

영한 기계번역에서의 전치사구 처리를 위한 격의미 체계와 의미속성 집합

강원석 서정연 김길창
한국과학기술원 전산학과, 인공지능 연구센터

A Semantic Case Scheme and a Feature Set for Processing Prepositional Phrases in English-to-Korean Machine Translation

WonSeog Kang JungYun Seo GilChang Kim
Department of Computer Science and CAIR, KAIST

E-mail: wskang@csking.kaist.ac.kr, seo@csking.kaist.ac.kr,
and gckim@csking.kaist.ac.kr

요 약

영한 기계번역에서의 전치사구 처리는 자연어 처리의 중요한 연구과제 중 하나이다. 특히 전치사구의 격의미는 그에 대한 한국어 표현에 중요한 실마리가 된다. 본 논문은 영어의 전치사구에 대한 한국어 표현을 선택하기 위한 격의미 체계를 정의하고 그 격의미 체계를 실험한다. 이 격의미 체계는 전치사구에 대한 한국어의 격의미를 분별함으로 보다 좋은 한국어 격조사를 선택하게 해준다. 본 논문에서는 각 격의미의 구분을 위한 의미속성 집합도 정의한다. 이 의미속성 집합은 WordNet의 시소러스에서 제공한 상하위 관계와 변환표를 근거로 자동적으로 추출된다. 의미속성 집합의 자동 획득은 격의미 체계의 표현과 기계번역의 성능 평가에 일반성을 부여한다. 격의미 체계와 의미속성 집합의 자동 획득은 영한 기계번역은 물론 한국어 생성과 자연어 처리에 중요한 기여를 할 것으로 보인다.

1. 서론

영한 기계번역에서의 영어 전치사구에 대하여 정확한 표현의 한국어 격조사를 선택하기 위하여 영어 전치사구의 격의미 체계를 정의할 필요가 있다. 격의미 체계의 설계에 대한 연구[6, 11]가 이루어져 왔지만 그 격의미들은 영한 기계번역에 적용하기에는 문제가 있다. 그 격의미들은 너무 일반적(general)이어서 정확한 한국어 표현의 격조사나 어미를 선택할 수가 없으므로 영한 기계번역에 적합한 격의미 체계를 설계해야 한다. 본 논문은 영어 전치사구에 대한 정확한 한국어 표현을 얻을 수 있도록 이와 같은 격의미 체계를 설계하여 번역 시스템의 질의 향상을 도모하였다.

영한 기계번역의 해석 시스템이 전치사구의 격의미를 파악하려면 입력 단어들에 대한 의미와 격의미 체계의 각 격의미들에 대한 제한 정보를 비교해야 한다. 주어진 입력이 제한 정보에 맞는지의 여부는 제한 정보를 표현하는 의미 속성과 입력의 의미 속성과의 비교로 이루어질 수 있다. 이 제한 정보와 입력을 표현하는 의미 속성 집합의 정의와 획득은 해석 시스템의 성능과 확장성에 중요한 역할을 한다. 이러한 이유로 객관

성 있고 일반적인 방법으로 의미속성을 구하는 것이 필요하다. 본 논문은 이를 위해 WordNet 시소러스와 변환표를 이용한다. WordNet는 프린스턴 대학에서 언어심리학자들과 어휘론자들에 의하여 만들어진 시소러스이다. 본 논문의 의미속성 집합의 정의도 WordNet의 상하위 관계의 개념들을 기초로 작성되었다. 이 격의미 체계와 의미속성 집합을 사용하여 영한 기계번역 시스템의 전치사구 해석에 적용하여 좋은 실험 결과를 얻었다. 이 격의미 체계와 의미속성 집합은 영한 기계번역 이외의 자연어 처리 영역에 기여를 할 것으로 생각된다.

본 논문의 2장은 영한 기계번역의 격의미 체계를 위한 문제의 예를 설명하고 3장은 격의미 체계의 정의와 실험, 그리고 4장은 결론을 기술한다.

2. 영한 기계번역에서의 전치사구에 대한 격의미 문제

영한 기계번역의 목적어인 한국어의 격표현은 영어의 격표현과는 많이 다르다. 한국어는 언어 분류학상 첨가어에 속하기 때문에 문장 성분의 격의미를 나타내는 첨가소가 성분의 후미에 부착된다. 그러나 영어는 문장 성분의 위치와 전치사구의

역할로 그 문장 성분의 격을 표현한다. 전치사는 구체적인 격 의미의 한국어의 조사보다 일반적이다. 따라서 전치사구에 대한 한국어의 번역을 위하여는 구체적인 격의미의 정의와 해석이 필요하다.

(1) He wrote the article *on the revolution*.

문장 (1)에서 전치사구 “on the revolution”은 명사구 “the article”의 내용을 의미하는 concerning 격의미를 표현한다. 이 격의미에 대한 한국어 표현은 “에관한”이다. 이 경우 영어의 전치사구가 concerning 격의미라는 것이 파악된다면 한국어 격조사 선정에는 문제가 없다.

(2) He put the parcel *in the table*.

(3) The earthquake occurred *in Japan*.

(4) The pencil is *in the drawer*.

그러나 문장 (2), (3), (4)의 경우는 그렇지 않다. 각 경우의 전치사 모두 장소를 나타내는 location 격의미에 속한다. 이에 대한 한국어 조사를 대역해보면 서로 다른 조사가 대역됨을 알 수 있다. 문장 (2)의 경우 “에”가 대역되고, 문장 (3)은 “에서”가, 문장 (4)는 “에”가 대역된다[8]. 이와 같은 한국어 조사의 차이는 그 전치사구와 관계짓는 동사의 의미에 따라 발생된다. 문장 (2)의 경우 전치사구 “in the table”의 피수식어인 동사는 “put”으로 동작 대상 장소를 요구하는 동사이고, 문장 (3)의 동사는 어떠한 사건의 발생을 표현하는 동사이고, 문장 (4)의 동사는 물건이 놓여 있는 장소를 묘사하는 상태동사이다.

전치사구가 장소를 나타내는 경우 그 격의미가 더 세분되어 있지 않으면 한국어 격조사를 선택할 수가 없다. 따라서 영한 기계번역에서는 장소를 의미하는 격의미를 구체적으로 세분하여 놓아야 한다. 본 논문에서 제안한 장소격의 구체적인 격의미는 location-goal, location-action, location-stative의 세 격의미이다. 문장 (2)의 전치사구의 경우는 location-goal 격의미에 해당하고 문장 (3)의 전치사구는 location-action 격의미에, 문장 (4)의 전치사구는 location-stative 격의미에 해당한다. 이로써 한국어 조사 선택의 문제를 해결할 수가 있다.

문장 (5)와 (6)도 이같은 문제를 가지고 있다.

(5) I sent the mail *to him*.

(6) I sent the mail *to the white house*.

문장 (5)와 (6)의 전치사구는 destination 격의미에 속한다. 그렇지만 한국어에서는 각 경우가 구분되어 다른 격조사가 사용된다. 전치사구의 명사구가 문장 (5)의 경우처럼 유정 물이면 “에게”가 사용되고 그렇지 않은 경우 “으로”가 사용된다. 역시 이를 구분하기 위하여는 구체적인 격의미의 선정이

필요하다.

3. 영한 기계번역에서의 전치사구 해석을 위한 격의미 체계

영한 기계번역에서 두 언어의 격의미 표현의 차이에서 오는 문제를 해결하기 위하여 구체적인 격의미 체계를 정의한다. 그리고 이 격의미의 변별을 위하여 의미 속성 집합을 정의한다.

3.1 격의미 체계

전치사구의 격의미는 전치사구와 전치사구의 피수식어와의 관계이다. 전치사구의 격의미는 전치사구의 전치사, 전치사구의 명사구, 그리고 전치사구의 피수식어의 특성으로 정의된다. 이를 기초로 영한 기계번역에서 세분되어야 하는 격의미를 정의한다.

앞 장에서 언급한 바와 같이 장소를 나타내는 격의미는 여러 개로 세분된다.

- location-goal : 에
He put the parcel *on the table*.
- location-stative : 에
The parcel is *on the table*.
- location-action : 에서
The earthquake occurred *in Japan*.
- location-noun : 의
the parcel *on the table*

첫번째의 격의미는 location-goal로서 전치사구가 동사의 대상 장소가 되는 관계를 표현한다. 이 격의미에 대한 한국어 격조사는 “에”이다. 이 격의미에 사용되는 전치사구의 피수식어는 장소를 목적으로 하는 하위범주와 정보를 가진 동사가 된다. location-stative 격의미는 동사가 상태동사로 물건의 위치 등의 상태를 묘사하는 경우에 사용된다. 이에 대해 한국어 격조사 “에”가 대응된다. location-action은 동사가 사건의 묘사 등의 행위 동사인 경우 그 사건의 장소를 나타내는 격의미이다. 이 때 한국어 조사 “에서”가 대응된다[8].

행위의 방향을 표현하는 전치사구는 destination 격의미에 속한다.

- destination-animate :에게
He sent the mail *to her*.
- destination-move :으로
He moves the table *to the second floor*.
- destination-arrival :에
He reaches *in the center*.

이 격의미는 세 개로 구분된다. destination-animate는 목적지를 나타내는 전치사구의 명사구가 유정체인 경우의 격의미이고 대역되는 한국어 격조사는 “에게”이다. 유정체가 아닌 경우는 destination-move 격의미로서 한국어 조사 “으로”가 대응된다. 이 두 개의 격의미는 이동의 방향동을 표현하고 destination-arrival 격의미는 이동의 최종 목적지를 표현한다. 이 경우 한국어 격조사는 “에”가 사용된다.

동사가 수여동사인 경우 수반되는 to-전치사구는 수여목적어를 표현한다. 이 때 한국어 표현은 역시 그 전치사구의 명사구가 유정체인가에 따라 표현되는 조사가 다르다. 유정체인 경우 recipient-animate 격의미로서 “에게”이고 그렇지 않은 경우 recipient-inanimate 격의미로서 “에”가 대응된다.

- recipient-inanimate : 에
The system gives water to the particle.
- recipient-animate : 에게
The mother offers the milk to the baby.

from-전치사구는 출처나 시작점을 나타내는데 사용된다. 출처를 의미하는 source 격의미는 출처의 대상이 유정체인 경우 source-animate 격의미로, 그렇지 않은 경우 source-inanimate 격의미로, 구간의 출발점을 나타내는 경우는 source-period 격의미로 세분된다.

- source-inanimate : 에서
The baby gets nutriments from the milk.
- source-animate : 에게서
I buy the desk from him.
- source-period : 부터
The man finds from the front to the rear.
- source-noun : 으로부터의
protein from milk

시간을 나타내는 전치사구도 여러 개로 세분된다. 이 구분은 동사의 의미에 따라 좌우된다. 시간의 시점을 표현하는 경우는 time-point 격의미로서 한국어 조사 “에”가 사용되고 시간의 기간을 표현하는 경우는 time-period 격의미로서 한국어 조사 “동안에”가 사용된다. 그리고 시간의 출발점은 time-start 격의미로서 한국어 조사 “부터”가 사용되고 시간의 도착점은 time-finish 격의미로서 한국어 조사 “까지”가 사용된다. 그리고 전치사구의 피수식어가 명사구일 때는 “의”가 사용된다.

- time-point : 에
The farmer fertilizes in the beginning of spring.

- time-period : 동안에
He wrote the paper for summer
- time-start : 부터
It is eight weeks from strewing.
- time-finish : 까지
He delays the meeting until 7 o'clock.
- time-noun : 의
the meeting at 7 o'clock

영어의 모든 전치사구에 대하여 이 절에서 언급한 격의미를 포함하여 64 개의 격의미를 정의하였다[2, 3]

3.2 의미 속성 집합

앞 절에서 언급한 격의미 체계의 격의미를 구분하기 위하여 의미속성 집합을 사용한다. 본 논문에서는 106개의 격의미 집합을 정의하고 자동으로 의미속성의 획득을 위하여 WordNet[7]과 변환표를 이용한다. 이 WordNet과 변환표에 의한 의미속성 추출은 의미속성 획득에 일반성과 객관성을 준다.

WordNet은 프린스턴 대학에서 언어심리학자들과 어휘론자들에 의하여 개발된 시스터리이다. 이 WordNet의 상하위관계를 근거로 하여 106개의 의미속성 집합을 정의하였다. 그리고 WordNet의 약 49000개의 개념 가운데 상위층에 속하는 1899개의 개념을 선택하고 이 개념을 106개의 의미속성으로 변환하는 변환표를 구성하였다.

이 변환표와 WordNet를 이용하여 주어진 단어의 의미속성을 획득한다. 그 과정은 다음과 같다. 먼저 WordNet의 기능을 이용하여 주어진 단어의 개념을 추출한다. 다음은 추출된 개념들이 변환표에 들어 있는가의 여부를 검사하고 들어 있는 개념들에 대하여 변환표를 통하여 해당하는 의미속성을 추출한다. 그렇지 않은 경우 그 개념은 상위층의 1899 개념에 속하지 않으므로 무시한다. 그리고 그 개념의 모든 상위 개념에 대하여 이와 같은 과정을 되풀이하여 최종 의미속성을 추출한다. 자동으로 획득한 의미속성들은 격의미를 구분하는 영한 기계번역 시스템의 확장에 큰 도움을 준다.

본 논문에서 제안한 격의미 체계와 의미속성 집합을 이용하여 영한 기계번역 시스템의 전치사구 격의미 해석에 사용하여 실험한 결과 90%의 성공율을 얻을 수 있었다[3].

4. 결론

본 논문에서는 영한 기계번역의 전치사구 격의미 해석을 위하여 격의미 체계를 정의하고 각 격의미의 변별을 위하여 의미속성 집합을 정의하였다. 격의미 체계는 전치사구에 대한 적합한 한국어 표현을 선택할 수 있도록 구체적으로 정의되었다. 시스

템의 확장을 위하여 의미 속성 집합의 자동 획득 방법을 사용하였다.

제안된 격의미 체계와 의미속성 집합, 그리고 자동 획득 방법은 영한 기계번역 시스템 이외에 한국어 생성과 자연어 처리 영역에 많은 기여를 할 것으로 생각된다.

참고 문헌

- [1] J. Allen, *Natural Language Understanding*, Benjamin/Cummings Co., 1987.
- [2] W. S. Kang, *Semantic Roles for English-to-Korean Machine Translation*, CSMEMO-93-01, KAIST, 1993.
- [3] W. S. Kang, J. H. Kim, J. Y. Seo, K. S. Choi, and G. C. Kim, "A Hybrid Method for the Semantic Analysis of Prepositional Phrases in English-to-Korean Machine Translation," *Proceedings of International Conference in Commemoration of 20th KISS Anniversary*, pp. 501-508, 1993.
- [4] 김 영희, "한국어 조사류어의 연구," 문법연구, 제 1집, 1974.
- [5] 김 광해, "{-의}의 의미," 문법연구, 제 5집, 1984.
- [6] J. L. McClelland and A. H. Kawamoto, "Mechanisms of Sentence Processing: Assigning Roles to Constituents of Sentences," *Parallel Distributed Processing*, Volume 2, pp. 272-325, 1986.
- [7] G. A. Miller, R. Beckwith, C. Fellbaum, D. Gross, K. Miller, *Introduction to WordNet: An On-line Lexical Database*, Report of WordNet, 1990.
- [8] J. M. Seo, "On the Meaning of the Postposition '-seo'," *Linguistic Journal of Korea*, Vol. 9, No.1, 1984.
- [9] 서 정수, "한국어와 일본어의 접속어미 연구," 문법연구, 제5집, 1984.
- [10] H. S. Shin, "A New Analysis of *-lul*," *Linguistic Journal of Korea*, Vol. 7, No. 1, 1982.
- [11] J. F. Sowa, *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley Co., 1983.
- [12] N. S. Yi, "The Syntax and the Meaning of *-ey* and *-lo*," *Linguistic Journal of Korea*, Vol. 8, No. 2, 1983.