

한국어 특수조사 ‘도’의 의미, 문맥적 기능 분석

박준경[○] 박종철

friel@nlp.kaist.ac.kr park@nlp.kaist.ac.kr

한국과학기술원 전산학과 및 첨단정보기술 연구센터¹

An Analysis of the Semantic and Discourse Functions of the Korean Special Marker ‘-to’

June K. Park Jong C. Park

Dept. of Computer Science and Advanced Information Technology Research Center(AITrc)
Korea Advanced Institute of Science and Technology

요 약

본 논문은 한국어의 특수조사, 특히 ‘도’의 의미, 문맥적 기능에 대하여 다루고 있다. ‘도’는 문맥의 자연스러운 연결에 있어서 중요한 역할을 수행한다. ‘도’가 쓰인 문장의 배경에는 반드시 일정한 전제가 존재한다. 전제는 그 문장의 의미 뿐만 아니라 기존 문맥과도 직접적으로 연관된다. 본 논문에서는 ‘같은’, ‘유사함’, ‘극한’, ‘첨가’ 및 병렬문에서 쓰이는 다섯 가지 ‘도’의 기능에 대하여 설명하고, alternatives semantics를 이용하여 이를 결합범주문법(CCG)에서 구현하는 방법을 제시한다.

1 서론

한국어와 같은 교착어에서는 조사가 가지는 기능이 굴절어에서의 어순과 같은 정도이거나 그 이상의 중요도를 가지며 이에 따라 조사에 대한 형태소 분석이 특히 많이 이루어져 있다. 한국어에서 사용되고 있는 조사는 격표지를 나타내는 격조사와 특수한 의미 기능을 수행하는 특수조사의 두 가지로 크게 나눌 수 있는데, 본 연구에서 ‘도’, ‘만’, ‘조차’, ‘은/는’, ‘까지’ 등 여러 특수조사 중에서 특히 ‘도’에 주목하게 된 이유는 이것이 부정극어와 가지는 관계에서 출발한다.² 부정극어(Negative Polarity Item, NPI)는 부정어와 함께 쓰여야 하는 특성을 가지는 어휘로 부정어와 부정극어간의 관계는 한국어가 가지는 언어적 표현력(linguistic expressiveness)의 정도를 가늠할 수 있게 한다[6, 7, 9, 10]. 이러한 한국어의 부정극어에는 특히 ‘도’와 관련된 것이 많은데, 예를 들어 ‘조금도’, ‘하나도’, ‘아무도’와 같이 극소량을 지칭하는 단어에 ‘도’가 연결되어 전체적으로 부정극어 역할을 하는 것 등이 있다.

자연언어처리 분야에서 이와 같은 ‘도’의 기능을 정확히 이해하고 사용하는 것이 중요한 이유는 문장에서의 ‘도’의 사용 여부가 문맥의 자연스러운 연결에 결정적인 역할을 한다는 데 있다. 다음의 예문을 보자.³

(1) (a) 철수는 베오올프를 샀다.

- (b) ?영희는 베오올프를 샀다.
- (c) 영희도 베오올프를 샀다.

여기서 (1b)의 경우는 (1a)가 앞에 있을 경우 문맥상 부자연스럽고, (1c)를 써야 자연스럽게 된다. 본 논문에서는 이를 포함한 여러 가지 ‘도’의 기능에 대한 분류 및 분석을 보인다. 이렇게 ‘도’와 같은 특수조사의 기능을 정확히 파악하는 것은 자연언어 생성에서나 파싱에서 공통적으로 중요한 문제이다.

‘도’에 대한 국어학적 연구 중 대표적인 것으로 홍사만[4]이 있는데 이에서는 ‘도’의 기능을 동류 지시, 극단 지시, 극단 부정 표시, 양보/허용 표시, 강조적 첨의 등 5가지로 나누고 있다. 동류 지시는 가장 기본적인 용법으로 ‘역시’, ‘마찬가지로’와 같이 동일하거나 유사한 성격을 지님을 나타내는 것이다. 이는 영어의 also, too, again이 가지는 용법과 같다. 극단 지시에서는 다른 특수조사, 특히 ‘은/는’과의 대조적인 성격이 드러나며 이는 척도(scale)의 개념으로 설명된다[1, 2, 4]. ‘도’에서의 척도는 ‘마지막-까지도’의 의미를 갖는 것으로 어떤 집합에서 모든 원소가 어떤 성질을 만족하고 그 극단에 존재하는 마지막 점까지도 그 성질을 만족한다는 것이다. ‘도’와 ‘은/는’은 이러한 척도의 개념을 포함할 수가 있으며 이 경우 서로 반대 방향의 척도를 갖게 된다. 즉, ‘도’는 ‘심지어 A조차도(even A)’, ‘은/는’은 ‘적어도 A만큼은(at least A)’의 의미를 각각 내포하는 것이다. 그러나 척도 개념이 들어가지 않은 상태에서도 ‘도’와 ‘은/는’은 서로 상반되는 성질을 나타낸다. ‘도’가 다른 것들도 P한 상태에서 ‘A 역시 P’의 의미를 가질 때 ‘은/는’은 다른 것들은 P한 상태에서 ‘A는 Q’라는 의미를 나타내기 때문이다. 이러한 성

¹본 연구는 첨단정보기술 연구센터를 통하여 과학재단의 지원을 받았음.

²특수조사는 보조사, 도움토씨 등의 다른 이름으로 불리기도 한다.

³베오올프는 8세기 경에 쓰여진 저자 미상의 영웅 서사시이다. 고대 영어로 쓰여 있으며, 시대적 배경은 6세기이다.

질에 관해서는 많은 연구 결과가 있으며 한국어뿐만 아니라 영어의 even, only 및 일본어의 'は(wa)', 'も(mo)' 등에 대한 연구도 많이 이루어져 있다[5, 8, 11, 14, 16].

이밖에 극단 부정 표시는 부정극어에 관련된 것이며, 양보/허용은 의미상 '-라도'와 같이 쓰이지만 앞의 기본적 용법과 크게 다르지 않고, 강조는 연결어미 및 부사 등에 연결되어 쓰이는 등 형태 자체가 독특한 양상을 띤다. 본 논문에서는 논의의 범위를 제한하기 위하여 이 세 가지에 대해서는 별도로 다루지 않기로 한다.

'도'에는 중요한 특징이 있다. '역시'의 의미를 가지기 때문에 반드시 그 뒤에는 같은 성질을 가지는 어떤 것, 즉 '전제(presupposition)'가 존재하고 있는 것이다. '전제'의 어떤 부분이 '도'가 보조하는 말과 같은 성질인가를 결정하는 것은 간단하지 않은 문제이다. 기본적으로 '전제'는 도의 보조를 받는 말과 완전히 같은 성격을 가질 수도 있는데, 이와 같은 기본적인 용법 이외에도 그 성격이 유사한 경우에도 자연스럽게 쓰이고 있는 경우가 있다. 이렇게 자연스러운 유사함의 정도가 어디까지인가를 계산적으로 정확히 판단하는 것은 컴퓨터를 통한 자연언어이해 및 생성에 있어서 매우 중요한 문제이다.

'도'는 명사, 부사 및 동사의 특정 형태까지도 보조할 수 있는 조사이지만, 본문에서는 논의의 영역을 제한하여 명사, 그중 특히 고유명사에 붙는 경우만을 다룬다.

2 관련 연구

Komagata[12]는 일본어의 조사 'は(wa)', 'が(ga)', 'を(o, wo)' 및 'に(ni)' 등에 대하여 논의하며, 특히 'wa'와 'ga'의 선택에 중점을 두고 있다.⁴ 이 연구는 조사의 성질을 계산적으로 밝혀 문맥상에서 적절한 조사를 선택할 수 있는 방법을 제시하기 위한 것으로 Table 1에서와 같이 조사의 성질을 정의하였으며, 이에 따라 Table 2에서와 같이 일본어에서의 조사 선택 방법을 제시하였다.⁵ Table 2는 선언적으로 기술되었으나 실제로는 순차적으로 실행된다. 여기에서 정보 구조(Information Structure)는 의미 구조 이외에 문맥에서 필요한 모든 정보를 포함하는 구조이다[3, 12, 15].

Komagata에 따르면 'wa'는 대조의 기능⁶을 가지고 있고, thematic 기능⁷은 long-distance fronting⁸의 결과로만 나타

⁴'は(wa)'는 한국어의 '은/는', '가(ga)'는 한국어의 '이/가', 'を(o, wo)'는 우리말의 '을/를', 'に(ni)'는 한국어의 '에' 또는 '에서'에 해당하는 조사이다.

⁵Table 1에서 Prominent는 일반적으로 고저 액센트(pitch accent)를 받았다는 것을 뜻한다.

⁶이하의 예문에서 제시한 것과 같이, 이는 '다른 무엇은 아니지만 그것 은'을 뜻한다.

⁷기은 문맥에서 주어진 것을 다시 언급할 때의 기능이다. 주제화 보조사(topic marker)로서의 역할이라 할 수 있다.

⁸대표적인 것으로 주제화(Topicalization)을 들 수 있다. 어절이 clause 밖으로까지 이동하는 현상이다. 다음의 예문을 보자.

● Banana_i-wa Naomi-ga [Erika-ga t_i tabeta-to] omotta.
banana-TOP Naomi-NOM Erika-NOM ate-COMP thought.

나기 때문에 matrix 또는 fronted position에만 쓰일 수 있고 내포문에는 쓰일 수 없다. 또한 embedded clause에서 쓰인 wa는 모두 대조 기능의 wa가 된다. 이하의 예문은 여기서 논의한 대조의 기능과 thematic 기능에 대한 예문들이다. (2)의 예문에는 thematic 기능, (3)에는 대조의 기능, (4)에는 이 두 가지 기능이 모두 나타나 있으며, (5)는 하나의 'wa'에 두 가지 기능이 공존하는 예이다.

- (2) (a) The person who came here was Ken.
(b) Ken-wa banana-o tabeta.
Ken-TOP banana-ACC ate.⁹
켄은 바나나를 먹었다.
- (3) (a) Among those people, who ate bananas?
(b) Ken-wa banana-o tabeta.
Ken-CNT banana-ACC ate.
켄은 바나나를 먹었다.
(다른 사람은 바나나를 먹지 않았다.)
- (4) (a) Did Ken eat something?
(b) Ken-wa banana-wa tabeta.
Ken-TOP banana-CNT ate.
켄은 바나나는 먹었다.
(Ken은 바나나 이외의 것은 먹지 않았다.)
- (5) (a) What did these people eat?
(b) Ken-wa banana-o tabeta.
Ken-TOP/CONT banana-ACC ate.
켄은 바나나를 먹었다.
(다른 사람은 바나나를 먹지 않았다.)

예문에서와 같이 일본어에 대한 조사의 예측은 한국어에서의 조사 예측에도 적용될 수 있다는 것을 알 수 있다. 그러나 이러한 방법에는 아직 문제점이 있다. 다음의 예문을 보자.

- (6) 대부분의 아이들은 바나나를 좋아하지만, 어떤 아이들() 사과를 좋아한다.

여기서 괄호 안에 들어갈 조사를 선택하면, Komagata의 방법으로는 '이'(일본어의 'ga')가 잘못 선택된다. 대부분의 아이들과 어떤 아이들이 대조된다는 것을 찾아낼 수 없기 때문이다.

Han[11]은 한국어의 '은/는'이 가지는 세 가지 기능을 설명하면서, 이를 하나의 '은/는'으로 볼 것인가, 세 가지 다른 '은/는'으로 볼 것인가에 대하여 논의하고 있다.

바나나는 나옴이가 에리카가 먹었다고 생각했다.

여기서 'banana-wa'라는 어절은 내포문의 목적어이지만, long-distance fronting에 의해 문장의 맨 앞으로 이동하였다. 한편 clause 내에서 이동하는 것은 local scrambling이라 한다.

⁹TOP은 topic, ACC은 accusative, CNT는 contrastive marker를 각각 지칭한다.

		Information Structure	
		Matric Clause	Embedded Clause
wa(adverbial particle)	Prominent	Theme or Rheme	Unspecified
	Non-Prominent	Theme	Not Available
ga(case particle)	Prominent	Rheme	Unspecified
	Non-Prominent	Rheme	Unspecified
o, ni(case particle)	Prominent	Rheme	Unspecified
	Non-Prominent	Theme or Rheme	Unspecified

Table 1: 일본어에서의 정보 구조

	Condition	Predict	
Embedded case	Strong contrastiveness is required	wa	
	Otherwise	ga	
Matrix case:	Parallel clause (subject contrast)	wa	
	Negative construction (one-place predicate)	wa	
	One-place stage-level predicate	ga	
	Otherwise	Theme	wa
		Rheme	ga

Table 2: 일본어에서의 조사 예측

(7) 존은 메리를 좋아한다.

(8) 존은 메리를 좋아한다.

(프랭크는 수잔을, 피터는 로라를 좋아한다.)

전제: $\exists x(x \neq \text{존})$

(9) 존이 메리는 좋아한다.

전제: $\exists x(x \neq \text{메리})$

implicature: $\forall x(x \neq \text{메리}) \rightarrow \neg \text{like}(\text{존}, x)$

(7)에서 보이는 ‘은/는’의 기능이 topic reading, (8)에서 보인 기능이 contrastive topic reading, (9)에서 보인 기능이 contrastive focus reading이다.

Han은 ‘은/는’이 전제를 제시하며, S-structure에서 VP-internal인 ‘은/는’이 contrastive focus reading, VP-external인 ‘은/는’이 topic 또는 contrastive topic reading임을 밝힘으로써 여기에서는 ‘은/는’이 세 가지 기능을 가지지만 세 개가 아닌 하나로 설명한다. 여기에서 ‘은/는’의 주요한 역할은 topic marker로서의 역할이 아니라 관련된 전제를 제시해주는 것이라는 점이 본 연구에 시사하는 바가 크다.

그러나 Han이 보인 전제에는 x에 대하여 별다른 제약이 없다는 문제점이 있다. 이 전제를 보다 구체화하면 다음과 같이 보일 수 있을 것이다.

(10) 존은 메리를 좋아한다.

(프랭크는 수잔을, 피터는 로라를 좋아한다.)

전제: $\exists x(x \neq \text{존}), (\text{like}(\text{존}, \text{메리}) = P(\text{존})) \wedge (Q \sqsubseteq P) \wedge Q(x)$.

여기서의 P는 like(메리), 즉 P(존)이라는 것은 (like(메리))(존)이 되며, $Q \sqsubseteq P$ 는 Q가 P의 non-empty restriction이라는 것으로 이러한 Q의 값에는 $\lambda x, y[\text{like}(x, y)]$, $\lambda x[\text{like}(x, \text{메리})]$ 뿐만 아니라 $\lambda f, x[f(x, \text{메리})]$ 도 포함한다.

Steedman[14]은 even과 only라는 focusing particle의 기능을 PSET이라는 함수를 이용하여 설명하였다.¹⁰ PSET은 open proposition을 인자로 받으며, 문맥상 가능한 그 open proposition의 모든 sentence의 집합을 나타낸다.¹¹

(11) (a) Freeman introduced even HARDY to Willis.

(b) Freeman introduced only HARDY to WILLIS.

이 경우 (11a), (11b) 둘 다 그 배경에는 전제가 되는 것이 있게 된다. 즉, (11a)의 even을 끌어낸 전제와 (11b)의 only를 끌어낸 전제가 있다는 것이다. 이것을 기호화하여 쓰면 다음과 같다.

(12) (a) $\forall p \in \text{PSET}(\lambda x[\text{Freeman introduced } x \text{ to Willis}]), p$.

(b) $\forall p \in \text{PSET}(\lambda x, y[\text{Freeman introduced } x \text{ to } y]),$

$\neg(p = \text{Freeman introduced Hardy to Willis}) \Rightarrow \neg p$.

(12a)의 경우 프리먼이 윌리스에게 소개하였을 가능성이 있는 사람을 하나씩 x 자리에 채워넣어(instantiate) 만들어진 문장들이 각각 p라고 하였을 때 이 p가 모두 참이라는 것이다.

¹⁰대문자로 쓰인 WILLIS와 HARDY는 강세가 들어가는 단어이지만 본 논문에서는 이러한 억양과 관련된 내용은 논외로 한다.

¹¹이는 Rooth[13]가 제안한 ‘alternatives semantics’에 따른다.

즉 even의 사용은 프리먼이 윌리스에게 소개했을 가능성이 있는 사람들은 모두 실제로 소개했다는 것을 전제로 한다. 한편 (12b)의 경우에는 ‘프리먼이 누군가에게 누군가를 소개하였다’에서 이 양쪽의 누군가(x,y)에 모두 가능성이 있는 사람들을 채워넣어서 만들어진 문장들이 p라고 할 때 원문의 경우를 제외하고는 모두 거짓이라는 것이다. 즉 ‘하디를 윌리스에게’가 아닌 경우 모두 거짓이라는 것이므로, 프리먼은 하디를 윌리스에게 소개했을 뿐 그 이외의 사람을 다른 사람에게 소개하지 않았다는 뜻이 된다.

한국어의 ‘은/는’ 및 일본어의 ‘は(wa)’에 대해서는 연구가 많이 이루어졌으며 현재 상당히 체계화된 단계에 있는 것으로 보인다. 그러나 ‘도’에 대해서는 아직 어디에서 어떤 의미로 쓰이는가 하는 국어학적 연구 이상은 활발히 연구가 진행되고 있지 않은 상태이다.

3 관련 현상 연구

- (13) 철수는 베오울프를 샀다.
영희는 무엇을 샀지?
영희도 베오울프를 샀다.

이것은 조사 ‘도’가 쓰이는 가장 기본적인 형태라고 할 수 있다. 앞서 오는 문장과 뒤에 오는 문장 사이에서 ‘도’가 보조하는 주어틀 제외한 다른 부분이 모두 일치하고 있다. 이 경우의 ‘도’에서는 가장 기본적인 ‘역시’라는 의미를 찾아볼 수 있다. 일본어의 ‘も’ 역시 같은 형태로 쓰인다.¹²

- (14) 철수는 베오울프를 샀다.
영희도 오디세이를 샀다.
돌이는 무엇을 샀지?
돌이도 일리아드를 샀다.

여기에서 보이는 것은 ‘사다’라는 동사의 일치와 목적어 대상의 유사성이다. 먼저 중간에 나온 ‘영희’를 보조하는 ‘도’가 목적어 사이에 존재할 수 있는 유사성을 제시해준다. 그리고 그 뒤에 다시 ‘도’가 나오으로써 다시금 그 유사성의 범위를 좁힐 수 있는 것이다. 이와 같이 ‘도’가 반복되어 사용될 수록 이러한 유사성의 범위는 점점 더 좁혀지는 것으로 보인다.

¹²다음은 일본어에서의 예이다.

- (1) Chulsuは Beowulfを買った.
Chulsu-wa Beowulf-wo katta.
철수는 베오울프를 샀다.
(2) Youngheeは 何を買ったのか?
Younghee-wa nani-wo katta-no-ka?
영희는 무엇을 샀지?
(3) Youngheeも Beowulfを買った.
Younghee-mo Beowulf-wo katta.
영희도 베오울프를 샀다.

- (15) 철수, 영희, 돌이는 모두 베오울프를 좋아한다.
철수는 베오울프를 샀다.
영희도 베오울프를 읽었다.
돌이도 베오울프를 암송했다.

이 경우 ‘베오울프’라는 공통의 대상에 대하여 서로 다른 행동을 하는 것으로, (14)의 경우와 마찬가지로 유사성이 존재하는 곳에 ‘도’가 쓰이고 있다.

- (16) 철수는 고전을 좋아한다.
영희도 고전을 좋아한다.
또 누군가 고전을 좋아하는 사람이 있니?
순이도 고전을 좋아한다.

이 경우 (13)과 가장 큰 차이점은 ‘순이’가 기존에 제시되지 않았다는 점이다. 앞서 질문의 ‘또 누군가’라는 것이 이런 대답을 이끌어낸다. 여기서의 ‘도’는 동일성/유사성보다 도리어 ‘첨가’의 의미가 강하다고 할 수 있다. 여기서의 동일성/유사성은 질문에서의 ‘또’라는 말이 이미 시사해주고 있기 때문이다.

- (17) 영희, 철수, 돌이는 책을 사서 읽어보는 것을 좋아한다.
특히 영희는 누구보다도 책을 사는 것을 좋아한다.
하지만 철수와 돌이는 베오울프를 읽는 것조차 싫어했다.
영희도 베오울프는 사지 않았다.

여기에는 ‘심지어’의 sense가 포함되어 있다. ‘책을 사서 보는 것을 매우 좋아하는 영희조차도 베오울프만큼은 사지 않았다(읽었는지 어떤지는 모름)’는 사실에 의해 ‘(생각할 수 있는 범위에 있는) 어느 누구도 베오울프는 사지 않았다’는 결론을 얻을 수 있다.

- (18) 영희도 장미를 샀고, 돌이도 수선화를 샀다.

병렬문이 아닐 경우 이런 형태는 나타나지 않는 것으로 보이는데, 성격적으로는 (14)의 예문과 유사하지만, 앞쪽에 ‘도’가 나오기 위해서는 병렬문의 형태여야 한다. 병렬문이 아닐 경우에는 (14)에서처럼 먼저 나오는 ‘도’를 이끌어내기 위한 문장이 그 앞에 있지 않은 경우 부자연스럽게 된다.

주어 이외에 목적어 등에 붙는 ‘도’ 역시 유사한 성격을 지닌다. 아래의 예문들이 그와 같은 경우이다.

- (19) 철수는 베오울프를 샀다.
철수는 또 무엇을 샀지?
철수는 오디세이도 샀다.

- (20) 올해 교재는 오디세이이다.
베오울프는 참고도서이다.
철수는 학생이다.
학생들은 모두 오디세이를 샀다.
철수는 베오울프도 샀다.

위의 예문들 중에는 질의/응답 형식으로 된 것과 그냥 문장의 나열 형식으로 된 것이 있는데 이들은 각자 그 성격이 약간 다르나, 본 논문에서는 더 자세한 논의는 생략한다.

이상에서 보인 ‘도’의 기능을 정리해보면 다음과 같다.

- ‘같은’을 나타내는 ‘도’
- ‘유사함’을 나타내는 ‘도’
- ‘십지어 -까지도’의 극한을 나타내는 ‘도’
- ‘첨가’의 ‘도’
- 병렬문에서의 ‘도’

다음 절에서는 이러한 기능을 구별하여 인식할 수 있는 결합 범주문법을 구현하는 방안에 대하여 논의한다.

4 결합범주문법을 통한 기능 해석

4.1 결합범주문법

결합범주문법(Combinatory Categorical Grammar, CCG)은 Application Rule만 존재하는 범주문법(Categorical Grammar)에 결합자(Combinator)에 해당하는 규칙(Rule)들을 추가한 것이다[15]. 한국어에 대하여 제안된 기본적인 규칙은 Table 3에 보인 것과 같다[3]. 자연스러운 고차원 표현을 위하여 변형된 λ -Calculus 표현을 사용하는데 이의 구현은 Terzo Prolog와 같은 Lambda Prolog를 이용한다. 이에 따르면 “철수가 책을 읽었다”라는 문장은 다음 Figure 1에서와 같이 분석된다.

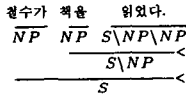


Figure 1: 철수가 책을 읽었다.

명사는 각자 property list를 가지고 있다고 볼 수 있다. 이 리스트는 WordNet에서와 같이 ‘책’이 ‘고문(古文)’이나 ‘서사시’보다 상위에, ‘고문(古文)’과 ‘서사시’는 같은 단계의 다른 기준하에 존재하는 것으로 보는 것이 타당하지만, 여기에서는 논의를 단순화하기 위하여 1차원의 리스트로 표현한다. 그리고 보다 구체적인 논의를 위하여 오디세이의 property list로 [책, 고문(古文), 서사시, 호메로스 저, 그리스문], 베오울프의 property list로 [책, 고문(古文), 서사시, 고대 영문], 일리아드의 property list로 [책, 고문(古文), 서사시, 호메로스 저, 그리스문]이 있다고 가정한다. 또 이를 위하여 prop(X)라는 임의의 명사 X에 대하여 property list를 돌려주는 함수를 정의한다.

한국어를 위한 결합범주문법에서 명사의 구문범주를 나타내는 방법에는 다음 두 가지가 있을 수 있다.¹³

- (21) $cat(\text{‘철수는’}, NP)$.
- (22) $cat(\text{‘철수’}, NP)$.
- $cat(\text{‘는’}, NP \backslash NP)$.

¹³다른 가능성에 대한 논의는 생략한다.

(21)는 어휘 사전(lexicon)에 조사와 연결된 형태로 모두 넣어 두는 것이고, (22)는 어휘 사전에는 조사와 명사를 따로 넣어두고 나중에 파싱 단계에서 결합하는 것이다. (21)의 경우는 처리는 간단하지만 어휘 사전의 entry가 많아지고, 병렬문의 자연스러운 처리에도 어려움이 따른다[3]. (22)의 경우는 어휘 사전이 간단해지는 대신에, 조사의 허용 가능한 sequence를 정확히 지정해주어야 하는 문제가 추가로 있고, 이를 위하여 모음 조화 및 자음접변등도 고려해야 하므로 결과적으로 복잡한 구현 기법을 필요로 한다. 또한 형태소 분석 단계를 별도로 거쳐야 하는 번거로움도 있다. 그러나 결과적으로 이 두 가지 방법은 동일한 결과를 나타내며, 여기에서는 설명의 편의상 (21)의 방법을 채택하여 논의한다.

명사 구문범주는 $NP[case:X, mark:Y]$ 형태로 나타낼 수 있다.¹⁴ 여기서 case는 NOM(nominative), ACC(accusative), mark는 ‘도’의 경우 TOO, ‘은/는’의 경우 TOP 또는 CNT 등이 된다. ‘도’와 같은 경우 case는 알려져 있지 않으므로 Prolog 변수 ‘_’를 할당하고, mark만 TOO의 값을 가져도록 한다. 반면 ‘을/를’은 mark에 NONE의 값을 가지게 하고 case에만 ACC의 값을 설정한다.

타동사는 $S \backslash NP \backslash NP$ 의 구문범주를 가진다[3]. 여기서 위의 feature가 들어간 NP를 적용하면 어순 뒤섞기 현상을 고려한 타동사의 구문범주는 $S \backslash NP[case:NOM, mark:_] \backslash NP[case:ACC, mark:_]$ 또는 $S \backslash NP[case:ACC, mark:_] \backslash NP[case:NOM, mark:_]$ 의 두 가지가 된다. ‘도’나 ‘은/는’과 같은 경우 [case:_]이므로 이것과 단일화(unification)가 가능하여 성공적인 파싱 결과를 낸다.¹⁵

4.2 ‘도’의 기능 해석

(13)의 예문을 다시 보도록 하자.

- (13) (a) 철수는 베오울프를 샀다.
 (b) 영희는 무엇을 샀?
 (c) 영희도 베오울프를 샀다.

전제가 없는 (a)의 형태는 Figure 2와 같이 된다. 이를 문맥으로 받는 (c) 문장의 경우 Figure 3과 같이 된다.

전제의 정확한 표현을 위하여 Steedman의 open proposition 대신에 다음과 같은 제한된(constrained) 형태를 정의한다.

$$(23) \lambda x[P(x):-Q].$$

이는 P(x)라는 open proposition이 항상 Q의 조건을 만족해야 함을 뜻하며 여기에서 ‘:-’는 Prolog의 operator와는 다른 기능을 가진다.

¹⁴mark는 그 NP가 가지는 특수한 성질을 나타내며, 일반적으로는 NONE이고, ‘도’나 ‘은/는’, ‘만’ 등의 특수조사를 통한 경우에만 그 값을 가진다.

¹⁵이하의 CCG derivation에서는 NP의 이러한 feature list는 생략하고 보인다.

X/Y	Y	\rightarrow	X	Forward Application ($>$)	
Y	$X \backslash Y$	\rightarrow	X	Backward Application ($<$)	
X/Y	Y/Z	\rightarrow	X/Z	Forward Composition ($> B$)	
$Y \backslash Z$	$X \backslash Y$	\rightarrow	$X \backslash Z$	Backward Composition ($< B$)	
X		\rightarrow	$T/(T \backslash X)$	Forward Raising ($> T$)	
X		\rightarrow	$T \backslash (T/X)$	Backward Raising ($< T$)	
X	<i>conj</i>	X	\rightarrow	X	Coordination ($< \phi^n >$)
X/Y	$Y \backslash Z$	\rightarrow	$X \backslash Z$	Forward Crossed Composition ($> B_x$)	

Table 3: 한국어를 위한 결합범주분법 규칙

$$\frac{\begin{array}{c} \text{철수는} \quad \text{베오울프를} \quad \text{샀다.} \\ NP : \text{철수} \quad NP : \text{베오울프} \quad (S \backslash NP \backslash NP) : \lambda x \lambda y. \text{past}(\text{buy}(x, y)) \\ \hline (S \backslash NP) : \lambda x. \text{past}(\text{buy}(x, \text{베오울프})) \\ \hline S : \text{past}(\text{buy}(\text{철수}, \text{베오울프})) \end{array}}{<}$$

Figure 2: 철수는 베오울프를 샀다.

$$\frac{\begin{array}{c} \text{영희도} \quad \text{베오울프를} \quad \text{샀다.} \\ NP : \text{영희} \quad NP : \text{베오울프} \quad (S \backslash NP \backslash NP) : \lambda x \lambda y. \text{past}(\text{buy}(x, y)) \\ \hline (S \backslash NP) : \lambda x. \text{past}(\text{buy}(x, \text{베오울프})) \\ \hline S : \text{past}(\text{buy}(\text{영희}, \text{베오울프})) \end{array}}{<}$$

Figure 3: 영희도 베오울프를 샀다.

(c)만을 보고 생각할 수 있는 전제는 다음과 같이 표현된다.¹⁶

$$(24) \exists p \in \text{PSET}(\lambda x [\text{buy}(x, \text{베오울프}) :- x \neq \text{영희}]), p.$$

$$(25) \exists p \in \text{PSET}(\lambda x \lambda y [\text{buy}(x, y) :- x \neq \text{영희} \wedge \text{sim}(\text{prop}(y), \text{prop}(\text{베오울프}))]), p.$$

$\text{sim}(\text{prop}(X), \text{prop}(Y))$ 는 대상의 유사성을 확인하는 것으로 $\text{prop}(X)$ 와 $\text{prop}(Y)$ 의 교집합이 공집합이 아닌 경우 참값을 돌려준다.

위의 두 전제 중 어느 쪽에 해당하는가 하는 것 역시 기존의 문맥을 살펴봄으로써 알 수 있다. 여기서 이전 문맥에는 ‘buy(철수, 베오울프)’가 포함된다.¹⁷ 이에 따라 여기서의 전제는 (24)에 해당함을 알 수 있다.

(14)에서의 전제는 앞서의 (25)에 해당한다. 그러면 대상의 유사성에 대하여 생각해볼 수도 하자.

‘buy(영희, 오디세이)’의 기존 문맥에는 ‘buy(철수, 일리아드)’가 포함된다. 여기서 일리아드와 오디세이의 성격이 유사하다는 것을 알아낼 수 있다. 첫 번째와 두 번째 문장에서 여기

¹⁶ 동사에 대해서는 그 행위자나 대상물의 성격 등이 동사에 따라 결정되며 이것 역시 각각 하나의 조건이 되지만 여기서는 따로 논의하지 않고 시제에 대한 논의 역시 여기에서는 생략하도록 한다.

¹⁷ 본 논문에서는 문맥을 정보 구조의 sequence로 생각하기로 한다. 정보 구조에는 Syntax(구문범주), Semantics 및 전제(Presupposition)가 모두 포함된다. 문장마다 생성된 이러한 IS를 계속 덧붙여가면 한없이 늘어날 수도 있지만, 주의를 환기시키는 접속사나 문단의 변경 등에 따라서 제한을 두게 되면 적절한 수준에서 유지할 수 있다. 문맥은 그 안에서 통일성이 유지되어야 하며, 이러한 여러 가지 처리에 대하여 현재 연구가 진행되고 있다.

서 이야기하는 property가 [책, 고문(古文), 서사시, 호메로스저, 그리스문] 중 하나 또는 복수의 property임을 알 수 있다. 여기에 마지막 문장이 더해짐으로써 베오울프 역시 유사한 성격을 가짐을 알 수 있고, 여기서 말하고 있는 대상의 유사성은 [책, 고문(古文), 서사시]로 줄여진다.

(15)에서는 명사가 아닌 동사가 변화하게 되며 이는 다음의 전제로 표현된다.

$$(26) \exists p \in \text{PSET}(\lambda f \lambda x [f(x, \text{베오울프}) :- x \neq \text{영희}]), p.$$

f 는 베오울프라는 것에 적용될 수 있는 모든 동사에 대한 predicate가 되며 문맥의존적으로 결정된다.

(24), (25), (26)의 전제를 함께 생각해 보자.

$$(27) \exists p \in \text{PSET}(\lambda x [\text{buy}(x, \text{베오울프}) :- x \neq \text{영희}]), p.$$

$$(28) \exists p \in \text{PSET}(\lambda x \lambda y [\text{buy}(x, y) :- x \neq \text{영희} \wedge \text{sim}(\text{prop}(y), \text{prop}(\text{베오울프}))]), p.$$

$$(29) \exists p \in \text{PSET}(\lambda f \lambda x [f(x, \text{베오울프}) :- x \neq \text{영희}]), p.$$

이 세 가지 전제에 앞서 Han의 연구에서의 제약 조건을 구체화하며 사용했던 방법을 적용하면, 다음과 같은 새로운 전제를 얻어낼 수 있다.

$$(30) \exists p \in \text{PSET}(\lambda x \lambda Q [Q(x) :- x \neq \text{영희} \wedge P(\text{영희}) = \text{buy}(\text{영희}, \text{베오울프}) \wedge \text{resem}(Q, P)]), p.$$

$\text{resem}(Q, P)$ 는 Q 와 P 가 ‘유사하다’는 것에 해당하는 조건으로, $P=f(x)$, $Q=g(y)$ 라 할 때, 다음과 같은 경우 참값을 돌려준다.

다.¹⁸

(31) $Q \sqsubseteq P$, $\text{sim}(\text{prop}(x), \text{prop}(y))$, f 와 g 는 모두 x 와 y 에 적용될 수 있는 동사에 해당하는 predicate

(30)의 전제를 살펴보면 ‘buy(영화,베오울프)’는 문장 전체의 semantics임을 알 수 있다. 따라서 이는 다음과 같이 표현할 수 있다.

(32) $\exists p \in \text{PSET}(\lambda x \lambda Q [Q(x) :- x \neq \text{영화}] \wedge P(\text{영화}) = \text{Sem} \wedge \text{resem}(Q, P))$, p .

여기서 Sem은 그 문장의 semantics를 나타낸다.

지금까지 CAT:SEM 형식으로 나타내고 있는 것을 CAT:[SEM, PRE] 형식으로 나타내기로 한다.¹⁹ PRE는 CAT에서 mark:NONE일 경우 [] (empty list)이다. mark가 TOO인 경우, 즉 ‘도’가 쓰인 경우, PRE는 (32)의 형태가 된다. 결합 과정에서, 이러한 PRE의 list를 덧붙이는 것으로 하자. (13)의 (c)에서 ‘영회도’를 제외한 나머지는 PRE가 []이다. 따라서 분석이 끝난 후 PRE에는 (32)가 남아있게 되고, 여기서 Sem의 값을 알 수 있게 되었으므로 (30)의 형태가 된다. 여기서 기존의 문맥을 살펴봄으로써 $\text{resem}(Q, P)$ 의 Q를 알아낼 수 있다. 예를 들어, 이전 문장의 semantics가 여기에서는 ‘buy(철수, 베오울프)’이므로, Q는 ‘ $\lambda x[\text{buy}(x, \text{베오울프})]$ ’가 된다.

병렬문의 경우, 양쪽에서 각각 PRE가 생성된다.²⁰ 병렬문의 전제의 범위를 좁힐 때에는 전반부는 후반부를, 후반부는 전반부를 문맥으로 보고 처리하도록 한다. 단, 전반부의 경우는 후반부에 같은 형태를 가진 전제가 없을 경우 전제가 성립하지 않는다. (18)의 경우 Coordination 규칙을 적용하기 전에 양쪽에서 생성된 전제는 각각 다음과 같은 형태가 된다.

(33) $\exists p \in \text{PSET}(\lambda x \lambda Q [Q(x) :- x \neq \text{영회}] \wedge P(\text{영회}) = \text{buy}(\text{영회}, \text{장미}) \wedge \text{resem}(Q, P))$, p .

(34) $\exists p \in \text{PSET}(\lambda x \lambda Q [Q(x) :- x \neq \text{둘이}] \wedge P(\text{둘이}) = \text{buy}(\text{둘이}, \text{수선화}) \wedge \text{resem}(Q, P))$, p .

이에 대하여 후반부의 전제를 전반부의 semantics를 보고 범위를 좁히면 다음과 같이 된다.

¹⁸ 다음과 같은 경우가 있을 수 있다.

방화 속제로 고문(古文)에 대한 것이 나왔다.
영회는 베오울프를 샀다.
둘이도 일리아드를 읽었다.

이 경우 동사와 명사 어느 쪽도 일치하지 않지만, ‘도’가 쓰여도 것도 자연스럽게 읽힌다. 이는 기존의 문맥에 ‘고문(古文)’이라는 성격이 나와 있기 때문에 좀 더 넓게 적용될 수 있었다고 보인다. 그러나 이에 대한 논의는 여기에서는 다루지 않기로 한다.

¹⁹ CAT은 구문범주, SEM은 그 semantics, PRE는 전제(presupposition)를 각각 지칭한다. 이 형태는 앞서 언급했던 정보 구조를 나타내는 방법이다.

²⁰ 병렬문과 그 이전 문장과의 관계에 대해서는 앞서 기술한 현상들과 병렬문에서의 현상 사이에 상호작용이 일어나게 되므로 논의의 편의를 위하여 여기서는 다루지 않기로 한다.

(35) $\exists p \in \text{PSET}(\lambda x \lambda y [\text{buy}(x, y) :- x \neq \text{둘이}] \wedge \text{sim}(\text{prop}(y), \text{prop}(\text{수선화})))$, p .

전반부의 전제는 후반부의 semantics를 보고 처리한 다음 후반부의 전제와 같은 형태인가를 확인한다. 같은 형태일 경우에만 Coordination 규칙을 적용하면서 두 개의 전제 list를 덧붙일 수 있다. 이 때 후반부에 전제가 존재하지 않는 경우에는 문장이 성립하지 않는다.

(36) 영회는 장미를 샀고, 둘이도 수선화를 샀다.

(37) ?영회도 장미를 샀고, 둘이는 수선화를 샀다.

(36)의 예문에서와 같이 후반부는 전반부에 전제가 없어도 성립할 수 있지만, (37)의 예문에서 보듯이 전반부는 후반부에 전제가 없을 경우 부자연스럽게 된다.

(38) 영회는 장미를 샀고, 둘이는 수선화를 샀다.

장미와 수선화 모두 [꽃]이라는 property를 가진다고 하자.²¹ (18)의 예문과 (38)의 예문을 비교해보면 ‘은/는’과 ‘도’의 유사성과 차이점을 알 수 있다. (38)의 경우 장미와 수선화가 다르다는 점이 초점이 되는 반면, (18)의 경우에는 영회와 둘이가 모두 꽃을 샀다는 점에 초점이 맞춰져 있는 것이다.

이상에서와 같이 전제와 의미구조를 병행하여 쓰는 방법으로 ‘도’의 기능을 정확하게 추적할 수 있다는 것을 알 수 있다.

4.3 추후 연구 과제

(17)에서는 ‘심지어’의 성격을 가지는 ‘도’에 대해서 다루고 있다. 가능한 인물을 철수, 둘이, 영회로 한정해보자. 여기서 마지막 문장을 봄으로써 생각할 수 있는 전제는 다음과 같다.

(39) $\forall p \in \text{PSET}(\lambda x [\neg \text{buy}(x, \text{베오울프})])$, p .

즉, 책을 사는 것을 좋아하는 경향이 있는 모든 사람이 베오울프 프만콤은 사지 않았다는 것을, 그 정점에 있는 영회를 통해 알 수 있게 되는 것이다. 이것이 ‘도’가 가지는 척도의 성격이라 할 수 있다. 이 전제는 지금까지 살펴본 것과는 다른 형태를 띄고 있다. ‘도’가 척도의 성격을 가지는가 아닌가를 판단하기 위해서는, 그 대상이 정점에 있는가 아닌가를 알아야 한다. 이것은 ‘심지어’가 나타났을 경우 바로 알 수 있는 것이지만 그렇지 않은 경우에는 기존의 문맥을 참조로 해야 한다. 여기서 기존 문맥에 영회가 정점임을 나타내는 내용이 있으므로, (32)의 전제가 아닌 (39) 쪽을 선택하게 된다. 문맥을 살펴보고 선택하기 이전에 (32)의 전제와 종합할 수 있는 방법에 대해서는 아직 연구 중이다.

다음과 같이 잇달아 나오는 ‘도’에서 서로 다른 종류의 전제를 제시하는 경우를 생각해 보자.

²¹ 더 자세한 분류는 여기에서는 생략한다.

- (40) 철수는 베오울프를 샀다.
순이도 베오울프를 샀다.
 ?영희도 일리아드를 샀다.

이런 경우 각 문장에서 각자 생성한 전제가 서로 충돌하게 된다. 세 번째 문장의 부자연스러움을 알아내기 위해서는 문맥에서 기존 문장의 semantics만이 아닌 전제도 살펴보아야 한다. 기존에 있던 문장의 전제가 더 넓은 범위를 포괄하는 것이었다면 전제의 범위가 축소된다는 면에서 허용될 수도 있지만, (40)에서와 같이 기존의 전제보다 더 넓은 범위의 전제를 요구하는 문장은 허용되기 힘들다. 이것은 문맥의 통일성을 유지하는 것으로 간주할 수 있다.

본 연구에서는 명사, 그 중에서도 특히 고유명사에 붙는 '도'에 대해서만 다루었기 때문에, 제안된 방법으로는 아직 다음과 같은 예문을 처리할 수 없다.

- (41) 영희는 베오울프를 사기도 했다.

이는 명사구 뿐만 아니라 동사 또는 동사구도 전제를 생성할 수 있다는 것을 뜻한다. 현재 이를 포함한 여러 문형에 대한 연구가 진행 중이다.

5 결론

한국어의 특수조사 '도'는 문맥의 자연스러운 연결에 있어서 중요한 역할을 수행한다. 이러한 '도'의 배경에는 '전제'가 존재하고 있는데, 이 '전제'의 성격을 밝히고 결합범주문법에서 구현함으로써 '도'의 기능에 대한 이해를 도울 수 있다. 본문에서는 고유명사에 붙어 사용되는 '도'만을 다루었으며 '같음', '유사함', '극한', '첨가' 및 병렬문에서 쓰이는 다섯 가지 의미를 보이고 이를 결합범주문법에서 전제와 관련지어 구현하는 방법을 제시했다. 현재 이러한 '도'의 생성에 대한 연구가 억양, 병렬문 등에 대한 생성의 연구와 병행하여 진행되고 있다.

References

- [1] 박시현 and 김명숙. '도'에 대하여. In *한글 197*, 1987.
- [2] 김윤신 and 이정민. 부정극어와 척도 개념. In *한글 및 한국어 정보처리*, 1996.
- [3] 조형준 and 박중철. 한국어 병렬문의 통사, 의미, 문맥분석을 위한 결합범주문법. Submitted, 1999.
- [4] 홍사만. *국어특수조사론*. 학문사, 1983.
- [5] 沼田善子. とりたて詞'も'のフォーカスとスコープ. 1995.
- [6] 남승호. 부정극어의 유형론. In *한글 및 한국어 정보처리*, 1995.
- [7] 시정근. 국어의 부정극어에 대한 연구. In *국어국문학 119*, 1997.
- [8] Hiroshi Aoyagi. Particles as adjunct clitics. *NELS*, 1998.
- [9] Cho, D.S. and Kwon, Y.J. LF movement and semantic interpretation of NPIs in English and Korean. *Korean Journal of Linguistics 20-1*, 1995.
- [10] Lee D.W. Negative polarity use and complementizer transparency in Korean bi-clausal structure. *Korean Journal of Linguistics 22-2*, 1997.
- [11] Chung hye Han. Asymmetric quantification: The case of the Korean topic marker -(n)un. In *ConSole4*.
- [12] Nobo Komagata. *A Computational Analysis of Information Structure Using Parallel Expository Texts in English and Japanese*. PhD thesis, University of Pennsylvania, 1999.
- [13] Mats Rooth. A theory of focus interpretation. In *Natural Language Semantics*, 1992.
- [14] Mark Steedman. *The grammar of intonation and focus*. 1994.
- [15] Mark Steedman. *The syntactic process*. Draft, 1998.
- [16] Karina Wilkinson. The scope of even. In *Natural Language Semantics*, 1996.