

한국어 병렬문의 통사, 의미, 문맥 분석을 위한 결합범주문법

(Combinatory Categorical Grammar for the Syntactic,
Semantic, and Discourse Analyses of Coordinate
Constructions in Korean)

조 형 준 [†] 박 종 철 ^{**}

(HyungJoon Cho) (Jong Cheol Park)

요 약 자연언어처리에 있어서 병렬구문은 분석의 복잡성, 단어의 애매성, 서술어 생략 등에 따른 처리의 어려움을 내포하고 있다. 본 논문에서는 한국어에서 발생하는 병렬문의 통사적 특징을 능력문법 (competence grammar)의 입장에서 접근하고 분석된 결과를 기반으로 하여 한국어 병렬문 해석을 위한 결합범주문법 (Combinatory Categorical Grammar)을 제안한다. 제안된 결합범주문법을 사용해서 병렬문에 대한 각각 다른 수준의 통사적, 의미적, 문맥적 정보들이 사전에 어휘적으로 통합될 수 있고 통합된 정보를 이용하여 통사적, 의미적, 문맥적 분석들이 각각 다른 수준의 처리를 거치지 않고 동시에 점진적으로 유도될 수 있음을 보인다. 유도된 정보들을 통해 일반적으로 한국어 병렬문의 주된 기능이라고 생각되는 두 문장이 가지는 정보를 대조, 비교하는 기능이 표현될 수 있음을 보인다. 말뭉치를 분석하여 병렬문이 한국어 처리에서 차지하는 비중과 제시한 문법으로 처리할 수 없는 문형들에 대한 논의를 제공한다.

Abstract Coordinate constructions in natural language pose a number of difficulties to natural language processing units, due to the increased complexity of syntactic analysis, the syntactic ambiguity of the involved lexical items, and the apparent deletion of predicates in various places. In this paper, we address the syntactic characteristics of the coordinate constructions in Korean from the viewpoint of constructing a competence grammar, and present a version of combinatory categorical grammar for the analysis of coordinate constructions in Korean. We also show how to utilize a unified lexicon in the proposed grammar formalism in deriving the sentential semantics and associated information structures as well, in order to capture the discourse functions of coordinate constructions in Korean. The presented analysis conforms to the common wisdom that coordinate constructions are utilized in language not simply to reduce multiple sentences to a single sentence, but also to convey the information of contrast. Finally, we provide an analysis of sample corpora for the frequency of coordinate constructions in Korean and discuss some problematic cases.

1. 서론

병렬구문이란 일반적으로 두 개 이상의 구나 문장이

서로의 의미에 영향을 끼치지 않고 동등한 관계로 연결되어 있는 문장을 지칭한다. 예를 들어 (1)은 명사구로 이루어진 병렬구문을 보인다. 그러나 (2)에서와 같이 병렬 관계를 가질 수 있는 단위는 구나 문장에만 국한되지는 않는다. (3)에서는 [영희에게 사과를]이라는 단위가 [길수에게 딸기를]이라는 단위와 병렬을 이루고 있고 (4)에서는 [중속절 + 문장성분]이 병렬을 이루고 있다.¹⁾ (5)에서는 원거리 종속관계가 병렬형태를 이루고

· 본 연구는 첨단정보기술 연구센터를 통하여 과학재단의 지원을 받았다.

[†] 비 회 원 : 한국과학기술원 전산학과
hjcho@nlp.kaist.ac.kr

^{**} 종신회원 : 한국과학기술원 전산학과 교수
park@cs.kaist.ac.kr

논문접수 : 1999년 7월 23일

심사완료 : 2000년 1월 11일

1) 본 논문에서는 A라는 문장성분과 B라는 문장 성분의 병렬구

있다.²⁾

- (1) [Mary] and [John] ate potatoes.
- (2) Mary gives [John flowers] and [a policeman a ring].
- (3) 철수가 [영희에게 사과를], [길수에게 딸기를] 주었다.
- (4) [고추값이 오르면 고추를], [돼지고기값이 오르면 돼지고기를] 수입한다.
- (5) 장미값이 [철수는 t_i 비싸다고 생각했고], [영희는 t_i 저렴하다고 생각했다].

문장에서 서술어는 각 성분들을 모아 그 문장의 의미를 결정짓는 functor의 역할을 한다. 병렬문에서 나타나는 서술어의 선택은 구문분석과정에서 문장의 의미를 유도하는데 어려움을 유발하고 일반적으로 문장성분으로 인정되지 않는 비정규문장성분들의 병렬을 일으킨다.

결합범주문법 (combinatory categorial grammar [1, 2])은 단일화 (unification)에 기반한 어휘문법 (lexicalized grammar)이다. 결합범주문법은 병렬구문 (coordination), 추출 (extraction), 원거리 종속 관계 (long distance dependency)를 별도의 복원 과정에 의존하지 않고 설명할 수 있다고 알려져 있고 최근에는 억양 (intonation), 어순뒤섞기 (scrambling), scope 등의 자연언어 현상에 대하여 역시 별도의 약정 (stipulation) 없이 설명을 제공하는 것으로 알려져 있다 [3-6].

[7]는 일차단일화 (first order unification)를 기반으로 병렬구문에서 나타나는 모든 현상들이 의미적으로 해석될 수 있음을 보여주었다. [8]은 결합범주문법을 기반으로 자유어순을 가진 터키어에서 논항 (argument)의 순서가 고정된 서술논항구조 (predicate argument structure)가 유도될 수 있음을 보이고 있다. 최근에는 정보구조 (information structure)의 개념을 도입하여 문장 내에서 나타내고자 하는 정보를 정형화하고 이를 응용하는 연구들이 있었다 [4, 9, 3].

본 논문에서는 한국어 병렬구문의 특징을 설명하고, 결합범주문법을 기반으로 한국어에서 발생하는 병렬구문을 분석한다. 분석된 결과로부터 한국어 병렬문 처리를 위한 결합범주문법을 제안하고 한국어를 처리하는데 필요한 결합자와 축약규칙을 제안한다. 제안된 결합범주문법을 사용하여 병렬문을 구문분석하는 과정에 정

보구조의 개념을 도입하여 한국어 병렬문의 정보구조가 구문분석하는 과정에서 의미구조와 함께 유도될 수 있음을 보인다. 마지막으로 말뭉치를 분석하여 병렬문이 한국어 처리에서 차지하는 비중과 제시한 문법으로 처리할 수 없는 문형들에 대한 논의를 제공한다.

언어 연구에 있어서 접근방법은 크게 내재적인 언어 능력의 한계를 찾고자 하는 능력문법 (competence grammar)과 실제적인 언어 수행을 중시하는 실행문법 (performance grammar)의 두 가지로 나뉘어진다. 후자의 경우 최근에 주로 말뭉치를 기반으로 한 연구가 진행되고 있으며, 비문과 정문의 구별보다는 문장이 말뭉치에 있는지 없는지에 따라 문법의 구성을 결정한다. 이런 접근 방법의 장점으로는 실용적인 시스템을 비교적 빠른 시간내에 구축할 수 있다는 데 있다. 하지만 동일한 이론에 기반한 문법이라도 주어진 말뭉치에 따라 서로 다른 결과를 주기 때문에 언어현상을 일관성 있게 표현하기 어렵다는 단점이 있다. 반면 능력문법의 경우는 인간이 내재하고 있는 언어능력의 한계를 정해 문법을 고안하기 때문에 정문, 비문의 한계를 정하는 것에 초점을 두며 말뭉치에 직접적인 관계없이 언어현상을 일관성 있게 표현할 수 있다는 장점이 있다. 반면에 정문, 비문의 한계를 정하기 어렵고 정형화하기 힘들다는 단점도 있다. 본 논문은 능력문법을 기반으로 한국어 병렬문에 접근하고 있으나 KAIST에서 개발한 말뭉치를 조사하여 이들간의 상대적인 관계에 대하여도 논의를 제공한다.

2. 관련연구

2.1 한국어 처리를 위한 결합범주문법

한국어를 처리하기 위해 제시되었던 범주문법에는 [10, 11, 12] 등이 있는데 본 논문에서와 같이 정보구조를 이용하려는 시도는 아직 없는 것으로 보인다. [11]는 한국어에서 발생하는 어순뒤섞기를 자연스럽게 처리하기 위해 집합표시를 사용하였고 구문분석시 type raising으로 인해 발생하는 중의성을 해소하기 위해 type raising 대신 typed lambda calculus를 사용해서 commutative productive category를 제안하였다. 제안된 문법은 병렬문, 원거리 종속관계, 어순뒤섞기 등의 다양한 문형을 처리할 수 있으며 [4]에서 제안한 중집합결합범주문법 [13]의 원형으로 볼 수 있다. 그러나 제안된 형태의 문법에는 과생성³⁾의 문제가 있다.

[12]에서는 한국어가 가지고 있는 특성에 적합하도록

조를 [A]와 [B]와 같이 꺾쇠 괄호를 써서 표시한다.

2) 원거리 종속관계에서 t_i 는 논항이 추출된 자리를 나타내고, '논항'은 추출된 논항을 나타낸다.

3) 나는 철수가 영희에게 착하다고 말했다.

기본적인 방향범주문법의 확장이 이루어졌다. 한국어의 특징들, 즉 비교적 자유로운 어순, 잦은 생략 등을 설명하기 위해 집합표시와 생략논항을 사용하였으며 방향범주문법에서 사용하는 축약규칙 (reduction rule)으로 좌축삭제와 우축삭제를 제시하였다. 이러한 방향범주문법을 한국어에 적용시킬 경우, 단순한 축약규칙에 의해 시스템이 간결해진다는 장점이 있지만 축약규칙이 서술어를 중심으로 적용되기 때문에 서술어가 생략된 형태를 지닌 (3)과 같은 문장을 처리할 수 없다는 단점이 있다. [10]에서는 이 문법에 type raising과 composition을 축약규칙으로 추가하여 서술어가 생략된 병렬문을 처리할 때 나타나는 문제점을 보완하였다. 그러나 제안된 규칙만으로는 한국어에서 발생하는 원거리 종속관계를 설명하지 못한다. 따라서 제안된 형태의 문법만으로는 원거리 종속관계가 나타난 문장들로 이루어진 병렬문을 설명하지 못한다.

2.2 결합범주문법에서 병렬문 처리

결합범주문법에서 형상언어 (configurational language)인 영어를 위하여 제안된 축약규칙은 표 1과 같다 [2].

축약규칙에서 사선의 방향은 functor가 필요로하는 논항의 방향을 나타낸다. 예를 들어 forward application의 경우, functor X/Y의 오른쪽에 논항 Y가 발견되면 축약규칙이 적용되어 결과로 Y를 내어놓는다는 의미이다. Composition은 functor와 functor를 합성하는 연산이고 type raising의 경우는 논항끼리 합성해야 할 필요가 있을 경우, 해당 논항을 논항으로 가지는 functor의 통사범주를 할당함으로써 논항끼리의 합성을 가능하게 만드는 축약규칙이다. 제안된 축약규칙을 사용하여 영어 병렬문을 처리하는 과정은 그림 1과 같다. <φ>는 접속사 'and'를 중심으로 좌우의 통사범주가 동일하기 때문에 coordination 축약규칙이 적용되는 부분이다. 이 연산(4)을 통해 동일한 통사범주 (syntactic category)를 가진 접속항들을 하나의 문법범주로 축약할 수 있게 된다. 결합범주문법을 사용해서 병렬문을 처리할 때 얻을 수 있는 이점은 다른 문법체계에서 처리할 수 없거나 특별한 약정 (stipulation)을 사용해서 해결하고 있는 문형들, 즉 비정규성분 (nonstandard constituent)이 포함된 병렬문이나 생략현상(gapping)⁵⁾

이 포함된 병렬문이 표 1에 포함된 기본 규칙만으로 해결될 수 있고 구문분석과 동시에 주어진 문장의 의미와 정보구조를 동시에 처리할 수 있다는 점이다 [1]. 결합범주문법이 비정규성분이나 생략현상이 포함된 병렬문을 처리할 수 있는 이유는 type raising과 composition을 사용해서 접속항을 하나의 문법범주로 축약할 수 있고 또 축약과정에서 접속항의 의미를 유도해낼 수 있기 때문인데 이는 결합범주문법의 associative nature에 근거한다. 이것은 결합범주문법이 비정규성분이나 생략현상이 포함된 병렬문을 처리할 수 있다는 장점을 갖기도 하지만 주어진 동일한 의미를 가진 통사범주에 대해서 여러 번의 분석을 하는 단점도 가지게 된다.⁶⁾ 이와 같은 현상을 spurious ambiguity라고 하는데 이것이 병렬문 등의 언어현상을 처리하기 위해서는 필요하지만 동일한 의미를 가진 분석을 여러 번 수행함으로써 시스템의 성능이 악영향을 끼치고 단순히 통사범주만을 가지고 문장을 처리할 때는 과생성을 유발하기도 한다.⁷⁾

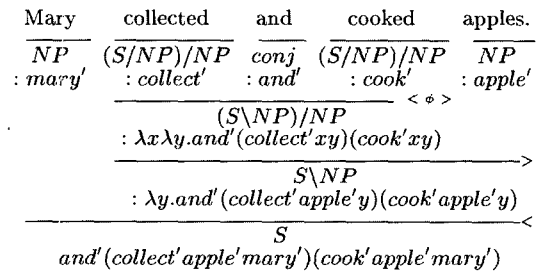


그림 1 Mary collected and cooked apples

표 1 결합범주문법의 축약규칙 (영어)

규칙	규칙이름	축약기호
X/Y Y → X	Forward Application	>
Y XY → X	Backward Application	<
X conj X → X	Coordination	<φ>
X/Y Y/Z → X/Z	Forward Composition	>B
YZ XY → XZ	Backward Composition	<B
X/Y (Y/Z)\$ → (X/Z)\$	Generalized Forward Composition	>B ⁿ
X → T/(TX)	Forward Type Raising	>T
X → T\'(T\'X)	Backward Type Raising	<T

4) 이 연산은 통상적으로 conjoin operation이라고 칭한다.

5) 본 연구에서 말하는 생략현상 (gapping)은 문장을 이루는 요소 (constituent)의 생략으로 인해 문장성분들이 불연속적으로 나열되어 있는 현상을 말한다.

John gave Mary flowers and Karl, Jane a ring.

6) [[철수가 사과를] 먹었다.]

[철수가 [사과를 먹었다.]]

7) 수 자질(number feature), 성 자질(gender feature), 의미 자질(semantic feature) 등을 사용하여 과생성을 방지할 수 있다.

2.3 한국어 병렬문

병렬문은 크게 체언연결과 용언연결로 나눌 수 있다. 이 때 체언연결에 대해 문연결로 보는 견해 [14], 구의 연결로 보는 견해 [15], 그리고 상황에 따라 구연결과 문연결을 모두 인정하는 견해 [16] 등이 있다. 반면 용언 연결에 대해서는 공통적으로 문접속으로 인정되고 있다.

서술어 중심 언어인 한국어에서는 서술어가 핵심성분(central constituent)이 되고, 주어나 목적어 등 나머지 성분은 주변성분(peripheral constituent)이 된다. 주변 성분들은 때에 따라서 탈락되기도 하지만 핵심 성분인 서술어는 탈락되지 않는다 [17]. 그러나 병렬문에서는 서술어가 탈락되는 경우가 있다. 예를 들어 (8)은 우접속항의 동사 '좋아한다'가 좌접속항에서 생략된 문장이다. 한국어 병렬문에서의 서술어 생략현상은 좌접속항에서만 일어난다. 우접속항에서 서술어 생략현상이 일어날 경우, 해당 문장은 비문(nonsentence)이 된다. (9)는 우접속항에서 서술어 생략현상이 발생한 경우이다.⁸⁾

- (8) 철수가 사과를, 영희가 딸기를 좋아한다.
 (9) *영희가 딸기를 좋아하고, 철수가 사과를.

서술어 생략의 범위는 하나의 서술어는 물론이고, 나아가서 확대된 구, 절까지도 범위가 될 수 있다 [18]. (10)은 '밥을 먹었다'라는 동사구가 좌접속항에서 생략된 형태의 문장이다.

- (10) 철수는 집에서, 영수는 식당에서 밥을 먹었다.

명사나 명사구의 생략에 대해서는 (11)에서와 같이 우접속항에서만 가능하다는 견해가 있다 [17]. 그러나 (12)와 같은 경우도 정문으로 간주하여야 하므로 이와 같은 제한은 없는 것으로 보인다. 명사구가 우접속항에서 생략될 때, 우접속항에서 생략된 명사나 명사구는 (13)과 같이 문두에 나타나는 경향이 있다.

- (11) 철수는 사과를 길수에게, 영희는 경희에게 주었다.
 (12) 철수는 길수에게, 영희는 경희에게 맛있는 딸기를 주었다.
 (13) 사과를 철수는 길수에게, 영희는 경희에게 주었다.

병렬구문이 내포되어 있을 경우에는, 내포문의 서술어

는 반드시 (14)에서와 같이 내포문의 끝에 위치해야 한다. 문장의 끝에 위치하지 않았을 경우, (15)에서와 같이 비문이 되거나 (17)에서와 같이 문장의 의미가 달라진다⁹⁾.

- (14) 길수가 철수가 밥을, 영희가 빵을 먹는다고 생각한다.
 (15) *길수가 먹는다고 철수가 밥을, 영희가 빵을 생각한다.
 (16) 철수가 영수는 착하고 영리하다고 생각한다.
 (17) 철수가 영리하고 영수는 착하다고 생각한다.

이상에서 논의한 한국어 병렬문의 문법적인 특징을 종합해 보면, 우접속항에서 동사 또는 동사구가 생략될 경우, 우접속항은 서술어가 생략된 형태의 문장이 되기 때문에 비문으로 간주되고, 명사나 명사구는 좌우 접속항 모두에서 생략이 가능하다고 본다. 내포문이 병렬을 이루고 있는 문장에서는 내포문의 서술어가 항상 내포문의 끝에 위치해야 한다. 그렇지 않을 경우 문장이 비문이 되거나 화자가 의도한 바와는 다른 의미의 문장이 될 수 있다.

2.4 한국어 대등접속사

접속문을 대등접속과 종속접속을 나누는 기준으로는 대칭성, 선행절 자리 바꾸기, 선행절 옮기기, 접속사 되풀이, 재귀 대명사 및 후행절 주제어 등과 관련된 고유 통사 특성이 제시된 바 있다 [19]. 또 대칭성과 교호성, 공통의미차원을 대등접속을 이루는 필수요건으로, 동일 통사구조를 수의적인 조건으로 간주해서 정립된 기준이 있다 [17]. 이렇게 제시된 기준들은 기본적으로 의미적 기준에서 설정된 것이다.

본 절에서는 교호성과 대칭성의 기준을 만족시키고, 영어에서 일반적으로 사용되는 대등접속사 (and, or, but)에 해당하는 의미와 기능을 가진 것으로 간주되는 대등접속사에 대한 설명을 다음과 같이 정리한다. '고'로 연결된 접속문은 그 선행절의 용언이 상태성을 드러낼 때는 대칭적 접속관계이다. (18),(19)와 같이 선행절과 후행절을 맞바꾸어도 의미 변화가 생기지 않는 대등접속관계를 이룬다.

9) (15)는 (14)에서 내포문의 서술어인 '먹었다'의 위치를 내포문의 문두로 옮긴 경우이다. (17)은 (16)의 서술어인 '영리하다'의 위치를 내포문의 문두로 보낸 예제이다. 한국어에서는 주어의 생략이 가능하기 때문에 (17)과 같이 화자의 의도와 다른 문장도 정문으로 취급된다.

8) *는 비문앞에 쓰인다.

- (18) 하늘은 푸르고 물은 맑다.
- (19) 물은 맑고 하늘은 푸르다.

(20)은 선행절의 용언이 동작성을 드러내는 경우이다. 이 경우 문장은 선행절의 동작성에서 비롯된 시간적 선후관계를 가지게 되고, 이로 인해 문장은 대등접속을 이루지 않고 있는 듯이 보인다. 하지만 이런 현상은 ‘고’ 자체가 순차성을 띄고 있기 때문에 발생한 것이 아니고, 선행 용언의 의미적 특질에 의한 것임을 (21)을 통해 알 수 있다. 즉 문장에서 ‘고’가 생략되더라도 문장은 시간적 선후관계를 띄고 있다. ‘고’의 접속문은 (22)와 같이 화용적 의미 (pragmatic meaning)로 말미암아 인과관계 등을 드러냄으로써 비대등적 접속관계가 이루어지는 경우도 있다. 그러나 이런 인과관계들도 ‘고’에 연유한 것이 아님을 알 수 있다. 즉 (23)과 같이 ‘고’를 제거하더라도 문장은 인과관계를 띄게 되고 이런 인과관계는 용언의 동작성으로 인해 발생한다.

- (20) 그는 전화를 받고 밖으로 나갔다.
- (21) 그는 전화를 받았다. 그는 이어 밖으로 나갔다.
- (22) 그 아이가 이 약을 먹고 병이 나았다.
- (23) 그 아이가 이 약을 먹었다. 그 아이가 병이 나왔다.

‘며’는 단순한 병렬성을 나타내는 경우와 동시성을 나타내는 경우가 있다. 전자는 ‘고’와 마찬가지로 순수한 대등접속 어미로 인정된다. 병렬성을 나타내는 경우의 ‘며’는 (24), (25)와 같이 상태를 나타내는 용언과 결합된 경우이다.

- (24) 봄에는 산과 들에 꽃이 피며 새가 운다.
- (25) 봄에는 산과 들에 새가 울며 꽃이 핀다.
- (26) 그들은 이야기를 나누며 차를 마셨다.

‘거나, ‘나’는 선택의 의미기능을 지닌 접속어미이다. 그 좌접속항과 우접속항 가운데 한 쪽을 선택하는 이접(disjunction) 관계를 보인다.

- (27) 우리는 신문을 읽거나 잡지를 읽는다.
- (28) 우리는 잡지를 읽거나 신문을 읽는다.
- (29) 가까운 거리는 차를 타나 비행기를 타나 마찬가지로이다.
- (30) 가까운 거리는 비행기를 타나 차를 타나 마찬가지로이다.

(27)과 (28)은 ‘거나’를 중심으로 접속을 이루는 좌접속항과 우접속항의 순서를 뒤바꾼 문장이다. (29)와 (30)은 ‘나’를 중심으로 접속을 이루는 각 접속항의 순서를 바꾼 문장이다. 접속항의 순서를 바꾸어도 원문의 의미가 그대로 살아있으므로 ‘거나’와 ‘나’가 대등접속어임을 알 수 있다.

한국어의 역접접속어미로는 ‘나’와 ‘지만’이 있다. 역접접속어미를 대등접속소로 간주할 수 있는지에는 여러 가지 의견이 있다 [17, 20, 21]. 그러나 (31), (32), (33), (34)를 통해 역접접속어미도 대등접속을 형성할 수 있음을 알 수 있다.

- (31) 그는 시장에 가지만 물건을 사지 않는다.
- (32) 그는 물건을 사지 않지만 시장에 간다.
- (33) 철수는 집에 있지만 경희는 밖에 나갔다.
- (34) 경희는 밖에 나갔지만 철수는 집에 있다.

‘그러나’와 ‘그렇지만’은 일반적으로 서로 대조되는 두 문장을 잇는 대등접속사이다 [21]. ‘나’와 ‘지만’이 대조되는 두 절을 연결하는데 사용되는 것과는 달리 ‘그러나’와 ‘그렇지만’은 (35), (36)과 같이 이미 끝맺은 앞 문장과 그 뒤 문장을 대등하게 연결한다.

- (35) 철수는 용감하다. 그러나 영리하지는 않다.
- (36) 그들은 열렬히 사랑하였다. 그렇지만 결혼은 할 수 없었다.

‘그리고’는 용언은 물론 단어, 구, 문장, 단락 등을 모두 대등연결시켜 주는 대등접속사로 볼 수 있다 [17]. ‘또’와 ‘또는’은 체언, 용언 양쪽에서 사용되는 것은 물론이고, 형태적으로도 특이하다. 순접접속사 ‘또’에 대조의 의미를 가지는 ‘는’이 붙음으로 해서 선택의 이접성을 가지게 된 것이다.

‘및’은 주로 명사구를 이어 주는 구실을 한다. 그 밖의 후치사구나 동사구를 연결하는 기능은 거의 없다 [21]. ‘와/과’와 ‘하고’는 주로 명사구끼리 나란히 접속하거나 나열하는 구실을 하며, 일부 후치사구 따위를 나란히 접속하는 데에도 쓰인다 [21]. (37)에서는 ‘와’가 명사 ‘철수’와 ‘영희’를 대등하게 연결하고 있음을 보인다. (38)처럼 ‘와’ 대신 ‘하고’를 사용해서 명사나 명사구들을 연결할 수도 있다.

- (37) 철수와 영희는 가끔 싸운다.
- (38) 철수하고 영희가 가끔 싸운다.

지금까지 논의된 내용을 근거로 정리한 대등접속사의 유형은 표 2와 같다 [17, 20, 21].

표 2 대등접속사의 유형

연결어미	순접대등소	이접대등소	역접대등소
체언대등소	와/과, 하고, 및, 이랑, 이고, 이며, 이니	이거나, 이든지, 이나	
용언대등소	고, 으며, 으니, 며	거나, 든지, 으나	나, 지만
문장대등소			그러나, 그렇지만
체언, 용언, 문장대등소	또, 그리고	또는, 혹은	

2.5 정보구조

정보구조 (information structure)는 발화된 말에 포함되어 있는 정보의 조직 (information organization)이라고 할 수 있다 [9]. 정보구조는 (39)와 같은 질의, 응답의 테스트로 명확히 분석될 수 있다.¹⁰⁾

(39) Q: What did John eat?
A:[He ate] [cookies].

질문에 대한 대답으로 제시된 정보구조는 질문을 할 때 도입된 가정에 의해 나뉘어 질 수 있으며 이렇게 나뉘어진 정보는 대답을 통해 문맥상에서 질문자가 필요로 하는 직접적인 정보를 제공한다. 이런 관점에서 보면 정보구조는 발화된 말을 이루는 단위 사이의 정보들간의 관계가 어떻게 이루어졌는지 보여준다고 할 수 있다. (39)에서 새로운 정보 'cookies'를 강조함으로써 질문을 받은 화자가 이미 알고 있던 정보 즉 'He ate'과 대조 (contrast)를 이루게 하고 이로 인해 질문자는 자신이 얻고자 하는 정보를 쉽게 얻을 수 있는 것이다. 하지만 정보구조의 형태에 대한 일반적인 함의는 아직 없으며, 다음과 같이 다양한 형태로 정보구조를 표현하고 있다 [4]:

Theme과 Rheme, Topic과 Comment, Topic과 Focus, Focus과 Presupposition, Focus과 Open-proposition, Link, Focus와 Tail, Topic, Focus와 Background.

10) 볼드체는 문장내에서 고저 악센트를 받는 부분을 나타낸다.

[3]은 (40)과 같이 하나의 문장을 구의 억양에 의해 theme과 rheme으로 나누고 각각의 고저 악센트 (pitch accents)에 근거하여 각 theme과 rheme 안에서 focus를 표시한다.¹¹⁾

(40) Q: I know that the **old** widget had a **slow** processor.
But what processor does the new widget include?

A: (The **new** widget includes) (a **fast** processor).
L+H* LH% H* LL%
Ground Focus Ground Ground Focus Ground
Theme Rheme

[4]는 자유어순으로 알려진 터키어에 대해서 하나의 문장을 topic과 comment로 나누고 나뉘어진 comment를 다시 focus와 ground로 나누는 정보구조를 제시하였고 정보구조를 이용하여 터키어의 어순변화에 따른 정보구조의 변화를 설명하였다. [9]는 영어에서 생성된 정보구조를 사용하여 영일 기계번역 시스템 등에서 일본어 번역문의 조사 'ga', 'wa'를 문맥에 맞게 생성하는데 사용하며 일본어에 대해서도 문장을 theme과 rheme으로 나누는 정보구조를 제시하였다. 정보구조를 나누는 기준은 (41)과 같이 조사 (예 ga, wa)의 의미 특성과 각 어절에 주어지는 고저 악센트를 사용하였다. .

(41) Q: Ken-wa nani-o tabeta-no?
Ken-TOP what-ACC ate-Q
“켄은 무엇을 먹었니?”

A:[Ken-wa] [banana-o] [tabeta].
theme rheme theme
Ken-TOP banana-ACC ate
“켄은 바나나를 먹었다.”

2.6 조사 '은/는'

한국어의 보조사 '은/는'은 일반적으로 문장내에서 주제 (theme)나 대조 (contrast)의 의미를 나타낸다 [22, 23].¹²⁾

11) H*는 날카롭게 올라가는 고저 악센트를 의미하고 L은 떨어지는 고저 악센트를 의미한다. L+H*는 약간 느리게 올라가는 고저 악센트를 나타낸다. %는 어절의 경계를 나타낸다.
12) 앞에서와 같이 볼드체는 문장내에서 고저 악센트를 받는 부분을 나타낸다.

- (42) i. 저 사람은 철수이다.
ii. 철수는 바나나를 먹었다.

(42)에서 '은/는'은 해당명사구가 주제 (theme)임을 나타낸다. (42)에서 사용된 '은/는'의 문맥적 (discourse) 의미는 이야기가 전개되는 앞뒤 관련의 측면에서 볼 때 일단 이야기에 등장하여 소개된 대상을 그 다음에 다시 나타내는 데에 쓰인다. 즉 옛 정보, 또는 문맥중심 (discourse center)을 나타낸다 [22, 23, 24]. 그런데 이 옛 정보는 반드시 언급된 것만을 지칭하지는 않고 유추 가능한 것들도 포함된다.¹³⁾

(43)에서 사용된 '은/는'은 대조의 의미를 나타내며 이미 언급된 어떤 요소와의 대조를 나타낼 때 쓰인다 [22]. (43)에 사용된 대답은 다른 사람들은 딸기를 먹지 않았다는 가정하에 오직 철수만 딸기를 먹었다는 사실을 나타낸다.¹⁴⁾

(43) Q: 여기에 있는 사람들 중에서 누가 딸기를 먹었나?

A: 철수는 딸기를 먹었어.

(44)는 철수는 다른 것은 먹지 않았다는 가정이 있는 문장이다. (46)은 철수가 다른 음식들은 먹었다는 가정이 깔려있는 문장이다. (44), (46)의 '철수는'에서 사용된 '는'의 역할은 문맥 상 앞에서 이미 언급되었던 것을 나타낸다.

- (44) 철수는 딸기는 먹었다.
(45) 철수는 딸기는 먹었다.
(46) 철수는 딸기는 먹지 않았다.

(45)는 여러 다른 사람들과 여러 다른 먹거리가 있는데 오직 철수만 다른 것은 먹지 않고 딸기를 먹었다는 가정에서 나타날 수 있는 문장이다. 이 때 사용된 '은/는'은 철수와 다른 사람을 대조시키고, 딸기와 다른 먹거리들을 대조시키는 역할을 한다. 이처럼 대조를 나타내는 '은/는'이 쓰이는 경우, 대조를 이루는 명사구는 일반적으로 고저 악센트 (pitch accent)를 받는다 [23].

(47)은 원거리종속관계 (long distance dependency) 또는 원거리어순뒤섞기 (long distance scrambling)의

예인때 '딸기는'에서 사용된 '는'은 주제화의 기능을 가지며, '철수가'와 대조시키기 위해서 문두로 나온 것이다. (48)은 다른 먹거리들이 존재하는 가운데 다른 사람들은 모르지만 철수만은 딸기를 먹은 사람이 영희라고 생각한다¹⁵⁾는 사실을 강조하는 의미로 쓰인다. 이 때 '딸기는'에서 쓰인 '는'과 '철수는'에서 쓰인 '는'은 각각 다른 먹거리와 다른 사람들과 대조하는 의미로 사용되었다. (49)는 평언 (rheme)인 '딸기'를 주제인 나머지 부분과 분리하기 위해서 문두로 보낸 것이다. 이 때 A2의 강조의 의미가 A1보다 강하다.

(47) Q: 누가 영희가 딸기를 먹었다고 생각했나?

A: 딸기는 철수가 [영희가 t_i 먹었다고] 생각했다.

(48) 딸기는 철수는 [영희가 t_i 먹었다고] 생각했다.

(49) Q: 무엇을 철수가 영희가 먹었다고 생각했나?

A1: 딸기,를 철수는 [영희가 t_i 먹었다고] 생각했다.

A2: 딸기,는 철수는 [영희가 t_i 먹었다고] 생각했다.

3. 한국어 병렬문 처리를 위한 결합범주문법

2절에서는 결합범주문법, 한국어 병렬구문의 특징, 한국어 병렬구문에서 사용되는 대등접속사, 정보구조, 보조사 '은/는'이 가지는 문장 상에서의 의미에 관해서 기존의 연구와 관련하여 정리, 논의하였다. 본 절에서는 앞에서 논의되었던 내용들을 기반으로 다양한 형태의 한국어 병렬문을 처리하기 위한 결합범주문법을 제안하고 한국어 병렬문 처리를 위한 결합자와 축약규칙을 정의한다.

3.1 기본적인 한국어 병렬문의 처리

(50)-(53)은 한국어에서 발생할 수 있는 기본적인 병렬문 문형들이고 (54)는 (51)을 어순뒤섞기¹⁵⁾한 문장이다. (56)은 [종속절 + 문장성분]이 병렬을 이루고 있는 문장이다. (57)은 내포문의 목적이 문두로 나와 원거리 종속관계를 이루고 있는 문장들이 병렬을 이루고 있는 문장이다.

- (50) 철수가 의사가, 영희가 선생님이 되었다.
(51) 철수가 사과를, 영희가 딸기를 먹었다.
(52) 철수가 영희에게 사과를, 영수가 길수에게 딸기를 준다.

13) 집 → 창문, 책 → 페이지 등

14) "다른 사람은 몰라도"라는 가정도 성립됨. 즉 다른 사람은 무엇을 먹었는지 모르지만 철수가 딸기를 먹은 사실에 대해서는 확신을 하고 있음.

15) 어순뒤섞기를 본 논문에서는 descriptive한 용어로만 사용한다. 특히 본 논문에서는 기본적인 한국어 어순인 '주어 목적어 서술어'가 변형하여 '목적어 주어 서술어'가 생긴 것이 아니라 이 두 어순 모두 기초생성 (base generate)되었다고 간주한다.

- (53) 클래식 음악이 영화를 즐겁게, 철수를 즐겁게 만든다.
- (54) 사과를 철수가, 딸기를 영화가 먹는다.
- (55) 사과를 철수가 영화에게, 딸기를 철수가 길수에게 주었다.
- (56) 고추값이 오르면 고추를, 돼지고기값이 오르면 돼지고기를 수입한다.
- (57) 장미값이 철수는 t_i 비싸다고 생각하고, 영화는 t_i 저렴하다고 생각한다.¹⁶⁾

결합범주문법에서 conjoin operation이 적용되기 위해서는 기본적으로 좌접속항과 우접속항의 문법범주가 동일해야 한다. 좌접속항에서 서술어가 생략된 한국어 병렬문 (50)-(55)의 형태는 논항몽치병렬 (argument cluster coordination)을 이루고 있다 [1, 25]. 이들 좌접속항을 이루고 있는 요소들은 그림 2¹⁷⁾같이 평서문에서 서술어라는 functor와 결합함으로써 문장을 구성한다. 하지만 좌접속항에서 서술어가 생략되었기 때문에 이들 범주로는 병렬문의 연산에 필요한 접속항을 구성할 수 없다. 이 경우 결합범주문법에서는 각 범주에 type raising된 문법범주를 할당하여 해결한다.

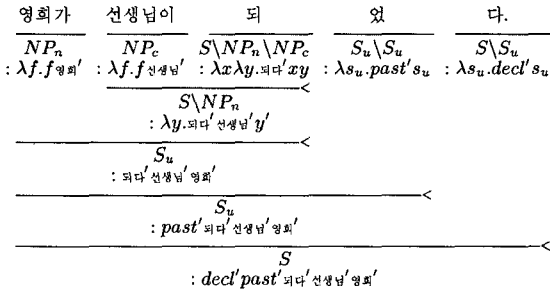


그림 2 영화가 선생님이 되었다.

영어의 2형식과 5형식에 해당하는 문장들 (50), (53)을 처리할 때, 보어로 쓰이는 범주들은 부사의 형태를 띄고 있는 경우도 있다. 한국어에서 부사는 보조적 요소로서 생략이 가능하지만, (53)에서 ‘즐겁게’와 ‘즐리게’를 생략한다면 문장이 성립되지 않는다. 따라서 부사의 형태로 쓰인 보어에도 동사나 동사구, 또는 문장 전체를 꾸며주는 일반적인 부사의 범주에 추가해서 NAC (주

격, 목적격 보어)라는 범주를 할당한다. 좌접속항에서 서술어가 생략된 형태의 한국어 병렬문을 처리할 때, NAC 범주도 다른 범주와 동일하게 type raising과 composition을 받는다.

표 3 사전에 들어있는 NAC 범주

어 절	내 용
즐겁게	klex(즐겁게, NAC:즐겁게).
	klex(즐겁게, ($s \setminus np[case:nom] \setminus np[case:acc]$)) / ($s \setminus np[case:nom] \setminus np[case:acc]$ \ NAC:즐겁게).
	klex(즐겁게, ($s \setminus np[case:acc] \setminus np[case:nom]$)) / ($s \setminus np[case:acc] \setminus np[case:nom]$ \ NAC:즐겁게).

표 3은 NAC 범주에 대한 사전의 내용이다.¹⁸⁾ 어떤 논항 n 이 가질 수 있는 type raising된 문법범주의 형태는 (58)과 같고 type raising에 쓰이는 schema는 (59)와 (60)과 같이 표현된다 [8].

- (58) $C(n) = T_r/T_a$ 또는 $T_r \setminus T_a$
- (59) $NP_n \rightarrow T/(T \setminus NP_n)$ 또는 $T \setminus (T \setminus NP_n)$
- (60) $NAC \rightarrow T/(T \setminus NAC)$ 또는 $T \setminus (T \setminus NAC)$

T_a 는 NP_n 또는 NAC를 논항으로 가지는 functor (verb)의 범주이고, T_r 은 T_a 에서 NP_n 또는 NAC를 생략한 형태의 범주이다. 결국 NP_n 에 type raising된 범주는 NP_n 또는 NAC를 논항으로 받아서 T_r 을 결과로 내어놓는 functor에 할당되는 범주와 동일하다. Type raising된 범주들에 composition 연산을 적용함으로써 composition 연산이 적용된 NP들을 논항으로 가지는 functor의 범주가 할당된다. 그림 3, 4, 5는¹⁹⁾²⁰⁾ (50), (51), (52)를 구문 분석한 과정을 나타낸 것이다. (53)은 영어의 5형식에 해당되는 문장이다. 일반적으로 한국어에서 필수요소가 아닌 부사어가 목적보어로 사용될 경우, 목적보어의 자리는 목적어의 뒤로 고정되고, 문장성립에 있어서 생략되어서는 안되는 필수요소가 된다. 그림 6은 (54)의 구문 분석과정을 나타낸 것이다. 좌접속

16) “장미값이 철수는 비싸다고 영화는 저렴하다고 생각한다.”가 좀 더 자연스러운 문장이다. 이에 대한 논의는 4절의 분석 및 토의에서 보인다.
 17) S_u 는 단독으로 완전한 문장을 만들지 못하는 통사범주이다.

18) klex는 사전에서 한국어 어휘임을 나타내는 predicate이다. 표현을 간략하게 하기 위해서 격을 제외한 다른 자질 (feature)들은 나타내지 않았다. 콜론 (:) 뒤의 ‘즐겁게’는 단어의 의미에 대한 표현을 나타낸다 [7].
 19) NP_n 은 주격, NP_n 은 목적격, NP_n 은 여격, NP_c 는 보격을 지칭한다.
 20) IV는 $S \setminus NP_n$, TV는 $S \setminus NP_n \setminus NP_n$, DV는 $S \setminus NP_n \setminus NP_n \setminus NP_n$ 를 지칭한다.

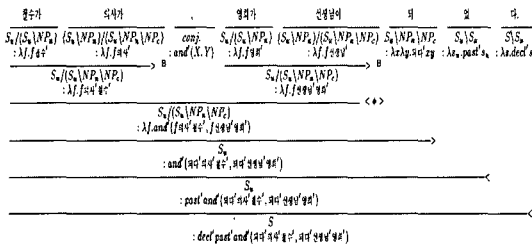


그림 3 철수가 의사가, 영희가 선생님이 되었다.

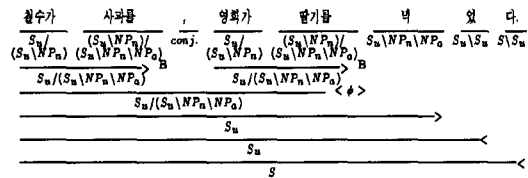


그림 4 철수가 사과를, 영희가 딸기를 먹었다.

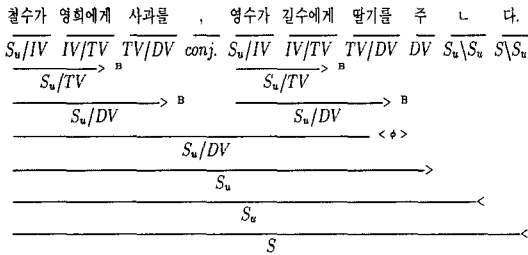


그림 5 철수가 영희에게 사과를, 영수가 길수에게 딸기를 준다.

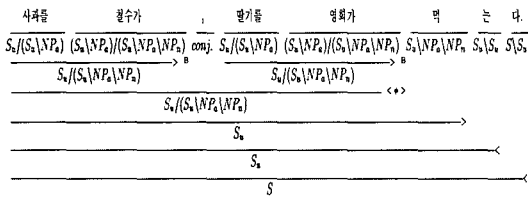


그림 6 사과를 철수가, 딸기를 영희가 먹는다.

항과 우접속항에서 서술어를 제외한 문장을 구성하는 요소의 생략이 없다면 (52)로부터 모두 3! (=6)개의 서술어로 끝나는 병렬문이 유도됨을 알 수 있다. 그림 3 - 그림 7은 각 병렬문이 유도되는 과정을 보인 것이다. 21)22)23) 그림 3을 구문분석하기 위해 필요한 사전의 내

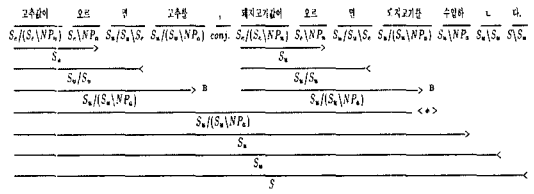


그림 7 고추값이 오르면 고추를, 돼지고기값이 오르면 돼지고기를 수입한다.

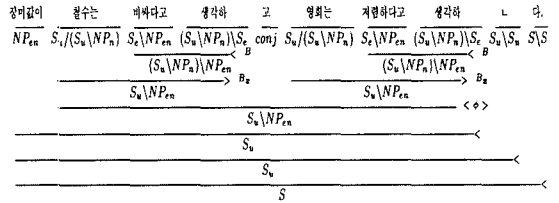


그림 8 장미값이 철수는 t1 비싸다고 영희는 t2 저렴하다고 생각한다.

용은 표 4와 같다. '철수가', '의사가', '영희가', '선생님이'는 각각 type raising된 통사범주이다. '통사범주: [자절]->의미'의 형태로 구성되어 있다. 표현의 간략함을 위해 표 4에는 격자절만 나타내었지만 과생성(overgeneration)을 막기 위해 격자절 외에도 여러 가지 자질이 필요하다. 본 논문은 능력 문법을 다루고 있기 때문에 조사가 붙은 명사의 경우 어절단위로 구현해서 문법의 표현상 정확성에 초점을 맞추었으나 실용 시스템을 위한 문법을 구현하기 위해서는 명사와 조사 등이 형태소 단위로 사전에 들어있어야 한다. 이 때 각 형태소들이 서로 불필요하게 결합해서 과생성이 발생하는 경우를 방지해야 하는데 이는 자질을 사용함으로써 일부 방지할 수 있다. 예를 들어 조사의 경우에는 왼쪽에 인접한 명사나 명사구와의 결합이 먼저 일어난 뒤 인접한 통사범주들과 결합이 일어나도록 해야 하는데 조사의 자질을 명사나 명사구와 결합하기 전에는 조사의 우측에 있는 통사범주와 결합하지 않도록 정의함으로써 범위를 제한한다.

- 21) 통사를 어간과 어미로 나누어서 처리하는 것은 문장의 의미를 서술논항구조로 유도하였을 때 명시적으로 문장의 시제와 상을 표현하기 위함이다.
- 22) S₀는 시제나 상 등의 정보가 완전히 제공되지 않은 문장의 상태를 의미하고 S_n는 내포문이나 조건절의 상태를 나타낸다.
- 23) '!'은 음성자료에서는 적절한 운율부 경계로 파악될 수 있다.

표 4 사전의 내용

어 절	내 용
철수가	klex('철수가', (s: PAS) / (s: PAS \ np: [case: nom] -> 철수)).
의사가	klex('의사가', (s: PAS \ np: [case: nom] -> X) / (s: PAS \ np: [case: nom] -> X \ np: [case: cmp] -> 의사)).
영회가	klex('영회가', (s: PAS) / (s: PAS \ np: [case: nom] -> 영회)).
선생님이	klex('선생님이', (s: PAS \ np: [case: nom] -> X) / (s: PAS \ np: [case: nom] -> X \ np: [case: cmp] -> 선생님)).
되	klex('되', s: 되다(X, Y) \ np: [case: nom] -> Y \ np: [case: cmp] -> X).
있	klex('있', s: past(S) \ s: S).
다	klex('다', s: decl(S) \ s: S).

그림 6은 어순뒤섞기 현상이 일어난 문장을 구문분석하면서 의미를 유도하는 과정을 보인 것이다. 동사 '먹'에 할당된 통사범주에 논항의 자리를 바꾸어 준 새로운 통사범주를 할당함으로써 어순뒤섞기가 일어나기 전의 문장과 동일한 서술논항구조를 유도할 수 있다. 어순뒤섞기 현상을 처리하기 위해서 각 명사의 격마다 가능한 어순을 고려해 type raising된 통사범주를 미리 사전에 입력해 두었고 동사의 경우도 논항의 순서를 고려해서 사전에 입력했다. 예를 들어 주어와 목적어를 논항으로 하는 '먹다'의 경우는 사전에 '주어, 목적어'의 어순과 '목적어, 주어'의 어순을 고려해 두 개의 통사범주를 할당하였고 세 개의 논항이 필요한 '주다'의 경우는 모두 6개의 통사범주를 할당하였다. 주격 명사의 경우는 기본적인 명사에 해당하는 통사범주를 제외하고 어순의 변화에 따라 11개의 type raising된 통사범주가 할당되었다. 각 명사의 type raising된 통사범주에 따라 조사의 통사범주도 별도로 할당된다.

(56)은 [종속절 + 문장성분] 형태의 비정규문장성분이 병렬을 이루고 있는 문장이다. (56)을 구문분석하는 과정은 그림 7과 같다. '면'에 대한 통사범주가 여러 개 있는데 단순히 통사범주만으로 구문분석을 시도할 경우 과생성의 위험이 있다. 능력 문법 차원에서 제시한 본 문법을 이용하여 실제로 운용되는 시스템을 구현할 때에는 이러한 점을 포함하여 과생성을 막기 위한 기법들을 개발하고 적용하는 노력이 필요하다.

(57)은 원거리 종속관계가 나타난 접속항들이 병렬을 이루고 있는 문장이다. (57)을 구문분석하는 과정은 그

림 8과 같다.²⁴⁾

본절에서는 다양한 한국어 병렬문을 분석해 보았다. 지금까지의 분석결과를 토대로 한국어 병렬문을 분석해 내기 위해 필요한 축약규칙은 표 5와 같이 정리된다.

표 5 한국어 병렬문처리를 위한 결합범주문법

규 칙	규칙이름	축약 기호
X/Y Y → X	Forward Application	>
Y XY → X	Backward Application	<
X conj X → X	Coordination	<φ ²⁵⁾ >
X/Y Y/Z → X/Z	Forward Composition	>B
YZ X\Y → X\Z	Backward Composition	<B
X/Y YZ → XZ	Forward Crossed Composition	>B _x
X → T/(TX)	Forward Raising	>T
X → T\ (T/X)	Backward Raising	<T

3.2 병렬문의 정보구조

3.1절에서는 다양한 형태의 병렬문을 구문분석하면서 의미구조를 서술논항구조로 유도하는 과정을 설명하였다. 한국어에서 두 문장이 병렬을 이루는 주된 이유는 두 문장이 가지고 있는 사실을 나열하기보다는 간결하게 비교, 대조해서 청자에게 전달하기 위함이다. 이는 일반적으로 (61)보다는 (62)를 좀 더 자연스럽게 느끼고 많이 사용하는 것을 통해서 알 수 있다.

한국어에서 대조는 '은/는'이라는 보조사를 통해 주로 나타낸다.²⁵⁾²⁶⁾ 본 절에서는 정보구조와 보조사 '은/는'에 대한 2.6절에서의 고찰 결과를 토대로 병렬문에서 화자가 의도하는 바를 정형화해 표현한다.

(62)와 (64)는 보조사 '은/는'을 사용한 병렬문이다.

- 24) NP_{cm}은 그림 8에서 설명의 편의상 내포문의 주어를 나타내는 범주로 쓴다. NP_{cm}으로 표시된 '장미값'의 경우 실제 사전에서는 일반적인 명사의 문법범주만 가지게 되며 이에 따라 그림 8에서 다루는 문장은 구문적인 애매성을 가지고 있는 것으로 분석되나 여기서는 '장미값이 비싸다'라는 의미적으로 적절한 derivation만 보였다.
- 25) 보조사 '은/는'은 서술어를 제외한 모든 성분을 문두로 이동시켜 주제화 시킬 수 있다.
 - 어제는 철수가 설악산에 갔다.
 - 철수는 어제 설악산에 갔다.
 - 설악산에는 어제 철수가 갔다.
- 26) '은/는' 외에도 특수조사 '도', '만'이 대조의 기능을 지닌 것으로 알려져 있으나 본 논문에서는 '은/는'으로 논의의 범위를 제한한다.

철수가	사과를	영희가	딸기를	먹	있	다.
$S_u/(S_u \setminus NP_n)$: $\lambda f.f\text{원수}'$ (-, $\lambda f.f\text{원수}'$)	$(S_u \setminus NP_n)/(S_u \setminus NP_n \setminus NP_e)$: $\lambda f.f\text{사과}'$ ($\lambda f.f\text{사과}'$, -)	conj. : $\text{and}'(X, Y)$ (-, -, (-, -))	$S_u/(S_u \setminus NP_n)$: $\lambda f.f\text{딸기}'$ (-, $\lambda f.f\text{딸기}'$, -)	$(S_u \setminus NP_n)/(S_u \setminus NP_n \setminus NP_e)$: $\lambda f.f\text{먹기}'$ ($\lambda f.f\text{먹기}'$, -)	$S_u \setminus NP_n \setminus NP_e$: $\lambda x\lambda y.\text{먹다}'xy$ ($\lambda x\lambda y.\text{먹다}'xy$, -)	$S_u \setminus S_u$: $\lambda S_u.\text{past}'S_u$ $S \setminus S_u$: $\lambda S_u.\text{decl}'S_u$
B			B			
$S_u/(S_u \setminus NP_n \setminus NP_e)$: $\lambda f.f\text{사과}'\text{원수}'$ ($\lambda f.f\text{사과}'$, $\lambda f.f\text{원수}'$)			$S_u/(S_u \setminus NP_n \setminus NP_e)$: $\lambda f.f\text{딸기}'\text{먹기}'$ ($\lambda f.f\text{딸기}'$, $\lambda f.f\text{먹기}'$)		< ϕ >	
$S_u/(S_u \setminus NP_n \setminus NP_e)$: $\lambda f.\text{and}'(f\text{사과}'\text{원수}'$, $f\text{딸기}'\text{먹기}'$) ($\lambda f.f\text{사과}'$, $\lambda f.f\text{원수}'$)($\lambda f.f\text{딸기}'$, $\lambda f.f\text{먹기}'$)						
S_u : $\text{and}'(\text{먹다}'\text{사과}'\text{원수}'$, $\text{먹다}'\text{딸기}'\text{먹기}'$) ($\lambda f.f\text{사과}'$ $\lambda x\lambda y.\text{먹다}'xy$, $\lambda f.f\text{원수}'$)($\lambda f.f\text{딸기}'$ $\lambda x\lambda y.\text{먹다}'xy$, $\lambda f.f\text{먹기}'$)						
S_u : $\text{past}'\text{and}'(\text{먹다}'\text{사과}'\text{원수}'$, $\text{먹다}'\text{딸기}'\text{먹기}'$) ($\lambda f.f\text{사과}'$ $\lambda x\lambda y.\text{먹다}'xy$, $\lambda f.f\text{원수}'$)($\lambda f.f\text{딸기}'$ $\lambda x\lambda y.\text{먹다}'xy$, $\lambda f.f\text{먹기}'$)						
S : $\text{decl}'\text{past}'\text{and}'(\text{먹다}'\text{사과}'\text{원수}'$, $\text{먹다}'\text{딸기}'\text{먹기}'$) ($\lambda f.f\text{사과}'$ $\lambda x\lambda y.\text{먹다}'xy$, $\lambda f.f\text{원수}'$)($\lambda f.f\text{딸기}'$ $\lambda x\lambda y.\text{먹다}'xy$, $\lambda f.f\text{먹기}'$)						

그림 9 철수가 사과를, 영희가 딸기를 먹었다.

(62)에서 사용된 '는'은 대조의 의미를 나타낸다. (61)이 단순히 철수가 사과를 먹었고 영희가 딸기를 먹었다는 사실을 나열하는 문장인데 반해 (62)는 '는'을 사용해서 사과를 먹은 사람이 철수라는 사실과 딸기를 먹은 사람이 영희라는 사실을 대조시키고 있다. 즉 (62)에서는 철수는 사과를 먹었지만 영희는 사과를 먹지 않았고, 영희는 딸기를 먹었지만 철수는 딸기를 먹지 않았다는 가정이 성립된다. 이는 (63)이 어색한 문장임을 통해 확인할 수 있다. (64)에서 사용된 '는'은 주제와 대조의 성격을 모두 나타낸다. 즉 문장이 발화되기 이전에 철수와 영희에 대한 언급이 이미 있었다는 가정에서 나타날 수 있는 문장이면서 동시에 '철수'와 '영희'를 대조시키고 있다.²⁷⁾

- (61) 철수가 사과를, 영희가 딸기를 먹었다.
- (62) 철수는 사과를, 영희는 딸기를 먹었다.
- (63) 철수는 사과를, 영희는 사과를 먹었다.
- (64) 철수는 사과를, 영희는 딸기를 먹었다.

(65)는 철수가 먹은 것과 영희가 먹은 것을 강조하는 형태의 문장이다. (66)는 '는'이 대조의 의미를 나타내는 문장이다. 화자가 말할 때의 상황은 사과와 딸기 또는 사과와 딸기 외의 여러 가지 먹거리가 존재하고 화자는 다른 사람들이 무엇을 먹었는지 모르지만 철수가 먹은 것이 사과라는 사실과 영희가 먹은 것이 딸기라는 사실은 확신하고 있는 상태이다. (65)와 (66)의 대조의 정도를 비교했을 때, (66)의 대조의 정도가 보다 강하다. (67)에서 사용된 '는'은 주제와 함께 대조의 의미도 동시

에 가지고 있다. 즉 문장이 발화되기 전에 화자는 사과와 딸기에 대해서 알고 있었고 사과를 먹은 사람이 철수라는 사실과 딸기를 먹은 사람이 영희라는 새로운 사실을 전달하는 기능도 하는 반면 '사과'와 '딸기'를 대조시키는 역할도 한다.

주제를 나타내는 '은/는'의 경우, '은/는'과 결합되는 명사구가 문맥상 언급되었던 내용을 나타냈을 자명하다. 반면에 대조를 나타내는 '은/는'의 경우는 문맥상 언급되었는지의 여부가 불확실하다.

- (65) 사과를 철수가 먹었고, 과자를 영희가 먹었다.
- (66) 사과는 철수가 먹었고, 딸기는 영희가 먹었다.
- (67) 사과는 철수가 먹었고, 딸기는 영희가 먹었다.

(68)에서 '철수는'은 다른 사람과 대조되는 기능으로 사용되었다. 이 때 대조되는 다른 사람들로부터 '철수'라는 개념이 추론 가능하다. 이 예문을 근거로 했을 때, 대조를 나타내는 '은/는'에서 대조의 대상이 되는 것은 문맥적으로나 개념적으로 이미 언급되었던 것이나 추론 가능한 것이어야 한다. 따라서 대조의 '은/는'과 결합된 명사구도 정보구조의 주제에 해당된다고 할 수 있다.

- (68) 철수는 사과를 먹었다.

단순히 표면구조로부터 주제를 유추할 수 있는 단서가 되는 것들은 다음과 같다.

- 문장에서 사용된 '은/는'과 결합한 명사 또는 명사구는 정보구조의 주제에 해당된다.

27) '철수와 영희는 무엇을 했나?'라는 문장에 대답으로 쓰인다.

- 지시 대명사나 'two-place noun' 은 주제로 정의한다.²⁸⁾

그리고 보다 정확한 정보구조를 유도해 내기 위해서는 문장내에서 각 어절의 고저 악센트에 대한 정보가 필요하다.

3.3 정보구조의 유도

3.2절에서는 기본적인 병렬문을 구문분석하면서 의미구조를 유도하는 과정과 병렬문의 정보구조에 대해서 논의하였다. 본 절에서는 논의된 내용을 바탕으로 구문분석과 동시에 의미구조와 정보구조를 유도하는 과정을 설명하고자 한다. 3.2절에서도 논의한 바와 같이 문장내에서 각 어절의 고저 악센트 (pitch accent)는 정보구조를 유도하는데 중요한 정보가 된다. 본 논문에서는 다양한 문장에 대해 정보구조를 논의하기 위해 예문에 고저 악센트에 대한 정보가 함께 주어지고 있다고 가정한다.²⁹⁾

- (69) 영화는 **철수**가 [t_i 예쁘다고] 생각한다.
- (70) 철수가 사과를, **영희**가 딸기를 먹었다.
- (71) 철수는 사과를, **영희**는 딸기를 먹었다.
- (72) 철수는 **멋있는 티셔츠**를, **영희**는 **예쁜 모자**를 샀다.

(69)는 내포문의 주어가 내포문의 밖으로 나온 문장이다. 이때 문두로 나온 내포문의 주어 '영희'는 문장의 주제가 되며 고저 악센트를 받는 '철수'는 화자가 말하고자하는 새로운 정보 즉 평언이 된다. (69)를 구문분석하기 위해서는 >B_x 축약규칙³⁰⁾이 사용되어야 한다. 그림 9는 (70)을 구문분석하면서 의미구조와 정보구조를 유도해 내는 과정이다. 이 때 고저 악센트가 주어진 '철수가'와 '영희가'는 문장에서 평언이 되고, 고저 악센트를 받지 않는 부분은 문장에서 주제가 된다.

정보구조의 유도는 <주제,평언>이라는 템플릿을 사용해서 유도한다³¹⁾. 고저 악센트를 받은 '철수가'와 '영희

가'는 문장에서 평언이 되고, 나머지 부분은 주제가 된다. 정보구조를 가지고 있지 않는 문법범주들에는 조사, 어미 등이 있다. 그림 9의 '었다'와 같이 정보구조를 가지고 있지 않는 문법범주들은 정보구조를 유도하는데 있어서 전혀 영향을 끼치지 않는다. 그림 10은 (71)의 의미구조를 유도하는 과정이다. '철수는'과 '영희는'은 비록 병렬문에서 대조의 의미를 지니며 고저 악센트를 받고 있지만 화자나 청자가 문맥적으로나 개념적으로 알고 있는 정보이기 때문에 주제로 취급되었다.

(72)는 '티셔츠'와 '모자'에 고저 악센트를 받고 있는 문장이다. 따라서 '티셔츠'와 '모자'는 평언이 된다. 구문분석과정에서 '티셔츠'는 '멋있는'이라는 수식어의 논항으로, '모자'는 '예쁜'이라는 수식어의 논항으로 사용되는데, forward application이 적용된 뒤에 '멋있는 티셔츠'와 '예쁜 모자'는 '티셔츠'와 '모자'가 가지고 있는 정보구조를 그대로 이어 받는다. (72)를 구문분석한 뒤 유도되는 정보구조는 <λf.f철수',λf.f'멋있는'(티셔츠')λxλy.사다'(x,y)><λf.f영희',λf.f'예쁜'(모자')λxλy.사다'(x,y)>와 같다.

본 절에서는 앞에서 소개되었던 개념들과 가정들을 토대로 한국어 병렬문을 점진적으로 구문분석하면서 동시에 문장의 의미구조와 정보구조를 유도하였다. 구문분석을 통해 유도된 의미구조와 정보구조는 기계번역에서 목적언어로 문장을 생성할 때 이용될 수 있다.

표 6 병렬문의 구성방법 및 발생빈도

	신문 (동아일보, 한겨레)	수필 (농민)	초등학교 교과서	기술 문서	소설 (겨울이 야기)	합계
대동적 연결어미 (문장수 / 총문장수)	21.4% (728 / 3403)	33.7% (1092 / 3239)	20.1% (1101 / 5485)	40% (46 / 115)	44.5% (2174 / 4881)	30.0% (5142 / 17123)
접속격 조사 (문장수 / 총문장수)	19.7% (671 / 3403)	10.7% (352 / 3239)	14.8% (812 / 5485)	20% (23 / 115)	6.6% (320 / 4881)	12.7% (2178 / 17123)
접속 부사 (문장수 / 총문장수)	2.7% (91 / 3403)	0.68% (22 / 3239)	1.1% (58 / 5485)	5.2% (6 / 115)	0.35% (17 / 4881)	1% (194 / 17123)
summary (문장수 / 총문장수)	4.5% (154 / 3403)	3.9% (126 / 3239)	3.9% (216 / 5485)	18.3% (21 / 115)	0.88% (43 / 4881)	3.5% (605 / 17123)

28) 'page: one side of a leaf of something printed or written...'과 같이 사전에 정의된 명사를 말한다. 이런 명사들은 의미를 정의하는데 사용된 명사들 ('book'에 해당된다)의 언급없이 갑자기 문맥상에 새롭게 나타날 수 없다 [25].

29) 일반적으로 문장에서 고저 악센트가 주어지는 어절은 새로운 정보 즉 평언 (theme)을 나타낸다. 정보전달의 효율성을 높이기 위해 전달하고자 하는 새로운 정보를 강조하는 경향이 있다.

30) XY YZ → XZ

31) 주제는 문맥적으로나 개념적으로 이미 나왔던 정보를 의미하고, 평언은 새롭게 문장에 나타난 정보를 의미한다.

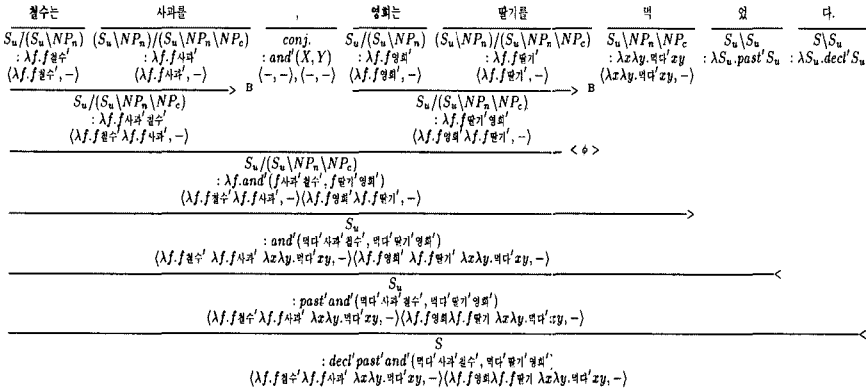


그림 10 철수는 사과를, 영희는 딸기를 먹었다

4. 분석 및 토의

3절에서는 한국어 병렬문의 통사적, 의미적, 문맥적 처리를 위한 결합범주문법을 제시하였다. 본 절에서는 한국어에서 병렬문이 차지하는 비중과 병렬문을 이루는 방법에 대해 알아보고, 3절에서 제시된 축약규칙으로 처리가능한 문형들과 처리할 수 없는 문형들에 대해서 알아본다.

한국어에서 병렬문이 차지하는 비중을 알아보기 위해 KAIST 말뭉치를 분석했다 [26]. KAIST 말뭉치의 특성을 살펴보면 신문기사가 40,428 어절, 수필 41,666 어절, 교과서 50,208 어절, 기술문서 2,729 어절, 소설 40,498 어절, 모두 175,524 어절로 구성되어 있으며 총 17,123 문장이다. 병렬을 이루는 방법을 조사한 결과는 표 6과 같다.³²⁾ 분석은 태깅된 말뭉치에서 기계적으로 조건에 부합되는 문장을 추출한 뒤, 병렬문의 여부를 수동으로 판별하였다³³⁾. 표 6을 통해 알 수 있듯이 전체 문장 중 약 47.2%의 문장에서 병렬현상이 나타났다. 이로써 양적으로도 병렬문이 한국어처리에 있어서 중요한 문제임을 알 수 있다. 병렬을 이루는 구성형태는 대등적 연결어미를 사용한 방법, 접속격조사를 사용한 방법, 삽표를 이용한 방법, 접속부사를 사용한 방법의 순으로

많이 나타났으며 말뭉치의 도메인과는 관계가 없었다.

지금까지 제시된 축약규칙으로는 한국어에서 나타나는 체언접속, 용언접속, 문장들간의 접속 등 정규문장성분들의 병렬이 설명가능하며 [종속절 + 문장성분], 논항뭉치들의 접속, 원거리 종속관계가 있는 문장 등 비정규문장성분이 병렬을 이루고 있는 문장에 대해서도 설명이 가능하며 이에 대한 결합범주문법 유도(derivation)를 구할 수 있다. (73)은 접속부사를 이용해서 명사구가 병렬을 이루고 있는 문장이고 (74)는 대등접속사인 '고'를 사용해서 동사구가 병렬을 이루고 있는 문장이다. (75)는 두 개의 문장이 병렬을 이루고 있는 문장이다. (76)은 [종속절 + 문장성분]이라는 비정규문장성분이 병렬을 이루고 있는 문장이고 (77)은 추출이 일어난 내포문의 주어 '장미값'을 공유하는 형태의 병렬문이다. (78)은 접속항내에 다시 명사구 접속이 나타난 문장이다.

- (73) 1십 판결에 대해 상당부분을 [검찰이] 그리고 [모든 피고인이] 항소하리라고 한다.
- (74) 북한은 [한반도비핵화선언의 의미를 정확하게 파악 하]고 [이를 실천해야 한다].
- (75) [사흘이 지나니 눈앞에 먹을 것만 아른거리]고 [밤이면 잠이 오지 않았다].
- (76) [고추값이 오르면 고추를], [돼지고기값이 오르면 돼지고기를] 수입한다.
- (77) 장미값이 [철수는 t_i 비싸다고 생각했고], [영희는 t_i 저렴하다고 생각했다].
- (78) [사람과 사람], [사람과 자연]이 어울려 사는 지구촌을 이루기 위해 할 일을 찾아보자.

32) 대등적 연결어미에는 '거나', '고', '든지', '며', '면서', '나', '지만' 등이 있으며 주로 용언의 접속을 이끄는 경우가 많다. 접속격 조사에는 '과/와', '하고', '및', '이나', '이거나', '이든지', '이나', 등이 있고 체언의 접속을 이끈다. 접속부사에는 '또', '또는', '그래서', '그리고', '혹은', '그러나', '그렇지만'과 같이 문장과 문장을 이어주거나 단어와 단어를 이어주는 역할을 한다.
33) 병렬문의 여부는 문장의 의미에 따라 판명되며, 품사태깅에만 의존해서는 이를 알 수 없다.

3장에서 제시한 축약규칙만으로는 해결할 수 없는 병렬문의 형태는 다음과 같다. (79)는 조건문의 주어와 제외한 부분과 주문의 문장성분이 접속항을 이루고 있는 문장이고 (80)은 원거리 종속관계를 이루는 문장들이 내포문의 추출된 주어와 모문의 주어와 접속항을 이루는 병렬문이다. 본 논문에서 제시된 축약규칙으로는 (79), (80), (81)의 접속항을 하나의 통사범주로 축약시킬 수 없기 때문에 처리가 불가능하며 현재 처리를 위한 방안이 연구 중이다.

(79) 두피가 [건조되기 쉬우면 헤어크림을], [끈적거리면 헤어토닉을] 곁하면 더욱 효과적이다.

(80) [사과는 철수가], [딸기는 길수가] t_i 맛있다고 생각한다.

(81) 장미값이 [철수는 t_i 비싸다고], [영희는 t_i 저렴하다고] 생각한다.

5. 결론

한국어 병렬문에서 발생하는 좌접속항 서술어 생략현상³⁴⁾을 처리하기 위해서 type raising과 composition 축약규칙을 사용했다. 어순뒤섞기 현상은 서술어에 할당될 수 있는 가능한 논항의 순서를 모두 고려해서 통사범주를 할당하였고, 이로 인해 어순이 바뀌어도 문장의 의미를 표시하는 서술논항구조의 각 논항의 위치는 변하지 않도록 하였다. 다양한 예문을 분석하고 결합범주문법으로 한국어에서 병렬문을 처리할 때 필요한 결합자와 축약규칙을 제안하였다.

병렬문이 단문의 연결임을 감안하면 제안된 축약규칙으로 한국어의 모든 문형을 처리할 수 있을 것으로 보인다.

구문분석과정에 정보구조의 개념을 도입하여 구문분석과정에서 동일한 의미구조를 가진 것으로 보이는 문장도 문장에 주어진 고저 악센트나 표현구조의 표지들 즉 조사, 두자리 명사, 대명사 등을 기반으로 상이한 정보구조를 유도해 낼 수 있음을 보였다.

결합범주문법을 사용함으로써 통사수준의 처리, 의미수준의 처리, 문맥수준의 처리를 여러 개의 수준으로 나누지 않고 각 어휘마다 정보를 통합하여 사전에 저장하였고 이를 이용하여 통사수준의 처리, 의미수준의 처리, 문맥수준의 처리를 동시에 점진적으로 할 수 있었다.

말뭉치 분석을 통해 병렬문이 한국어처리에서 차지하는 비중을 살펴보고 제안된 결합자와 축약규칙만으로는 해결할 수 없는 문형들에 대해 논의했다.

구문분석을 통해 유도된 의미구조는 [27]에서 논한 바와 같이 목적언어를 생성할 때 직접 사용될 수 있다. 시스템에 입력되는 원시언어 (source language)에 고저 악센트에 대한 정보가 주어진다면 구문분석과정에서 정보구조를 유도해 낼 수 있고 이를 다시 목적언어 (target language)의 생성에 이용한다면 화자가 의도한 바를 좀 더 정확히 나타내는 문장을 생성해 낼 수 있다. 현재 음성언어와 결합범주문법을 결합하는 연구가 본 연구와 병행하여 진행 중이다.

참고 문헌

- [1] Mark Steedman, "Gapping as constituent coordination," *Linguistics and Philosophy* 13 pp. 207-263, 1990.
- [2] Mark Steedman, "Surface structure and interpretation," *Linguistic Inquiry Monograph* 30, 1997.
- [3] Scott Prevost, "A semantics of contrast and information structure for specifying intonation in spoken language generation," PhD thesis, University of Pennsylvania, 1995.
- [4] Berryl Hoffman, "The computational analysis of the syntax and interpretation of free word order in Turkish," PhD thesis, University of Pennsylvania, 1995.
- [5] Jong C. Park, "Quantifier scope and constituency," *Proceedings of 33rd ACL*, pp. 205-212, 1995.
- [6] Jong C. Park, "A lexical theory of quantification in ambiguous query interpretation," PhD thesis, University of Pennsylvania, 1996.
- [7] Jong C. Park, "A unification-based semantic interpretation for coordinate constructs," *Proceedings of 30th ACL*, pp. 209-215, 1992.
- [8] Cem Bozsahin, "Deriving the predicate-argument structure for a free word order language," *Proceedings of 36th COLING-ACL*, pp. 167-173, 1988.
- [9] Nobo N. Komagata, "The computational analysis of information structure using parallel expository texts in English and Japanese," PhD thesis, University of Pennsylvania, 1999.
- [10] Jeongwon Cha, Wonll Lee, Geunbae Lee, Jong-Hyeok Lee, "Morpho-syntactic modeling of Korean with K-CCG," *Proceedings of 18th ICCPOL*, pp. 67-73, 1999.
- [11] Sung-Kook Han, Chan-Gon Park, "Combinatory categorial grammar for Korean," 제 2회 한글 및 한국어정보처리 학술발표논문집, pp. 164-171, 1990.

34) 영어에서의 gapping과 혼동되기 때문에 서술어 생략현상이라 칭하였다.

- [12] Wonll Lee, Geunbae Lee, Jong-Hyeok Lee, "Morpho-syntactic modeling of Korean with a categorial grammar," Proceedings of NLPERS, pp. 545-548, 1997.
- [13] 이진복, 박종철, "MCCG에 기반한 한국어 Quantifier Floating", 한국정보과학회 99년 봄 학술발표논문집 (B), 제 26권, 1호, pp. 330-332, 1999.
- [14] 최재희, "국어 명사구 접속의 연구", 한글 제 188호, pp. 91-115, 1985.
- [15] 김진수, "국어 접속조사와 어미연구", 탑출판사, 1987.
- [16] 김영희, "와의 양상". 국어국문학, 1974.
- [17] 이관규, "국어 대등구성 연구", pp. 53-136, 서광학술자료사, 1992.
- [18] 이관규, "대등구성의 요건과 유형", 한글 제 217호, pp. 65-81, 1992.
- [19] 김영희, "등위접속문의 통사특성", 한글 제 201호, pp. 83-117, 1988.
- [20] 서정수, "국어의 접속어미 연구(I)", 한글, 제 189호, pp. 51-86, 1985.
- [21] 서정수, "국어문법", pp. 1129-1178, 한양대학교 출판원, 1996.
- [22] 이회자, 이종희, "텍스트분석적 국어 조사의 연구", pp. 32-37. 한국문화사, 1998.
- [23] 홍사만, "국어특수조사론", pp. 113-147, 학문사, 1983.
- [24] Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein, "Centering: A Framework for Modeling the Local Coherence of Discourse," Computational Linguistics 21(2), pp. 203-225, 1995.
- [25] Mark Steedman, "The Syntactic Process," MIT Press, 1999.
- [26] 김재훈, "오류-보정 기법을 이용한 어휘 모호성 해소", 박사학위 논문, 한국과학기술원, 1996.
- [27] Jong C. Park, Hyun Ah Lee, Gil Chang Kim, "Predicate-argument structure and lexical translation," Manuscript, 1999.



박 종 철

1984년 서울대학교 컴퓨터공학과 (공학사). 1986년 서울대학교 대학원 컴퓨터공학과 (공학석사). 1996년 Univ. of Pennsylvania Computer & Info. Science (Ph.D). 현재 한국과학기술원 전자전산학과 전산학전공 조교수. 관심분야는 계산언어학, 자연언어처리, 인공지능, 인지과학.



조 형 준

1998년 숭실대학교 컴퓨터학부 (학사). 2000년 한국과학기술원 전산학과 (공학석사). 현재 SK 텔레텍 근무. 관심분야는 자연언어처리.