).

여기에서 ‘이’ 형은 서술의 주체를 나타내는 표지로 파악되고 본 논의에서는 이러한 주격표지가 실현되는 주어의 의미역들을 밝히려 하고 있다.

기계번역에서 유용하게 쓰일 수 있도록 의미역을 중심으로 논항 체계를 연구한 업적으로는 최호철 외(1998)를 들 수 있다. 이 연구에서는 ‘이, 을, 로, 에, 와’를 대상으로 39개의 의미역을 정리하였다.

격형태	내면격
이	범위, 감각체, 행위자, 감정자, 피위자, 수동자, 사역자, 지각자, 작용자, 존재자, 장소, 경험주, 소재, 지정, 대상
을	동작, 결과, 대상, 재료, 목적, 범위, 경로, 기점, 피해자, 수혜자, 피위자
로	범위, 이유, 재료, 도달, 경로, 도구, 자격, 선정, 수단, 결과, 방향
에	빈도, 시간, 원인, 행위자, 동인, 피사역자, 기점, 도달, 장소, 대비, 범위, 수혜자, 기준, 소재
와	상대, 공동

〈표 4〉 최호철 외 (1998)에 제안된 39개의 의미역

21세기 세종계획 1998~2002년도 보고서에서는 의미역을 다음과 같이 할당하고 있다.

주어의 의미역 연구
3. 주어의 의미역 설정

1998~2000년도	2001년도	2002년도
행위주 (AGT)	대상행위주 (AFA)	행위주 (AGT)
	공헌주 (DON)	
	소속주 (RCP)	
	공동행위주 (JAG)	
	원인주 (CAU)	
	행위주 (AGT)	
	상태주 (POS)	
	비의도행위주 (EFF)	
	소유주 (PSS)	
	피해주 (PAT)	
대상 (THM)	동작대상 (MOT)	대상 (THM)
	공동대상 (JMO)	
	상태대상 (MTH)	
	공공상태대상 (JMA)	
	대상 (THM)	
	창조대상 (CRE)	
	중립대상 (NTR)	
경험주 (EXP)	경험주 (AGE)	경험주 (EXP)
	경험주 (EXP)	
수반주 (COM)	수반행위주 (COA)	수반주 (COM)
	수반대상 (COT)	
	기준치 (CRT)	
처소 (LOC)	처소 (LOC)	처소 (LOC)
	범위 (RAN)	
	장면 (SCN)	
출발점 (SRC)	출발점 (SRC)	출발점 (SRC)
	원시향태 (INT)	
종점 (GOL)	종점 (GOL)	종점 (GOL)
	결과상태 (FIN)	
방향 (DIR)	방향 (DIR)	방향 (DIR)
도구 (INS)	도구 (INS)	도구 (INS)
	경로 (PTH)	
	재료 (MAT)	
이유 (RSN)	비의도행위주 (EFF)	이유 (RSN)
	목적 (OBJ)	
기준치 (CRT)	기준치 (CRT)	기준치 (CRT)
자격 (APP)		
정도 (DGR)	정도 (DGR)	정도 (DGR)

<표 5> 21세기 세종계획(1998~2002)의 의미역 분류

‘21세기 세종계획’ 2001년도 보고서에서 의미역을 검정하는 기본은 통사기반원칙이라 할 수 있다. 이에 대해 2002년도 보고서에서는 의미역 목록을 다시 정리하여 12개의 의미역 체계를 내세웠으나 그 하위 부류에 대한 세부 명칭을 부여하지 않은 문제점을 안고 있다.

국립국어원에서 말뭉치를 구축함에 따라 점차 말뭉치를 대상으로 한 의미역 연구가 이루어지기 시작하였는데 이희자(1998)에서는 말뭉치 분석을 통해서 국어 조사의 특성을 밝힌 연구가 그것이다. 그러나 조사가 가질 수 있는 의미역은 기존 연구를 기준으로 재해석하는 데 그치고 있다.

의미역은 통사적 위치에 있는 명사의 의미자질에 대한 고도로 추상화되고 개괄적이며 동사에 의해 지배되는 명사 의미자질의 집합이다. 이상에서 살핀 대로 의미역의 명칭이나 분류는 논의에 따라 다양한 양상을 보여 왔다. 이러한 이유는 앞서 언급한 대로 의미 기준이 갖는 근본적인 모호성에서 출발한다. 하나의 논항에 대해 연구자의 관점에 따라 ‘행위’로 인식되기도 하고 ‘경험’이나 ‘수혜’로 인식되기 때문이다. 이러한 문제를 일거에 해결할 수는 없지만, 합리적인 의미역 체계를 수립하고 이를 성분어의 고유 자질을 통해 상호 연산함으로써 기계에 적용 가능한 알고리즘을 확립하는 일은 한국어 정보처리 시스템의 개발과 같은 응용적 필요성 외에도 한국어 동사나 명사의 의미 분류 체계 확립이나 격 체계 확립과 같은 근본적인 언어학적 문제들을 풀어가는데도 도움을 줄 수 있다.

3.2. 자질연산문법에서의 의미역 체계

3.1 장에서 본 바와 같이 기존의 의미역 체계 설정 방법은 몇 가지 문제를 안고 있음을 확인하였다. 이에 따라 본 연구에서는 새로운 설정 방법을

시도하려 한다. 이것은 기존의 학자들의 논의를 부정하려는 것이 아니라 기존의 논의에 나타났던 의미역의 구분점이 불분명하고 실제 말뭉치 적용 과정에 맞지 않는 부분들을 다시 고찰해 봄으로써 보다 합리적인 한국어 의미역 체계를 모색해 보기 위함이다. 더 나아가 체계 상 가장 많은 분포를 보이면서도 세분화되지 못하였던 주어의 의미역을 전체의 체계 안에서 보다 효율적으로 설명하고자 한다. 본 연구에서는 이 분류 체계를 기저로 두고 자질연산문법을 이용하여 기존의 말뭉치를 적용·연산해서 의미역 자동 배당 알고리즘을 제안하였다.

이 연구에서는 성광수(1974)에 기반을 두었던 김원경(2000)의 의미역 체계의 기준을 따를 것이다. 성광수(1974)에서는 의미역을 위격, 여격, 구격, 객격, 공격, 원격, 달격, 처격, 시격, 노격으로 분류하였다. 김원경(2000)에서는 이러한 10개의 내면격에 소유격을 더하여, 다음과 같이 11개의 격에 대한 정의 및 배당 기준을 정리하였다. 그것은 다음과 같다.

주어의 의미역 연구
3. 주어의 의미역 설정

격 명칭	의미
위격 (Agent, A)	행위나 사건의 교사자로 인지되는 유생물의 격, 상태의 주체가 되는 유생물의 격
여격 (Dative, D)	상태나 행동의 영향을 받는 유생물의 격
구격 (Instrument, I)	행위나 상태에 원인적으로 내포된 무생적인 힘이나 대상의 격, 자격이 되는 유생물의 격
객격 (Object, O)	행위나 상태의 영향을 받는 대상의 격
공격 (Comitative, C)	동사에 대해 참여나 대칭 관계를 표시하는 격
원격 (Source, S)	출발점을 뜻하는 장소의 격
달격 (Goal, G)	도달점을 뜻하는 장소의 격
처격 (Locative, L)	상태나 행위의 위치나 공간적 방향을 표시하는 격
시격 (Time, T)	상태나 행위의 시간을 표시하는 격
노격 (Path, P)	이동의 경로를 표시하는 격
소유격 (Possessive, Poss)	소유자를 표시하는 격

〈표 6〉 김원경(2000)의 내면격(의미역) 명칭과 정의

표에서 보여주는 의미역 분류를 실제 문장에 적용해 보면, 적용의 일률성은 만족할 수 있지만 의미 기준보다는 통사적 기준에 가까울 수 있기 때문에 특정 문장들에서는 화자의 언어 직관이 기대하는 것과 다른 결과를 낼 수 있다. 따라서 보다 의미기준에 충실한 것으로 보이는 Longacre(1976)의 의미역 기준과 절충하는 것이 필요하였다.

주어의 의미역 연구
3. 주어의 의미역 설정

의미역	의미
Experience 경험자역	감각 기관에 대한 제어가 서술과 관련되는 유정적 실체
Patient 수동자역	상태나 장소가 서술되거나 변화를 겪는 무정적 실체; (물리적) 상태나 장소의 변화를 겪는 유정적 실체
Agent 행위자역	과정을 유발하거나 동작을 수행하는 유정적 실체; 동작을 수행하는 무정적 실체
Range 보충역	서술을 완성시키거나 더 특정화하는 어떤 표면구조의 명사에 부여되는 역할; 서술의 동작의 산물
Measure 수량역	수량화하여 서술을 완성시키는 어떤 표면적 명사에 부여되는 역할; 전달에서의 값
Instrument 도구역	(유정적) 행위자가 동작을 성취하거나 과정을 유발시키기 위해 의도적으로 사용하는 무정적 실체 또는 신체의 부분; (감정적) 상태에 조건을 부여하거나 감정적 또는 신체적 상태에서 변화를 유발하는 (유정적이면서 비의도적인) 실체
Locative 처소역	서술의 처소
Source 원천역	서술의 기원이 되는 곳으로 추정되는 장소; 신체적 감각이 발산되는 실체; 전달에서 본래의 소유자인 유정적 실체
Goal 목표역	서술에 대해 종착 지점이 되는 장소; 상태의 필수적인 변화 없이 서술의 방향을 가리키는 실체; 비전이적이거나 종착적인 소유자가 되는 유정적 실체
Path 경로역	이동 등의 서술에서 통과되는 장소 또는 장소들; 전이적인 소유자

〈표 7〉 Longacre(1976)의 의미역 기준

두 연구자의 의미역 정의를 절충하여 이 연구에서 제안하는 한국어의 의

주어의 의미역 연구
3. 주어의 의미역 설정

미역 체계를 세웠다. 이를 적용한 자질연산 결과 샘플은 다음과 같다(고창수·김원경·박선옥·권정현·이현주 2006).

주어의 의미역 연구
3. 주어의 의미역 설정

내면격	구조격	논항의 고유자질	격형태	서술어 자질
Experiencer 경험자역	sub	animate	-가	cognitive [인지성]
Agent 행위자역	sub	animate or group	-가	action [동작성]
	adjunct	animate	-에게 ²⁶⁾	passive [수동성]
		group	-에서 ²⁷⁾	action [동작성]
Beneficiary 수혜자역	sub	animate or group	-가	passive [수동성] benefit [수혜성]
	adjunct	animate	-에게	benefit [수혜성]
Instrument 도구역	sub		-가	causer [사역성]
	adjunct	space	-에	psych [심리성] passive [수동성]
Theme 주제역	sub		-가	stative [상태성] passive [수동성] copula [계사성]
				action [동작성]
	obj		-를	
Comitative 공동역	obj		-와	comitative [공동성] ²⁸⁾
Source 원천역	adjunct	space or animate	-에서 -로부터	
Goal 목표역	adjunct	space or position	-로 -까지 -에	move [이동성]
Path 경로역	adjunct	space	-로	move [이동성]
Time 시간역	adjunct	time	-에	
Locative 처소역	adjunct	space	-에	being [존재성]
Possessive 소유역	gen		-의	
Complement 보충역 *떨이함 ²⁹⁾ (wastebasket)	sub		-가	

<표 8> 한국어의 의미역 체계 및 배당 규칙(고창수 외 2006)

이 연구를 위한 주어의 의미역 개념 범주를 나눌 때 가장 문제가 되었던 것은 <주제역>의 설정에서 ‘의인화’ 된 문장일 때 [유정성] 자질을 주어야 하는가의 문제였다. 이것은 ‘심리적 실재성’의 문제와도 연관이 된다. 실제 말뭉치를 살펴보면 무시할 수 없는 양의 이러한 구문들이 나온다. 단순히 형태정보와 동사와의 자질만으로는 주어가 가지는 이러한 특징으로 인하여 설정자체에 혼란을 가져올 수 있다. 따라서 이 연구에서는 <주제역>과의 구분에서 <행위자역>의 경우 Longacre(1976)의 정의를 따르는 것이 말뭉치 샘플에서 적용이 보다 안정되며 심리적 직관에 부합할 수 있다고 판단하였다.

또한 하나의 논항에 오직 하나의 의미역만 배당된다는 의미역 기준을 만족하기 위하여 떨어함(wastebasket)으로서 <보충역(Complement)>의 설정이 불가피하였다. 떨어함은 격의 목록을 폐쇄적으로 정의하기 위한 하나의 방편이다. 만약 격의 목록이 폐쇄적이라면, 그들 중의 어느 하나는 떨어함(다른 어느 격으로도 배당되기 어렵다는 의미에서)이 될 수밖에 없다는 것은 중요한 이론적 문제를 제기한다. 비록 그러한 처리가 직관적으로는 받아들이기 어렵긴 하지만, 이론적 방편으로는 가능한 설정이라는 주장이다(Somers 1987). 이러한 문제들은 앞으로의 연구에서 더 많은 실제 문장을 적용해 봄으로써 자연스럽게 정리될 수도 있을 것이다.

서술어에서 기본적으로 지정해 두어야 할 자질은 [동작성]과 [상태성]

26) ‘에게’는 유정명사와만 공기하기 때문에 [유정성] 자질이 잉여적이지만 표체계의 일관성을 위해 표시하였다.

27) ‘에서’는 구조격 조사는 아니지만 [단체성] 자질을 가진 명사에 결합될 경우, 그 문장에서 다른 주어 후보가 없으면 이 논항만이 주어로 해석될 수 있기 때문에 이 표에서 다루었다. 예) 철수가 학교에서 놀았다. (부사격)

학교에서 급식중단을 선언했다. (주어)

28) 예) 영희와 철수가 싸웠다.

영희와 철수가 만났다. 등의 서술어의 [공동성] 자질의 경우만 해당한다.

29) Somers(1987)에서 제안된 wastebasket의 용어 번역인 ‘떨어함’은 ‘다른 격으로 배당되기 어려운 것들의 모음함’이라는 의미에서 명칭을 정하게 되었다.

(stative)]이다. 표본화(sampling) 과정에서 보여주듯이 두 부류의 동사가 실제로 많은 논항들의 의미역 배당에 직접 관계한다는 것은 흥미로운 사실이다. 또한 말뭉치에 나타난 많은 예문들은 전접하는 명사구 혹은 동사구와의 ‘구뭉음’ 처리를 통해 보다 효율적인 결과값을 산출하는 것으로 보인다. ‘구뭉음’에 의한 자질연산은, 예를 들어 서술어에서는 보이지 않는 [인지성] 자질이 전접구에서 보이면 이 자질을 서술어가 상속받아서 서술어에 [인지성] 자질을 부여하는 것이다. 이를 통하여 통사적인 기준보다 실제의 언어 직관에 근접하는 의미역 배당이 가능하다면 구뭉음은 우리의 생각보다 중요한 언어 현상으로 간주할 필요가 있다.

이 장에서 의미역 체계를 설정하는 데에 있어서 자질연산문법의 두 가지 원리가 적용된다.

1) 자질포화원리

: 특정한 어휘 항목의 자질이 폐쇄 집합을 구성하면, 이 자질들은 포화(saturated)된다.

이 원리는 어휘부를 구성하는 각 개체들이 체계를 이루기 위해 목록을 미리 ‘최적화(optimization)’ 할 필요가 없음을 의미한다.³⁰⁾ 기존에 논의되었던 체계 설정에서는 항상 체계를 먼저 만들어 놓고 현실 세계의 대입을 하려고 했었다. 그러나 이것은 실재하는 현실 세계의 변화성과 불완전성을 간과하는 것이 되어 오히려 전체를 조망하지 못하는 체계를 낳게 된다. 실재 이 의미역 배당 체계를 세울 때에도 말뭉치를 분석하여 같은 항목에 넣을 수 없는 반증이 나오면 자질을 추가하는 방식으로 이루어졌다.

30) 체계는 전체 안에서 개체가 각각의 기능을 하며 서로 관계를 가지는 것을 말한다.

2) 동일자질 결합원리

: 단어 형성을 이루는 어간과 어미는 품사 자질이 동일해야 한다.

이 원리는 이 연구에서 확장 적용시켰다. 문장에서 동사가 전접구와의 자질 승계를 이루는 구뭉음을 통해 자질이 결합하여 주어와 자질연산을 이루는 것이다. 또한 동사가 합성 동사이거나 본-보조 동사의 경우도 해당된다.³¹⁾

이 연구를 위한 사전 준비로 샘플링한 100문장에 대한 자질 연산 결과 데이터베이스는 별첨에 제시하였다.

3.3. 주어의 의미역 설정

앞장에서는 의미역 체계 설정의 문제점과 그것의 방법론적인 해결 이론으로 자질 연산 문법의 기본원리와 한국어 의미역 체계의 알고리즘을 제안하였다.

기존의 주어의 의미역은 주로 <행위자역>과 <주제역>으로 나누는 것이 일반적이었다. 논의에 따라서는 <수혜자역(Beneficiary)>과 <피행위자역>을 주어의 의미역으로 설정하기도 했다(Jackendoff 1990:261). 우리나라에서는 오충연(2001:210)에서 주어의 의미역을 동사의 문법적 자질인 자동/타동성과 피/사동성, 형용성/계사성으로 나누어 결합가와 함께 제시한 시도가 있었다.

31) 이 연구에서 이 논의는 깊게 다루지 않는다. 연결 동사(serial verb)에 관한 논의는 Sebba(1987)를 참조. 국내 논의는 강현화(1998)가 있다.

주어의 의미역 연구
3. 주어의 의미역 설정

동사	항가	주어의 의미역	비고
형용사	1	대상(Theme)	
자동사	1	대상(Theme), 행위주 (Agent)	
타동사	2-3	행위주 (Agent)	
‘이다’ 등	1	대상(Theme)	결합가를 2항으로 볼 수도 있음
피동사	1	대상(Theme)	피행위자는 대상의 하위부류
사동사	2-3	행위주 (Agent)	사동주는 행위주의 하위부류

〈표 9〉 오충연(2001)에 제안된 주어의 의미역

그러나 주어의 의미역을 설정함에 있어서 앞서 제시한 격 형태 분포를 볼 때 <행위자역>과 <주제역>으로만 나누고 하위 부류를 나눈다는 점은 하위부류를 <행위자역>과 <주제역>이 포함하고 있다는 점에서 결국 하위 부류의 설정 자체가 불필요하게 된다.

따라서 이 연구에서는 자질 연산 문법이라는 방법론적인 적용으로 내면 격 부여 알고리즘을 제안한 김원경(2000)의 연구를 검증 보완하는 시도를 하고자 한다.

구조격 정보	선행 명사	서술어	내면격
Sub	animate		Agent
Sub	animate	passive	Object
Sub	animate		Agent
Sub		cause	Instrument
Sub	L	stative	Locative

〈표 10〉 김원경(2000)에 제안된 주격 명사구의 내면격 부여 알고리즘

위의 표를 자세히 보면 내면격(의미역)은 <행위자역>, <대상역>

주어의 의미역 연구
3. 주어의 의미역 설정

(Object)〉, 〈도구역〉, 〈처소역(Locative)〉으로 나뉘었다. 이 알고리즘을 100개의 가공된 말뭉치 문장으로 점검한 결과 의미상 〈대상역〉을 〈주제역〉으로 전환하였고, 〈행위자역〉은 〈경험자역〉과 〈수혜자역〉로 확대하였다. 또한 〈처소역〉의 경우는 격 형태가 주는 혼동으로 인하여 부사격(adjunct) 설정으로 전환하였다. 그리고 4장에서 더 자세히 논할 예정이지만 〈보충역(Complement)〉을 새로이 설정하여 이중 주어 구문에서의 문제를 보완하였다. 정리하여 주어의 의미역을 표로 제시하면 아래의 진한 표시로 제시된 5개의 의미역으로 나타낼 수 있다. 이들의 연산 절차식은 4장에서 각각의 의미자질의 설명과 함께 논의할 것이다.

주어의 의미역 연구
4. 주어의 의미역 부여 방법

내면격	구조격	논항의 고유자질	격형태	서술어 자질
Experiencer 경험자역	sub 주격	animate	-가	cognitive [인지성]
Agent 행위자역	sub 주격	animate or group	-가	action [동작성]
	adjunct 부사격	group	-에서	action [동작성]
Beneficiary 수혜자역	sub 주격	animate or group	-가	passive [수동성] benefit [수혜성]
Instrument 도구역	sub 주격		-가	causer [사역성]
Theme 주제역	sub 주격		-가	stative [상태성]
				passive [수동성]
				copula [계사성]
Complement 보충역 *떨어짐 (wastebasket)	sub		-가	* 주격중출문에서 주 어가 Theme인 경우

〈표 11〉 주어 의미역의 자동 배당 알고리즘

4. 주어의 의미역 부여 방법

이 장에서는 주어의 의미역 설정 방법론을 제한하였다. 3장에서 밝힌 바

와 같이 객관적인 의미역 체계를 설정하기 위해서는 의미역을 산출할 수 있는 규칙이 필요하다. 이 규칙은 말뭉치를 기반으로 동사와 논항의 자질 설정을 통해 이루어질 수 있다. 자질의 설정은 말뭉치를 가공한 의미역 데이터베이스를 통해 결정할 수 있으며 이를 기반으로 의미역을 상정할 수 있다. 이 장에서는 가장 중심적인 작업이 되었던 말뭉치의 가공 절차를 제안하였고 동사와 논항의 결정 자질들을 제시하였다.

4.1. 의미역 데이터베이스 구축 방법

이 연구에서는 앞서 제시된 자질 연산 문법을 이론적 기반으로 하였으며 세종 말뭉치를 가공하여 데이터베이스를 구축하였다. 이 데이터베이스는 1000만 어절인 대규모 말뭉치에서 단문 분할한 1000문장 중에서 가공 절차를 거쳐 정규화된 100문장을 대상으로 적용함으로써 알고리즘의 신뢰성을 입증할 뿐 아니라 말뭉치의 태깅 정보를 통해 원시 말뭉치에서 이용할 수 없는 깊은 언어 정보를 데이터베이스를 통해 쉽게 접근할 수 있다. 또한 이미 구축된 구문 분석 말뭉치 정보와 비교함으로써 상호 신뢰성을 제고하는 효과도 얻을 수 있다.

이 연구에서는 말뭉치에 대한 의미역 배당 절차 확립을 단면적으로 해석하는 것이 아니라, 일정한 문장들을 그룹으로 관리함으로써 의미역의 출현 빈도에 대한 통계를 입체적으로 해석하는 것이 가능하다. 즉, 1그룹에서 알고리즘의 신뢰성이 90%라고 가정할 때 70% 이하의 신뢰성을 보이는 것이 2그룹 이상 출현한다면 이 알고리즘의 신뢰성은 전체 문장의 평균치 이하로 해석 가능하다. 또한 대상이 되는 10그룹에서 비슷한 신뢰도를 보인다면, 비록 그 신뢰도가 낮다고 하더라도 알고리즘을 개정하여 신뢰도를

주어의 의미역 연구
4. 주어의 의미역 부여 방법

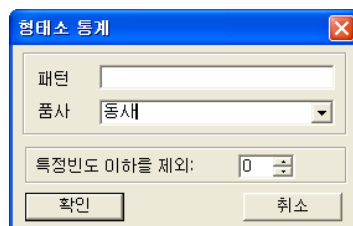
연속적으로 높일 수 있는 방법론을 확립하는 것이 가능하다.

이러한 방법론의 확립은 언어학자들이 정보처리의 알고리즘을 특별한 기계 장치의 도움 없이 예측 가능하게 해 준다. 만약 위와 같은 방법으로 전체 문장의 의미역 배당 알고리즘이 매우 높은 신뢰성을 보인다면, 이제까지 실제의 기계 구현에서 보인 바와 같이 생산 가능한 임의의 문장 집합에 대해서도 비슷한 신뢰도를 기대할 수 있다. 즉, 어떤 기계가 대상 문장에서 정확률 95% 수준의 분석값을 산출한다면, 임의의 문장 집합을 몇 배로 확장해도 비슷한 결과값을 낼 수 있다는 뜻이다. 따라서 이 연구에서 대상으로 한 말뭉치 그룹들의 문장에 적용 가능한 알고리즘이 확립되면, 한국어의 어떤 문장도 만족스럽게 의미역 배당이 이루어질 수 있게 된다.

세종 말뭉치 가공의 세부 절차는 다음과 같다.

먼저 국립국어원에서 말뭉치와 함께 제공하는 분석도구인 글잡이Ⅱ로 대상 말뭉치를 분석한다. 세종 말뭉치는 자료의 용량이 워낙 커서 용례추출까지는 이 분석도구가 반드시 필요하다. 이 자료는 www.sejong.or.kr에 들어가면 분석도구와 함께 자료실에 공개되어 있다.

이번 연구의 말뭉치 자료처리 과정은 다음과 같다. 첫 번째로 글잡이Ⅱ에서 세종 말뭉치 구어(spoken)와 문어(written)를 연다. 그 다음으로 형태소 통계처리를 하여 상위최빈도 순위별로 동사를 추출한다. 형태소 통계창은 아래 <그림1>과 같고 결과값은 <그림2>와 같이 나온다.



<그림 1> 형태소 통계

주어의 의미역 연구
4. 주어의 의미역 부여 방법

형태소	빈도	전체비율(%)	상대비율(%)
하/동사	88943	0.44418263	7.02196360
되/동사	44708	0.22327241	3.52965331
보/동사	23502	0.11736933	1.85546017
말하/동사	21216	0.10595302	1.67498267
가/동사	16584	0.08282074	1.30929077
위하/동사	16567	0.08273584	1.30794859
틀/동사	16142	0.08061338	1.27439523
받/동사	15944	0.07962457	1.25876331
보이/동사	14506	0.07244318	1.14523458
따르/동사	14211	0.07096994	1.12194467
알/동사	13780	0.06881752	1.08791769
쓰/동사	13301	0.06642539	1.05010104
나오/동사	12440	0.06212554	0.98212594
모르/동사	11832	0.05908918	0.93412495
만들/동사	11668	0.05827016	0.92117727
대하/동사	11218	0.05602286	0.88565022
살/동사	11210	0.05598290	0.88501865
오/동사	9893	0.04940579	0.78104275
나/동사	9257	0.04622959	0.73083115
이하/동사	8742	0.04365768	0.69017243
통하/동사	8631	0.04310334	0.68140906
먹/동사	8307	0.04148528	0.65582961
가지/동사	7945	0.03967744	0.62725008
내/동사	7587	0.03788959	0.59898627
있/동사	7345	0.03668104	0.57988065
이루/동사	6839	0.03415406	0.53993243
주/동사	6699	0.03345490	0.52887958
것/동사	6602	0.03297048	0.52122152
몸/동사	6390	0.03191175	0.50448430
찾/동사	6382	0.03187180	0.50385273
느끼/동사	6320	0.03156217	0.49895787
밝히/동사	6068	0.03030368	0.47906271
나타나/동사	6029	0.03010891	0.47598371
틀/동사	5984	0.02988418	0.47243100
죽/동사	5949	0.02970939	0.46966779
틀어가/동사	5878	0.02935482	0.46406239
알/동사	5491	0.02742213	0.43350914
서/동사	5183	0.02588398	0.40919283
부르/동사	5099	0.02546448	0.40256110
잡/동사	5050	0.02521977	0.39869261
말대나/동사	5012	0.02503000	0.39569253
그리/동사	4839	0.02416604	0.38203436
미르/동사	4829	0.02411610	0.38124487
맞/동사	4674	0.02334202	0.36900777
나서/동사	4605	0.02299744	0.36356029
내리/동사	4470	0.02232325	0.35290217

통계결과 총 4163 건 정렬방법
준비 색인디렉토리 - 색인

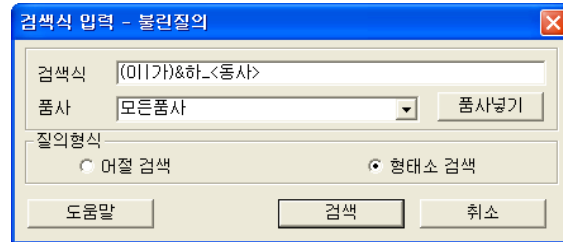
<그림 2> 최빈도 동사 통계창

이 상위최빈도수 동사를 통계 낸 데이터는 김홍규·강범모(2000)의 ‘한국어 형태소 및 어휘 사용 빈도의 분석의 빈도’ 데이터와 비교하였는데, 이 비교는 데이터의 신뢰도를 높이는 선행 작업으로 최빈도 동사들의 상위 분포를 제한한다.³²⁾ 상위빈도 개별동사가 설정되면 각 개별동사별로 한국

32) 김홍규·강범모(2000)와 실제 국립국어원에서 배포용으로 제공한 말뭉치(색인)를 글잡이II를 통해 분석한 결과값이 상이하게 나왔다. 예를 들어 말뭉치 최빈도 동사 ‘하-’는 실제 분석에서 ‘102413’에 달하는 빈도수를 보였으나, 김홍규·

주어의 의미역 연구
4. 주어의 의미역 부여 방법

어 의미역 산출을 위한 불린 검색식을 만들어서 추출하였다.



〈그림 3〉 불린 검색식

그러나 위낙 대용량 말뭉치를 다루는 글잡이Ⅱ에서는 개별동사별 용례추출에서 추출률 15%가 넘으면 연산 오류가 반복적으로 발생하였다. 이것은 전체 말뭉치를 추출하는 과정에서 문장 추출 상에 발견되는 오류로 해결할 수 없는 근본적인 문제임이 발견되었다. 따라서 오류를 피하기 위한 방법으로 용례추출률을 15% 이하로 제한하였다. 이는 각 동사별로 약 3000문장 이하이며 무작위 추출이므로 자료의 선정에 객관성에는 문제를 미치지 않는다. 말뭉치 분석도구는 글잡이Ⅱ 이외에도 분석의 정확성을 위해서 각각 말샘41과 배포용 감썩새33)를 함께 사용하였다. 이 분석도구들은 글잡이Ⅱ가 동사의 동음이의어나 표기의 오타로 인한 추출오류를 방지하기 위한 보조적 수단으로 재검증 차원에서 사용하였다.

강범모(2000)에서는 ‘18464’에 그치고 있다. 또한 실제 분석에서 4번째로 높은 빈도수를 보였던 동사 ‘말하-’는 김홍규·강범모(2000)에서는 상위 100위권 순위에 나타나지 않는다. 이는 다루었던 말뭉치의 텍스트 종류와 어절수의 차이에서 기인하겠으나, 본 연구에서 분석한 말뭉치는 spoken과 written을 모두 분석대상으로 삼았음을 미리 밝혀 둔다. 말뭉치 대상 빈도 분석 목록은 [부록]에 자세히 제시 하겠다.

33) 감썩새는 전주대 소강춘팀에서 제작한 통합형 한글 자료 처리기(SYNthesized Korean Data Processor)이다.

주어의 의미역 연구

4. 주어의 의미역 부여 방법

The screenshot shows the '말샘41' (Malsam41) software interface. The main window displays a list of words and their associated grammatical information. The table below represents the data shown in the screenshot.

단어	형태	품사	빈도수	사용자 정의
====전제말====				
문장: 715	[(이, <보인>)]	보인	1	
문장: 15293	사	조사	6	
어간형태: 17046	보	보	2	
어간형태: 12062	보	보	2	
보합어: 1469	보	보	111	
문장: 10806	보	보	2	
문장: 37	보	보	2	
문장: 282	보	보	2	
문장: 4320	보	보	2	
문장: 306	보	보	17	
문장: 345	보	보	2	
문장: 14	보	보	2	
문장: 342	보	보	695	
문장: 35	보	보	480	
문장: 302	보	보	210	
문장: 6835	보	보	1	
문장: 5831	보	보	983	
문장: 189	보	보	642	
====중립====				
어간형태: 사: 6	보	보	4	
어간형태: 사: 2	보	보	104	
	보	보	23	
	보	보	5	
	보	보	8	

〈그림 4〉 「말샘41」의 어휘 빈도 처리 결과

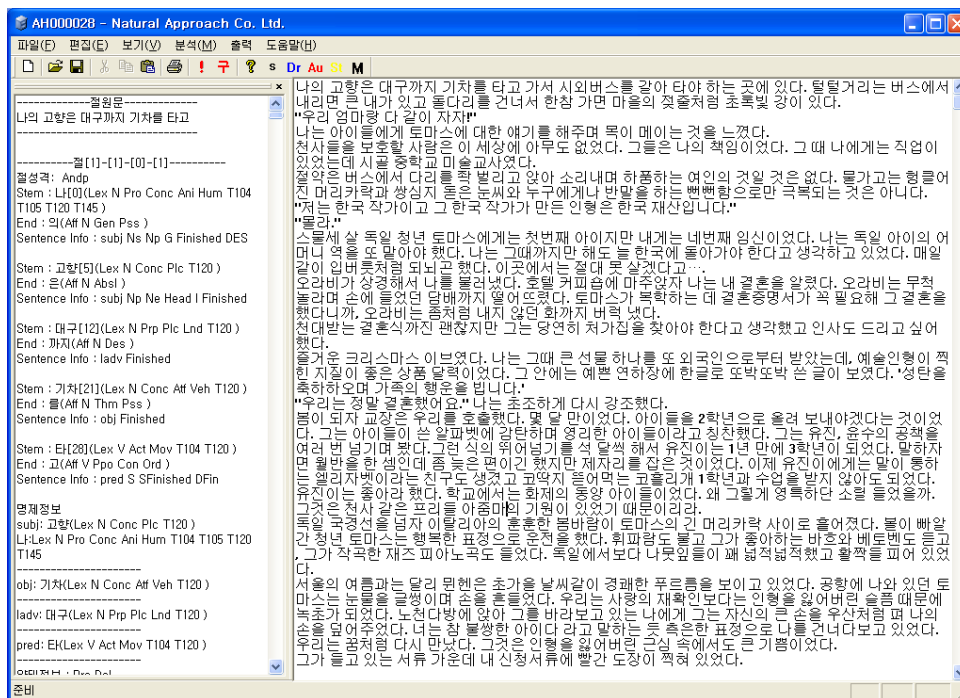
그러나 이 말샘41과 감작새를 사용하기 위해서는 세종 말뭉치의 태그 제거가 필수적이다. 태그가 장착된 말뭉치는 다른 분석도구에서 사용할 수 없기 때문이다. 따라서 본 연구에서는 TEI Markup을 제거하기 위하여 ‘chk_tei2.exe’³⁴⁾를 사용하였다. TEI Markup은 각 어절 단위마다 컴퓨터가 품사 등등의 정보를 인식하게 하기 위에 부착한 말뭉치 태그 표지이다. 따라서 TEI Markup 제거 프로그램인 ‘chk_tei2.exe’를 사용하여 명령 프롬프트에서 제거해야만 다른 분석도구에서 사용이 가능하다.

글잡이II로 분석된 각 동사별 15%(약 3000여개의 문장)추출 용례는 텍스트파일 형태로 저장하여 구문분석기로 분석하는 절차가 필요하나 현재 국립국어원에서 배포한 구문분석 말뭉치는 최운호(2005)³⁵⁾에서 지적한

34) 21세기 세종계획 홈페이지에 들어가면 공개자료실에 ‘국어 정보 처리 프로그램’ 게시판이 있다. 여기에서 ‘세종 말뭉치의 TEI Markup 제거 프로그램’을 다운로드 할 수 있다.

주어의 의미역 연구 4. 주어의 의미역 부여 방법

바와 같이 성분어로 태깅된 말뭉치가 30만 어절 정도만 구축되어 있어서 아직 구문분석 연구 자료로써 상용화하기 힘든 단계이다. 여기에서 보다 많은 어절을 다루는 구문분석기의 분석도구용 말뭉치 제작이 시급하다는 것을 보여주고 있다. 다만 이 단계에서 단문분할 수작업의 신뢰도를 올리기 위해 <그림5>의 고창수 · 김원경이 개발한 구문 분석 및 어휘통계처리 시스템인 ‘구문분석 시스템 워크벤치(workbench)’로 성분어 구문분석을 병행하였다.

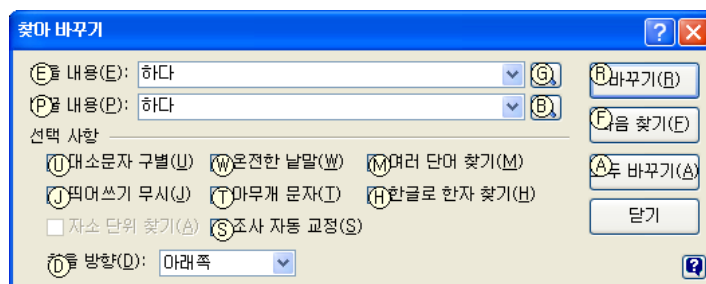


<그림 5> 단문분할 작업과 병행하는 구문분석기 워크벤치(workbench)

35) 최운호 · 박병선(2005)에서는 ‘21세기 세종계획’의 ‘국어기초자료분과’에서 2001년도에 구축을 위한 기초 연구를 바탕으로 2002년부터 2004년까지 3년에 걸쳐 구축한 30만어 어절 규모의 ‘구문분석 말뭉치’의 활용 방법을 논하고 있으나 규모가 30만 어절에 불과하기 때문에 성분어 분석에 부족한 면을 지적하였다.

주어의 의미역 연구
4. 주어의 의미역 부여 방법

이렇게 분석된 용례파일은 보다 정밀한 연구자 수작업을 위해 <그림6>과 <그림7>에서 보이는 바와 같이 한글2005 프로그램에서 색표시 작업이 다시 필요하다. 한글 2005에서는 매크로 기능을 활용하여 단문분할작업의 전처리로서 용례파일을 각 격 표지와 서술어별로 색표시 처리한다. 이것은 한글의 ‘찾아 바꾸기’ 기능에서 색지정³⁶⁾을 한 후 ‘모두 바꾸기’를 설정하는 아주 간단한 기능으로 연구의 수작업에서 가장 큰 효율을 올리게 해 준다.



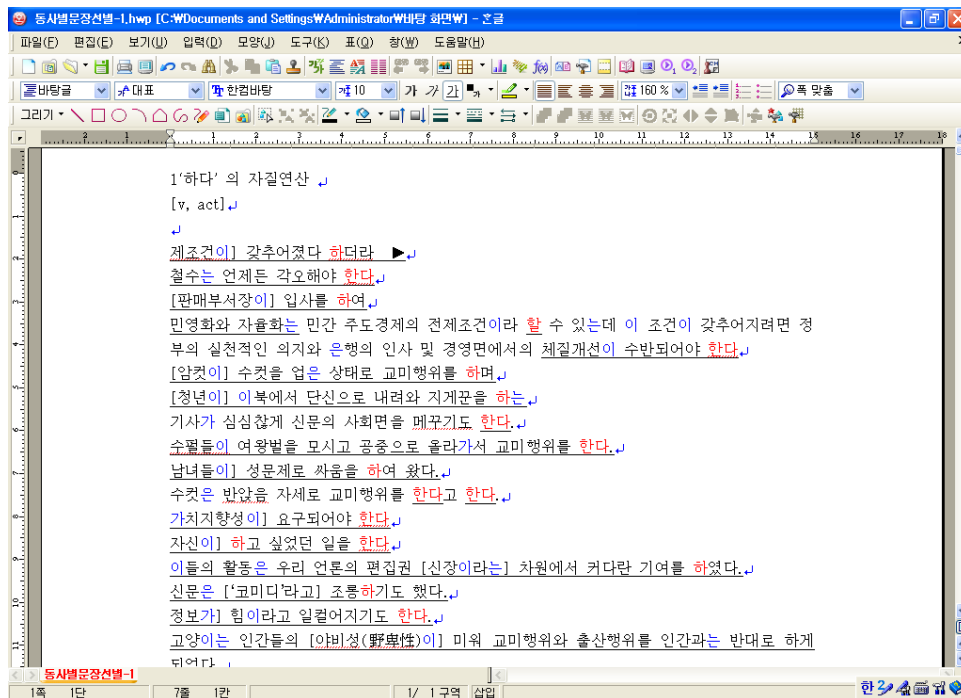
<그림 6> 주어와 서술어의 「색표시 처리」 최초과정

이 연구의 자료 처리 중에서는 단문 분할 부문이 가장 중요하다. 왜냐하면 말뭉치는 현재 사용되는 무작위의 문어와 구어를 대상으로 하기 때문에 적형의 문장이 아닌 것들이 많다. 이들은 모두 문장의 성분위치보다는 표지를 치환하여 호응관계를 따져보아야 보다 신뢰성 있는 자료를 얻을 수 있다.

36) ‘글자모양’에서 글자색 지정을 하면 된다.

주어의 의미역 연구

4. 주어의 의미역 부여 방법



<그림 7> 용례의 단문분할을 위한 「색표시 처리」

이러한 과정에서 얻은 단문들은 알고리즘 작업을 위해 엑셀 프로그램을 사용하여 정리해야만 한다. 아래 <그림8>과 같이 의미역 자동배당 알고리즘을 구조화하기 위해서는 우선 각 문장별로 수작업으로 분석한 조건들을 모두 입력한다. 입력 후 각 개별 의미역별로 필터링하여 정렬시키면 의미역의 조건들의 점검이 용이해진다.

주어의 의미역 연구

4. 주어의 의미역 부여 방법

Microsoft Excel - 의미역 분류

	A	B	C	D	E	F	G
	의미역	구조격	논항자	술어자	문유	서술어	문장
1	Experiencer	subject	animate	cognitive		읽다	사람이 책을 읽는다
2	Experiencer	subject	animate	cognitive		읽다	어린 아이가 책을 읽는다
3	Experiencer	subject	animate	cognitive		읽다	경제인들은 [68기] 신문을 직접 읽고
4	Agent	subject	group	action		치다	[일본이] 자기 안방 차지하듯 곳곳에 진을 치고 있기 때문이다.
5	Agent	subject	animate	action		치다	[후배들이] 박수를 친 것
6	Agent	subject	animate	action		치다	군인 두 사람이 민트를 치고
7	Agent	subject	animate	action		치다	저들은 자기네들이 성공하였다고 큰 소리를 쳤다.
8	Agent	subject	animate	action		보내다	경연위원회 위원장이나 사장이 직접 답신을 보내기도 한다.
9	Agent	subject	animate	action		보내다	그러나 삼척기획처럼 [사원들이] 취미활동만으로 하루를 보내는 것은 아니다.
10	Theme	subject				생기다	여성의 [얼굴이] 남성의 얼굴에 비하여 일반적으로 [잘 생겼다]
11	Theme	object	animate			생기다	현주는 [욕심이] 생겼다.
12	Experiencer	subject	animate	cognitive		생기다	현주는 [욕심이] 생겼다.
13	Source	object			에서	생기다	수치심은 도덕관념이나 윤리의식에서 생긴다.
14	Source	object			에서	생기다	(미스터 강의 호텔예약에서 [문제가] 생겼다.
15	Locative	object	space		에서	생기다	우리나라에서는 남자는 '반지' 여자는 '치마'라는 [고정관념이] 생겼다.
16	Theme	subject		copular		이다	정책목표는 당연히 미운국대화에 두어지게 [마련이다.]
17	Agent	subject	animate	action		두다	[TV프로듀서들이] 시청자들의 조급성을 염두에 두는 건 당연하고 필요한 일임에 틀림없다
18	Locative	object			에	두다	시민정치는 조직보다는 토론의 내용에 더 의미를 둔다
19	Theme	subject				떨어지다	지지가 할만큼 떨어졌다.
20	Theme	subject				떨어지다	방송의 신뢰도는 양에 떨어지고 말았다.
21	Theme	subject				떨어지다	감기가 떨어지는
22	Locative	object			에	남다	이념에는 [편향성이] 강하게 남아 있다
23	Theme	subject				남다	이념에는 [편향성이] 강하게 남아 있다

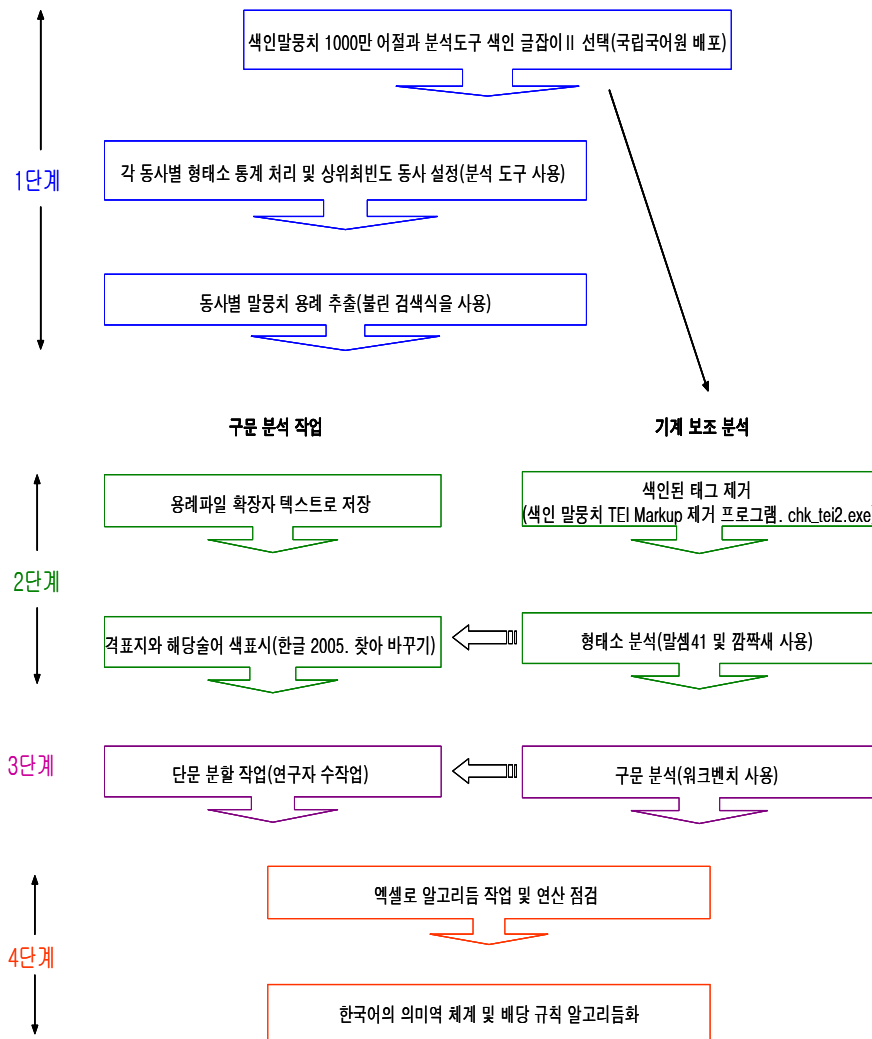
〈그림 8〉 엑셀을 이용한 의미역 자동배당 규칙 「데이터베이스」

이 연구 작업 진행 과정에서 말뭉치 대상 연구의 어려움을 몇 가지 확인할 수 있었다. 첫 번째로 세종 말뭉치는 다른 분석도구에서 사용하기 위해서는 태그 제거를 해야 하는 번거로움이 있다는 것이다. 게다가 명령 프롬프트를 사용하여 제거해야 하기 때문에 명령 프롬프트 프로그램에 능숙하지 않은 입장에서는 말뭉치 활용의 기회가 제한된다. 두 번째로는 불린 검색한 문장의 신뢰도가 너무나 낮다는 데에 있다. 오류 문장이 많을 뿐만 아니라 추출도중 프로그램 오류가 일정 추출률에서 반복적으로 나타났다. 글잡이Ⅱ는 앞으로의 수정판 작업에서 이와 같은 문제점을 해결할 것을 기대한다. 세 번째로는 구문분석 말뭉치 구축의 아쉬움이 있다. 말뭉치 계량 분석에 있어서 보다 많은 자료의 분석처리가 연구에 선행되어야 하는데 연구자의 수작업에는 한계가 있기 때문이다. 또한 연구의 단문분할 수작업을

주어의 의미역 연구
4. 주어의 의미역 부여 방법

하면서 구문분석기 연산 규칙의 연구도 모색되면 앞으로의 연구 진행에 많은 도움이 되리라고 본다.

이러한 작업의 순서도를 시간적 순서에 따라 자세히 기술하면 다음과 같다.



<그림 9> 말뭉치 가공 절차 순서도

4.2. 논항과 동사의 의미자질

논항은 명사구로 이루어져있다. 명사구의 의미자질은 곧 명사의 의미부류에 의해 결정되며 본 연구에서도 기존의 명사 의미부류에서 크게 벗어나지 않고 자질을 설정하였다. 한국어 명사의 의미 분류는 최경봉(1996)에서 논의되었다. 최경봉(1996)에서는 의미 영역의 분류에 대한 견해로 Nida(1975)와, Lyons(1977)을 따르고 있으며 나은미(2004)에서는 이러한 최경봉(1996)의 연구를 수용하여 존재론적 분류방식³⁷⁾을 적용하였다. 이들 논의에서 분류하는 명사는 “실체”라는 네 가지 영역 중 하나로 설정되어 있다. 이 실체 명사는 먼저 ‘인간’과 ‘사물’로 하위분류를 하고 있는데 ‘인간’을 사물과 양립시킨 것은 언어에 나타나는 인식 과정의 특수성을 반영한 결과이다.

그러나 이 연구에서는 김원경(2000)에서와 같이 논항의 의미자질을 설정함에 있어서 [유정성]을 먼저 제안한다.

1) 주어의 [유정성/animate)] 자질

주어의 의미역을 배당하는 결정자질인 [유정성]의 설정은 ‘인간’과 ‘동물’을 하나의 영역으로 하위분류 한다. 다만 식물은 스스로 ‘의도성’을 가지지 못하므로 유정성에 함께 분류하지는 않는다. [유정성]을 갖는 주어의 경우 동사가 상태성 또는 동작성을 가질 때 <행위자역(Agent)>으로 설정된다.

37) Lyons(1977)과 Nida(1975)에서 의미 영역은 기본적으로 ‘실체, 사건, 상태, 관계’로 나누고 있다.

(15) 가. 은행나무가 마주 보다. → <경험자역(Experiencer)>

식물

나. 벌레가 집을 짓는다. → <행위자역(Agent)>

동물

다. 친구가 문을 연다. → <행위자역(Agent)>

사람

2) 주어의 [단체성(group)] 자질

주어의 의미역을 상정할 때 논항의 의미자질 중 그 설정이 가장 광범위한 것이 바로 [단체성(group)]이다. 여기에서 말하는 [단체성(group)]이란 단순히 복수형을 의미하지는 않는다. 아래의 예를 통해 보듯이 [유정성]을 가지고 있지는 않지만 ‘의인화’가 가능하여 유정물과 공기 가능한 동사들이 사용된다. 실재 말뭉치에서 이러한 [단체성]을 가지는 문장들의 비중이 상당히 높다.

(16) 가. 언론들은 대개가 침묵을 지켰다.

나. 국가가 넘어지다.

이미 서론에서 논항에 의미역을 부여하는 결정 인자는 서술어인 동사에 있음을 상술했다. 그러한 동사의 의미자질을 설정하기에 앞서서 한국어의 용언 분류 체계를 살펴보면 두 가지 방향으로 연구되었음을 알 수 있다. 하나는 의미적·기능적 속성에 따라 용언을 분류하여 계층적 구조로 갈래를 제시하였고, 다른 하나는 의미적 범주에 따라 비슷한 의미 부류로 묶이는 용언들을 하나의 집합으로 구성하여 병렬적 구조로 제시하였다. 본 연구에서는 홍재성 외(1997)을 참조하여 말뭉치와 병행하며 동사의 의미

주어의 의미역 연구
4. 주어의 의미역 부여 방법

자질을 설정하였다. 동사의 의미자질을 제시하기에 앞서 기본적으로 수용한 남경완(2005:114)에서 참고한 용언의 의미적 속성을 밝혀 둔다.

용언의 의미 분석 기준	의미적 속성(의미 자질)
분류 체계의 측면	[상태성]과 [동작성], [자동성]과 [타동성]
의미 부류의 측면	[물리성]과 [정신성], [이동성]과 [비이동성]
문법적 범주의 측면	[사동성]과 [주동성], [피동성]과 [능동성]

〈표 12〉 남경완(2005)에서 제안된 용언의 의미 분석 기준과 의미적 속성

1) 동사의 [동작성(action)] 자질

흔히 학교 문법이나 용언을 분류하는 연구에서 보면 동사와 형용사의 구분은 [동작성]과 [상태성]으로 구분한다. 그러나 본 연구에서 다루었던 동사들에는 ‘상태 동사’들의 모습이 보인다. 이제까지 많은 논의가 있어왔듯이 동사와 형용사를 가르는 기준이 [동작성]의 유무만을 따지는 것에는 문제가 있다.

하지만 앞서 제시하였다시피 이 [동작성]은 주어에 <행위자역>을 부여하는 대표적 자질이다. 아래의 (17)처럼 ‘타다’나 ‘열다’는 전접구의 자질승계만 이루어지지 않으면 동사 자체가 가지는 [동작성] 자질로 인하여 <행위자역>을 자동 배당 받게 되는 것이다.

(17) 가. 젊은 여자가 폼내면서 차를 탄다.

나. 아내의 친구가 문을 열다.

2) 동사의 [상태성(stative)] 자질

동사 중에서도 [상태성] 자질을 갖는 것들이 있다. 이는 전접구의 [상태성] 자질을 승계 받아서 이루어지기도 하고 자체가 상태성 동사인 것들도 있다. 이들은 주어에 <주제역>을 자동 배당한다.

(18) 가. 경기가 장기적인 [과도를 탄다.]

나. 주장이 설득력을 [지니게 된다]

(18가)의 경우, 전접구와 묶이면 연어적인 표현이 된다. 따라서 [상태성] 자질을 승계 받는다고 볼 수 있다. (18나)의 경우는 ‘-게 되다’의 문형이 [상태성] 자질을 가지고 있기 때문에 구 묶음이 되면서 [상태성]을 갖는다.

3) 동사의 [인지성(cognitive)] 자질

[인지성]은 <경험자역>을 결정짓는 자질로써 감각, 지각, 인지 동사류를 포괄하여 자질을 부여할 수 있다. 또한 전접구의 경우는 (다)와 같이 ‘마음의 감정상태’를 나타내는 명사 의미자질로 승계가 이루어질 수 있다.

[인지성]은 논항이 [유정성] 또는 [단체성]을 가지고 있을 때 상호 연산하여 <경험자역>을 부여 한다.

동사의 의미부류에서도 이 [인지성]은 주어가 기본적으로 가지는 ‘주체’라는 위치 때문에 결정적 자질로 연산되는 것이다.

(19) 가. 사람이 책을 읽는다.

나. 김현식씨는 일편단심 철썩같이 믿었다.

다. 현주는 [욕심이 생겼다.]

4.3. 주어 의미역 연산 절차

앞서 제시하였던 논항과 의미자질들에 의해서 주어의 의미역이 연산된다. 본 장에서는 지금까지 제시하였던 주어 의미역 알고리즘을 통해 의미역이 자동 배당되는 절차를 보여줄 것이다.

주어의 의미역 연산 절차는 개념적인 정의를 적용하여 사전 배당하는 것이 우선이다. 이것은 연산의 절차에서 기존의 의미역 이론들의 틀을 배제하는 것이 아니라 그대로 적용하고 문제점들은 해결하는 과정에서 자질연산을 도입하여 검증하는 방식을 취하는 것이다.

예를 들어 술어가 가지는 자질이 [동작성]과[상태성] 등등이 있다고 가정했을 때 자질이 더 추가될 것인지의 문제부터 직면하게 된다. 하지만 자질연산문법에서 제안하는 ‘자질포화원리’는 항상 연산을 적용하는 시점에서 자질이 더 추가될 것인지를 판단하기 때문에 미리 자질을 전부 체계화하지 않는다. 또한 해당 술어가 가지는 자질이 전접구에서 승계 받아 자질삭제가 일어날 수도 있고, 상위 체계표에서 보여주듯이 논항자질과, 술어 자질, 격 형태, 구조격의 조건들이 전부 필요할 수도 있고 한가지만으로 변별이 일어날 수도 있다. 이는 다음의 선별된 예문에서 자세히 살펴보고자 한다.

(20) 가. 여성이 물에 빠지다.

나. 국가가 넘어지다.

다. 암수가 [짝을 짓다.]

- 라. 개발도상국이 피해를 받는다.
마. 빛이 6 조원이 넘는다.
바. 우리들은 다시 명량한 [기분으로 돌아왔다.]
사. 우리나라에서는 남자는 ‘바지’ 여자는 ‘치마’ 라는 고정관념이
생겼다.
아. 미스터 강의 호텔예약에서 문제가 생겼다.
→ 호텔에서 문제가 생겼다.

(20가)은 우선 주어가 [유정성]을 갖는다. 자질의 부여는 각 조건들에서 확정적인 것에서 모호한 순서로 진행한다. 이것은 결국 다른 의미역과의 구분을 자질의 총합인 자질복합체로 결정하는 방식이라고 설명할 수 있다. (20가)에서 술어가 [동작성]이 확인되면 주어가 갖고 있는 자질과 범주 부합 여부를 보고 주어에 <행위자역>을 부여한다. 하지만 (20나)과 같이 주어가 [단체성]을 가진다면 이것은 [유정성]의 대유적인 개념으로 상정하여야 범주 논란에서 자유로워질 수 있다.

(20다)의 경우는 동사의 전접구를 ‘구뭉음’을 통해 ‘관용어 및 연어 처리’를 하고 [상태성]을 부여한다. (20라)에서 ‘주다/받다’의 [수혜성(benefit)]을 가진 동사는 주어에 ‘<수혜자역>’을 부여하고 주어는 [단체성] 자질을 가지므로 ‘<행위자역>’과 ‘<주제역>’의 구분을 조건 부여 절차를 통해 지을 수 있다.

(20마)의 경우는 주어가 ‘<주제역>’이 될 때 따라오는 조건으로 목적어 자리는 ‘<보충역>’ 처리를 하게 된다. 이는 앞에서 이미 상술한 ‘떨이함’의 설정으로 이중주어의 중출문에서 적용하게 되며 체계의 빈자리를 설명하려는 필수 불가결한 설정이다.

(20바)는 전접구의 구뭉음을 통해 [인지성]을 자질 승계 받는다. 반면에 주어의 격 형태로 ‘에서’가 보이는데 (20사)에서는 주어의 설정에서 ‘에

서’가 주어로 해석될 수 있으나 다른 주어 후보인 ‘고정관념이’가 존재하기 때문에 부사격으로 다루며, <처소역>으로 설정한다.

마지막으로 (20아)는 (20사)와 같은 관점에 놓고 주어로 보지 않는다. 하지만 (20아)의 두 예문은 의미적 모호성을 보여주고 있다. 부사격에 오는 자질을 둘 다 [공간성]로 볼 것인가가 문제되기 때문이다. 하지만 이것은 동사구의 ‘문제가 생긴’ 장소의 의미가 더 강하므로 ‘<원천역 (Source)>’을 설정한다.

이제까지의 예문들을 종합해 보면 우선 자질부여절차는 다른 의미역과의 구분점이 나뉘는 선까지 자질을 추가하여 진행한다. 본 연구에서 제시한 13개의 의미역들은 포화 가능한 최소한의 자질들로 합산되어 있다. 그리고 주어진 조건의 자질들을 전접구와 묶어서 자질 승계 여부를 알아보며 연어 및 관용구를 처리할지를 본다.

따라서 의미역 배당 연산 절차는 아래와 같은 전제 조건을 필요로 한다.

(21) 한국어의 의미역 체계 및 배당 규칙을 위한 전제 조건

가. 구조격 배당

나. 명사구 묶기

다. 동사구 묶기

라. <공동역(Comitative)>인 경우 ‘와/과’가 나오면 묶기

마. 연어 및 관용어 처리

바. 자질 승계

사. 주격 중출문의 경우, 1주어가 <주제역>이 배당되면 2주어에 <보충역>을 배당한다.

이러한 절차를 적용하여 데이터베이스화한 100개의 문장은 [부록]에 제시하며 연산절차 과정에서 나타난 남은 과제들은 결론에서 제시할 것이다.

5. 결론

지금까지 주어의 의미역 자동 배당 알고리즘의 연산 절차를 제시해 보았다. 이 연구는 기존에 논의되었던 의미역의 체계 설정 방법을 살펴보고 자질연산문법을 통하여 새로운 방법론을 적용하여 한국어정보처리를 위한 기계적 절차를 제안할 수 있었다.

의미역(Theta-role, θ -role)은 술어(predicate)의 논항이 갖는 의미상의 역할이다. 이런 의미역의 배당 규칙은 기계번역이나 대화형 시스템의 사용을 위해서 반드시 거쳐야 하는 매우 중요한 과제이다. 주어는 아직 많은 부분 주격 중출이나 표지 연구에 머물렀었다. 이에 주어의 의미역을 논의한 본 연구는 이러한 주어에 대하여 서술어와의 관계를 밝힘으로써 {를} 논항에만 집중되었던 의미역 연구에 균형을 주는 효율적인 방법이 될 것이다. 아울러 객관적인 의미역 자질 설정 방법을 제시하여 기계적 절차를 마련하였다. 이러한 절차는 구문분석의 한계점과 해결점을 동시에 보여주었다. 그리고 주어 설정 방법을 객관화하고 자동화함으로써 모어 화자가 아닌 사람들에게 효과적으로 성분어 개념과 의미역 개념을 교육할 수 있도록 한국어 교육 현장에서 도구적으로 활용될 수 있을 것이다.

이제까지 논의한 내용을 간단히 되짚어 보면 다음과 같다.

1장에서는 이 연구의 목적인 주어의 의미역 체계 설정 방법이 새롭게 모색되어야 하는 필요성을 제시하였다. 한국어정보처리를 위해서는 의미역

설정의 알고리즘화가 필수적이다. 이에 이 연구에서 자질연산문법의 이론을 소개하고 말뭉치를 가공한 자료를 대상으로 한 처리절차를 제안하였다.

2장에서는 한국어에서 그동안 논의되었던 주어와 관련된 연구에서 주어의 의미역 연구의 현 위치를 제고해 보았다. 주격 중출 현상은 주어의 의미역 논란의 그 중심에 있으며 여러 논의들을 종합해 보고 주어의 의미역 설정에 기여되는 하나의 논점을 제시하였다.

3장에서는 주어의 의미역을 설정하기에 앞서서 그동안 진행되어 왔던 한국어 의미역 체계의 한계점을 방법론적인 측면에서 지적하였다. 이 장에서는 자질 연산 문법 이론을 토대로 의미역 체계 설정 방법을 모색하여 새로운 의미역 체계를 제시하였다.

4장에서는 3장에서 제시된 의미역 체계를 설정하는 결정 자질들에 대하여 논의하였다. 세종 배포 균형 말뭉치를 가공하는 활용 방안과 그 한계점을 지적함으로써 앞으로 많은 분야에서 적용될 말뭉치의 효용성을 높일 수 있도록 제안하였다. 그리고 주어의 의미역 부여 방법을 제시해서 논항과 동사의 의미자질을 기존의 명사와 용언의 의미 분류를 하였고 실제 말뭉치에 적용하여 연산 절차를 알고리즘화하였다.

주어 의미역의 자동 배당 알고리즘은 본 연구에서 완벽히 구현된 것은 아니다. 모든 언어가 그러하듯이 하나의 체계는 지속적으로 변화하고 있고 늘 불완전성을 가지고 있다. 본 연구에서는 1000만 어절을 대상으로 30만 문장을 각각 100개의 동사별로 추출하였고 이 중에서 다시 1000문장을 추출하여 100문장의 알고리즘 적용 데이터베이스를 얻었다. 하지만 이것도 어디까지나 통계적인 확률을 높이는 것일 뿐 앞으로 더 많은 실제 예문 중심의 적용이 필요하다.

규칙을 세우거나 체계를 설정하는 것은 곧 ‘예측력’을 갖기 위함이다. 하지만 언어는 변화무쌍한 유기체와도 같아서 그 변화의 방향을 파악하기가 쉽지만은 않다. 이에 이 연구에서는 실제 말뭉치를 가지고 규칙을 세우

기 위한 검증을 시도하였고 그것만으로도 기계처리를 위한 의미역 알고리즘을 구축하는 데 기여할 수 있다. 더 나아가 본 연구가 한국어정보처리의 궁극적 목표인 대화형 시스템에 적용되기 위해서는, 더 많은 데이터베이스를 대상으로 의미역 자동 배당 알고리즘의 정밀성을 확인할 수 있어야 할 것이다.

참고문헌

- 강범모·김성도 역. 1998. 「언어의 과학」. 민음사.
- 강승식. 1993. 「음절 정보와 복수어 단위 정보를 이용한 한국어 형태소 분석」. 서울대 컴퓨터공학과 박사학위논문.
- 강신재·박정혜. 2003. “대규모 말뭉치와 전산 언어 사전을 이용한 의미역 결정 규칙의 구축.” 「정보처리학회논문지」. 한국정보처리학회.
- 강현화. 1998. 「국어의 동사연결 구성에 대한 연구」. 한국문화사.
- 고광주. 1997. “능격 동사 구성에 대한 고찰.” 「어문논집」 26. 안암어문학회.
- 고광주. 2002. “국어의 논항구조와 그 유형.” 「한국어학」 18, 389-408, 한국어학회.
- 고석주. 2001. 「국어 격조사에 대한 연구 : ‘이/가’ 와 ‘을/를’ 을 중심으로」. 연세대 박사학위논문.
- 고창수·김원경. 1998ㄱ. “‘이다’ 는 동사인가?” 「한성어문학」 17. 한성대학교 한국어문학부.
- 고창수·김원경. 1998ㄴ. “한국어 선어말어미의 정보처리.” 「한국어학」 8, 한국어학회.
- 고창수. 1992. 「고대 국어의 구조격 연구」. 고려대 박사학위논문.
- 고창수. 1997. “한국어 조사 결합에 대한 연구.” 「한국어학」 5. 한국어학회.
- 고창수. 2002. “언어정보산업의 미래와 언어학.” 「한국언어문화」 22. 한국언어문화학회
- 고창수. 2002. 「자질연산문법이론」. 도서출판 월인.
- 고창수. 2003. “자연언어의 본질과 기능.” 「이중언어학」 23. 이중언어

학회

- 고창수. 2003. “정보추출과 한국어 구문 분석.” 「한성인문학」 1. 한성대학교 인문과학연구원
- 고창수. 2004. “구문분석을 위한 자질연산문법.” 「한국어 정보화와 구문분석」. 고려대학교 민족문화연구원.
- 고창수·김원경·박선옥·권정현·이현주. 2006. “한국어 의미역의 자동 배당 방법론 연구.” 수고본.
- 국립국어연구원. 1999. 「표준국어대사전」. 두산동아.
- 권재일. 1989. “조사의 성격과 그 생략 현상에 대한 한 기술 방법.” 「어학연구」 25-1. 서울대 어학연구소.
- 김귀화. 1988. 「국어의 격연구」. 한신문화사.
- 김미령. 2006. 「국어의 격표지 교체와 의미역 연구」. 고려대 박사학위논문.
- 김민수. 1971. 「국어문법론」. 일조각.
- 김민수. 1981. 「국어의미론」. 일조각.
- 김승곤. 1980. “한국어의 격이론.” 「인문과학논총」 13. 건국대학교.
- 김양진. 1999. 「국어 형태 정보 연구」. 고려대 박사학위논문.
- 김영택. 1994. 「자연언어처리」. 신아사.
- 김영희. 1974. 「한국어의 격문법 연구」. 연세대 석사학위논문.
- 김원경. 1993. 「국어 접사 피동의 생성론적 연구」. 고려대 석사학위논문.
- 김원경. 1997. “‘에게’와 격.” 「한국어학의 이해와 전망」. 도서출판 박이정.
- 김원경. 1999. “처격 조사의 자질 연산.” 「국어의 격과 조사」. 도서출판 월인.
- 김원경. 2000. 「한국어 격 정보와 자질 연산 문법」. 고려대 박사학위

논문.

김유정. 1996. “기계번역에서의 시제처리.” 「한국어학」 4.

김의수. 2003. 「국어의 격과 의미역 연구: 명사구의 문법기능 획득론」.
고려대 박사학위논문.

김중복 외 . 2004. 「한국어 정보화와 구문분석」. 도서출판 월인.

김지은. 1998. 「우리말 양태용언 구문 연구」. 한국문화사.

김진해. 2000. 「국어 연어 연구」. 경희대 박사학위논문.

김홍규·강범모. 1996. “고려대학교 말모듬 1 : 설계 및 구성.” 「한국어학」 3.

김홍규·강범모. 2000. 「한국어 형태소 및 어휘 사용 빈도의 분석」 (컴퓨터와 인문학 총서). 고려대 민족문화연구원.

나은미. 2004. 「연결주의 관점에서 본 단어 형성 연구」. 한성대 박사학위논문.

남경완. 2005. 「국어 용언의 의미 분석 연구」. 고려대 박사학위논문.

남기심. 1986. “ ‘서술절’ 의 설정은 타당한가?” 「국어학신연구」. 탑출판사. pp.191-198.

남기심. 1993. 「국어 조사의 용법」. 서광학술자료사.

남윤진. 2000. 「현대국어 조사에 대한 계량언어학적 연구」. 태학사.

도원영. 2002. “교차성 용언의 다의성과 사전 처리.” 「한국어학」 15.
25-76. 한국어학회.

류구상. 1995. “국어 격조사에 대하여.” 「한국어학」 2. 한국어학회.

문영호. 1994. 「응용언어학」. 한국문화사.

문유진. 1996. 「의미론적 어휘개념에 기반한 한국어 명사 Word Net의 설계와 구축」. 서울대 컴퓨터공학과 박사학위논문.

민현식. 1999. 「국어 문법 연구」. 도서출판 역락.

박병수·윤혜석·홍기선. 1999. 「문법이론」. 한국문화사.

- 박선옥. 2005. 「국어 보조동사의 통사와 의미 연구」. 도서출판 역락.
- 박성배 · 김영택. 1997. “Semantic Role Determination in Korean Relative Clauses using Idiomatic Patterns” In Proceedings of the 17th International Conference on Computer Processing of Oriental Languages. 1-6.
- 박성배 · 김영택. 2000. “한영 기계번역에서 결정 트리 학습에 의한 한국어 부사격 조사의 의미 중의성 해소” 「한국정보과학회지」 제27권 제 6호 한국정보과학회. pp.668-677.
- 박영순. 1999. “조사의 인지 의미론적 고찰 : ‘는’ 을 중심으로.” 「국어의 격과 조사」. 도서출판사 월인.
- 박영준 · 최경봉. 1996. 「관용어사전」. 태학사.
- 배광희. 1991. “기계번역에서의 의미 모형화 이론에 대하여.” 「조선어 문」 2.
- 서상규 · 한영균. 1999. 「국어정보학 입문」. 태학사.
- 서상규 편. 1999. 「언어 정보의 탐구 1」. 연세대 언어정보개발연구원.
- 서상규. 1998. “말뭉치 분석에 기반을 둔 낱말 빈도의 조사와 그 응용 ‘연세 말뭉치’ 를 중심으로.” 「한글」 24. 한글학회.
- 서영훈. 1991. 「의미 정보를 이용하는 중심어 주도의 한국어 파싱」. 서울대 컴퓨터공학과 박사학위논문.
- 서정수. 1994. 「국어문법」. 뿌리깊은 나무.
- 성광수. 1974. “국어 격문법 시론 I.” 「인문논집」 19. 고려대학교.
- 성광수. 1993. “어휘부의 형태/통사론적 접근.” 「어문논집」 32. 안암어문학회.
- 성광수. 1999. 「격표현과 조사의 의미」. 도서출판 월인.
- 손석락. 2004. 「국어의 주제격·주격 부여자 연구」. 경북대 박사학위논문.
- 송복승. 1995. 「국어의 논항구조 연구」. 보고서.

- 시정곤 · 고광주 · 유혜원 · 김미령. 2000. 「논항 구조란 무엇인가」. 도서출판 월인.
- 시정곤 · 김원경 · 고창수. 2000. “영-한 기계번역 성능평가 방안 연구.” 「언어와 정보」 4. 한국언어정보학회
- 시정곤. 2001. “{를}의 정체는 무엇인가?” 「한국어의 목적어」. 도서출판 월인.
- 신한승. 1984. “주격표지(主格標識)의 중출현상에 대하여.” 「한성어문학」 3.
- 양단희 · 송만석. 1998. “기계학습에 의한 단어의 격 원형성 자동 획득” 「한국정보과학회지」 제25권 제7호 한국정보과학회. pp. 1116-1127.
- 양동휘. 1994. 「문법론」. 한국문화사.
- 양동휘. 1998. “최소주의의 이해와 전망 : 한국어에의 적용” 「한국어학」 7 한국어학회. pp.5-35.
- 양정석. 1995. 「국어동사의 의미 분석과 연결 이론」. 도서출판 박이정.
- 엄정호. 1993. “‘이다’의 범주규정.” 「국어국문학」 110. 국어국문학회.
- 오길록 · 최기선 · 박세영. 1995. 「한글공학」. 대영사.
- 옥철영. 1993. 「한영 기계번역을 위한 구단위 변환 사전」. 서울대 컴퓨터공학과 박사학위논문.
- 왕문용 · 민현식. 1993. 「국어 문법론의 이해」. 개문사.
- 유형선. 1999. “이중주격 구문의 논항구조에 대한 연구.” 「국어의 격과 조사」. 도서출판 월인.
- 유형선. 2003. “한국어 주어 연구사.” 「순천향 인문학논총」 12.
- 유혜원. 2002. 「국어의 격 교체 구문의 연구」. 고려대 박사학위논문.
- 유혜원. 2004. 「한국어정보처리의 이론과 실제」. 제이앤씨.
- 윤덕호. 1993. 「숙어 정보를 활용한 한국어 파싱」. 서울대 컴퓨터공학과

박사학위논문.

이동혁. 2000. “한국어 부사의 자질 연산.” 「한국어학회 여름연구회 발표집」. 한국어학회.

이동혁. 2004. 「국어 연어관계 연구」. 고려대 박사학위논문.

이석주. 2004. “현대 국어의 유추적 변화.” 「이중언어학」 26. 이중언어학회

이석주·이주행. 1994. 「국어학개론」. 대한교과서주식회사.

이선희. 2004. 「국어의 조사와 의미역: 조사{-를}과 논항 실현을 중심으로」. 한국문화사.

이수현 외. 1994. 「중간언어 기반의 한-영 기계번역 기술 개발에 관한 연구」. 과학기술처.

이은희. 1993. 「접속관계의 텍스트 언어학적 연구」. 서울대 박사학위논문.

이은희. 1997. “언어 이해과정에서의 텍스트 연구.” 「한성어문학」 16. 한성대 한국어문학부.

이익섭·임홍빈. 1983. 「국어문법론」. 학연사.

이주행. 1992. 「현대국어문법론」. 대한교과서주식회사.

이초식. 1993. 「인공 지능의 철학」. 고려대 출판부.

이희자·이종희. 1998. 「텍스트분석적 국어 조사의 연구」. 한국문화사.

임동훈. 1991. “격조사는 핵인가.” 「주시경학보」 8.

정보사회학회 편. 1998. 「정보사회의 이해」. 나남.

정유진. 1995. 「국어의 보어 연구」. 고려대 석사학위논문.

정주리. 2004. 「동사, 구문, 그리고 의미」. 국학자료원.

조경임. 1993. 「국어 주격표지의 기능 연구」. 고려대 석사학위논문.

조정미·김길창. 1996. “한국어 의미 해석시 중의성 해소에 대한 연구” 「정보과학회지」 제14권 제7호 정보과학회. pp. 71-83.

- 최경봉. 1996. 「국어 명사의 의미 구조 연구」. 고려대 박사학위논문.
- 최수영. 1984. “주제화와 주격조사-조사 ‘는’ 과 ‘가’ 를 중심으로.” 「어학연구」 20-3.
- 최운호. 2005. 「한국어 처리에서 ‘구뭉음’ 을 위한 명사의 특성 연구」. 서울대 박사학위논문.
- 최운호 · 박병선. 2005. “구문분석 말뭉치 활용 방법 연구 구축” . 제36차 한국어학회 학술대회발표집.
- 최호철 · 홍중선 · 조일영 · 송향근 · 고창수. 1998. “기계 번역을 위한 한국어 논항 체계 연구.” 「한국어 의미학」 3.
- 최호철. 1993. 「현대 국어 서술어의 의미」. 고려대 박사학위논문.
- 필옥덕. 2004. 「현대 한국어 동사의미결합관계 연구」. 도서출판 역락.
- 한국어학회. 1999. 「국어의 격과 조사」. 도서출판 월인.
- 한송화. 2000. 「현대 국어 자동사 연구」. 한국문화사.
- 한정환. 2001. “의미역 계층 이론과 국어의 주격, 대격.” 「한국어학」 13. 한국어학회. pp. 303-321.
- 홍재성 외 저. 1997. 「현대 한국어 동사 구문 사전」. 두산동아.
- 홍중선 · 고광주 .1999. “ ‘-을’ 논항의 의미역 체계 연구.” 「한글학회」 243.
- 홍중선 · 고창수 · 시정곤. 1993. 「장벽이후의 생성문법」. 집문당.
- 황화상. 1998. “자연 언어 처리를 위한 형태소 분석 방법론.” 「어문논집」 37. 안암어문학회.
- Baker, M. 1988. *Incorporation : a Theory of Grammatical Function Changing*. University of Chicago Press.
- Blake, B. J. 1994. *Case*. Cambridge University Press , 고석주 역.
1998. 「격」. 한신문화사.
- Chafe, W.L. 1970. *Meaning and structure of language*. Chicago:

University Press.

Chomsky, N. 1970. "Remarks on Nominalization." *Readings in English Transformational Grammar*. Ginn · Waltham · Mass.

Chomsky, N. 1987. "Transformational Grammar : Past-Present-Future." Talk delivered at Kyoto.

Chomsky, N. 1986. *Knowledge of Language: Its Nature, Origins, and Use*. Praeger Publishers, 이선우 역. 1990. 「언어에 대한 지식」. 민음사.

Chomsky, N. 1995. *The Minimalist Program*. MIT.

Cruse, D, A. 1986. *Lexical Semantics*. Cambridge University Press.

Fellbaum, C. 1998. *WordNet*. The MIT Press.

Fillmore, C. J. 1968. *The case of case*. In E. Bach & R. T. Harms (eds.), *Universals in linguistic theory*, 1-88. London: Holt, Rinehart & Winston

Fillmore, C. J. 1971. "Some Problems for Case Grammar." *Working Papers in Linguistics* 10.

Fukui, N. & Speas, M. 1986. "Specifiers and Projection." *MIT Working Papers in Linguistics* 8. MIT, Cambridge, Mass.

Grimshaw, J. 1990. *Argument Structure*. Cambridge. MA : MIT Press.

Halliday, M. A. K. 1985. *An introduction to functional grammar*. London: Edward Arnold

Jackendoff, R. 1972. *Semantic interpretation in generative grammar*. Cambridge : MIT Press.

Jackendoff, R. 1987. "The Status of Thematic Relations in

- Linguistic Theory.” *Linguistic Inquiry* 17.
- Jackendoff, R. 1990. *Semantic Structures*. Cambridge, MA : MIT Press.
- Jakendoff, R. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press
- Jespersen, O. 1968. *philosophy of grammar*. New York : Norton,
- 이환묵·이석무 역. 2002. 「문법철학」. 한신문화사.
- Longacre, R. E. 1976. *An anatomy of speech notions*. Lisse : Peter de Ridder Press
- Nida, E. 1975. *Componential Analysis of Meaning*. Mouton Publishers.
- Radford, A. 1997. *Syntax -A minimalist introduction*. Cambridge University Press.
- Somers, H. L. 1987. *Valency and Case in Computational Linguistics*. Edinburgh University Press.

[부록]

1. “글잡이 II” 를 이용한 세종21 균형 말뭉치 통계 최빈도 동사

순위	형태(품사)	빈도수	상대비율(%)	전체비율(%)
1.	하/동사	102413	0.45942813	7.24321651
2.	되/동사	50701	0.22744638	3.58585668
3.	보/동사	28455	0.12765008	2.01249576
4.	말하/동사	22880	0.10264044	1.61820078
5.	가/동사	19353	0.08681820	1.36875176
6.	들/동사	18689	0.08383948	1.32178998
7.	받/동사	18433	0.08269105	1.30368423
8.	위하/동사	18356	0.08234563	1.29823840
9.	알/동사	16168	0.07253019	1.14349079
10.	보이/동사	15985	0.07170925	1.13054812
11.	따르/동사	15299	0.06863183	1.08203030
12.	나오/동사	14778	0.06629460	1.04518235
13.	쓰/동사	14708	0.06598058	1.04023159
14.	모르/동사	13171	0.05908554	0.93152636
15.	대하/동사	12913	0.05792815	0.91327918
16.	만들/동사	12811	0.05747057	0.90606517
17.	살/동사	12577	0.05642084	0.88951534
18.	오/동사	11800	0.05293519	0.83456159
19.	나/동사	10473	0.04698223	0.74070877
20.	먹/동사	9634	0.04321845	0.68137002
21.	통하/동사	9450	0.04239301	0.66835654
22.	가지/동사	9434	0.04232124	0.66722494

부 록

23. 의하/동사	9069	0.04068384	0.64141011
24. 내/동사	8567	0.03843185	0.60590589
25. 있/동사	8130	0.03647145	0.57499880
26. 갖/동사	7960	0.03570883	0.56297547
27. 주/동사	7730	0.03467704	0.54670858
28. 이루/동사	7469	0.03350618	0.52824920
29. 찾/동사	7071	0.03172074	0.50010043
30. 느끼/동사	6948	0.03116896	0.49140120
31. 그리/동사	6944	0.03115102	0.49111828
32. 밝히/동사	6872	0.03082802	0.48602605
33. 들어가/동사	6839	0.03067998	0.48369211
34. 묻/동사	6796	0.03048708	0.48065090
35. 듣/동사	6785	0.03043774	0.47987291
36. 죽/동사	6398	0.02870164	0.45250213
37. 나타나/동사	6382	0.02862987	0.45137051
38. 앓/동사	6254	0.02805565	0.44231763
39. 잡/동사	5690	0.02552553	0.40242842
40. 서/동사	5643	0.02531469	0.39910433
41. 맞/동사	5596	0.02510384	0.39578024
42. 부르/동사	5592	0.02508590	0.39549732
43. 일어나/동사	5536	0.02483468	0.39153668
44. 만나/동사	5343	0.02396888	0.37788665
45. 이르/동사	5032	0.02257372	0.35589102
46. 내리/동사	5020	0.02251989	0.35504231
47. 나서/동사	5012	0.02248400	0.35447651
48. 나가/동사	5000	0.02243017	0.35362780
49. 얻/동사	4736	0.02124585	0.33495626
50. 관하/동사	4727	0.02120548	0.33431971

부 록

51. 보내/동사	4644	0.02083314	0.32844949
52. 들어오/동사	4543	0.02038005	0.32130623
53. 생기/동사	4526	0.02030379	0.32010388
54. 치/동사	4506	0.02021407	0.31868938
55. 두/동사	4480	0.02009743	0.31685051
56. 읽/동사	4443	0.01993145	0.31423366
57. 떨어지/동사	4441	0.01992247	0.31409222
58. 남/동사	4359	0.01955462	0.30829272
59. 짓/동사	4199	0.01883685	0.29697663
60. 잇/동사	4164	0.01867984	0.29450122
61. 알리/동사	4129	0.01852283	0.29202583
62. 열/동사	4044	0.01814152	0.28601417
63. 타/동사	4018	0.01802488	0.28417531
64. 떠나/동사	3990	0.01789927	0.28219497
65. 사/동사	3986	0.01788133	0.28191209
66. 웃/동사	3924	0.01760319	0.27752709
67. 믿/동사	3923	0.01759871	0.27745637
68. 넣/동사	3921	0.01758974	0.27731493
69. 다니/동사	3845	0.01724880	0.27193978
70. 지키/동사	3555	0.01594785	0.25142935
71. 빠지/동사	3501	0.01570560	0.24761018
72. 기다리/동사	3482	0.01562037	0.24626639
73. 나누/동사	3445	0.01545438	0.24364956
74. 돌아오/동사	3368	0.01510896	0.23820369
75. 념/동사	3326	0.01492055	0.23523322
76. 놓/동사	3262	0.01463344	0.23070678
77. 입/동사	3259	0.01461998	0.23049460
78. 바라보/동사	3240	0.01453475	0.22915082

부 록

79. 끝나/동사	3236	0.01451680	0.22886790
80. 지나/동사	3231	0.01449437	0.22851428
81. 걷/동사	3213	0.01441363	0.22724122
82. 열리/동사	3175	0.01424316	0.22455364
83. 지니/동사	3087	0.01384838	0.21832980
84. 향하/동사	3066	0.01375418	0.21684456
85. 배우/동사	3022	0.01355679	0.21373264
86. 들리/동사	2991	0.01341773	0.21154015
87. 못하/동사	2985	0.01339081	0.21111579
88. 마시/동사	2955	0.01325623	0.20899403
89. 돌리/동사	2889	0.01296015	0.20432614
90. 오르/동사	2876	0.01290183	0.20340671
91. 그리/동사	2863	0.01284351	0.20248728
92. 비하/동사	2799	0.01255641	0.19796084
93. 안되/동사	2786	0.01249809	0.19704141
94. 걸리/동사	2775	0.01244874	0.19626343
95. 당하/동사	2688	0.01205846	0.19011031
96. 맡/동사	2685	0.01204500	0.18989813
97. 세우/동사	2656	0.01191490	0.18784709
98. 바꾸/동사	2625	0.01177584	0.18565460
99. 돌아가/동사	2588	0.01160985	0.18303774
100. 늘/동사	2577	0.01156051	0.18225977

2. 김홍규 · 강범모(2000)의 ‘한국어 형태소 및 어휘 사용 빈도의 분석의 빈도’ 데이터

품사	타입 (TYPE)	토큰 (TOKEN)
VV(동사)	3312	234303
계	3312	234303

순위	형태	품사	빈도	백분율	누적백분율
1	하	VV	18464	7.880	7.880
2	있	VV	18283	7.803	15.684
3	되	VV	7700	3.286	18.970
4	보	VV	5393	2.302	21.272
5	대하	VV	5363	2.289	23.561
6	위하	VV	3695	1.577	25.138
7	가	VV	3071	1.311	26.448
8	받	VV	3067	1.309	27.757
9	보이	VV	2748	1.173	28.930
10	알	VV	2715	1.159	30.089
11	들	VV	2554	1.090	31.179
12	따르	VV	2541	1.084	32.263
13	살	VV	2364	1.009	33.272
14	오	VV	2235	0.954	34.226
15	쓰	VV	2073	0.885	35.111
16	모르	VV	2000	0.854	35.965
17	나오	VV	1973	0.842	36.807
18	통하	VV	1815	0.775	37.581
19	나	VV	1778	0.759	38.340
20	가지	VV	1755	0.749	39.089
21	만들	VV	1731	0.739	39.828
22	의하	VV	1671	0.713	40.541
23	먹	VV	1486	0.634	41.175
24	찾	VV	1285	0.548	41.724
25	주	VV	1266	0.540	42.264

부 록

26	듣	VV	1185	0.506	42.770
27	밝히	VV	1184	0.505	43.275
28	나타나	VV	1158	0.494	43.769
29	갓	VV	1078	0.460	44.229
30	느끼	VV	1035	0.442	44.671
31	부르	VV	991	0.423	45.094
32	짓	VV	979	0.418	45.512
33	내	VV	959	0.409	45.921
34	내리	VV	954	0.407	46.328
35	묻	VV	935	0.399	46.728
36	이루	VV	927	0.396	47.123
37	죽	VV	881	0.376	47.499
38	들어가	VV	876	0.374	47.873
39	만나	VV	840	0.359	48.232
40	서	VV	834	0.356	48.588
41	두	VV	819	0.350	48.937
42	잡	VV	819	0.350	49.287
43	이르	VV	813	0.347	49.634
44	앉	VV	797	0.340	49.974
45	타	VV	786	0.335	50.309
46	읽	VV	777	0.332	50.641
47	남	VV	766	0.327	50.968
48	지나	VV	755	0.322	51.290
49	보내	VV	743	0.317	51.607
50	관하	VV	730	0.312	51.919
51	일어나	VV	730	0.312	52.230
52	맞	VV	701	0.299	52.529
53	떨어지	VV	684	0.292	52.821
54	다니	VV	680	0.290	53.112
55	치	VV	648	0.277	53.388
56	나가	VV	643	0.274	53.663
57	얻	VV	643	0.274	53.937
58	나누	VV	596	0.254	54.191
59	떠나	VV	590	0.252	54.443

부 록

60	사	VV	585	0.250	54.693
61	기다리	VV	583	0.249	54.942
62	생기	VV	580	0.248	55.189
63	열	VV	580	0.248	55.437
64	알리	VV	577	0.246	55.683
65	넣	VV	544	0.232	55.915
66	돌아오	VV	533	0.227	56.143
67	흐르	VV	530	0.226	56.369
68	민	VV	522	0.223	56.592
69	들어오	VV	519	0.222	56.813
70	오르	VV	514	0.219	57.033
71	빠지	VV	510	0.218	57.250
72	놓	VV	510	0.218	57.468
73	세우	VV	509	0.217	57.685
74	지니	VV	507	0.216	57.902
75	넘	VV	501	0.214	58.115
76	입	VV	497	0.212	58.327
77	열리	VV	483	0.206	58.534
78	바꾸	VV	477	0.204	58.737
79	낱	VV	476	0.203	58.940
80	그리	VV	474	0.202	59.143
81	지키	VV	471	0.201	59.344
82	끝나	VV	469	0.200	59.544
83	이루어지	VV	468	0.200	59.744
84	돌	VV	461	0.197	59.940
85	비하	VV	453	0.193	60.134
86	향하	VV	451	0.192	60.326
87	돌리	VV	450	0.192	60.518
88	나서	VV	448	0.191	60.709
89	움직이	VV	442	0.189	60.898
90	올리	VV	427	0.182	61.080
91	벗어나	VV	424	0.181	61.261
92	당하	VV	422	0.180	61.441
93	바라보	VV	422	0.180	61.621

부 록

94	웃	VV	410	0.175	61.796
95	걸리	VV	404	0.172	61.969
96	달리	VV	400	0.171	62.140
97	드리나	VV	400	0.171	62.310
98	끌	VV	397	0.169	62.480
99	들리	VV	395	0.169	62.648
100	삼	VV	395	0.169	62.817

3. 연산결과 의미역 데이터베이스

의미역	구조격	논항자질	술어자질	격형태	서술어	문 장
Agent	subject	animate	action		들어오다	지성이 들어왔다.
Agent	subject	animate	action		빠지다	여성이 물에 빠지다.
Agent	subject	group	action		넘어지다	국가가 넘어지다.
Agent	subject	group	action		치다	일본이 자기 안방 차지하듯 곳곳에다 진을 치고 있기 때문이다.
Agent	subject	animate	action		치다	후배들이 박수를 치다.
Agent	subject	animate	action		치다	군인 두 사람이 텐트를 치다.
Agent	subject	animate	action		치다	저들은 자기네들이 성공하였다고 큰 소리를 쳤다.
Agent	subject	animate	action		보내다	경영위원회 위원장이나 사장이 직접 답신을 보내기도 한다.
Agent	subject	animate	action		보내다	그러나 삼희기획처럼 사원들이 취미활동만으로 [하루를 보내는] 것은 아니다.
Agent	subject	group	action		걷다	언론이 현재나 미래에 정권과의 관계에서 바람직한 길을 [걷고 있다.]
Agent	subject	animate	action		마시다	남편이 친구를 집에 데려와 늦게까지 술을 마시다.
Agent	subject	animate	action		나누다	두 사람이 대화내용을 나누었다.
Agent	subject	animate	action		들다	제주도 조랑말은 유량하는 말 광대에 팔려 나가 이빨을 드러내어 웃고 발을 [들어 보인다.]
Agent	subject	animate	action		들다	증거로서 그는 당시 새로 나온 케플러의 시력에 관한 이론을 들었다.
Agent	subject	animate	action		들다	포르투갈 사람들이 서양 문명을 [들고 갔다.]
Agent	subject	animate	action		들다	남들은 잡혀가는 판인데 당신은 구속되긴커녕 보너스 봉투를 들었다.

Agent	subject	group	action		두다	시민정치는 조직보다는 토론의 내용에 더 의미를 둔다.
Agent	subject	animate	action		두다	TV프로듀서들이 시청자들의 조급성을 염두에 두는 건 당연하고 필요한 일임에 틀림없다.
Agent	subject	animate	action		들다	혜주는 별 거부감없이 메모지를 들었다.
Agent	subject	animate	action		가다	전두환 씨가 백담사로 가다.
Agent	subject	animate	action		말하다	독자들이 흔히 “그게 그거” 라고 말한다.
Agent	subject	animate	action		말하다	학생이 웃는 학생들을 향하여 크게 말했다.
Agent	subject	animate	action		말하다	청년이 처녀에게 말했다.
Agent	subject		action		돌아가다	탄소는 식물이 불타거나 썩을 때, 그리고 동물이 호흡할 때, 이산화탄소의 형태로 대기권으로 돌아온다.
Agent	subject	animate	action		오다	여성이 남성과 결혼하면 생부모(生父母)를 떠나 남성의 집으로 온다.
Agent	subject	animate	action		하다	판매부서장이 [입사를 하다.]
Agent	subject	animate	action		향하다	경수는 어머니가 작성해 준 혼수감 주문서를 들고 약속 장소로 향했다.
Agent	subject	animate	action		말하다	던져 교장선생님을 난감케 했던 학생이 웃는 학생들을 향하여 크게 말했다.
Agent	subject	animate	action		짓다	남편들이 하나같이 실없는 [웃음을 짓는다.]
Agent	subject	group	stative		짓다	암수가 [짜를 짓다.]? 관용어처리
Agent	subject	animate	action		짓다	이 벌레가 집을 짓는다
Agent	subject	group	stative		지키다	여타의 언론들은 대개가 [침묵을 지켰다.]
Agent	subject	animate	action		돌아가다	제비가 돌아간다.
Agent	subject	animate	action		들어오다	팔애가 약을 사가지고 들어왔다.
Agent	subject	animate	action		열다	양지선씨가 서울전을 연다.

Agent	subject	group	action		건다	여러 나라들이 탈사회주의의 길을 [견고 있다.]
Agent	subject	group	action		넣다	공진청이 KSC-5842에 잘못된 규정을 넣었다.
Agent	subject	animate	action		지키다	군인이 총장을 세우고 성문을 [지키고 있습니다.]
Agent	subject	group	action		바라보다	신문이 우리 사회를 바라보다.
Agent	subject	animate	action		기다리다	거미는 뱀벌이 움직이기를 기다린다.
Agent	subject	animate	action		기다리다	한 무리의 젊은 남녀가 모여 무엇인가를 기다린다.
Agent	subject	animate	action		알리다	프로메테우스는 세계가 신 중심에서 인간 중심으로 전환할 것을 알렸다.
Agent	subject	animate	action		사다	사람들이 땅을 산다.
Agent	subject	animate	action		떠났다	가운데 65명만을 신규 보임해 13명이 자리를 떠났다.
Agent	subject	animate	action		기다리다	네 모녀는 한참 동안 도둑이 마루문을 따는 인기척을 기다렸다.
Agent	subject	animate	action		타다	젊은 여자가 품내면서 차를 탄다.
Agent	subject	animate	action		열다	아내의 친구가 문을 연다.
Agent	subject	animate	action		다니다	일본인들이 가까운 우리 나라는 물론 동남아국으로 이른바 섹스(Sex)여행을 다녔다.
Agent	subject	animate	action		웃다	그녀는 역마살이 끼었다며 웃는다.
Agent	subject	animate	action		넣다	천자가 달맞이꽃 화분을 희빈의 방에 넣어 둔다.
Agent	subject	animate	stative		지키다	호위병으로서의 수필은 여왕벌을 지킨다.
Agent	subject	animate	action		나누다	우리는 보통 피로를 육체적인 것과 정신적인 것으로 나눈다.
Agent	subject	animate	action		웃다	새댁이 나를 보고 웃었다.
Agent	subject	group	passive		빠지다	언론은 가격경쟁을 할 수 없고 망하거나 한계적인 상태에 [빠지게 된다].

Agent	subject	animate	action		나누다	조씨는 이들과 인사를 나누다.
Agent	subject	animate	action		다니다	어릴 때 남자 아이가 고추를 내놓고 다닌다.
Agent	subject	animate	action		사다	과거 사람들은 자신이 원하는 물건을 산다.
Agent	subject	animate	action		사다	임신부가 왼쪽으로만 누워 자고 타인이 낀 아들 태몽을 사기도 하였다.
Agent	subject	group	action		받다	언론이 사회의 다양성을 반영하고 공정하고 객관적인 언론이 아니라는 비판을 받는다.
Beneficiary	subject		passive		받다	피해는 개발도상국이 받는다.
Theme	¹ subject		stative		되다	¹ 발행부수는 판매뿐만이 아니라, 편집·제작이나 광고에 있어서도 업무적인 중요한 ² 기준이 [되기도 한다.]
Complement	² subject					
Theme	subject		stative		지니다	주장이 설득력을 [지니게 된다.]
Theme	¹ subject		stative		넘다	¹ 빛이 6 조원이 넘는다.
Complement	² subject					빛이 ² 6 조원이 넘는다.
Experiencer	subject	animate	cognitive		하다	광래는 언제든 [각오해야 한다.]
Experiencer	subject	animate	cognitive		돌아가다	우리들은 다시 명랑한 [기분으로 돌아왔다.]
Experiencer	subject	group	cognitive		빠지다	자유진영은 1백 년 동안의 경쟁에서 최종적으로 승리했다며 [홍분에 빠졌다.]
Experiencer	subject	group	cognitive		빠지다	언론은 가격경쟁을 할 수 없고 망하거나 한계적인 [상태에 빠지게 된다.]
Experiencer	subject	animate	cognitive		믿다	장철중씨는 일편단심 철썩같이 믿었다.
Experiencer	subject	animate	cognitive		믿다	여성이 이 내부의 벽을 보호벽으로 믿고 있다.
Experiencer	subject	animate	cognitive		생기다	현주는 [욕심이 생겼다.]
Experiencer	subject	animate	cognitive		빠지다	여성이 [자학에 빠지다.]
Experiencer	subject	animate	cognitive		읽다	어린 아이가 책을 읽다.
Experiencer	subject		cognitive		보다	은행나무는 마주 보다.
Experiencer	subject	animate	cognitive		보다	대만사람들이 한국을 보다.

Experiencer	subject	animate	cognitive		읽다	사람이 책을 읽는다.
Experiencer	subject	animate	cognitive		읽다	경제인들은 68가 신문을 직접 읽는다.
Theme	subject		stative		하다	씨앗이 함께 종을 잇는 [기능을 한다.]
Theme	subject		stative		넘어지다	큰 나무가 넘어지다.
Theme	subject		stative		지니다	사유재산 자체가 공동목적에 바탕을 둔 사회적 성격을 [지니고 있다.]
Theme	subject		stative		있다	대기업의 부도사태는 금융기관의 부실화와 이들 기관에 대한 신뢰도 하락으로 이어졌다.
Theme	subject		stative		지니다	주장이 설득력을 [지니게 된다.]
Theme	subject		stative		타다	경기가 장기적인 [과도를 탄다.]
Theme	subject		stative		있다	전기가스업이 8 등으로 그 뒤를 이었다.
Theme	subject		stative		지나다	한보철강의 부도가 어느덧 두 달이 지났다.
Theme	subject		stative		기여하다	이들의 활동은 우리 언론의 편집권 신장이라는 차원에서 커다란 [기여를 하였다.]
Theme	subject		stative		있다	경쟁이 반드시 상품의 질 향상으로 이어지는 것은 아니라는 점 때문이다.
Theme	subject		stative		되다	발행부수는 판매뿐만이 아니라, 편집·제작이나 광고에 있어서도 업무적인 중요한 기준이 되기도 한다.
Theme	subject		stative		되다	발행부수가 일단 결정이 되었다.
Theme	subject		stative		되다	구독료가 결정이 되었다.
Theme	subject		stative		남다	이념에는 편향성이 강하게 남아 있다.

Theme	subject		stative		돌아가다	소변량이 정상으로 돌아온다.
Theme	subject		stative		나타나다	행정부패는 무사안일, 뇌물수수, 부정축재 등이 주요한 유형으로 나타나고 있으며,
Theme	subject		passive		보이다	뇌물수수의 액수가 커져가는 경향을 보이고 있다.
Theme	subject		stative		지니다	문화는 그 사회구성원이 그 안에서 제안되고 시험되는 새로운 관찰 및 의미라는 두 가지 측면을 지닌다.
Theme	subject		passive		떨어지다	감기가 떨어지다.
Theme	subject		passive		떨어지다	[방송의 신뢰도는] 땅에 떨어지고 말았다.
Theme	subject		passive		떨어지다	지지가 형편없이 떨어졌다.
Theme	subject		stative		가다	자본주의가 [혼란으로 갔다.] (연어처리/의인화)
Theme	subject		stative		가다	이미지는 어디 가다. (연어처리)
Theme	subject		stative		남다	아쉬움이 남는다.
Theme	subject		copular		이다	정책목표는 당연히 이윤극대화에 두어지게 [마련이다.]
Theme	subject		stative		남다	그러면 마지막으로 다음과 같은 질문이 남는다.
Theme	subject		stative		수반되다	[민영화와 자율화는] 정부의 실권적인 의지와 은행의 인사 및 경영면에서의 체질개선이 [수반되어야 한다.]
Theme	subject		passive		받다	인간의 성행위는 많은 제약을 받는다.

Theme	subject		stative		생기다	[여성의 얼굴이] 남성의 얼굴에 비하여 일반적으로 [잘 생겼다.]
Theme	subject		stative		생기다	수치심은 도덕관념이나 윤리의식에서 생긴다.
Locative	adjunct	group		에서	알리다	한국에서는 이승복 사건의 대표적인 조작사건으로 알려져 있다.
Locative	adjunct				남다	앞으로의 변화에서는 지역들이 어떻게 적응할 것인가가 문제로 남는다.
Locative	adjunct			에서	생기다	우리나라에서는 남자는 '바지' 여자는 '치마' 라는 [고정관념이 생겼다.]
Locative	adjunct			에서	들어오다	외국에서 투자물결이 쏟아져 들어온다.
Locative	adjunct			에서	생기다	미스터 강의 호텔예약에서 문제가 생겼다.

* 위와 같이 ‘에서’ 가 붙은 명사구는 구조적 표지가 있는 다른 주어 후보가 존재하기 때문에 부사격으로 설정하고 <처소역>이 배당되었다.

주어의 의미역	백분율
Agent	50 %
Theme	34 %
Experiencer	13 %
Complement	2 %
Beneficiary	1 %

ABSTRACT

A Study on the Thematic role of the Subject

kwon, Jung-hyun

Major in the Information Processing of Korean Language

Dept. of Korean Language and Literature

Graduate School of Hansung University

This research is to suggest automatic allotment algorithm of theta-role for the subject of sentences in Korean. First, I examined the concept of the subject in the system of the cases in Korean and discussed about the whole items of the systemic theta-role. So I proposed the operation process of automatic allotment algorithm of theta-role of the subjects based on the these theta-role these. This research can be a suggestion of a mechanical process for data process of Korean by applying a new methodology through Feature Computation Grammar(FCG) and reviewing methods to establish the theta-role system that has been discussed before.

The theta-role means a role of meaning that has in the paragraph of predicate. The allotment rule of this theta-role is a very important task that have to be pass through for mechanical translation or using of conversational system. Mostly researches on the subject are still remained as the phase of cover studies or the subjective collision. Thus, this study that discussed about the theta-role can be a effective

way to balance theta-role study that had placed too much emphasis on arg of {reul(를)} by establishing the relation between the subject and predicate.

Summary of studies that has been discussed until now is :

In chapter 1, I clarified the goal and range of this research and studied preceding researches of the structure case and theta-role with classifying as the field of the study of the Korean language and data processing of Korean.

In chapter 2, I examined researches that has been discussed on the subject in Korean and reviewed position of theta-role research of the present. The subjective case collision phenomenon is the main part of controversy of the subject's theta-role, so I put these discussions together and suggested a point that can be helpful to establish theta-role of the subject.

In chapter 3, I examined the system of theta-role of Korean that has been processed until now, before constructing the theta-role of the subject and pointed the limitations in methodological view. In this chapter, I presented the new system of the theta-role by grouping establishment methods of theta-role on the basis of the FCG theory.

In chapter 4, I discussed about the decision dispositions that establish the system so the theta-role that showed in chapter 3. I inspected meaning dispositions of argument and predicate through meaning classification of nouns and declinable words of old in presenting the method of the theta-role marking to the subject and the algorithm which process as applying to bundles of words in effect. Especially, I suggested the way to make it more useful that will be applied to many fields as pointing using methods that processing the Sejong corpus(distribution balancing bundles of words and limitations of

bundles of words).

In this study, automatic allotment algorithm of the subject's theta-role is not embodied perfectly. The system is changing continuously and has defects constantly, like all other languages. In this research, I extracted three hundred thousand sentences that classified into a hundred verbs, and draw out a thousand sentences again so I got a hundred sentences of data base that can be applied to algorithm among objects of ten million clauses. But, from now on we need more applications of real example sentences because this sampling process is for just enhancing statistical probabilities.

Constructing rules and establishing systems are to have 'the power of prediction' which in the explanatory adequacy. But, it is worth that to concern about directions of its changes because language is like the kaleidoscope of organic system. I presume the subject's automatic allotment algorithm of the theta-role can make more advance for applying to conversational system, the ultimate goal of data processing of Korean.

Abstract

Thematic Role	Structural Case	Argument Feature	Case Mark	Predicate Feature
Experiencer	sub	animate	-ka	cognitive
Agent	sub	animate or group	-ka	action
	adjunct	group	-eseo	action
Beneficiary	sub	animate or group	-ka	passive benefit
Instrument	sub		-ka	causer
Theme	sub		-ka	stative
				passive
				copula
*Complement (wastebasket)	sub		-ka	

Automatic Allotment Algorithm of Theta-role for the Subject of sentences in Korean