

일-한 기계 번역 시스템의 한국어 생성에서  
양상류 의미자질을 이용한 술부 처리

\*김 은 자, \*\*허 남 원, \*이 종 혁

\*포항공과대학 전자계산학과, \*\*계명전문대학교 전산정보처리과

Synthesis of Korean Predicates in Japanese-Korean MT System  
Using Semantic Features for Modality

\*EunJa Kim, \*\*NamWon Heo, \*Jong-Hyeok Lee

\*Dept. of Computer Science, Pohang Institute of Science & Technology

\*\*Dept. of Computer & Information Science, Keimyung Junior College

본 논문에서는 일-한 기계 번역 시스템에서 한국어 생성에 대한 효율적인 방법을 제안한다. 일본어와 한국어는 대부분의 문법 체계가 비슷하지만 어절 내, 특히 술부 내에서는 문법 형태소의 어순이 일치하지 않고 형태소 간에도 일대일 대응이 불가능한 경우가 많다. 또한 일본어 용언에 부정의 조동사가 사용된 문장을 한국어로 번역할 때에도 부정의 뜻을 내포하는 한국어 용언이 존재하면 부정적 역어로 생성하여야 한다. 이러한 점으로 인하여 형태소 대 형태소를 일대일로 대응시켜 생성하면 자연스럽게 올바른 한국어 문장을 얻기 어렵다. 따라서 기본적으로 직접 번역 방식을 이용하면서 술부의 생성을 위하여 양상류 의미자질을 이용하는 방법을 제안한다. 본 논문에서 제안한 한국어 생성 방법은 기존의 방법보다 적은 사전 작업으로 간단하게 자연스러운 한국어 표현을 얻을 수 있었다.

## 1. 서론

한국어와 일본어는 대부분의 문법 체계가 비슷하고 특히 어절 단위에서는 양언어의 어순이 거의 일치한다. 따라서 일-한 번역 시스템에서는 어절(문절) 단위로 양언어를 일대일(1:1) 대응시키는 직접 번역 방식이 많이 이용

되고 있다. 직접 번역 방식의 번역 과정은 형태소 분석, 변환, 생성의 3 단계로 나누어질 수 있다.

그러나 한·일 양언어의 어절 간에는 어순이 일치하지만, 술부내의 문법 형태소 간에는 많은 어순의 차이점이 있다. 또한, 일본어의 문법 형태소에 대한 부정 표현이 한국어에서 문법 형태소가 부정적 의미의 본용언으로 대응되는 경우가 종종 있다. 그리고, 양언어 형태소 간의 일대일 대응이 되지 않는 경우가 많다. 이러한 점으로 인하여 직접 번역 방식의 시스템에서는 자연스러운 한국어 술부를 생성하기가 어렵다. 본 논문에서는 한국어의 술부에 나타나는 문법 형태소들의 부분 순서 관계를 표현한 술부 양상류 어휘화 테이블과 한국어 접속 정보를 이용하는 한국어 생성 방법을 제안한다.

## 2. 기존의 연구

직접 번역 방식의 일한 번역 시스템에서의 한국어 생성은 형태소 간의 일대일 대응에 의해 이루어져 왔다 [1]. 그러나 술부 내에서의 일본어-한국어의 어순 불일치와 양언어 간에 형태소가 일대일 대응이 되지 않는 점으로 인하여 자연스러운 한국어 술부를 생성하기 어렵다. 다음은 [4][5]에서 지적한 문제점들이다.

1) 형태소 간에 일대일 대응이 되지 않는다.

① 野球を見るが,

야구를 보지만



야구를 보지만

② これは 何ですか.

이것은 무엇입니까?



이것은 무엇입니까?

③ 野球を見た.

야구를 보았.



야구를 보았다.

2) 술부 내에서 어순이 일치하지 않는다.

④ 野球を見ません.

야구를 봅니 + 않다.



야구를 보지 않습니다.

3) 일본어 동사의 부정형에 해당하는 한국어가 존재한다.

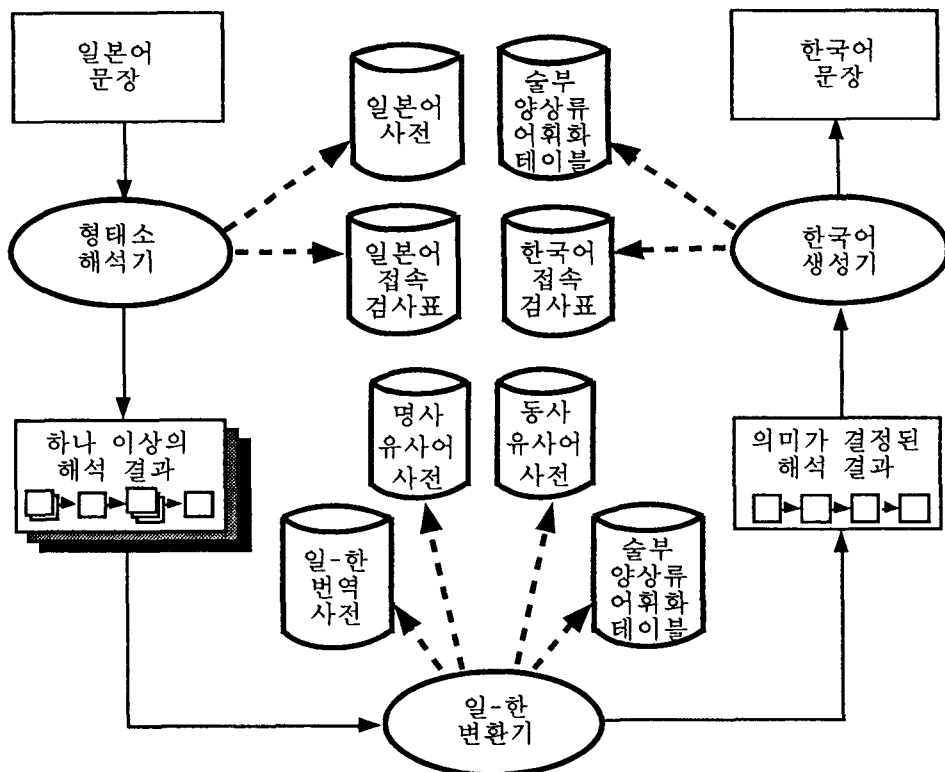
⑤ 私は知りません.

나는 알지 않습니다.      ■      나는 모릅니다.

이 문제들의 해결 방법으로 [4]에서는 용언 테이블을 이용하였다. 용언테이블이란 각 용언이 가질 수 있는 가능한 모든 의미자질의 조합에 따라 한국어 역어를 등록한 것을 말한다. 그러나 이 방법은 용언마다 테이블을 가지고 있어야 하고 테이블마다 가능한 의미자질에 따른 모든 용언의 형태를 등록해 두어야 하기 때문에 사전 크기도 매우 커질 뿐만 아니라 사전 작업에 많은 노력이 든다.

그러나, 한국어 슬부에서 양상류에 관한 정보 (시제, 상, 양상, 법 등) 를 표현하는 방법이 용언마다 다른 것이 아니고 보조용언, 선어말어미, 어미로 표현하기 때문에 용언마다 표현할 필요는 없다. 우리는 일본어 슬부의 양상류에 관한 의미자질을 추출하여 이를 기반으로 한국어 슬부를 생성하는 방법을 구현하였다.

### 3. 시스템 구조



[그림 1] PTrans 시스템의 전체 구조도

일한 기계번역 시스템(PTrans)의 주요한 구성은 [그림 1]과 같다. 전체 시스템은 형태소 분석기와 변환기 그리고 생성기로 구성되어 있다. 시스템에서는 일-한 번역사전, 일본어 형태소와 한국어 형태소에 대한 접속검사표, 의미자질 어휘화 테이블(PFOLT)과 명사와 동사의 관련어 계층 구조를 이용한다.

PTrans 는 일본어 문장을 입력으로 받아서 먼저 형태소 해석을 한다. 형태소 해석은 CYK 법과 일본어 형태소간의 접속가능성 검사로 이루어진다. 이 방법으로 가능한 모든 해석 결과들을 얻고 형태소 수 최소법과 형태소간의 접속 강도를 이용하는 휴리스틱[3]으로 해석 결과들을 우선 순위에 따라 정렬한다. 이 정렬된 형태소 해석 결과들은 변환기에게 넘겨져서 여러가지의 의미를 가지는 조사, 어미, 용언의 의미 결정이 이루어진다. 변환은 일본어 형태소의 각 의미(역어)가 선택되기 위한 공기 패턴(collocational pattern)을 형태소 분석 결과와 매칭(matching)하여 하나의 의미를 결정한다. 이 공기 패턴은 예외사례를 가지고 있어서 관용구의 처리를 쉽게한다. 이 때, 관련어 계층 구조(thesaurus hierarchy)를 통한 매칭(matching)이 이루어진다. 이 의미 결정의 결과 중에서 가장 선호도가 높은 해석 결과가 생성기로 넘어간다. 생성기는 모호성이 제거된 결정된 해석 결과에서 술부의 의미자질을 PFOLT 에 활성화 시키고 부정어, 사역어 등의 처리를 한후 PFOLT 의 순서관계를 따라 한국어 술부를 생성한다. 어미 등의 이형태 처리와 음운 축약, 불규칙 처리가 이때 이루어진다. 술부의 생성이 끝나면 한국어의 접속정보를 이용하여 조사의 이형태 처리를 하여 전체 한국어 문장을 생성한다.

#### 4. 한국어 생성

생성기는 출력 언어인 한국어를 생성하는 부분으로 특히 고려되어야 할 점은 술부의 생성으로 보조용언의 나열 순서, 어미의 이형태처리, 불규칙 용언의 처리, 음운 변화의 처리이다. 한국어의 술부에 나타나는 문법 형태소들의 부분 순서 관계를 표현한 술부 양상류 어휘화 테이블과 한국어 형태소의 접속 정보를 이용하여 한국어를 생성한다.

##### 4.1. 사전의 내용과 의미자질

일-한 번역 시스템의 사전 내용으로 일본어 표제어, 일본어 접속정보, 번역될 한국어, 한국어 접속정보가 등록된다. 술부에 시제, 양상 등의 의미를

더해 주는 조동사나 어미의 경우에는 한국어 역어 대신에 의미 자질이 등록된다. 의미자질로는 과거, 겸양, 추측, 피동, 소망, 가능, 부정 등의 의미가 사용된다. 용언의 경우에는 한국어 용언의 사역 형태와 부정어가 존재할 경우의 부정어, 부정어의 접속정보가 추가로 등록된다. [그림 2]는 동사 '知る'에 대한 사전 내용을 보여준다.

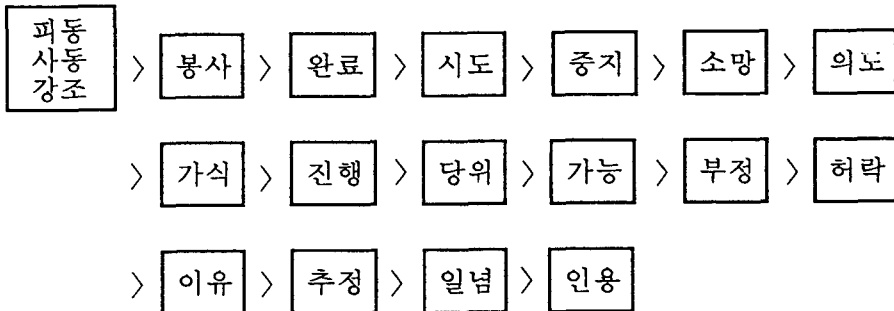
일본어 표제어	知		좌접속정보	200	우접속정보	226	
한국어 역어	알	접속정보	dRYa	사역형태	Caus8	피동형태	Pass5
부정적 역어	모르	접속정보	dImI R	사역형태	Caus8	피동형태	Pass5

[그림 2] 일-한 번역 사전의 내용

#### 4.2. 술부 양상류 어휘화 테이블 (PFOLT: Predicate Feature Ordering and Lexicalizing Table)

일본어에서는 조동사가 용언에 결합되어 술부의 의미를 표현하는 반면, 한국어에서는 보조 용언, 선어말 어미, 어말 어미로써 용언에 의미를 더해준다.

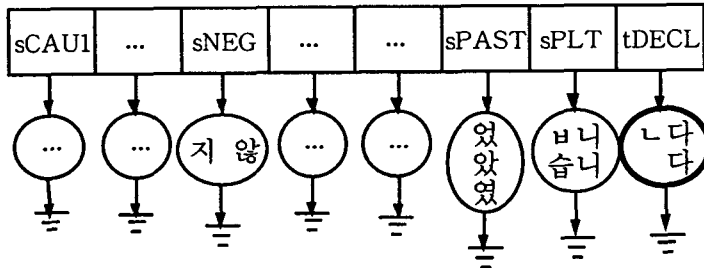
이 보조용언들은 아래의 19 가지로 구분되고, 여러 개가 동시에 사용될 경우에는 일정한 순서를 가진다 [2].



어미는 선어말 어미와 어말 어미로 나누어진다. 선어말 어미는 어미의 분류에 속하고 표충문에 나타나는 순서가 일정하게 정해져 있다. 선어말 어미는 보조용언, 어말 어미와 함께 용언의 의미를 더해주는 역할을 하는데 그 순서 관계는 다음과 같다.

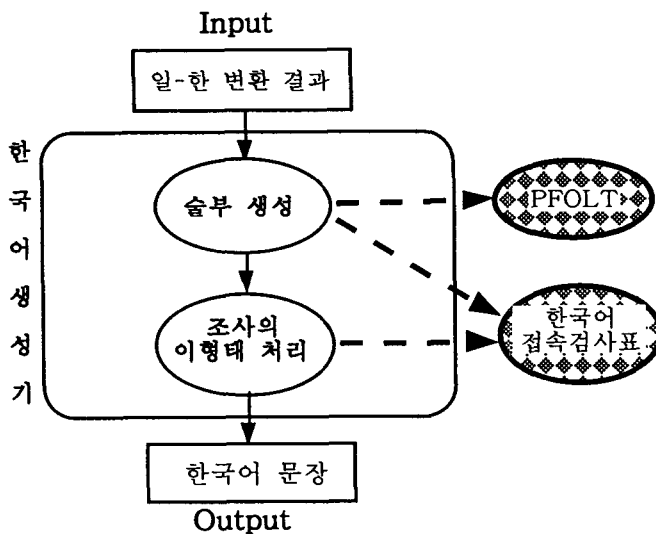


PFOLT (Predicate Feature Ordering and Lexicalizing Table) 이라고 불리는 자료 구조는 이러한 부분 순서 관계를 표현한다. PFOLT 는 한국어의 문법적인 순서에 따라 시제와 양상 등의 의미자질을 표현하고, 각 자질은 해당하는 한국어 형태소를 그 내용으로 가지고 있다. [그림 3]은 PFOLT 를 보여주고 있다. 그림에서 'Caus1', 'Neg' 등은 의미자질을 나타내고, 각 자질은 한국어 술부내의 순서 관계에 따라 왼쪽에서 오른쪽으로 배열되어 있다. 각 자질은 한국어 형태를 가지고 있는데 예를 들어 'Neg' 자질의 경우에는 한국어 보조 용언 '지 않'을 가지고 있다.



[그림 3] PFOLT

#### 4.3. 생성기의 구조



[그림 4] 한국어 생성기

한국어 생성은 의미자질을 이용한 술부의 생성과 한국어 형태소 간의 접속 가능성에 따라 한국어 조사의 이형태를 처리하는 두 부분으로 되어있다 [그림 4].

#### 4.3.1. 술부 생성

술부 생성은 다음의 5단계로 이루어 진다 :

- ① 의미자질 활성화 - 제일 먼저 용언을 찾아서 그 용언의 의미를 더하는 조동사, 어미의 의미자질들을 활성화 한다. 하나의 형태소가 하나이상의 의미자질을 가지는 경우에는 그 모두를 PFOLT 에 활성화한다.
- ② 부정어 처리 - PFOLT 에 부정 자질이 활성화 되었는지를 확인하여 부정어 처리를 한다. 부정 자질이 활성화 되었으면 용언이 가지고 있는 부정어를 새로운 용언 어간으로 바꾸어 주고 부정자질을 지운다. 이것은 부정어가 존재하는 경우에 자연스러운 번역이 되게 한다.
- ③ 사역 표현, 피동 표현의 처리 - 용언의 일-한 대역 사전에 한국어 역어와 그 용언의 사역 형태와 피동 형태가 같이 등록되어 있다. PFOLT 의 사역 자질을 활성화할 때, 사전에 등록된 형태의 사역, 피동 자질을 활성화한다. 사역 자질이 활성화된 경우는 용언의 사역어 유형에 따라 사역 표현을 만든다.
- ④ 보조용언, 선어말어미, 어미의 이형태 처리 - PFOLT 의 순서를 따라 활성화된 자질이 가지고 있는 한국어 형태소를 연결하여 술부를 생성한다. 이 때에 각 형태소간의 접속 검사에 따라서 이형태가 결정된다.
- ⑤ 음운축약, 불규칙 처리 - 용언과 보조 용언의 접속정보에 표현된 불규칙 정보에 따라 불규칙 처리를 하고 음운 조건을 검사하여 축약을 처리한다.

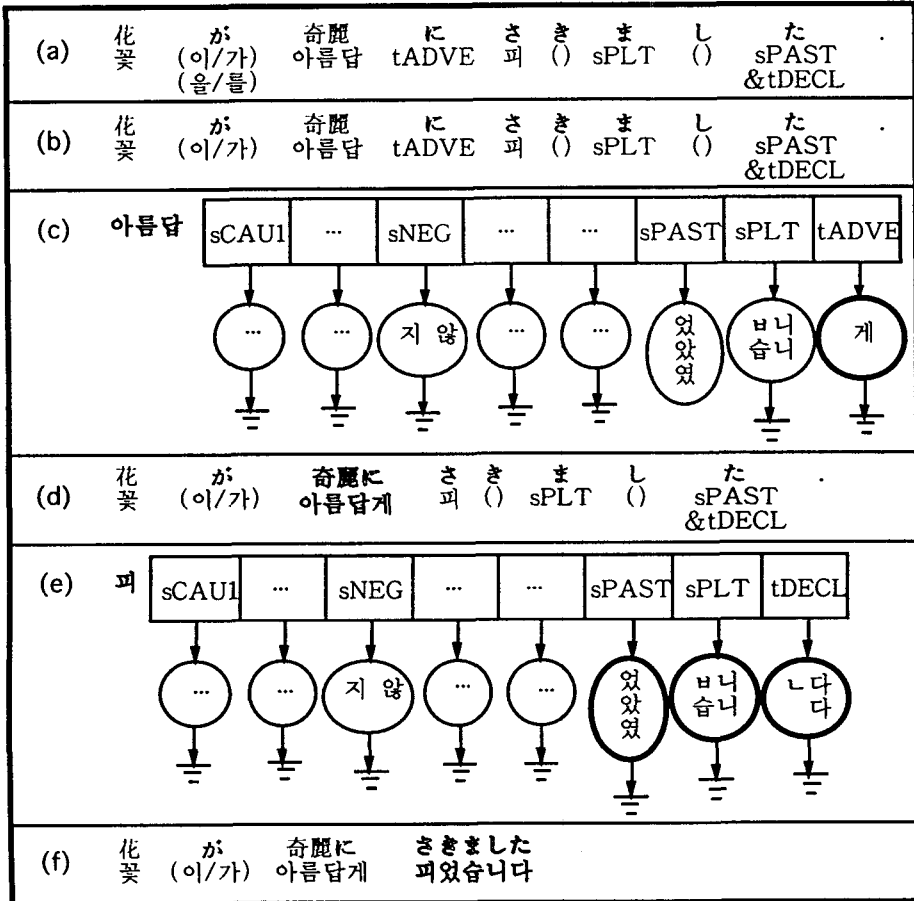
#### 4.3.2. 조사의 이형태 처리

조사의 이형태 처리는 접속 검사를 통해 이루어진다.

4.4. 생성의 예

다음은 일본어 문장 "花が奇麗にさきました."의 처리 과정이다 [그림 5].

이 문장의 형태소 분석 결과는 [그림 5(a)]와 같다. 조사 "が"는 (이/가)와 (을/를)로의 변환 가능성을 가진다.



[그림 5] "花が奇麗にさきました."의 처리 과정

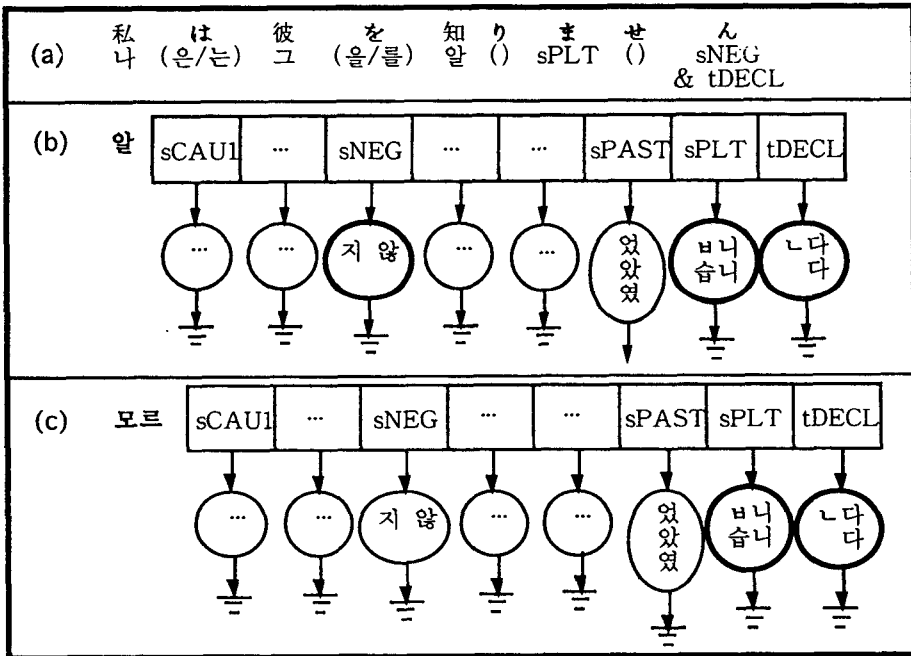
변환 과정에서는 조사 "が"의 의미가 (이/가)로 결정되어 [그림 5(b)]가 생성기로 전해진다. 생성에서는 먼저 "奇麗(아름답)"과 그 의미자질을 처리한다. 어간 "아름답"과 의미자질 "tADVE"가 PFOLT에 활성화된다 [그림 5(c)]. 이제 "아름답"과 "게"는 접속이 가능하며 "아름답게"가 생성된다 [그림 5(d)]. 동사 "피"의 경우에는 "ま"에서 정중의 의미자질인 "sPLT"가, "た"에서 과거의 의미자질 "sPAST"와 종결 어말의 의미자질인 "tDECL"가 활성화되어 PFOLT가 [그림 5(e)]와 같이 된다. "피(다)"와 "sPAST"의



이형태 "있/았/였" 간의 접속 검사를 하면 "피"는 "있"과만 접속이 가능하며 "피었"이 생성된다. "피었"과 "보니/습니"에서는 "피었습니다"가 생성되고, "니다/다"와 접속을 검사하여 "피었습니다"가 생성된다 [그림 5(f)].

마지막으로 조사의 이형태 "꽃"과 "이/가"의 접속 검사로 "꽃이"가 생성되어 최종 결과인 "꽃이 아름답게 피었습니다."가 생성된다

다음은 부정의 의미자질을 포함하고 있는 예문 "私は彼を知りません."의 처리이다. 이 문장은 [그림 6(a)]와 같이 분석된 결과가 변환기를 거쳐 생성기에 주어진다.



[그림 6] 부정문 "私は彼を知りません."의 처리 과정

생성에서는 "알"에 대해 "마"에서 정중의 의미자질 "sPLT"가, "ん"에서 부정의 의미자질 "sNEG"와 종결 어말의 의미자질 "tDECL"이 활성화된다 [그림 6(b)]. PFOLT 에 부정의 의미자질 "sNEG"가 활성화되어서 "알"의 부정적 역할 "모르"로 용언의 어간이 바뀌고, 부정에 해당하는 자질이 비활성화된다 [그림 6(c)]. 이 PFOLT 를 생성하면 "모릅니다"가 되고 전체 문장 "나는 그를 모릅니다."가 생성된다.

## 5. 실험 결과

[Table 1]에서는 일본어 입력 문장에 대한 번역 결과를 보여 준다. 실험에 사용된 문장들은 앞에서 제기된 문제점들을 포함하는 문장들로서 하나의 일본어 형태소가 여러 개의 의미 자질을 가지는 문장, 형태소 간에 일대일 대응이 되지 않는 문장, 형태소 간에 어순이 일치하지 않는 문장, 그리고 부정적 역어로 번역되어야 하는 문장들이다. 이 방법으로 제기되었던 문제점들을 모두 해결할 수 있었다.

입력문장	기존의 번역 시스템의 결과	본 시스템의 결과
彼を見るが,	그를 보지만	그를 보지만
これは 何ですか.	이것은 무엇입니까?	이것은 무엇입니까?
彼を見た.	그를 보았	그를 보았다.
彼を見ません.	그를 봅니않다.	그를 보지 않습니다.
お金がありません.	돈이 있지 않습니다.	돈이 없습니다.
彼を知りません.	그를 알지 않습니다.	그를 모릅니다.
彼は知らせなくなかつた と言う.	그는 알게 하고 싶지 않았다고 말한다.	그는 모르게 하고 싶었다고 말한다.

[Table 1] 실험 결과

## 6. 결론

본 논문에서는 의미자질을 이용하여 직접 번역 시스템의 문제점이었던 한국어 슬부의 생성을 해결하였다.

일본어 슬부와 한국어 슬부 간의 문법 형태소들의 순서의 문제는 PFOLT에 표현된 순서 관계로 해결하였고, 하나의 일본어 형태소가 여러 개의 의미를 표현하는 경우에는 PFOLT에 여러 개의 의미자질들을 활성화함으로써 해결하였다. 한국어 용언의 불규칙 처리는 한국어 형태소의 접속 정보에 나타난 불규칙 정보를 이용하여 처리하였다. 조사와 어미의 이형태 처리는 한국어 접속 정보 간의 접속 가능성을 조사하여 처리하였다. 본 논문에서 제안한 방법으로 일본어 문장에서 한국어를 생성한 결과 문제점으로 지적한 문

장을 모두 해결할 수 있었다. 그리고 기존의 방법보다 적은 사전 작업으로 자연스러운 한국어 표현을 얻을 수 있었다.

## 7. 참고 문헌

- [1] 강원석, "일한 직접 기계번역에서의 동형어의어 해결 기법에 관한 연구", 한국과학기술원, 전산학과, 석사논문, 1988.
- [2] 고영근, 남기심, 박경조, "고등학교 문법 자습서", 탑출판사, 1985.
- [3] 김은자, 이종혁, "일-한 기계 번역 시스템 구현: 휴리스틱을 이용한 일본어 형태소 해석 기법", 정보과학회 춘계학술대회 논문집, pp.797-800, 1993.
- [4] 金泰錫 外 1人, "日韓機械翻譯における意味接續關係を用いた韓國語の生成方法", 日本情報處理學會論文誌, pp.1578-1587, vol.33, No.12, 1992.
- [5] 金泰錫 外 1人, "日韓機械翻譯における否定文の處理", 日本情報處理學會論文誌, pp.892-904, vol.34, No.5, 1993.
- [6] 황찬호 외 3명, "한일어 대조 분석", 명지출판사, 1988.