

말뭉치로부터 격틀 구축에 필요한 학습 데이터 추출

양 단 희, 송 만 석

연세대학교 컴퓨터과학과

서울시 서대문구 신촌동 134 우: 120-749,

<http://december.yonsei.ac.kr/~dhyang,~mssong>

Extraction of the Training Data for Building Case Frames from a Corpus

Dan-Hee Yang, Mansuk Song

Department of Computer Science, Yonsei University

요약

실용적인 격틀(Case frame) 정보를 말뭉치로부터 자동구축하기 위해서는 대량의 홀문장¹이 필요하다. 그리고 국어 문장 형식은 영어와 많은 차이점이 있다. 또한 기존의 격틀 구축 연구에서 진제했던 광범위한 학습 데이터와 언어학적 지식은 국어에 대해 현재 존재하지 않는다.

그러므로 본 연구는 그러한 문제점들을 밝히고 현실적인 접근 방법을 제시한다. 그리고 겹문장을 홀문장 형태의 문장들로 바꾸기 위한 알고리즘을 제시한다.

1 서론

자연어처리에서 올바른 구문분석이나 의미 분석을 위해 격틀 정보는 매우 중요하다. 격틀이란 용어는 Fillmore의 격문법 체계에 근간을 두고 있다. 현재 그 용어는 HPSG 류의 문법에서 어휘사전 항목(lexicon entry)의 구문 정보 부분만을 칭하기도 하고 의미 정보 부분까지를 총칭하기도 한다. 본 연구에서는 국어 격틀은 서술어가 취하는 논항 구조(argument structure), 의미역 구조(thematic structure), 격 구조(Case structure), 그리고 선택제약(selectional restriction) 정보의 모

임으로 정의한다. 논항은 서술어가 의미적으로 필요로 하는 문장 성분을 의미한다. 논항 구조는 서술어에 대한 논항²들의 나열 혹은 논항 수에 대한 정보이다. 의미역 구조는 논항 구조에 따른 각 논항의 의미역에 대한 정보이다.³ 격 구조는 논항 구조에 따른 각 논항의 구문격(syntactic Case) 즉 격 조사의 나열이다. 선택제약 정보는 의미역 구조에 따라 각 논항 자리를 차지할 수 있는 명사상당어구(nominals)의 의미적 자격에 대한 정보를 말한다.

국내에서 주요 용언을 대상으로 수작업에 의해 국어 격틀을 구축하는 작업이 국립국어연구원, 서울대, 연세대 언어정보연구원, SERI 등에서 진행되고 있다[16,17]. 영어권에서는 이미 수작업에 의해 FrameNet (Charles Fillmore et al - UC Berkeley), COMLEX corpus (distributed by the Linguistic Data Consortium - LDC), LDOCE (Longman Dictionary of Contemporary English) 등과 같이 상당 수준의 격틀이 이미 구축되어 있다. 지금은 자연어처리 분야에서 수작업 연구 결과의 문제점을 해소하기 위해 이의 자동 구축에 대한 연구가 주 관심사가 되어 있다[1,3,4,5,6,7].

영어를 대상으로 [13]의 연구가 있으나 격틀 학습을 위한 학습 데이터로서 홀문장을 준비

* 본 연구는 정보통신부 '98 국책연구개발사업 [과제번호: 98-86] 연구비 지원으로 수행된 것임.

¹ 규범 문법에 따른 용어로 홀문장은 기존의 단문, 겹문장은 기존의 복문을 말한다.

² 이 때 논항은 번호에 의해 명명된다. 논항에는 필수적 논항과 수의적 논항이 있다.

³ 논항 구조와 의미역 구조에 대한 정의는 Chomsky의 GB 이론에 따른 것이다.

하는 과정은 논의로 하고 있다. [11]은 문헌으로부터 정보를 자동 추출하기 위해 구문패턴을 자동 습득시키려는 연구이다. 그 연구는 기본 문형의 개념 없이 영문을 대상으로 문장 단위로 일정 패턴을 찾는다. [14]는 수작업으로 품사와 의미 속성 정보가 부가된 교과서 말뭉치를 대상으로 문장 패턴을 자동 추출하는 연구를 수행하였다. 연구 [4]는 구문 분석된 문장을 학습 데이터로 사용하고 있다.

위와 같이 국내외적으로 격틀 정보를 자동 구축하기 위한 연구들은 주로 수작업 혹은 구문 분석기의 결과로 생성된 구문 분석 정보가 부가된 말뭉치를 학습 데이터로 사용하고 있다. 기계 학습에 의한 소량의 실험적 연구에는 학습 데이터를 수작업에 의해 준비할 수 있으나 한 언어 전체를 대상으로 한 연구에는 한계가 있다.⁴

그리고 구문 분석기를 사용하는 방식은 정확도 높은 구문분석을 하기 위한 정보를 구축하기 위해 기존의 정확도 낮은 구문분석기를 그대로 사용하는 것이기 때문에 교착상태(deadlock) 현상을 가져온다. 특히 국어에 있어서 현 구문 분석기들은 별로 신뢰할 만한 수준에 있지 않다.

국어의 전 용언에 대한 격틀 구축에 필요한 학습 데이터는 다음아닌 대량의 홀문장이다. 그러나 일반적인 문장은 홀문장 형태로 존재하기 보다는 거의 겹문장 형태를 취하고 있다. 그러므로 말뭉치로부터 홀문장 형태로 이미 존재하는 것만을 추출하는 경우 현재 구축되어 있지 않은 너무 광대한 말뭉치를 필요로 한다[15].

그러므로 겹문장으로부터 일렬의 홀문장을 분리해내야만 한다. 이와 관련하여 [10]의 연구는 국어의 거의 모든 서술어에 대해 이미 논항 구조와 격 구조에 대한 정보를 갖고 있다고 가정하였다. 그리고 안은 문장으로부터 홀문장을 분리하는 연구를 수행하였는데 본 연구의 궁극적인 목표점을 현존하는 기본 지식으로 상정하였다.

본 연구는 격틀 정보를 구축하기 위해 필요한 학습 데이터를 말뭉치로부터 자동으로 추출하는 알고리즘 개발에 관한 것이다. 이를 위해 NM-KTS 형태소 분석기⁵의 출력 결과를 이용한

다. 그러나 관심의 초점은 격틀 정보 중 통사 정보에 속하는 논항 구조와 격 구조에 있다.

앞에서 언급한 것처럼 완전 구문 분석기를 사용하면 현재 그 분석기가 가지고 있는 내부적 알고리즘과 격틀 정보를 일관되게 반영하는 결과를 얻게 된다. 그러므로 본 연구에서는 부분 구문 분석 알고리즘을 제안한다. 또한 분석 정확도를 높이기 위해 구문 분석 도중 오류를 야기시킬 수 있는 문장은 걸러내어 배제시킨다. 부분 구문 분석기 개발은 한 문장을 올바르게 분석해야 하는 구문 분석기 개발과는 다른 접근 방법을 사용할 수 있다. 왜냐하면 불완전하게 분석된 대량의 홀문장을 대상으로 기계학습 기법을 사용할 수 있기 때문이다.

2 홀문장화의 문제점

영어에 비해 국어가 형태소 분석기 개발에 고려해야 할 점이 훨씬 많은 것처럼 홀문장화하는 작업에도 고려해야 할 점이 많다. 다음은 고려되어야 할 현 국어사건의 정보 수록 상태 및 국어의 언어학적인 주요 특성이다.

- (1) 국어사전에서 각 동사가 자동사인지 타동사인지에 대한 정보는 명확하다. 그러나 보어를 필요로 하는지에 대한 정보는 아예 없다.
- (2) 국어는 서술어가 문장 끝에 오는 SOV 구조이기 때문에 (3)의 생략과 도치 현상과 결합될 경우 서술어 연결 구성이 매우 복잡하다[9].
- (3) 생략과 도치가 매우 자유롭다.
- (4) 관계대명사와 관계부사가 없다.
- (5) 주어의 겹침⁷과 목적어의 겹침 현상이 있다.

(1)과 관련하여 영어의 경우 보어 요구에 대한 정보는 Hornby 영어사전 등에 각 동사에 대한 문형 정보의 일부로 수록되어 있다. 그러나 국어에 대해 그런 광범위한 정보는 아직 구축되어 있지 않다. 다만 현재 수작업으로 제한된 서술어에 대해 작업이 진행되고 있다[8,17]. 또한 그러한 정보가 존재하더라도 서술어에 동철이의

⁴ 이에 대한 자세한 논의는 [7,15] 참조

⁵ 분석 정확률은 96%, 미등록어 추정 확률은 0.75

⁶ 여기서 보어는 주어, 목적어가 아니면서 서술어가 논항으로서 필요로 하는 문장 성분을 말한다.

⁷ 주격 중출 혹은 다중 주어

어가 18.74%가 존재하기 때문에 상당한 중의성(ambiguity)이 발생한다. 그뿐 아니라 필수적 보어인지 수의적 보어인지를 구분하는 것 자체가 어려운 문제이다[12]. 이에 대한 해법으로 보어 요구도를 확률적으로 표기해야 바람직하다.

(2)는 각 서술어의 지배 영역(governing domain)을 파악하기 어렵게 만든다. 가령 '그는 책을 그녀로부터 받아 가로챘다.'에서 선행 동사 '받아'는 '책을'과 '그녀로부터'를 논항으로 취하는 반면 후행 동사 '가로챘다'는 '책을'만을 논항으로 취한다[8]. 영어식으로 표현하면 '그는 책을 그녀로부터 받아 그것을 가로챘다.'가 된다. 즉 영어에서는 목적어 '그것을'을 생략할 수 없고 명시적으로 표현해 주어야 한다. 또한 '진회는 빨래를 빨아서 줄에 널었다.'와 같은 문장도 국어에서는 성립된다.

연구 [2]에서는 한 문장 내에서 대명사가 생략된 경우(pro-drop)와 어순이 뒤섞인 경우(scrambling)에도 서술어에 해당하는 주어를 찾아주는 알고리즘을 제시하였다. 그 알고리즘은 항상 주어는 대응되는 서술어의 왼편에 놓인다고 묵시적으로 전제하고서 다음의 절차를 적용시킨다.

- [1] 왼쪽에서 오른쪽으로 문장을 훑어 나간다.
- [2] 서술어가 발견되면 그에 대응되는 주어는 다음을 만족시키는 주격조사를 취하고 있는 명사이다.
 - [2a] 그 서술어의 왼편에 위치하고 있다.
 - [2b] 그 서술어와 가장 근접해 있다.
 - [2c] 이전에 다른 서술어와 대응된 적이 없다.
- [3] 만약 [2c]를 만족시키지 못하는 경우 그 서술어의 왼편으로 가장 근접해 있는 서술어에 대응되는 것이 된다.

그 연구는 인용절만을 안은 문장 '철회가 순옥이 영수가 그 영화를 보았다고 말했다고 생각한다.'를 예시문으로 사용하여 논리를 전개하였다. 그러나 안은 문장에서 문제가 되는 것은 인용절을 안은 문장에서가 아니라 짧은 관형절을 안은 문장에서 주로 발생한다. 가령 '에인이 떠나 슬퍼하는 그를 보았다.'라는 문장에 위 알

고리즘을 적용하면 '떠나', '슬퍼하는', '보았다'의 주어를 모두 '에인이'로 파악하게 된다.

또한 명사절을 안은 문장 '철수가 영희를 때리는 것을 보았다.'에서 '때리는'의 주어를 '철수가'로 보면 '보았다'의 주어가 생략된 것이며, '보았다'의 주어를 '철수가'로 보면 '때리는'의 주어가 생략된 것이다. 이와 같은 반례에서 보여진 것처럼 그 연구는 첫째, 관형절을 안은 문장에서는 주어가 서술어 오른편에 놓일 수 있다는 점을 고려하지 않았다. 둘째, 주어가 생략될 수 있다는 점을 고려하지 않았다.

(3)에서의 생략 문제는 위에서 이미 언급되었다. 영어인 경우 생략은 문장 내에서 중복되는 부분이 있는 경우만으로 한정되기 때문에 그 생략의 복원이 용이하다. 도치의 경우⁹ '나는 학교로 철수가 가는 것을 보았다.'에서 '학교로'가 서술어 '가는'의 지배 영역에 속하는지 '보았다'의 지배 영역에 속하는지를 의미나 문맥 정보 없이 구문 정보만 가지고 단언할 수는 없다. 영어인 경우 도치는 주어나 목적어가 길어서 도치되는 경우는 'it'를 사용하여 그 도치 여부를 명확히 표시해준다. 그 외 강조를 위해 도치되는 경우도 있으나 일반적인 언어 현상은 아니다.

(4)와 관련하여 영어는 일반적으로 어순에 의해 묵시적 격 표지를 하나 관계대명사와 같이 대명사인 경우는 어형이 변하여 명시적 격 표지를 한다. 관계부사인 경우는 선행사가 시간적인 것인지 장소적인 것인지 등을 알 수 있도록 명시적 표지를 한다. 또한 'that'과 'what'을 사용하는 경우도 바로 다음에 오는 단어가 동사인지 아닌지에 따라 주격, 목적격 여부를 확실히 알 수 있다.

그러나 국어에서는 영어와 반대 현상이 일어난다. 일반적으로 격조사에 의해 명시적 격표지를 하나 관형절을 안은 문장에서는 하위문¹⁰ 복구를 위한 관형절의 피수식어¹¹의 격 조사가 소멸되고 상위문을 위한 격 조사만 남는다. 예를 들어 '그는 그가 공부를 했던 학교로 도망쳤다.'에서 하위문에서의 '학교에서'가 상위문에서 '학

⁸ 주어, 목적어와 같이 문장 성분에 해당하는 어휘를 지칭할 때 격조사를 포함하여 어절 단위로 칭하겠다

⁹ 혹은 일반적인 어순에서 벗어난 경우

¹⁰ 계층적 구문 분석 구조에 따라 안은 문장에서 밖에서 안고 있는 문장을 상위문, 안에 안긴 문장을 하위문이라고 칭하겠다.

교로' 바뀌었다.

그러므로 구문 정보만 가지고 하위문을 위한 격을 파악할 수 없고 의미적 분석을 통해 격을 유추해 내야 한다. 또한 관계부사도 존재하지 않아 보이인지 여부를 결정하기 어렵다. 이는 국어에서 관형절이 관계대명사와 관계 부사의 역할을 함께 수행하고 있음을 의미한다.

(5)의 현상은 관형절을 분리해야 할 경우 문제가 된다. 가령 목적격이 두 개 온 경우 각각 다른 서술어에 속하는 것인지 혹은 목적격 겹침 현상인지 구문 정보만 가지고 파악하기 어렵다.

3 접근 방법

본 연구에서는 문제를 간략화시키고 분석 정확률을 높이기 위해 다음과 같은 기본 원칙을 세우고 이에 따른 세부 원칙을 정한다. 그리고 현재 상태에서 언어적 현상이 더 규명되어야 할 부분은 본 연구에서 제외시킨다.

3.1 기본 원칙

(1) 실험에 사용될 말뭉치는 필수 성분 생략이 거의 발생하지 않는 교과서나 신문 말뭉치 등을 대상으로 한다. 생략이 빈번한 구어체 말뭉치는 제외시킨다.

표 1. 처리의 우선 순위표

우선 순위	문장 유형
1	홀문장
2	이어진 문장
3	명사절, 서술절, 부사절, 인용절, 긴 관형절
4	짧은 관형절

(2) 처리의 우선순위는 부분 구문 분석이 용이한 순서대로 다음 표 1을 따른다. 표의 우선순위에 따라 말뭉치에서 완전 문장만을 먼저 처리하여 필요한 정보를 취한다. 그런 후 마지막으로 우선 순위 4를 대상으로 말뭉치에서 일정빈도 이상 발생하면 그 정보를 신뢰하는 확률적 방법을 취한다.

(3) 국어와 영어를 비교하여 국어 문법 현상을 전산처리에 도움이 될 수 있다고 판단되면 영어 문법의 테두리에 맞춰 해석한다.

(4) 격 겹침 현상과 같이 서술어의 특성이 아닌

화용론적 특성은 배제한다.

(5) 일반 격조사 중 주격, 목적격, 부사격 조사가 붙은 어절만을 대상으로 한다. 보조사와 격조사가 결합된 형태에서는 보조사를 제거하고 격조사만을 취한다.

3.2 세부 원칙

(1) 관형절의 피수식의 특성에 따라 하위문은 완전 문장 성립 여부가 가변적이다. 그러므로 상위문이 필요한 논항을 먼저 취하게 한 후 하위문을 고려한다.

(2) '-를 위해, -에 의해, -에 대해' 등과 같이 '조사+용언'이 영어의 한 전치사에 대응되는 경우 단일 조사로 취급한다.

(3) 보어 요구에 대한 정보는 표 1의 우선 순위 3까지의 처리 결과로부터 구한다. 보어 요구에 대한 정보가 빈약한 서술어에 대한 문장은 처리에서 배제시킨다.

(4) '좋은 점은 그가 공부를 잘 한다는 것이다.'와 '사실은 그가 시계를 잃어버렸다.'와 같은 문장은 서술절과 동일하게 취급한다.

(5) 완전 문장이든 불완전 문장이든 목적어가 도치되는 경우는 서술어가 자동사인지 타동사인지 여부로 복원이 가능하다. 그러나 보어가 도치되면 그 보어가 상위문에 속하는 것인지 하위문에 속하는 것인지 구문적으로 판별할 수 없다. 이 때는 그 사실을 기록해 두고 그 서술어에 대한 다른 문장에서의 정보를 보고 결정한다.

(6) '아름다운 꽃을 좋아한다.'에서 형용사나 자동사가 수식으로 사용된 경우 '아름다운 꽃'이나 '여행하는 사람'을 관형절로 간주하지 않는다. '학자인 소크라테스는 불행했다.'와 같이 서술격조사의 관형형도 마찬가지다. 이러한 단독 관형절은 모두 제거시킨다.

(7) '철수와 영희는 공부한다.', '철수와 영희가 싸웠다.'와 같이 '와/과'에 의한 명사간의 결합 구조에서 앞의 명사와 조사(여기서는 '철수와')는 제거한다.

(8) '나는 그녀의 손을 잡았다.'와 같이 격조사 '의'와 결합된 '그녀의'는 서술어의 논항이 아니기 때문에 제거한다.

3.3 본 연구 범위 밖

(1) 서술어를 제외한 격조사가 붙지 않은 어절

11 영어의 선행사에 해당한다.

(제 10회 한글 및 한국어 정보처리 학술대회)

들은 문장 성분에서 제외된다. 그러므로 '그녀는 귀엽게 생겼다.', '그는 기뻐서 날뛰었다.'와 같이 '귀엽게'와 '기뻐서'는 실질적인 논항이나 격조사가 없는 부사어이기 때문에 본 연구 대상에서 제외된다.

- (2) '언덕을 넘어 가고 있다.', '달라 붙어 썩어 버렸다'와 같이 연이은 서술어를 포함하는 문장은 지배 영역을 파악하기 어렵기 때문에 제외시킨다.

4 안은 문장 분리 알고리즘

국어의 안은 문장 처리에서 문제가 되는 것은 짧은 관형절을 안은 문장에서 발생하기 때문에 알고리즘의 기술에서는 이에 대한 논의로 한정하겠다.

4.1 형태소 분석과 용어 정의

형태소 분석기를 사용하여 문장을 형태소 분석한다. 불완전 동사는 보어를 필요로 하는 동사를 하는데 본 연구에서는 부사격 조사가 붙은 어절을 보어로 본다. 상위문을 S_0 그리고 그 서술어를 P_0 , 하위문을 S_1 그리고 그 서술어를 P_1 이라고 한다. 그리고 S_i 이 관형절인 경우 그 관형절의 피수식어를 $M(modifiee)$ 이라고 한다. 탐색 방향이 forward라는 것은 S를 왼쪽에서 오른쪽 방향으로 훑어가는 것을 말하고 backward는 그 반대 방향을 말한다.

형태소 분석된 문장의 일반형을 비공식적으로 다음과 같이 정의된다. 여기서 N은 '체언+격조사'를 말한다.

$$S = N_1 N_2 N_3 N_4 P_1 M N_5 N_6 P_0$$

그러면 분리 결과는 대충 다음과 같은 형태를 취하게 된다.

$$S_0 = N_1 N_2 M N_4 N_6 P_0$$

$$S_1 = N_3 M P_1 \quad : \text{관계 관형절인 경우}$$

$$S_i = N_3 P_1 \quad : \text{관계 관형절이 아닌 경우}$$

4.2 알고리즘

표 2와 3 알고리즘을 예문 '철수가 그린 풍경화가 전람회에서 특선으로 뽑혔다.'를 가지고 추적해보자.

- (1) P_1 부터 P_0 사이에 있는 M , N_5 , N_6 를 S_0 소속으로 표시한다.

- (2) P_0 에 대해 먼저 표 1의 격 처리하기 알고리즘을 수행하고 다음에 P_1 을 가지고 수행한다.
- (3) 형태소 분석 결과 '철수가[주격] 그린 $[P_1]$ 풍경화가 $[M/주격]$ 전람회에서[부사격] 특선으로[부사격] 뽑혔다 $[P_0]$.'가 된다.
- (4) 소속 표시에 의해 '풍경화가 $[M/주격/S_0]$ 전람회에서[부사격/ $S_0]$ 특선으로[부사격/ $S_0]$ '가 모두 S_0 소속이 된다.

표 2. 격 처리하기 알고리즘

```

1 격 처리하기(Input: 문의 동사)
2 {
3   if (문 동사 = P0) { 탐색시작위치 = N1; 탐색끝위치 = N6;
4     탐색방향 = forward; }
5   else { 탐색끝위치 = 탐색시작위치; 탐색시작위치 = N4;
6     탐색방향 = backward; }
7
8   if ((문 동사 = 타동사) and (목적격 미발견))
9     격 발견하기(목적격);
10  else if ((문 동사 = 불완전동사) and (부사격 미발견))
11    격 발견하기(부사격);
12  else if (주격 미발견) 격 발견하기(주격);
13 }
    
```

표 3. 격 발견하기 알고리즘

```

1 격 발견하기(Input: 격 유형)
2 {
3   for (탐색시작위치부터 탐색방향으로 탐색끝위치까지) {
4     탐색하고 있는 문 소속을 표시해간다;
5     Case = 발견된 격;
6
7     if (Case = 격 유형) {
8       탐색시작위치 = 발견 위치;
9       return(OK);
10    }
11  }
12  if ((문장 = 짧은 관형절) and (M 사용 = NO))
13    M을 취한다;
14  M 사용 = YES;
15  }
16  else if (격 유형 != 목적격)
17    return(ERROR);
18  else return(OK);
19 }
    
```

- (5) P_0 가 자동사이기 때문에 목적격을 찾을 필요가 없다.
- (6) 불완전 동사라는 정보가 이미 있는 경우 (4)에서 부사격이 이미 찾아졌기 때문에 더 이상 찾을 필요는 없다.
- (7) 주격도 M 이 주격조사를 취하기 때문에 이미 찾아져 있다.
- (8) 다음으로 P_1 에 대해 목적격 조사를 back-

ward로 찾아보나 없다. 아직 M 이 S_1 에 의 해 사용된 것이 아니기 때문에 M 이 목적격을 취한다고 가정하여 '풍경화[목적격/ M/S_1]'이 된다.

- (9) P_1 이 불완전 동사가 아니기 때문에 부사격을 찾지 않는다.
- (10) P_1 에 대한 주격을 backward로 찾아 '철수가 [주격/ S_0]'로 표시한다.

5 결론

대량의 홀문장은 자연어처리에서 매우 가치있는 자료이다. 본 연구 결과로 생성될 홀문장의 집합은 논항 구조나 격 구조를 자동 구축하기 위해서 필수적이다. 그리고 의미역 파악을 위한 학습 데이터 구축을 위해서도 사용될 수 있다. 또한 유사 단어 군집화(clustering) 작업에서 단어간 유사도(similarity)를 측정하기 위해서도 홀문장 안에서의 단어거리를 계산하는 것이 정확도를 크게 향상시켜 주기 때문에 큰 도움이 된다.

본 연구는 기본적으로 현재 존재하지 않는 정보나 지식이 실험을 위해 존재하는 것으로 가정하지 않았다. 즉 현실적인 실험 환경을 설정하고 그 토대하에서 실용적인 자연어처리 시스템 개발에 필요한 정보를 구축하려고 했다. 그러나 본 알고리즘은 간단한 실험을 거쳤을 뿐 광범위한 데이터를 대상으로 검증되지는 않았다. 그러므로 보다 광범위한 실험 및 분석을 통해 개선되어질 여지가 상당하다.

감사의 글

자료조사에 도움을 주신 국어정보학 석사과정 이경혜 선생님께 감사드리며, 언어학적 용어 사용에 도움을 주신 독어학 박사 신효식 선생님께 감사 드립니다.

참고문헌

- [1] Brent, "Automatic Acquisition of Subcategorization Frames from Untagged Text," In *Proceedings of the 29th ACL*, 1991
- [2] Dong-Young Lee, "A Computational Search for a Verb and its Corresponding Subject in the Korean Sentence Containing Embedded Clauses", In *Proceedings of the Pacific Rim International Conference on AI.*, Vol. 2, pp. 219-225, 1992

- [3] Hang Li and Naoki Abe, "Generalizing Case Frames Using Thesaurus and the MDL Principle," In *Proceedings of ACL*, 1995
- [4] Hideki Tanaka, "Verbal Case Frame Acquisition from a Bilingual Corpus: Gradual Knowledge Acquisition," In *Proceedings of COLING*, 1994
- [5] Kemal Oflazer and Okan Yilmaz, "A Constraint-based Case Frame Lexicon," In *Proceedings of COLING*, 1996
- [6] Manning, "Automatic Acquisition of a Large Subcategorization Dictionary from Corpora," In *Proceedings of ACL*, 1992
- [7] Ted Briscoe and John Carroll, *Towards Automatic Extraction of Argument Structure from Corpora*, Technical Report, MLTT-006, Rank Xerox Research Center, Grenoble, 1994
- [8] 강은국, 조선어 문형 연구, 서광 학술 자료사, 1993
- [9] 강현화, 동사 연결 구성의 다단계성에 관한 연구 - 'V'어 'V' 구조를 중심으로 -, 연세대학교 국어국문학과 박사학위 논문, 1995
- [10] 김광진, 송영훈, 이정현, "한국어 내포문을 단문으로 분리하는 시스템의 구현", 한글 및 국어정보처리 학술발표논문집, 1994
- [11] 김준태, "자연언어 문서에서의 정보추출을 위한 구문패턴의 자동습득", 정보과학회 논문지(B), 제 23권 제 1호, 1996
- [12] 남기심, 국어조사의 용법, 서광학술자료사, 1993
- [13] 박재득, 점진 확률적 스키마 학습과 격역할 배정, 한국과학기술원 전산과학과 박사학위 논문, 1993
- [14] 송재관, 홍성웅, 박찬곤, "기계 번역을 위한 한국어 문장 패턴에 관한 연구", 한글 및 국어 정보처리 학술발표발표집, 1996
- [15] 양단희, 송만석, "한국어 기계학습과 말뭉치 구축", 정보과학회 봄 발표 논문집, 1998
- [16] 이정민, 강범모, 남승호, "한국어 술어의 의미구조 연구", 소프트과학 워크샵, 1997
- [17] 홍재성 외, 현대 한국어 동사 구문 사전, 두산동아출판사, 1996