

일한기계번역에서 진행형 “ている”의 번역처리

김 정 인[†] · 문 경 희^{††} · 이 종 혁^{†††}

요 약

본 논문은 일한기계번역에서 일본어 진행표현 “ている” 형태로부터 한국어 대역어의 선정과정에서 발생하는 애매성을 해소하기 위하여 연구한 것이다. 대부분의 일한 기계번역 시스템은 양국어의 문법적인 유사성에 기초하여 어휘적인 단계의 처리만으로 고품질의 번역이 가능한 직접 번역방식을 채용하고 있다. 그러나, 직접 번역방식에 기초한 일한 기계번역에서는 술부에 존재하는 “ている”형태의 상적인 의미를 구별할 수 있는 방법론이 아직 제안되지 않았다. 일본어에서 “ている”형태는 동작진행과 상태진행을 모두 나타내지만 한국어에서는 “고 있다”와 “어 있다”로 나누어 표기한다. 양 언어간의 상적인 의미 대응은 간단하지 않지만, 술부의 의미정보, 부사와 부사어의 의미정보 등을 이용하여 “ている”형태의 상적인 의미를 결정하는 것이 가능하다. “ている”형태의 적절한 대역어 선정을 위하여, 사전 속의 모든 일본어 술어에 다섯 종류의 의미코드를 입력한다. 즉, “1 : 동작진행만으로 사용되는 술어”, “2 : 일반적으로는 동작진행으로 사용되지만, 수동인 경우에는 상태진행의 형태로도 사용되는 술어”, “3 : 상태진행으로만 사용되는 술어”, “4 : 동작진행, 상태진행의 구별이 애매한 술어”, “5 : 기타” 등의 상적인 의미분류코드를 술어별로 입력한다. 그리고 “2”, “4”형태의 술어로부터 진행형은 구별하기 위하여 부사와 부사어를 사용하는 방법을 제안한다. 실험에는 아사히 신문의 기사 중에서 임의로 약 15,000문을 추출하여 이용하였다. 제안한 방법은 실험문장에서 83.6%의 성공률을 보였으며, 단순히 동작진행과 상태진행, 기타만으로 나누어 처리하던 종전 시스템에 비하여 약 5.7% 정도 더 좋은 결과가 얻어졌다.

A Processing of Progressive Aspect “te-iru” in Japanese-Korean Machine Translation

Jung-In Kim[†] · Kyong-Hi Moon^{††} · Jong-Hyeok Lee^{†††}

ABSTRACT

This paper describes how to disambiguate the aspectual meaning of Japanese expression “-te iru” in Japanese-Korean machine translation. Due to grammatical similarities of both languages, almost all Japanese-Korean MT systems have been developed under the direct MT strategy, in which the lexical disambiguation is essential to high-quality translation. Japanese has a progressive aspectual marker “-te iru” which is difficult to translate into Korean equivalents because in Korean there are two different progressive aspectual markers: “-ko isssta” for “action progressive” and “-e isssta” for “state progressive”. Moreover, the aspectual system of both languages does not quite coincide with each other, so the Korean progressive aspect could not be determined by Japanese meaning of “-te iru” alone. The progressive aspectual meaning may be partially determined by the meaning of predicates and also the semantic meaning of predicates may be partially restricted by adverbials, so all Japanese predicates are classified into five classes: the 1st verb is used only for “action progressive”, 2nd verb generally for “action progressive”, but occasionally for “state progressive”, the 3rd verb only for “state progressive”, the 4th verb generally for “state progressive”, but occasionally for “action progressive”, and the 5th verb for the others. Some heuristic rules are defined for disambiguation of the 2nd and 4th verbs on the basis of adverbs and adverbial phrases. In an experimental evaluation using more than 15,000 sentences from “Asahi newspapers”, the proposed method improved the translation quality by about 5%, which proves that it is effective in disambiguating “-te iru” for Japanese-Korean machine translation.

키워드 : 한국어(Korean), 일본어(Japanese), 기계번역(Machine Translation), 진행형(Progressive aspect), 술부(Predicates), 부사(Adverbs), 부사어(Adverbial phrases)

1. 서 론

일본어와 한국어는 언어 계통 상 알타이어족에 속하여

문법체계가 유사하며 또한 동일 한자문화권에 속한다. 따라서, 일한 기계번역 시스템의 개발은 영어나나 중한 번역 시스템처럼 서로 다른 계통의 언어간을 번역하는 시스템보다는 상대적으로 용이하다고 할 수 있다[13-15]. 그러나, 현재, 개발된 시스템이나 개발중인 많은 일한 기계번역 시스템은 다의성, 술부 어순의 차이, 경어처리, 대화체 번역 등 아직 해결해야 할 상당 부분의 과제를 가지고 있다[15].

※ 본 논문은 한국학술진흥재단의 신진교수 연구지원(과제번호 1999-003-E00403)에 의한 연구결과임.

† 종신희원 : 동명정보대학교 컴퓨터공학과 교수

†† 문 경 희 원 : 포항공과대학교 대학원 전자계산학과

††† 정 희 원 : 포항공과대학교 전자계산학과 교수

논문접수 : 2001년 5월 14일, 심사완료 : 2001년 11월 6일

본 논문에서는 일한 기계번역 시스템이 가지고 있는 과제 중에서 일본어 문장에 진행형으로 자주 사용되어지는 “ている”형태가 한국어로 번역될 때 나타나는 동작진행과 상태진행의 상적인 애매성에 대하여 그 해결 방법을 생각해 본다. 일본어에서 “ている”형태는 한국어로 번역될 경우 “고 있다”와 “어 있다”의 두 형태로 번역이 가능하다. “고 있다”와 “어 있다”는 언어학적인 관점에서 각각 동작진행과 상태진행이라는 의미의 차가 있으며, 직접 번역방식에 기초한 일한 기계번역에서는 상적인(Aspectual) 의미를 구별하여 번역하는 방법론이 아직 제안되어 있지 않다. 본 논문에서는 일한 양국어의 진행형 표현에 대하여 상적인 의미를 언어학적으로 고찰하고 그 결과로부터 실제로 일한 기계번역 시스템에 적용할 수 있는 방안을 제시한다.

2. 진행의 언어학적인 해석

대부분의 일한 기계번역 시스템은 조사와 용언 등의 다의성 해소 측면에서 많은 노력을 기울여 왔다[13]. 그러나, 상(aspect)에 대한 문제 해결의 의지는 상대적으로 미약했다고 할 수 있다. 이것은 문제 해결에 필요한 노력에 비하여 전체의 번역률에 미치는 영향이 적기 때문으로 생각되어진다. 그러나, 아사히 신문기사로부터 추출한 실험용 문장에서 슬부에 “ている”형태가 나타난 문을 헤아려 본 결과, 그 비율이 15%이상에 달하였다. 이는 실제 문장에서 “ている”의 사용 빈도가 상당히 높으며, 한국어로의 번역시 자연스러운 진행상의 표현을 위해서는 “ている”형태의 대역어 선정에서 발생하는 애매성을 무시하기에는 너무 큰 수치이다.

우선 양 언어에서 진행상으로 표현되는 “ている”와 “고 있다”, “어 있다”에 대하여 언어학적인 성질을 살펴보기로 한다. 한국어 “어(아/여) 있다”는 상태진행의 의미를 가지며 전접하는 단어와의 음운법칙에 의해 3가지의 형태로 나타나지만 편의상 이하 “어 있다”로 표기한다.

2.1 일본어 “ている”의 의미

일본어에서 진행상의 개념은 시제를 빼고는 생각할 수 없다. 일본어에서 상은 <표 1>에서 나타내는 것처럼 동작 완료형, 동작 진행형, 상태 진행형으로 구분한다[10]. 그 때 “ている”형태는 동작 진행형, 상태진행형의 여러 가지 시제에 나타난다.

<표 1> 일본어에서 상과 시제에 의한 종결어미

	질대시제			상대시제		
	과거	현재	미래	전	동	후
동작 완료형	た		る	た		る
동작 진행형	ていた	ている	ている		ている	
상태 진행형	ていた	ている	ている		ている	

“ている”형태가 동작 진행형으로 사용되었는지 상태 진행형으로 사용되었는지의 구분은 학자별로 다소 이견을 보이고 있지만, [10]에서는 이 구분을 위하여 동사를 상태동사,

경과동사, 변화동사, 경과/변화동사로 나누고 있다. 상태동사의 경우는 본질적으로 상태, 즉 상태 진행형을 나타내므로 이미 “ている”의 의미를 가지고 있다. 따라서, 상태동사에는 “ている”를 붙이지 않는다. 나머지 경과동사, 변화동사, 그리고 경과/변화동사의 구분은 다음과 같다.

경과동사

움직임의 기술 등으로 이용되는 동사로서 “ている”와의 결합은 동작 진행형을 의미한다.

예 : 眺める。うようよする

변화동사

표현하려는 현상, 그것은 동작이지만 상의 관점에서는 상태의 변화에 초점을 맞춘 동사로서 “ている”와의 결합은 상태 진행형을 의미한다.

예 : 知る, くつつく

경과/변화동사

“ている”와 결합시, 동사의 주체가 유정, 혹은 움직임의 의미 정보를 가진 경우는 동작 진행형, 그 외의 주체에 대해서는 상태 진행형을 나타낸다.

예 : のぞく, さえきる

한편, [11]은 [10]와 달리, 술어를 개괄적으로 계속동사와 순간동사로 구분하고 있으며 이 방법에서는 계속동사의 경우 동작 진행형을, 순간동사의 경우는 다시 움직임의 부사구를 취하지 않으면 상태 진행형을 의미하지만, 많은 술부가 양쪽의 성질을 가지고 있다고 기술하고 있다.

2.2 한국어 “고 있다”의 의미

“고 있다”는 진행·지속을 나타내는 보조용언으로서 기본의미는 동작의 진행을 나타내며 대부분의 동사에 넓게 사용되고 있다[2]. 한국어에서 진행은 “는다”에도 표시되지만 “고 있다”를 사용하여 보다 명확한 의미의 표시가 가능하다. 이것은 “고 있다”의 기본 의미가 동작의 진행이며 문맥의 의미를 파악하지 않아도 언제나 진행의 의미를 가지고 있기 때문이다.

a 사나이가 공중전화를 걸고 있었다.
b 해선이 은뿔을 굶고 있다.

(예 1) “고 있다”의 사용 예

(예 1)은 진행 중의 동사를 표현한다. 시간 부사어를 사용하지 않고 동작의 진행으로 해석된다. (예 1)의 a는 과거 시점에서의 진행, (예 1)의 b는 현재시점에서의 진행이며 “고 있다”가 표시하는 진행상의 각 시점에 있어서 넓게 사용되고 있는 것을 알 수 있다.

또한, “고 있다”는 문맥에 따라 지속과 반복, 습관, 결과, 상태의 지속 등의 의미로 해석된다.

2.2.1 지속과 반복, 습관

시간 혹은 회수 관련의 부사어가 나타난 경우, “고 있다”

는 지속과 반복, 습관과 같은 의미를 가진다. 이것들의 구별은 대부분이 시간 부사어에 의해 명확해진다. 진행의 기본 의미로부터 문맥의 의미가 어떻게 나타나는지를 (예 2)에 나타낸다.

- a. 두 사람이 지금 만나고 있다.
- b. 두 사람이 가끔 만나고 있다.
- c. 두 사람이 주일마다 만나고 있다.
- d. 두 사람이 계속 만나고 있다.

(예 2) “고 있다”가 “지속, 반복, 습관”의 의미로 사용된 예

(예 2)의 a는 현재 진행중의 동작을 나타낸다. 부사어 “지금”의 첨가는 동작의 현재진행을 보다 명확하게 한다. (예 2)의 b는 반복적인 형태를 표현하며, (예 2)의 c는 습관적인 모양을 나타내고 (예 2)의 d는 지속적인 행위를 나타낸다. (예 2)에서 반복, 습관, 지속의 구분은 부사어 “가끔, 주일마다, 계속”에 의해 결정된다. (예 2)의 a는 하나의 동작의 진행을 나타내는 것뿐이지만 (예 2)의 b,c,d는 동작이 반복하고 있는 모양을 나타내는 점이 다르다. 반복되는 모양이 어느 정도 자주 나타나는가에 따라 부사어가 다르게 사용된다.

2.2.2 완결상태의 지속

“고 있다”는 동작의 완결된 후의 상태가 지속되는 것을 나타낸다.

- a. 그는 오늘 한복을 입고 있다.
- b. 저기 가서 손들고 있어.
- c. 그는 아까부터 공을 차고 있다.

(예 3) “고 있다”가 완결상태의 지속으로 사용된 예

(예 3)의 a,b에서 “고 있다”는 동작 완결 후의 상태가 지속되는 것을 나타낸다. (예 3)의 a는 한복을 입은 후, 입고 있는 모양의 지속으로 설명할 수 있다. (예 3)의 b는 손을 든 후, 그 상태가 지속되는 것을 나타낸다. (예 3)의 c는 볼을 차는 행위가 지속적으로 성립하는 것을 나타낸다.

즉, “고 있다”는 대부분의 동사에 넓게 사용되며, 동작의 진행, 반복, 습관, 그리고 완결상태의 지속을 의미하며 시간 또는 회수를 나타내는 부사어와 결합한다.

2.3 한국어 “어 있다”의 의미

“어 있다”는 넓은 의미의 진행을 기본의미로 하는 보조용언으로, 이것이 표현하는 진행은 동작의 진행이 아니라 상태의 진행이다. 상태의 진행이란 동작의 완결로부터 시작하여 상태가 지속되는 모양 혹은 동작의 변화가 없는 상태가 지속되는 것을 의미한다.

- a. 아직도 피로가 남아 있다.
- b. 인문관은 언덕 위에 서 있다.

(예 4) “어 있다”가 사용된 예

(예 4)는 동작의 변화가 없는 지속을 표시한다. 즉, “남

다”와 “서다”의 용언은 ‘동작진행’의 전형적인 보조용언인 “고 있다”와 결합할 수 없다. [3]에서는 지속형에는 동작하지 않고 그대로 계속되는 “동작지속형”과 정지상태가 계속되는 “상태지속형”으로 구분하고 “어 있다”가 “상태지속형”을 나타낸다고 지적하고 있다. 이것을 다시 “단순상태지속형”과 “완료상태지속형”으로 구분하여 전자는 동작의 변화가 없는 동사로부터, 후자는 동작 후 결과가 지속되는 동사로부터 표시된다. 또한 “어 있다”의 용법을 보면, 시간성을 가지는 부사어와는 같이 상용되지만, 짧은 순간을 표시하는 부사어, 도달점, 혹은 완료를 나타내는 부사어와는 같이 사용되지 않는 특성을 가지고 있다[2].

- a. 계속 잠이 들어 있다.
- b. 여태까지 누워 있다.
- c. 아직 일어서 있어서 다행이다.

(예 5) “어 있다”와 결합할 수 있는 부사의 예

(예 5)의 a,b,c에서 “계속, 여태까지, 아직” 등과 같은 시간성을 가지는 부사어는 사용된다. 그러나, “지금 바로”와 같은 부사어와의 결합에는 제약을 받는다. 즉, “어 있다”의 의미는 상태의 진행이며, 긴 시간에 걸치는 시간성 부사어와 같이 나타나고 현재, 과거, 미래의 계속을 모두 나타낼 수 있다고 하겠다.

2.4 “고 있다”와 “어 있다”의 비교

한국어에서 “고 있다”와 “어 있다”의 구분은 많은 경우 문맥적 상황에 따라 구분할 수 있다. 즉, 동사에 따라 전형적인 상적 특징을 가지는 경우는 그 구분이 곤란하지 않다. 예를 들어, 타동사의 경우는 상태진행의 “어 있다” 형태의 보조 용언을 취할 수 없다. 그러나, 자동사와 일부 타동사의 수동형이 사용된 문에서는 문맥에 따라 “고 있다”와 “어 있다”의 양 형태가 함께 사용된다.

- a. 정원에 눈이 내려/내리고 있다.
- b. 생각에 잠겨/잠기고 있다.

(예 6) 동작진행과 상태진행의 구별이 곤란한 경우

(예 6)과 같이 양 형태가 모두 사용되는 경우에는 보조용언의 선택이 간단하지 않다. 즉, (예 6)의 문만으로는 동작진행과 상태진행의 구분이 불가능하다. 그러나 동작진행과 상태진행의 애매성이 있는 문이 상적인 전형성이 강한 부사어가 함께 사용되어지면 부사어에 의한 진행상의 구별은 어느 정도 가능하다고 생각될 수 있다.

2.5 양국어의 진행상에 대한 비교

일한 양국어는 용언의 의미정보에 따라 상의 구분이 어느 정도 가능하며, 술부의 문맥적인 의미의 상당 부분을 부사어에 의존하고, 동시에 시제를 구분하지 않으면 상의 정의가 불가능한 점 등, 상에 관한 언어학적인 성질은 거의 차이가 나지 않는다.

그러나, 일본어에는 동작진행의 개념과 상태진행의 개념이 문장에 나타날 경우, 외형적인 표현이 “ている”로서 같은 모양이나, 한국어에서는 사용자가 “고 있다”와 “어 있다”의 선택을 문맥적 상황으로부터 직관적으로 구분해서 사용하고 있다. 앞 절에서 언급한 일본어 동사 분류 방법에서 경과동사로 분류된 술어의 “ている”형은 동작진행형으로, 변화동사로 분류된 술어의 “ている”형은 상태진행형으로 분류되면 각각 한국어의 동작진행과 상태진행으로 대응시키는 것이 가능하다. 그러나, (예 7)을 보면 “知っている”는 한국어 “알고 있다”로 대응되어 동작진행으로 구분되지만, 일본어에서는 “知る”라는 술어가 “知らない”라는 상태에서부터의 변화로 해석하여 변화동사로 분류되어 상태진행이 되어버려, 한국어의 대역어 “알아 있다”를 선택해 버린다. 이것은 한국어에서의 “알다”가 목적어를 취하는 타동사이므로 무조건 “고 있다”를 취하는 관점과는 거리가 있다.

だいたい日本人も知っている。
대개의 일본인도 알아 있다. (X)
대개의 일본인도 알고 있다. (O)

(예 7) 일본어와 한국어의 상에 대한 관점의 차이

또한, “ている”형은 진행형의 의미를 제거하는 편이 한국어에서는 오히려 자연스러운 표현이 되는 경우가 있다. 진행형을 없애고 번역해야 하는 동사의 예를 (예 8, 9)에 보인다.

a. 彼は疲れている。
그는 피곤하다.
b. この棒は長すぎている。
이 막대는 너무 길다.

(예 8) 진행형을 없애고 번역한 예

a. 彼女は結婚している。
그녀는 결혼했다.
b. 彼は3回カナダを訪問している。
그는 3번 캐나다를 방문한 적이 있다.

(예 9) “ている”형을 과거나 경험으로 번역한 예

결국, 일본어 동사와 한국어 동사의 상적인 구분이 일정한 대응 관계를 취하고 있지 않으므로 동사마다 상적인 대응 처리에 관한 정보를 기술할 필요가 있다. 이를 위하여 “ている”형태의 의미구분은 술어별로 대응하는 한국어로부터 상적인 의미정보를 취하여 그것을 진행형의 구분 기준으로 이용한다.

3. 진행상의 구분방안

“ている”에서 동작진행과 상태진행의 구분은 깊은 의미 해석을 실행하지 않고는 판단하기 어렵다. 그러나, 한국어 진행상의 특징과 일본어와의 유사성에 기초하면 술어별로 진행상의 구분이 어느 정도 가능하다. 종래의 일한 기계 번역 시스템에서는 단순히 사전에 정보를 입력하여 동작진행

과 상태진행의 문제를 부분적으로 해결해 왔다. 즉, 동사부분의 사전 정보를 구축할 때, 동사의 상적인 의미정보에 따라 상의 항목에 고정된 값으로 동작진행, 상태진행, 기타를 설정한다. 만약, 동작진행, 상태진행이 모두 가능한 경우는 사전 등록자의 직관에 의존하여 한쪽을 선택한다. 이와 같은 고정된 사전 정보를 이용하여 1개의 역어를 결정해도 “ている”의 애매성이 약 78% 정도 해결되었다. 그러나 “ている”의 대역어를 사전정보에 전반적으로 의존함에 따라 상적 의미의 유연성이 결여되어, 대역어 선정의 오류가 당연한 결과로 받아들여지게 된다. 본 장에서는 “ている”의 애매성을 해소하기 위하여 구체적으로 사용 가능한 정보를 제시하고, 그 정보를 사용하는 방법을 제시한다.

3.1 술어의 분류에 의한 진행상의 구분

“ている”의 한국어 대역어를 선정하는 과정에서 나타나는 술어별 선택경향에 따라 일본어 술어를 다음과 같이 분류하였다.

3.1.1 제1동사

“ている”형태의 한국어 대역어가 전형적으로 “고 있다”로 선택되는 동사로 대부분의 타동사 등이 여기에 속한다.

예 : <走る(달리다), 乗る(타다), 読む(읽다), 問い合わせる(문의하다), 見送る(배웅하다)>

3.1.2 제2동사

제1동사에 분류되는 동사 가운데에 수동태로 사용되는 경우 “고 있다”와 “어 있다” 양쪽으로 대역 가능한 동사를 제1동사와 구별하였다.

예 : <下がる(내리다), 上がる(오르다), 食べる(먹다), 削る(깎다), 刻む(새기다), つめる(채우다)>

3.1.3 제3동사

“ている”형의 한국어 대역어가 전형적으로 “어 있다”로 선택되는 동사로 대다수의 자동사가 여기에 속한다.

예 : <横になる(눕다), 立つ(서다), なれる(익숙하다), 倒れる(넘어지다)>

3.1.4 제4동사

일반적으로 “ている”형의 한국어 대역어가 “고 있다”와 “어 있다” 모두 가능한 동사이다.

예 : <降る(내리다), 増える(늘다), ぼける(둔해지다), 差し替える(바꾸어 넣다), 交える(섞다)>

3.1.5 제5동사

“ている”형의 한국어 대역어가 “고 있다”와 “어 있다” 둘 다 자연스럽게 않은 동사이다. 이와 같은 동사는 가장 무난한 의미인 과거로 취급하여 “쌌다/었다”로 대응시킨다.

예 : <卒業する(졸업하다), 結婚する(결혼하다), 難航する(난항이다), 疲れる(피곤하다), 掃る(돌아가다)>

이 정보를 사전 속의 모든 동사 엔트리에 진행형의 번역 정보로 입력한다. 이 정보는 주로 입력하는 사람의 주관적인 판단에 따르지만, 기존의 3개 값을 입력하는 경우보다 판단이 용이하므로 보다 정확한 분류가 가능하다. 5개의 동사 타입에 대응하는 기본적인 한국어 대역어를 아래에 표시한다.

- 제 1동사 : 고 있다
- 제 2동사 : 고(어) 있다
- 제 3동사 : 어 있다
- 제 4동사 : 어(고) 있다
- 제 5동사 : 쓰다/었다

제 2동사와 제 4동사의 기본적인 대역어가 다른 것은 제 2동사의 기본형은 전형적인 동작진행형을 의미하는 제 1동사와 같은 부류이므로 “고 있다”가 보다 자연스럽다고 생각하여 “고(어) 있다”로, 제 4동사는 대략 자동사이므로 일반적으로 “어 있다”가 많이 사용되는 동사의 본성을 고려하여 “어(고) 있다”로 하였다. 양쪽 모두 판단 기준이 명확하지는 않으므로 괄호 속의 후보 대역어를 준비하였지만, 실험의 결과를 나타낼 때는 괄호를 무시하고 계산한 수치를 사용한다.

이와 같은 동사의 분류를 사용하면 제 1동사, 제 3동사, 그리고 제 5동사는 “ている”형태의 한국어 대역어로서 각각 “고 있다”, “어 있다”, “쓰다/었다”를 취하는 것이 가능한 반면, 제 2동사와 제 4동사의 경우는 진행상의 의미정보만으로는 상의 구분이 불가능하다.

위의 정보 이외에도 본 시스템은 동사마다 결합가 정보를 두고 있다. 결합가 정보는 4자리 숫자로 표기하며, 첫 번째 자리수는 주격 “가”를 필수로 가지는 동사인가의 여부에 따라 0/1을, 두 번째 자리수는 목적격 “을”를 필수로 가지는가의 여부에 따라 0/1을, 세 번째 자리수는 대상격 “に”를 필수로 가지는가의 여부에 따라 0/1을 표기하고 마지막 네 번째 자리수는 “ている”처리의 5가지 동사분류를 나타내기 위하여 0/1/2/3/4를 입력하고 있다.

앞의 세자리 숫자는 격형식을 이용한 번역처리에 사용되며 마지막 숫자가 진행형 처리를 위하여 사용되는 것이다.

3.2 부사어를 이용한 상태진행·동작진행의 구분

[7]에서 부사어는 서술어의 의미가 확실히 나타나도록 용언을 한정시키는 부속 성분의 말이라고 정의하고 있다. 이와 같은 부사어에는 순수부사와 체언에 부사격 조사가 붙어 있는 말로 나누어진다. 여기에서 부사어를 상격 의미에 따라 분류하여, 애매성이 생기는 “ている”형태 술부의 문맥적인 의미를 파악하는 기준으로 한다.

3.2.1 순수부사에 관한 고찰

부사마다 그것이 동작을 나타내는 부사인지, 상태를 나타내는 부사인지를 언어학적인 측면에서 구분한 문헌은 많이 출판되어 있다. 그러나, 많은 동사 표현 부사들과 상태표현

부사들은 상에 의한 표현의 제약이 없다[5]. 즉, 문의 술부가 동작진행형에도 상태진행형에도 넓게 사용된다. 그러나, 다음의 부사들은 진행형의 구분이 가능하며, 진행형 문맥의 이해를 도울 수 있다.

災害時にする案がともに検討されている。
 제해시의 피난소로 하는 안이 함께 검토되고 있다.

(예 10) 동작성 부사를 통한 상의 구분

(예 10)에서 “検討されている”은 “검토되고 있다”와 “검토되어 있다”라는 2종류의 대역어로 대응될 수 있다. 이와 같은 상황에서 “検討されている”의 문맥적인 의미는 부사 “ともに”에 의해 명확하게 동작진행형으로 구분된다. 이와 같이 동작진행의 전형성이 강하다고 생각되는 부사들은 “すぐ、いますぐ、直に、ともに、だんだん、ますます、しばしば、たまに”등이며, 이것들을 진행형의 구분에 사용하고자 한다.

또한, (예 11)처럼 “すでに、もはや、とうに、もう”등의 상태진행의 성질이 강한 부사를 이용하여 애매성이 있는 진행상을 구별한다.

工場はすでに取り壊しに入っている。
 공장은 벌써 해체에 들어가 있다.

(예 11) 상태성 부사를 이용한 상의 구분

이와 같이 상적인 전형성이 강한 부사들이 문 속에 나타난 경우, 술부에 나타난 “ている”형태의 진행형은 부사에 의해 각각 동작진행과 상태진행으로 결정할 수 있다.

3.2.2 일반적인 부사어를 이용한 상의 구분 방안

체언과 부사격 조사로 구성되어 있는 일반적인 부사어(이하: 부사어)를 이용한 상의 구분은 일본어의 부사격 조사와 한국어의 부사격 조사 사이에 다의성이 발생하므로 상적인 의미를 추출하는 것이 간단하지 않다. 실제의 문에서 부사어를 작성하는데 자주 사용되는 조사 “に”와 “で”는 아래와 같이 복수의 한국어 조사에 대응한다.

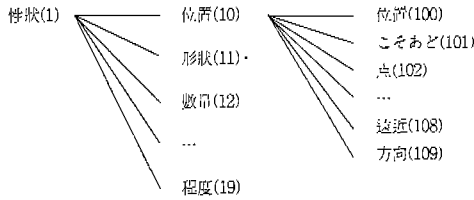
- に : 에, 에게, (으)로, (으)러, 와/과, 게
- で : 에서, (으)로

그러나, 조사의 다의성 해소에 관한 연구가 활발하게 행하여지고[13], 실제의 시스템에서도 상당한 부분이 개선되어 있으므로 본 논문에서는 조사의 다의성에 의해 발생하는 문제는 일어나지 않는 것으로 한다.

상의 구분을 위하여 부사격 조사 가운데에서도 움직임의 대상, 일어나는 장소, 그리고 시간적 관계 등을 의미하는 “에, 에서, 에게, (으)로, 까지” 등의 조사와 결합한 체언의 의미를 참조한다. 이와 같은 범주의 부사격 조사와 결합한 체언 가운데 장소와 시간을 나타내는 체언과 결합한 부사어를 문으로부터 뽑아, 진행상의 구분에 이용한다. 상의 구분에 도움을 주는 부사어가 되는 대표적인 일본어의 조사

로서 “に,で,から,まで”등이 있다.

또한, “체언+부사격 조사”의 형태를 취하는 부사어의 경우는 부사어가 포함된 체언의 의미적 특질에 따라 그 성질이 결정된다. [8]에서는 일본어 단어를 의미에 따라 (그림 1)처럼 1,000개로 분류하고, 현재 우리의 일한 기계번역 시스템은 이 의미분류 체계에 따라 체언을 분류하고 있다. 사전에 등록되어 있는 모든 체언은 그 분류에 따라 의미속성을 가진다.



(그림 1) [오우노]가 제안한 체언의 분류방법의 예

“ている”의 애매성을 해소하기 위하여 부사어를 이용할 때 필요한 체언의 속성은 다음과 같다.

- 시간성 체언류 : 시간(150), 시기(151), 시각(152), 기간(153)...
- 회수성 체언류 : 수(121), 회(122), 얼마(129)...
- 장소성 체언류 : 시제(030), 육지(031), 지역(700-709)...
- 유정성 체언류 : 인물(500-599), 식물(050-059), 동물(060-069)

체언의 개별적인 의미와 결합한 조사는 부사어가 되어, “ている”가 나타날 때의 문맥적인 상황을 유추하는 것이 가능하다. 예문과 예문에 적용하는 패턴을 (예 12)에 보인다.

この雑誌は若い女性によく読られている.
 <유정성 체언 + “에게”>
 彼とは公園で待ち合っている.
 <장소성 체언 + “에서”>
 彼に3回まで機会を興えている.
 <회수성 체언 + “까지”>
 あのドアは3時まで開いている.
 <시간성 체언 + “까지”>
 公園に大きい松ノ木が立っている.
 <장소성 체언 + “에”>

(예 12) 부사어를 이용한 “ている”의 애매성 해소

이것들은 순수부사와 함께 진행형을 구별하는 판단기준으로 이용할 수 있다.

순수부사와 부사어를 이용하여 “ている”의 상적인 애매성을 해소하는 방안을 알고리즘으로 정리하여 (그림 2)에 보인다.

```

If (제2동사 이면서 수동태가 아닌 경우)
    Return (고 있다)
/* 순수부사 */
else if (동작성 부사가 있다)
    return (고 있다)
else if (상태성 부사가 있다)
    return (어 있다)
        
```

```

/* 부사어 */
else if (유정성체언류 + “에게” 가 있다)
    or (시간성체언류 + “에” 형태의 부사어가 존재)
    or (회수성체언류 + “까지” 형태의 부사어가 존재)
    or (장소성체언류 + “에서, (으)로” 형태의 부사어가 존재)
    return (고 있다)
else if (장소성체언류 + “에” 형태의 부사어가 존재)
    or (시간성체언류 + “까지” 형태의 부사어가 존재)
    return (어 있다)
/* 문맥 의미파악 불가 : default value */
else if (제2동사)
    return (고(어) 있다)
else
    return (어(고) 있다)
        
```

(그림 2) 제 2, 4동사를 대상으로 한 “ている”형의 상적인 애매성을 해소하는 알고리즘

4. 실험 및 검토

실험에서 이용한 일한 기계번역 시스템은 크게 나누어 해석, 변환, 생성의 3단계로 구성되어 있다. 형태소 해석 부분은 차등 접속규칙을 이용한 2문절 최장 일치법을 적용하고 있으며, 변환 부분에서의 용언과 조사의 다의성 해소는 언어 패턴을 이용한 패턴 매칭 기법으로 처리하고 있다[13]. 생성부분은 자연스러운 번역문 출력을 위하여 한국어 생성에 필요한 술부의 변환 테이블을 이용한다[14]. “ている”의 상적인 애매성 해소는 변환 부분에서 처리하는 추가부분에 해당한다.

4.1 실험의 대상

본 논문에서 제안한 “ている”의 애매성 해소 방안이 어느 정도 효과가 있는지를 확인하기 위하여, 아사히 신문기사 중에서 사회, 경제, 국제, 스포츠, 인물, 칼럼 등 6개면에 대한 기사를 대상으로 약 8만개의 문으로부터 “ている”가 포함되어 있는 15,018개의 문을 추출하였다. 또한, 실험을 행한 시스템의 사전에는 현재 약 12만 단어가 등록되어 있으며, 그 중에서 동사의 수는 17,776개였다. 이것들을 대상으로 미리 사전 속의 동사를 5타입으로 분류하였다. 종래의 분류에 의한 빈도와 본 논문에서 제안한 분류에 의한 빈도수를 <표 2>에 나타낸다.

<표 2> 실험용 데이터의 분포

구분	분류	사전에서의 빈도
종래의 3타입 분류	제 1동사	14,673
	제 2동사	1,891
	제 3동사	912
	소 계	17,776
본 논문에서 제안한 5타입의 분류	제 1동사	5,395
	제 2동사	7,515
	제 3동사	1,208
	제 4동사	2,746
	제 5동사	912
소 계		17,776

4.2 실험 결과 및 검토

“ている”를 3타입으로 나누어 번역한 종래의 방식에서는 <표 3>과 같이 약 77.9%의 성공률을 보였다. <표 3>에서 문의 수가 15,072개 인 것은 하나의 문에 “ている”가 두개 이상 나타난 경우가 있어서 실제 15,018개의 문보다 높게 표시되었다. 즉, 15,072개의 진행형 동사가 등장한 것을 편의상 문수로 표시한 것이다.

<표 3> 종래의 방법에 의한 실험 결과

구분	문수	성공수	성공률
제 1동사	12,172	9,220	75.7
제 2동사	2,754	2,382	86.5
제 3동사	147	143	97.3
합 계	15,072	11,745	77.9

술어를 5타입으로 나누어 “ている”의 대역어를 선정해 본 결과, 약 82.6%의 번역률이 얻어졌다. 5타입으로 나눈 분류만으로도 약 4.7%의 번역률이 올랐다. <표 4>에 5타입으로 나누어 처리한 결과를 나타낸다.

<표 4> 술어의 5타입에 의한 “ている”의 실험결과

구분	문수	성공수	성공률
제 1동사	2,860	2,860	100
제 2동사	8,397	6,531	77.8
제 3동사	785	785	100
제 4동사	2,883	2,128	73.8
제 5동사	147	143	97.3
합 계	15,072	12,447	82.6

<표 5> 실험결과

구분	문수	성공수	성공률	
제 1동사	2,860	2,860	100	
제 2동사	수동형이 아닌 경우	2,565	2,565	100
	순수부사	79	72	91.1
	부사어	310	246	79.4
	기타	5,443	3,829	70.3
제 3동사	785	785	100	
제 4동사	순수부사	67	61	91.0
	부사어	237	196	82.7
	기타	2,579	1,848	71.7
제 5동사	147	143	97.3	
합 계	15,072	12,605	83.6	

우선, 제 1동사, 제 3동사는 술어의 분류에 따라 진행상이 분명히 결정되므로 100%의 성공률을 얻는다. 제 5동사의 경우는 진행상을 제거하고 과거의 의미로 번역하였으며 문맥의 의미가 연결되는 정도를 성공으로 보았을 때 97%의 성공률을 보였다.

여기서 우리는 술어의 분류만으로 진행상의 구분이 불가능한 제 2, 제 4동사를 대상으로 순수 부사와 부사어를 이용하여 (그림 1)의 알고리즘을 이용하여 성공률을 올리는 실험

을 행하였으며 그 결과는 <표 5>와 같다.

제 2동사로부터는 수동이 아닐 경우 “고 있다”의 일률적인 해석이 가능하므로 진행상의 구분이 100%가능하다. 그리고 상적인 전형성이 강한 순수부사를 이용한 “ている”의 진행상적 구별은 91%의 성공률을 보여 애매성 처리에 상당한 효과가 있음을 입증하였다. 그러나, 전형적인 동작진행과 상태진행을 의미하는 부사의 출현빈도가 너무 낮아서 (애매성이 남는 문의 1.7%), 이 방법은 상적인 전형성이 강한 순수부사가 그다지 나타나지 않는 문을 대상으로 하는 경우에는 효과를 보기가 어렵다. 즉, 순수부사를 비교적 사용하지 않는 신문기사 보다는 순수부사를 자주 사용하는 수필이나 소설류 쪽에서는 효과를 볼 수 있을 것이다.

또한, 부사어를 이용한 “ている”의 애매성 해소는 약 81%의 성공률을 보였지만, 부사어의 출현수도 생각보다 적었다[약6.3%]. 그러나, 이 처리로 부사어의 출현빈도가 낮은 데도 불구하고 성공률은 1% 정도 더 높아진 것에 주목할 필요가 있다.

제 2동사와 제 4동사에서 “기타”는 본 논문에서 제안한 수법으로는 어떠한 특징도 추출할 수 없는 문이었으며, 전체 문 수의 약 50%까지 이르고 있었다. 본 시스템에서는 “ている”에 대한 예비 대역어를 팔호를 사용하여 같이 표시하는 방식을 취하고 있지만, 만약 “기타”에 포함된 문으로부터 팔호를 사용하지 않고 대표적인 대역어를 취한다면 성공률은 70.6%가 되어 “기타”의 약 30%가 잘못된 대역어를 선정해 버린다.

본 논문에서 제안한 순수부사와 부사어를 이용한 진행형의 처리는 아무 처리도 하지 않을 경우의 70% 성공률과 비교하면, 순수부사 이용시의 91%나 부사어를 이용한 81%의 성공률은 상당히 향상된 수치라 하겠다.

앞으로 본 실험에서 참조된 부사들에 비해 상적인 전형성이 다소 약한 순수부사와 부사어에 관해서도 연구를 병행하여 “기타”의 비율을 낮출 필요가 있다고 생각된다.

전체적으로 보면, “ている”의 상적인 애매성을 해소하기 위해 제안한 본 방법은 83.6%의 성공률이 얻어져, 단순히 동작진행과 상태진행으로 나누어 처리한 종래의 방법보다 진행형의 대역어 선정에서 성공률이 5.7% 향상된 것이다.

다음은 잘된 번역과 잘못된 번역의 예를 유형별로 나타내고 원인을 분석한 것이다.

搜索は難航している。

수색은 난항이고 있다.(X) 수색은 난항이었다.(O)

- “난항하다”라는 번역어는 어색하므로 “난항이다”로 번역한 결과 “난항이고 있다”라는 어색한 표현이 되어서 정확한 번역이 되도록 제 5동사로 분류하였다.

志望者が 300人も増えている。

300명도 넘(어)고 있다.(X) 300명도 넘었다.(O)

- 제 4동사로 분류하였지만 진행형이 아닌 과거형으로 번역하는 것이 더욱 자연스럽다.

すでに韓国で市販されている。

이미 한국에서 시판되어 있다.(X) 이미 한국에서 시판되고 있다.(O)

- 상태진행의 부사가 등장함에 따라 동작진행의 번역이 행해지지 않았다.

5. 결 론

일본어에는 동작진행과 상태진행을 나타내는 진행형의 형태가 “ている”형을 취하므로 그 형태가 같다. 또한, 진행상의 구분이 문맥에 의존하므로 직접번역을 채택하고 있는 일한 기계번역 시스템에서 “ている”에 대한 대역어를 선정하기는 쉽지 않다. 그러나, 다행히도 일한 양국어에서 나타나는 동사의 진행형은 유사성을 가지고 있고 동사별로 진행의 형태가 고정된 경우가 많아서, 단순히 동사사전에 동작진행형이나 상태진행형을 고정하여 번역하여도 약 78%의 번역 성공률을 얻을 수 있었다. 본 논문에서는 거기에 더하여 진행형을 가지는 동사를 조금 더 세분화하였다. 즉, “1: 동작진행만으로 사용되는 술어”, “2: 일반적으로는 동작진행으로 사용되지만, 수동인 경우에는 상태진행의 형태로도 사용되는 술어”, “3: 상태진행으로만 사용되는 술어”, “4: 동작진행, 상태진행의 구별이 애매한 술어”, “5: 기타” 등의 상적인 의미분류코드를 술어별로 입력하여 세분화한다. 그리고 “2”, “4”형태의 술어로부터 진행형을 구별하기 위하여 부사와 부사어를 사용하는 방법을 제안하였다.

그 결과, 약 83.6%의 번역 성공률을 보여, 단순 3가지 분류에 의한 진행형의 번역처리에 비해 약 5.7% 정도 향상된 것을 알았다. 앞으로 진행형의 대역어 선정 성공률을 더욱 더 높이기 위하여 상적인 전형성이 다소 약한 순수부사와 부사어등에 대해서도 관심을 가지고 실험을 계속할 계획이다.

참 고 문 헌

[1] 長尾 真, “自然言語処理”, 岩波出版, 1996.
 [2] 손세모돌, “국어보조용언연구”, 한국문화사, 1996.
 [3] 김성덕, “우리말 형태론”, 탐출한사, 1996.
 [4] 益岡隆志 “基礎日本語文法”, くろしお出版, 1995.
 [5] 손남익, “국어부사연구”, 박이정출판사, 1995.
 [6] 황관호, “한일어 대조분석”, 명지출판사, 1993.
 [7] 교육부, “고등학교교문법”, 성균관대학교 대동문화연구원, 1991.
 [8] 金政仁, 大駒誠一, “日韓機械翻訳における拡張翻訳テーブルを用いた韓国語の生成方法”, 日本情報処理学会論文誌, Vol. 37, No.9, pp.1697-1707, 1996.
 [9] 田辺洵一, “くわしい国文法”, 文英堂, 1987.
 [10] 水谷静夫, “文法と意味”, 朝倉書店, 1983.
 [11] 金田一, “日本語動詞のアスペクト”, むぎ書店, 1976.
 [12] Vendler, Z., “Linguistics in Philosophy,” Ithaca, Cornell University Press, NewYork, 1967.
 [13] 朴哲濟, 李鐘赫, 李根培, “日韓機械翻訳における連語パターンを用いた変換手法”, 日本情報処理学会論文誌, Vol.38, No.

4. pp.707-718, 1997.

[14] 金政仁, 李鐘赫, 李根培, “日韓機械翻訳における構相テーブルに基づいた韓国語述部の生成処理”, 日本言語処理学会論文誌, Vol.5, No.2, pp.3-24, 1998.
 [15] 崔杞鮮, 金泰完, “日韓機械翻訳における現状および分析”, 日本言語処理学会第2回年次大会, 発表論文集, pp.433-443.
 [16] 김정인, 문경희, 이종혁, “일한기계번역시스템에서 진행형의 번역처리”, 정보과학회 추계학술발표대회, Vol.27, No.2, pp. 185-187, 2000.
 [17] 한명수, 문경희, 이종혁, “일한기계번역의 성능 향상을 위한 조동사 れる・られる의 애매성 해결”, 정보과학회 추계학술 발표대회, Vol.27, No.2, pp.188-190, 2000.
 [18] 김정인, 문경희, 이종혁, 이근배, “일한기계번역에 있어서 한국어 술부의 생성과 평가”, 한글 및 한국어 정보처리학술대회는문집, pp.329-337, 1996.
 [19] 정중락, 김정인, 문경희, 이종혁, 이근배, “일한기계번역시스템 COBALT-J/K의 성능평가”, 한글 및 한국어 정보처리학술대회는문집, pp.338-345, 1996.



김 정 인

e-mail : jikim@tmic.tit.ac.kr

1986년 계명대학교 통계학과(이학사)

1993년 일본 게이오대학 계산기과학전공 (공학석사)

1996년 일본 게이오대학 계산기과학전공 (공학박사)

1986년~1988년 국제컴퓨터엔지니어링 개발부

1996년~1998년 포항공과대학교 정보통신연구소 연구원

1998년~현재 동명정보대학교 컴퓨터공학과 조교수

관심분야 : 기계번역, 자연어처리, 정보검색 등



문 경 희

e-mail : khmoon@postech.ac.kr

1993년 부산대학교 전자계산학과(이학사)

1995년 부산대학교 전자계산학과(이학석사)

1995년~1998년 포항공과대학교 정보통신 연구소(연구원)

1998년~현재 포항공과대학교 컴퓨터공학과 (박사과정)

관심분야 : 기계번역, 자연어처리, 정보검색 등



이 종 혁

e-mail : jhlee@postech.ac.kr

1980년 서울대학교 수학교육과(이학사)

1982년 한국과학기술원 전자계산학과 (이학석사)

1988년 한국과학기술원 전자계산학과 (공학박사)

1989년~1991년 일본 NEC C&C 정보연구소 (초빙연구원)

1991년~현재 포항공과대학교 컴퓨터공학과(부교수)

관심분야 : 자연어처리, 한국어처리, 기계번역, 자동통역, 정보 검색, 문서요약 등임.