

# 한일 기계번역에서 일본어 슬부생성

곽 중근, 김 은자, 이 종혁, 이 근배  
포항공과대학교 전자계산학과

## Japanese Predicate Generation in Korean-Japanese Machine Translation

Jong-Geun Kwak, Eun-Ja Kim, Jong-Hyeok Lee, Lee Geunbae Lee  
Dept. of Computer Science & Engineering  
Pohang University of Science & Technology

### 요 약

서로 다른 언어를 사용하는 사람들 사이에 언어의 장벽을 넘기위한 수단으로 컴퓨터를 이용한 기계번역이 각광 받고 있다. 본 논문에서는 한일 기계번역 시스템에서 한국어로 부터 일본어를 생성하는 과정에서 고려해야 할 사항에 대해 다룬다. 일본어는 한국어와 문법적 특성이 거의 유사하여 기계번역의 방법 가운데 직접 번역 방식이 선호되고 있다. 그러나 일본어의 한 슬부내의 문법형태소들은 한국어와 비교해 볼 때, 상이한 점이 많아서 다른 부분에 비해 자연스러운 생성이 어렵다.

본 논문에서는 한일 기계번역에서 일본어의 생성시 발생하는 문제점들을 해결할 수 있는 방안을 제시한다.

### 1. 서론

한국어와 일본어는 양언어 공학 같은 알타이 어족에 속하며 형태론적으로는 교착어에 속한다. 어순의 관점에서 살펴볼 때에도 두 언어 모두 SOV 언어에 속하며, 수식어가 피수식어 앞에 위치한다는 제약물 제외하면 어순에 큰 제약은 없다. 양언어가 가진 이러한 유사점으로 인하여 한일 기계번역 시스템에서는 직접번역방식이 변환번역방식이나 피복번역방식에 비해 상대적으로 유리하다고 볼 수 있는데, 이는 구문분석이나 의미분석을 행하지 않고 단어 수준에서 번역이 이루어지므로 시스템 구현에 있어서 그 구현 범위가 축소될 수 있고, 상대적으로 고품질의 번역 결과를 얻을 수 있기 때문이다[1].

그러나 한일 양언어가 서로 비슷하고, 문법적 특성이 비슷하다고는 하더라도 일본어에 있어서 빈번한 조사의 생략 문제, 한국어 슬부와 일본어 슬부의 순서 관계 및 문법 형태소들의 불일치, 그리고 조사의 번역 문제들이 직접번역방식을 이용한 한일 기계번역 시스템에서 해결해야 할 과제로 남는다. 본 논문에서는 이러한 문제점들 중에서 한국어 슬부를 일본어 슬부로 생성하는 과정에서 고려해야 할 문제점들에 관하여 논한다.

### 2. 일본어 슬부 생성시의 문제점

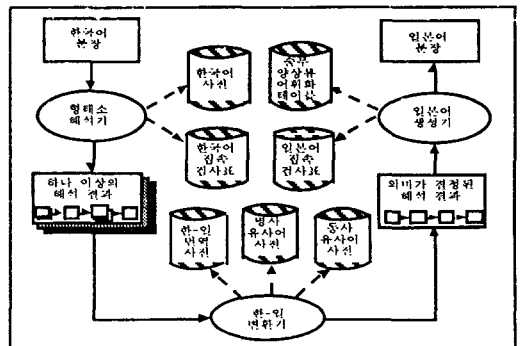
슬부 표현에 있어서 한일 양언어의 차이점은, 첫째 문장내의 단어 수준에서는 양언어의 순서 관계가 거의 같지만, 한 슬부내에서의 문법형태소 간 순서 관계가 일치하지 않고, 둘째 한국어 내에서 어간과 양상이 결합되어 표현된 슬부가 일본어에서는 양상이 가지는 의미가 어간에 포함되어 하나의 어간으로만 표현되는 경우가 존재하고, 역으로 한국어에서 하나의 어간으로 표현되는 슬부가 일본어에서는 어간과 양상의 결합형태

로 표현되는 경우가 존재한다. 마지막으로 한국어에서 “하다”가 명사와 결합하여 용언이 되는 것과 마찬가지로 일본어에서도 명사에 “즈ル”가 접속하여 용언이 될 수 있는데, 이 경우, 대부분은 한국어와 일본어의 표현이 일치하지만, 일부 용언은 한국어와 일본어가 서로 다르게 표현되는 경우가 존재한다.

양언어에 이러한 차이점들이 있으므로, 이들을 잘 처리하면 자연스러운 번역문을 얻을 수 있다.

### 3. 시스템 구조

한일 기계번역 시스템은 크게 세가지 주요 모듈로 분할하여 생각할 수 있다. 우선 입력문인 한국어를 형태소 분석하는 형태소 분석기, 그리고 분석된 결과로부터 한국어에서의 의미를 결정하는 변환기, 마지막으로 변환 결과로부터 일본어를 생성하는 생성기로 나누어 볼 수 있다. 다음은 전체적인 시스템의 구조를 나타낸 것이다.



[그림 1] 시스템 구조도

\* 본 논문은 한국통신 장기초 연구과제 "자동통의 신화개발을 위한 대화체 기계번역에 관한 연구(94-37)"에 의한 것이다.

형태소 분석기에서는 한국어를 형태소로 분할하는 일을 행하고, 변환기

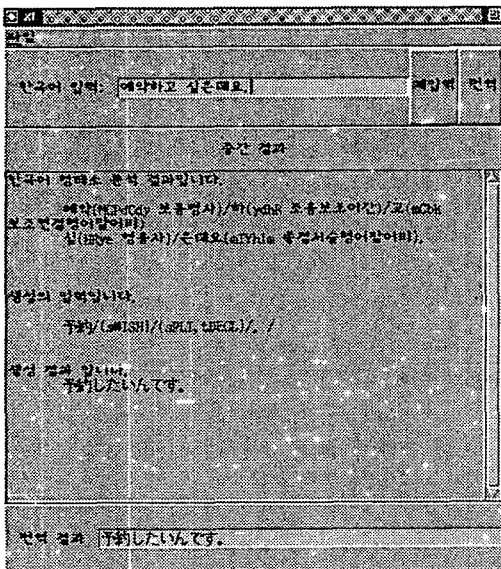
에서는 분석 가능한 여러개의 형태소 분석 결과 중에서 의미적으로 가장 적절한 하나를 선택한다. 이 결과를 토대로 얻어진 슬부의 의미자질을 이용하여 생성기에서는 일본어를 생성하게 된다.

다음은 한국어 슬부에 대한 형태소 분석결과 및 여기서 얻어진 의미자질을 이용하여 일본어를 생성할 때에 생성기의 입출력이 무엇인지를 보이고 있다.

슬부 표현	형태소 분석 결과
생성의 입력	어간 + 의미자질
생성 결과	해당 일본어
갑니다	가+ ㅂ니다 行/POLITE 行きます.
갔습니다	가+았+습니다 行/PAST, POLITE 行きました.
가고 싶습니다	가+고+싶+습니다 行/WISH, POLITE 行きたいんです.
갈 수 있습니다	가+ㄹ+수+있+습니다 行/POSSIBLE, POLITE 行くことができます.

[표 1] 형태소 분석 결과 및 생성기의 입출력

다음 그림은 실제 시스템이 입력문을 처리하는 예를 보이고 있다.



[그림 2] 실행예

한국어에서 필요한 의미자질을 뽑아내기 위해서 우선 한국어에서 양상류들이 어떻게 표현되는가를 알아보자.

### 3.1 한국어의 슬부의 구조

한국어에서의 양상류는 크게 본용언과 보조 용언의 결합에 의해서 표현되는 형태와 의존명사에 의해서 표현되는 두가지 형태가 존재한다[3].

#### 본용언 어간 + {보조적 연결어미 + 보조용언 어간}\* + 어미

여기서 보조적 연결어미와 보조용언 어간의 쌍은 한번도 쓰이지 않을 수도 있고, 여러 번 반복하여 쓰일 수 있다. 예를 들어 '먹고 싶지 않았습니' 라는 한국어의 슬부는

본용언 어간(먹) + 보조적 연결어미(고)  
+ 보조용언 어간(싶) + 보조적 연결어미(지)  
+ 보조용언 어간(않) + 았(시제선어말어미) + 습니다(어말어미)

의 결합형태로 볼 수 있다. 한국어에 있어서의 시상(aspect) 표현도 본용언과 보조용언의 결합형태로 나타난다[5]. 예를 들어 '먹고 있다', '먹어 버리다' 등의 시상 표현에서 나타나는 것처럼 보조적 연결어미 '고', '아어' 와 시상을 표현하는 보조용언 어간 '있다', '버리다', '말다' 등이 결합하여 표현되므로 그 형태는 동일하다고 볼 수 있다.

그리고 의존명사에 의해서 표현되는 양상류의 형태는 다음과 같다.

#### 본용언 어간 + 전성어미 + 의존명사 + 조사 + 있/없 + 어미

의존명사와 결합되어 양상류를 표현할 때 쓰이는 의존명사는 그 종류가 매우 한정되어 있다.

예를 들어 '간 적이 있습니다' 는

본용언 어간(가) + 전성어미(ㄴ)  
+ 의존명사(적) + 조사(이) + 있 + 습니다.

의 결합형태로 볼 수 있다. 이 두가지 형태 모두에서 시제의 표현은 어미 부분에서 시제를 나타내는 선어말어미에 의해서 표현된다. 이상에서 한국어 슬부에 나타나는 양상류의 구조를 살펴보았다. 한국어 슬부에서 뽑아낸 양상류를 이용하여 일본어 슬부를 생성하기 위해서, 일본어 슬부의 구조를 살펴보기로 하자.

### 3.2 일본어 슬부의 구조

일본어의 슬부는 다음의 형태로 표현된다.

#### 용언어간 + 어미류 + {조동사}\* + 종조사

조동사에 의한 슬부 양상류 표현은 せる, させる, れる, られる, できる 등과 같이 동사의 활용을 따르는 조동사와, たい, ない, らしい 등과 같이 형용사의 활용을 따르는 조동사, そうだ, ようだ 등과 같이 형용동사의 활용을 따르는 조동사 및 복수활용하는 조동사와 변화형이 없는 조동사가 있다.

여기서 동사, 형용사, 형용동사의 활용을 따르는 조동사는, 한국어 슬부의 표현과 거의 같은 형태를 띄게 된다. 즉 본용언 활용과 보조용언 활용이 결합된 형태와 유사하다.

일본어 슬부에서 시제의 표현은 특수활용형 조동사에 의해서 표현되고, 종조사에 의해서 표현되는 양상류로는 ぶん(か), 역접(が), 원인(から), 나열(たり), 감탄(ね), 조건(ば), 동시(ながら), 심층(ほど) 등이 있으며, 이것들은 슬부의 맨 마지막 부분에 위치한다[6].

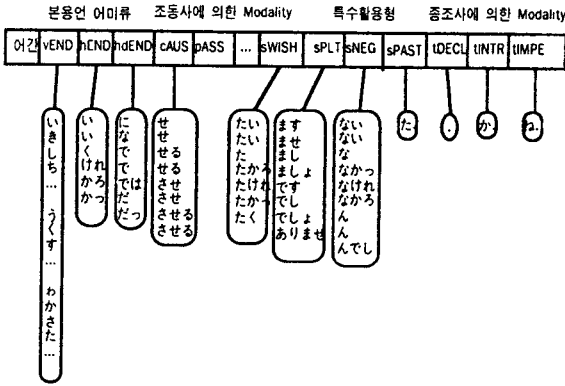
이제 한국어 슬부로부터 얻어낸 양상정보를 이용하여 일본어를 생성하는 과정에 대해서 살펴보기로 하자.

#### 4. 일본어 슬부 생성

양언어의 슬부내의 순서관계를 바로 잡아주고, 어간과 양상류를 분리하여 처리하기위해서, 본 시스템에서는 MFOLT<sup>1</sup>라 불리는 자료구조를 이용한다. MFOLT는 슬부내에서 표현되는 양상류들을 상대적인 선행순서관계를 고려하여 늘어놓고, 각 양상 자질에 대응하는 일본어의 표층 형태소들을 접속정보와 함께 가지고 있는 자료구조이다[2] [1].

기존에 연구된 시스템에서는 한국어와 일본어의 슬부에서는 상당한 차이가 있음에도 불구하고 형태소 사이에서 직접 번역을 시도한 경우가 있는데, 이는 번역문의 품질이 낮아질 뿐만 아니라 상대적으로 사전의 크기도 커지는 단점이 존재한다[7].

현재 본 단일 기계번역 시스템에서 MFOLT에 등록된 양상류의 의미자질은 실험 수준으로 약 30개 정도이며, 이것은 앞으로 더 확장시킬 것이다. 이것을 개략적으로 보이면, 다음의 그림과 같다.



[ 그림 3 ] MFOLT 구조

여기서 일본어 슬부를 생성하는 과정은 다음의 3 단계로 나누어 질 수 있다.

#### 4.1 슬부 생성 과정

##### 1. 의미 자질 활성화

한국어에서 추출한 양상류에 해당하는 의미자질을 MFOLT 상에서 활성화시킨다.

##### 2. 특수 대역어 처리

한국어와 일본어 사이에 서로 다른 양상의 표현이 존재하는 경우에 이를 처리하는 과정이다. 이것은 한국어나 일본어의 용언과 양상류의 결합이 하나의 용언으로 번역되어야 할 경우에 처리된다.

예를 들어 한국어에서는 '주시다'가 '주다'의 존칭표현이지만, 일본어에서는 주다에 해당하는 동사인 'やる'와 주시다에 해당하는 동사 'あげる'가 따로 존재한다. 이렇게 어간과 양상류의 쌍이 양언어에서 서로 대응되지 않는 상황에서 한국어의 용언 '주다'에 양상류인 겸양을 더해 '주시다'가 되는 것처럼, 일본어에서 'やる'에 겸양자질을 더해 'あげる'를 생성할 수는 없으므로 이 경우 처음부터 'あげる'를 생성을 위한 용언으로 선택하고, 겸양의 의미자질을 없애야 한다. 이것은 MFOLT 자료구조 상에서 겸양의 의미자질을 비활성화시키고 용언어간을 'やる' 대신 'あげる'로 바꿔줌으로써 간단히 해결할 수 있다.

이와는 반대의 경우로, 한국어 용언 '모르다'는 일본어로 표현될 때에 일본어에 '모르다'라는 용언이 없으므로 '알다'('知る')에 부정의 양상이 결합된 형태로 표현해 주어야 한다. 이것은 생성을 위한 용언을 '知る'로 바꾸고, MFOLT 상에서 부정의 의미자질을 활성화시킴으로써 구현이 가능해진다.

#### 3. 활성화된 MFOLT 상의 형태소 추출

MFOLT 상에서 활성화된 각각의 의미자질들에 대해서 한개 이상의 일본어 형태소들이 달려있는데, 활성화된 모든 의미자질들에 대해서 접속이 가능한 하나의 path를 이루는 형태소들을 모두 골라낸다.

각 형태소들은 좌접속 정보와 우접속 정보값을 갖고 있다. 그리고 각각의 좌우 접속정보를 이용하여 두개의 형태소가 서로 접속이 가능한지를 매트릭스를 이용하여 검사한다. 이를 이용하여 각기 활성화된 의미자질의 형태소중에서 접속가능한 것들을 서로 연결함으로써, 일본어 슬부가 생성된다. 또, 용언어간과 의미자질의 결합형태가 양언어에서 서로다른 경우의 문제점도 MFOLT를 이용함으로써 해결할 수 있다.

#### 4.2 “하다”가 붙는 슬부처리

한국어에 있어서 명사와 조용보조어간인 “하다”가 결합하여 슬부로 쓰이는 경우가 많다. 이 결합의 대부분은 한자어 명사이지만, 일부 고유명사에도 “하다”가 붙어서 용언이 되는 경우가 존재한다. 예를 들면 '말하다'가 고유명사 '말'에 조용보조어간 “하다”가 결합한 형태이다.

이처럼 명사에 “하다”가 접속하여 용언이 되는 경우, 명사와 “하다” 결합형 용언으로 둘 다 등록하는 방법과, 명사만을 등록하고 “하다”가 결합한 형태로 나오는 경우에 형태소 분석기가 절차적으로 명사와 조용보조어간 “하다”를 분리해 낼 수 있도록 하는 두가지 방법이 존재한다. 전자의 경우에 사전의 크기가 매우 커지므로, 명사만을 등록하고 접속정보로 “하다”의 결합 가능성을 표시하게 된다. 한국어에서 “하다”가 접속하여 용언이 될 때 이 용언이 가질 수 있는 품사는 동사나 형용사 두가지로 한정된다.

한국어에서 많은 명사에 “하다”가 접속하여 용언이 되는 것과 마찬가지로 일본어에서도 한자어 명사에 대해서 “する”가 접속하여 슬부를 이루는 경우가 한국어에서처럼 빈번하게 발생한다.

그런데 한국어에서의 “하다” 접속형 슬부가 일본어에서 “する” 접속형 슬부로 대응되지 않거나, 그 반대로 일본어에서의 “する” 접속형 슬부가 한국어에서 “하다” 결합형 슬부로 대응되지 않는 경우가 발생한다.

#### 4.3 “하다” 접속 용언의 일본어 표현

다음은 “하다” 접속하는 한국어 슬부와 일본어 슬부와의 관계를 보인 것이다.

##### 1. “하다” 접속 용언이 일본어에서 “する” 결합형 동사로 대응 되는 경우.

이 경우는 가장 일반적인 경우에 해당하며, 사전에 등록할 때 한국어 명사를 등록하고 대역어인 일본어 명사의 형태소 정보를 등록할 때 “する”가 접속가능하다는 정보를 함께 등록한다. 예를 들어 “예약”에 해당하는 일본어 “予約”는 “예약하다”에 대해서 “する”를 결합시킨 형태인 “予約する”로 대응한다.

이경우에 양언어를 대응시키기 위해서 사전에 입력하여야 할 정보의 내용은 한국어의 “예약”에 대해서 품사가 명사라는 것과 이것이 하다와 접속하여 용언이 될 수 있다는 정보를 등록한다. 대역어인 “予約”에 대한 정보로도 마찬가지로 품사정보가 명사라는 것과 “する”가 접속가능하다는 것을 명시한다.

사전 정보

<sup>1</sup> Modality Feature Ordering & Lexicalizing Table

예약 {명사, “하다”접속 가능}

→ 予約{명사, “する” 접속가능}

2. “하다” 접속 동사가 일본어에서 일반 용언으로 대응되는 경우.

다음의 예는 한국어에서 “하다” 접속 동사가 일본어에서는 “する” 결합형 용언이 되지 않는 것을 보이고 있다.

차지하다 → 占める

말하다 → 言う

이 경우에 사전에 등록하는 방법은

차지 {명사, “하다”접속 가능},

→ 占有{명사},

→ 占める{동사}

자랑 {명사, “하다”접속 가능},

→ 自慢{명사},

→ 誇る{동사}

로 등록하는 방법과

차지 {명사} → 占有{명사}

차지하 {동사} → 占める{동사}

자랑 {명사}, → 自慢{명사}

자랑하 {동사} → 誇る{동사}

로 한국어의 형태소 자체를 나누어서 등록하는 2 가지 방법을 생각할 수 있다.

3. “하다” 접속 형용사가 일본어에서 형용동사로 표현되는 경우.

소란하다 → 騒乱だ

불리하다 → 不利だ

4. “하다” 접속 형용사가 일본어에서 형용사로 표현되는 경우.

야속하다 → つれない

2번, 3번, 4번의 경우는 한국어에서 “하다” 접속형 용언이 일본어의 “する” 접속형 동사로 대응되지 않는다는 동일한 문제점을 갖고 있으므로, 사전에 등록하는 방법도 2번에서 제기한 방법과 마찬가지로 두가지의 방법을 생각할 수 있다. 사전 등록시, 2번과의 차이점은 단지 일본어 대역어에 대한 접속정보로 2번은 일반동사의 접속정보를 등록하지만, 2번이나 3번의 경우에는 형용동사, 혹은 형용사의 접속정보를 각각 등록하여 주면 된다.

따라서 한국어의 “하다” 접속동사의 대역어를 찾을 때에는 대역어가 1번에서 4번 중의 어느 형태로 표현되는지에 관한 정보를 대역사전에 기술하여 두고, 한국어에서 “하다” 접속형이 나오면 일본어를 생성할 때 이 정보를 토대로 “する” 를 이용하여 일본어를 생성할 것인지, 아닌지의 여부를 결정한다.

5. 결론

이상에서 직접 번역 방식을 채택한 한일 기계번역 시스템을 구현할 때 술부내 문법형태소들의 순서 관계가 서로 일치하지 않는다는 것과, 한국어나 일본어중의 한 언어에서 양상이 결합된 형태의 용언이 다른 언어로 표현될 때 동일한 의미의 용언이 존재하지 않아서 용언과 양상류를 따로 처리해야할 필요성이 있다는 것, 그리고 한국어에서 명사와 “하다”가 결합하는 용언과 일본어에서 명사와 “する”가 결합하여 용언이 될 때 이 둘이 서로 일치하지 않는 경우를 살펴보았다. 그리고 이러한 차이점으로 발생할 수 있는 직접 번역 방식의 문제점들을 해결할 수 있는 방법론에 관해 정리하였다.

참고 문헌

- [1] 김 은자, “언어 패턴에 기반한 일한 기계 번역 시스템”, 포항 공대 전산학과 졸업논문, 1994
- [2] 김 은자, 이 종혁, “일-한 기계 번역 시스템의 한국어 생성에서 양상류 의미 자질을 이용한 술부 처리”, 한글 및 한국어 정보처리 논문집, 1993, pp 547 ~ 557.
- [3] 이 수현, 한 광록, “한국어의 사상과 양상정보추출에 관한 연구” 인지과학 제1권 제2호, 1989, pp 255 ~ 275.
- [4] 고 영근, 남 기심, “고등학교 문법 자습서” 탐출판사
- [5] 안 동언, 이 종혁, “영한 기계번역 시스템: 일본어 생성 시스템 환경에서의 한국어 생성 시스템의 개발” 인공지능소식지, pp 41~45.
- [6] 김 영진, “재미있는 김 영진 일본어 문법”, 일본어뱅크
- [7] 김 태석, “A Study of Japanese-Korean Machine Translation Based on Connection Forms of Semantics”, 1992
- [8] 이 한섭, “일어학 개설”, 한신문화사, 1989