

# 일한번역시스템에서 진행형의 번역처리

김정인\* 문경희\*\* 이종혁\*\*

동명정보대학교 컴퓨터공학과\*, 포항공과대학교 전자계산학과\*\*  
jikim@tmc.tit.ac.kr, {khmoon, jhlee}@postech.ac.kr

## A Processing of Progressive Aspect in Japanese-Korean Machine Translation

Jung-In, Kim\* Kyung-Hee, Moon\*\* Jong-Hyeok Lee\*\*

Dept. of Computer Engineering, Tongmyong University of Technology\*  
Dept. of Computer Science, POSTECH\*\*

### 요 약

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 진행형을 구별하기 위하여 부사와 부사어를 사용하는 방법을 제안한다.

### 1. 서론

일본어와 한국어는 언어 계통 상 알타이어족에 속하여 문법체계가 유사하며 또한 동일 한자문화권에 속한다. 따라서, 일한 기계번역 시스템의 개발은 영한이나 중한 번역 시스템처럼 서로 다른 계통의 언어 간을 번역하는 시스템보다는 상대적으로 용이하다고 할 수 있다 [13-15]. 그러나, 현재, 개발된 시스템이나 개발중인 많은 일한 기계번역 시스템은 아직 해결해야 할 상당 부분의 과제를 가지고 있다[15].

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, 언어학적으로 고찰하고 그 결과로부터 실제로 일한 기계번역 시스템에 적용할 수 있는 방안을 제시한다.

### 2. 양언어의 진행형 구분

일한 양국어는 용언의 의미정보에 따라 상의 구분이 어느 정도 가능하다. 또한 술부의 문맥적인 의미의 상당 부분을 부사어에 의존하며, 시체에 따른 상적인 의미 변화 등, 상에 관한 언어학적인 성질은 거의 차이가 나지 않는다[2-6]. 그러나, 일본어에는 동작진행과 상태진행의 표현이 “ている”로서 같은 모양이나, 한국어에서는 사용자가 “고 있다”와 “어 있다”의 선택을 문맥적 상황으로부터 직관적으로 구분해서 사용하고 있다.

“ている”형태가 동작 진행형으로서 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 진행을 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, きえきる

그러나, <예1>을 보면 “知っている”은 한국어로 “알고 있다”로 대응되어 동작진행으로 구분되지만, 일본어에서는 “知る”라는 술어가 “知らない”라는 상태에서부터의 변화로 해석하여 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.  
대개의 일본인도 알고 있다.

<예1> 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 <예2>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.

<예2> 진행형을 없애고 번역한 예

결국, 일본어 동사와 한국어 동사의 상적인 구분이 일정한 대응 관계를 취하고 있지 않으므로 동사마다 상적인 대응 처리를 기술할 필요가 있으며, 대응관계의 규칙을 만들 필요가 있다.

**3. 진행상의 구분 방안**

“ている”에서 동작진행과 상태진행의 구분은 깊은 의미해석을 실행하지 않고는 판단하기 어렵다. 그러나, 한국어와 일본어의 유사성에 기초하면 술어별로 진행상의 구분이 어느 정도 가능하다. 종래의 일한 기계 번역 시스템에서는 사전의 정보를 이용하여 동작진행과 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 의존한다. 이와 같은 최소한의 사전 정보를 이용하여 1개의 역어를 결정해도 “ている”의 애매성이 약 78% 정도 해결되었다. 그러나 “ている”의 대역어를 사전정보에 의존함에 따라 상적 의미의 유연성이 결여되어, 대역어 선정의 오류가 당연한 결과로 받아들여지게 된다. 본 절에서는 “ている”의 애매성을 해소하기 위하여, 기계 번역 시스템에 적용 가능한 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.

**제1동사**

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

**제2동사**

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

**제3동사**

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

**제4동사**

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

**제5동사**

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

이 정보를 사전 속의 모든 동사 엔트리에 진행형의 번역정보로 입력한다. 5 개의 동사 타입에 대응하는 기본적인 한국어 대역어를 아래에 표시한다.

- 제1동사: 고 있다
- 제2동사: 고(어) 있다
- 제3동사: 어 있다
- 제4동사: 어(고) 있다
- 제5동사: 쓰다/었다

이와 같은 동사의 분류를 사용하면 제1동사, 제3동사, 그리고 제5동사는 “ている”형태의 한국어 대역어로서 각각 “고 있다”, “어 있다”, “쓰다/었다”를 취하는 것이 가능한 반면, 제2동사와 제4동사의 경우는 진행상의 의미정보만으로는 상의 구분이 불가능하다.

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

**3.2 순수부사에 관한 고찰**

부사마다 그것이 동작을 나타내는 부사인지, 상태를 나타내는 부사인지를 언어화적인 측면에서 구분한 문헌은 많이 출판되어 있다. 많은 동사 표현 부사들과 상태표현 부사들은 상에 의한 표현의 제약이 없으나, 다음의 부사들은 진행형의 구분이 가능하며, 진행형 문맥의 이해를 도울 수 있다.

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

<예3> 동작성 부사를 통한 상의 구별

<예3>에서 “検討されている”은 “검토되고 있다”와 “검토되어 있다”라는 2 종류의 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 구별에 사용하고 자 한다.

또한, <예4>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.

<예4> 상태성 부사를 이용한 상의 구별

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.

3.3 일반적인 부사어를 이용한 상의 구분 방안

체언과 부사격 조사로 구성되어 있는 일반적인 부사어(이하:부사어)를 이용하여 상을 구분한다. 다음의 알고리즘은 지금까지의 진행형 구분을 위하여 사용한 방법을 정리한 것이다.

<알고리즘1> 진행형 구분을 위한 알고리즘

```

If(제2동사 이면서 수동태가 아닌 경우)
    Return(고 있다)
/* 순수동사 */
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)
/* default value */
else if(제2동사)
    return(고(어) 있다)
else
    return(어(고) 있다)
    
```

4. 결론

일본어에는 동작진행과 상태진행을 나타내는 진행형의 형태가 “ている”형으로 표면적인 형태가 같다. 또한 그 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 표시하여 번역하여도 약 78%의 번역 성공률을 얻을 수 있었다. 본 논문에서는 거기에 더하여 진행형을 가지는 동사를 조금 더 세분화하였다. 그리고 동작진행과 상태진행을 구별할 수 있도록 술어의 문맥적인 의미를 파악하기 위하여 순수부사와 부사어를 사용한 알고리즘을 사용하였다.

문장에 포함된 부사나 부사어를 사용하는 제안 알고리즘을 사용할 경우, 진행형의 번역에서 나타나는 번역의 오류를 상당부분 해결할 수 있을 것으로 기대되며, 실제로 일한번역시스템에 적용해 볼 예정이다.

참고문헌

- [1] 나가오 다다시, “자연언어처리”, 이와나미 출판, 1996.
- [2] 손세모들, “국어보조용언연구”, 한국문화사, 1996.
- [3] 김성덕, “우리말 형태론”, 탐출한사, 1996.
- [4] 다다오카, “기초일본어 문법”, 구로시오출판, 1995.
- [5] 손남익, “국어부사연구”, 박이정출판사, 1995.
- [6] 황찬호, “한일어 대조분석”, 명지출판사, 1993.
- [7] 교육부, “고등학교문법”, 성균관대학교 대동문화연구원, 1991.
- [8] 오오노, “유어신사전”, 가꾸가와서점, 1991.
- [9] 다나베, “쿠와시이고꾸분뽀우”, 문영당, 1987.
- [10] 미즈다니, “문법과 의미I”, 아사꾸라서점, 1983.
- [11] 김전일, “일본어 동사의 아스펙트”, 무기서점, 1976.
- [12] Vendler, Z, “Linguistics in Philosophy”, Ithaca, Cornell University Press, NewYork, 1967.
- [13] 박철제, 이종혁, 이근배, “일한기계번역에서 언어패턴을 이용한 변환수법”, 일본 정보처리학회 논문지, 제38권 4호, pp.707-718, 1997.
- [14] 김정인, 이종혁, 이근배, “일한기계번역에서 양상태이블에 기초한 한국어 술부의 생성처리”, 일본 언어처리학회 논문지, 제5권 2호, pp.3-24, 1998.
- [15] 최기선, 김태완, “일한기계번역시스템의 현황과 분석”, 언어처리학회 제2회 연차대회, 발표논문집, pp.433-443.