

한<=>일 양방향 번역 시스템의  
기본 설계

○한 경 복                      장 두 영                      이 주 근  
인 하 대학교 전자공학과

A Basic Implementation of Bidirectional  
MT System between Korean and Japanese

K.R.Han                      D.Y.Chang                      J.K.Lee  
Dept. of Electronics Inha Univ.

ABSTRACT

This paper proposes a basic implementation of bi-directional translation system between Korean and Japanese.

The syntactic structure of Korean language is very similar to that of Japanese. In order to use the same main program for bi-directional translation, system is classified into two parts. One is directly translatable part and the other is untranslatable part by direct matching because of different structure.

Solving these problems, the syntactic and semantic informations control bi-directional interpreting process and make equivalent syntactic level between two languages.

I. 서 본

현재 기계번역 시스템의 연구는 대부분 단방향 시스템이다. [1,2] 저자들이 과거에 발표한 것은 2 가지 type이 있다. [3,4,7,8,9] 하나는 단방향 번역 시스템이고, 다른 하나는 양방향 동시 번역 시스템이다. 단방향 시스템으로서 두 언어간의 번역을 하기 위해서는 두 개의 시스템이 필요하다. 그러나 본 논문은 한 개의 시스템으로서 양방향 동시 번역하기 때문에 사전을 비롯한 전체 시스템이 거의 반으로 줄어들었다. 양방향 시스템은 이미 개발한 한,일 번역 시스템과 같은 기능의 능력을 갖는 시스템이고 따라서 시스템의 조작성은 의미정보를 도입하여 양방향 번역에 구별없이 사용할 수 있도록 한다. 언어 번역에서 문제가 되는 애매한 의미저리는 종래의 방식에 단어의 수식관계를 보강하고 양 언어 모두 동사를 중심으로 한 의존관계이므로 동사와 명사와 조사로 구성된 의미 frame을 이용한다.

II. 시스템 구성과 해석 과정

A. 양 언어의 구조

한국어와 일본어는 문절 간의 관계를 의존관계 구조로서 취급하는 것이 의미저리 면에서 적합하다 의존관계 구조는 두 문절 사이의 의미적 측면에서 본 수식, 피수식 관계를 표시하는 것도 있지만 구문 구조적 측면에서 보아도 명확한 구조로 되어 있고 간결한 구문 규칙을 설정하여 단순한 알고리즘으로 구문 해석을 할 수 있다. 구조상으로 본 두 언어의 특징은 다음과 같다.

- a. 문장이나 구절의 의미 중심은 용언부가 담당한다.  
제언부는 용언부의 의미 수식을 담당한다.
- b. 의존관계는 기본적으로 다음과 같이 정리된다.  
\* 관형적 결합(제언부)  
제언부 → 제언부  
용언의 관형형 → 제언부  
\* 각관계, 부사적 결합(용언부)  
제언부 → 용언부  
용언의 부사형 → 용언부
- c. 문장 형태의 구별은 문장 끝에 있는 용언부의 어미에 의해서 결정된다.

B. 시스템의 구성

그림1은 양방향 번역 시스템의 구성도를 나타낸 것이다. 시스템은 입력된 문장을 분석하여 일대일로 직접 번역되는 부분과 해당 언어의 특징 때문에 직접 번역되지 않는 부분으로 나누어 처리를 행하고 애매한 해석을 제거하기 위하여 형태소 해석된 결과를 구문과 의미해석부를 거쳐서 source언어의 입력에 대응하는 target언어를 생성해 낸다. [5,6]

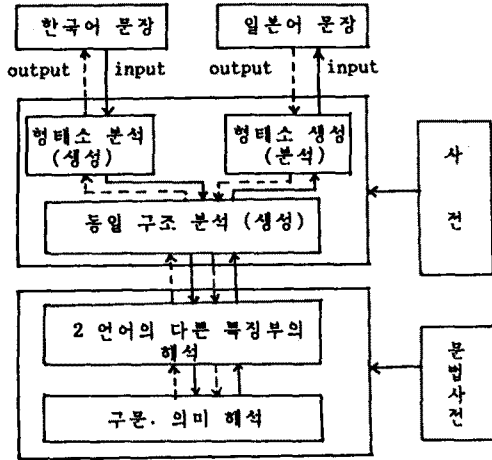


그림1. 양방향 시스템의 구성도

C. 해석 과정

한국어와 일본어는 같은 구문 구조를 가지고 있기 때문에 그림2와 같은 동일한 routine으로 해석된다. 해석 과정은 형태소 분석과 의존관계 해석을 행하면서 문장을 left-to-right 방향으로 각각 구절을 인식하고 대응하는 의미 구조를 생성, 확장해 가면서 문장 전체의 의미해석을 진행해 나간다.

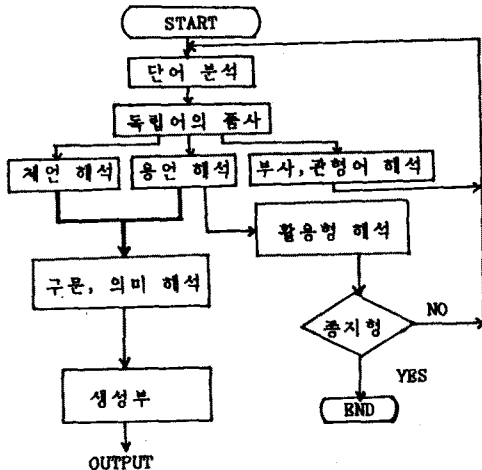
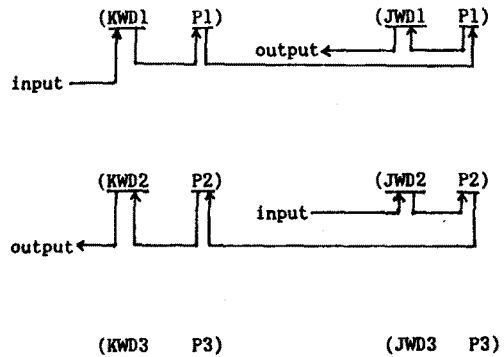


그림2. 해석 순서도

D. pointer방식 사전에 의한 일대일 번역

다의어가 아닌 대부분의 단어들은 사전에서 해당 번역어를 선택해 낸다면 같은 구문 구조를 가지고 있기 때문에 쉽게 번역문을 생성할 수 있다. 이 때 사전을 한(=)일 양방향 번역에서 동시에 사용할 수 있도록

구성한다면 database의 양을 줄일 수 있다. 그림3은 일대일 대응 관계에 있는 사전 구성과 번역의 예를 보인다.



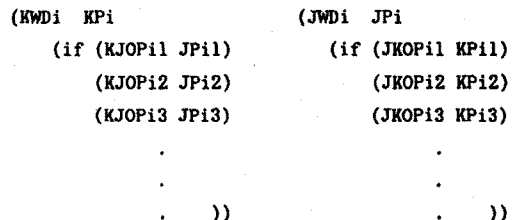
(KWD3 P3) (JWD3 P3)

KWDi : 한국어 단어  
JWDi : 일본어 단어  
Pi : pointer

그림3. 일대일 대응에 의한 번역

E. 의미정보에 의한 상호 간의 선택 방식

다의어를 번역할 때는 필요한 정보를 가지고 있는 의미정보를 test하여 조건에 맞는 해당 상대어를 선택해 낸다. 그림4에서 의미정보인 KJOP 와 JKOP는 입력된 단어의 종사에 따라서 달라지는데 만일 입력된 단어가 조사라면 조사를 선택할 수 있는 의미정보가 동작되고 동사라면 어간과 어미를 분석하여 상대어 어간을 선택하는 의미정보와 활용 어미를 찾은 의미정보를 동작하여 번역을 해 낸다.



KWDi : 한국어 단어, KPi : 한국어 pointer  
KJOP : 한국어에서 일본어 단어를 선택하기 위한 의미정보  
JWDi : 일본어 단어, JPi : 일본어 pointer  
JKOP : 일본어 단어에서 한국어 단어를 선택하기 위한 의미정보

그림4. 의미정보에 의한 상호 선택

1. 한국어와 일본어 조사 관계

한국어 조사 <====> 일본어 조사

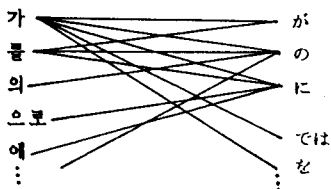


그림5. 2 언어의 조사에 대한 many to many 대응 관계

그림5와 같이 한국어 조사와 일본어 조사 사이에는 many to many 대응 관계가 있기 때문에 한 개의 source 조사에 대응 하는 여러 개의 target 조사들 중에서 해당하는 조사를 선택하는 일은 구분적으로나 의미적으로 여러가지 처리를 해야 한다.

1.1 한국어 조사에서 일본어 조사의 선택

한국어 조사에서 일본어 조사를 선택하는 과정은 한국어 조사를 t1, t2, t3, t4 type로 분류하고 해당 조건에 따라서 환경에 맞는 일본어 조사를 선택한다. [8,9]

1.2 일본어 조사에서 한국어 조사의 선택

하나의 일본어 조사가 여러 개의 한국어 조사와 대응할 때는 조사를 기준으로 해서 앞의 명사의 의미소성이나 구분 상의 역할과 뒤에 나오는 동사 등을 참조하여 한국어 조사를 선택해 낸다. 그림6은 일본어 조사 'に'에 대하여 한국어 조사를 선택하는 의미정보를 나타낸다.

( に Procedure

- ((COND ((COMP = MLC) 에)
- ((COMP = HUM) 에게)
- ((COMP = MS) 러)
- ((COMP = MB) 히)
- ((BEWORD = POFIX) 으로)
- ((AFWORD = NARU) 가))

그림6. 일본어 처리 procedure

2. 양언어에서 용언부 분석과 생성

양언어의 용언부는 비슷하게 구성되어 있으며 활용되는 형태도 유사하다. 용언을 구성하는 단어는 동사와 형용사 그리고 명사에 용언적 기능을 하는 본범소가 결합하는 경우가 있다. 구조적 분석과

생성을 위해서 한국어는 규칙 용언과 불규칙 용언, 체언의 용언화 기능부로 분류하고 일본어는 5단, 상일단, 하일단, 변형동사 등에 따라서 14개 type으로 분류한다.

2.1 한국어 용언의 구조 해석과 생성

\* 규칙 용언

어간의 음절 구성에 따라서 같은 기능을 하는 어미가 다른 형태로 결합 하므로 어간의 음절 구조에 의하여 규칙 용언을 표1과 같이 5개 group으로 분류한다.

표1. 규칙 용언의 group 분류

group	내 용
group1	어간의 끝이 모음 "ㅏ, ㅣ, ㅓ, ㅕ"로서 받침이 없다
group2	어간의 끝이 모음 "ㅓ"로서 받침이 없다
group3	어간의 끝이 모음 "ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ, ㅟ, ㅠ"로서 받침이 없다
group4	어간의 끝이 모음 "ㄱ, ㅋ"로서 받침이 있다
group5	어간의 끝이 "ㅈ, ㅊ" 이외의 모음으로서 받침이 있다

\* 불규칙 용언

사전에는 변형된 단어는 수록되어 있지는 않고 원형들만 있으므로 기계 번역 시에 변형된 어간으로 인하여 애매성을 초래한다. 따라서 한국어 분석 시에는 변형된 어간에서 원래의 규칙 어간을 찾고 생성 시에는 불규칙에 맞게 변형을 하여 생성해 내야 한다. 이들 불규칙 용언들은 변화하는 성질에 따라서 +, 0, - type로 분류하여 처리한다.

\* 체언의 용언적 기능

$$N + \begin{pmatrix} V_k \\ q \end{pmatrix} + Afx \begin{pmatrix} 다, ㅓ, 구나 \\ ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ, ㅟ, ㅠ \\ 고, 지, 어서 \end{pmatrix}$$

N : 체언                      V<sub>k</sub> : 본범소(하)  
q : 본범소(이)                Afx : 어미

그림7. 체언의 용언적 기능

그림7과 같이 체언에 용언적 기능을 하는 본범소가 결합하여 새로운 어간을 생성하고 그 뒤에 각종 어미가 연결되어 활용되므로 이 성질을 이용하여 분석과 생성을 해 낸다.

이와 같은 처리 과정을 그림8의 상태 전이도로 나타내었다.

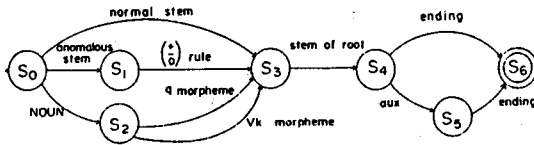


그림8. 용언 처리의 상태 전이도

2.2 일본어 용언부의 구조 형식

일본어 용언부를 구조적으로 분석과 생성하기 위하여 first group, second group, irregular group으로 분류한다. first group은 어간과 어미가 규칙적으로 결합하는 형태이고 second와 irregular group은 불규칙적인 형태들이다. 따라서 second와 irregular group에서도 first group과 동일한 어간과 어미의 활용 규칙을 적용할 수 있도록 하였으며 각 group을 전체 14개 type으로 세 분류하였다.[8,9] 형용사도 동사와 같이 어미 활용을 하므로 이것도 하나의 동사 type으로 취급하면 별도의 처리 과정을 거치지 않고 동사와 같은 routine에서 처리한다.

(VS TYPE

- ((MEMBER SUFTY '(1 2 6))
- ((JVEB = (TAK JSF)))
- ((SUFTY = 3)
- ((JVER = (TAK 'RU JSF)))
- ((SUFTY = 4)
- ((JVER = (TAK 'RE JSF)))
- ((SUFTY = 5)
- ((JVER = (TAK 'YO JSF)))

그림9. VS type 동사의 생성 procedure

따라서 일본어 동사의 생성 시에는 어간의 type을 결정하고 어미의 활용 순서가 결정되면 해당하는 음절을 선택하여 일본어 동사를 생성한다. 분석 시에는 반대 과정으로 처리하면 된다. 그림9은 second group인 VS type 동사의 어간과 어미가 생성되는 처리 과정을 나타낸다.

III. 구문 및 의미 해석

양언어는 관용적인 표현 등의 일부를 제외하면 문장 중의 동사를 중심으로 여러 개의 명사들이 조사와 더불어 격을 형성하는 의존 관계 구조를 이룬다. 따라서 문장의 구문 및 의미 해석을 위해서 명사의 의미 소성과 동사와 조사를 고려한 의미 frame을 구성하고[5,10] 구문 및 의미 상의 역할을 파악하여 문맥을 대응 시킨다. 의미 frame의 예를 그림10에 나타내 있고 처리 모델은 그림11과 같다.

V2R1	[HUM:AGT]가 [PRO:OBJ]를 ((PAR:SPC)에)
V3U1	[HUM:AGT]가 [MKT:OBT]를 ((MLC:SPC)에서 [TIM:TIM]에)
V4W1	[HUM:AGT]가 [PLA:OBJ]를 ((PRO:TOO)로 [MLC:SPC]에서 [TIM:TIM]에)

그림10. 의미 frame의 예

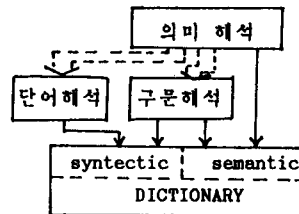


그림11. 처리 모델

그림10에서 "=>"는 의미 해석을 위해 구동되는 부분을 나타내고 "→"는 처리 과정에서 사전을 참조하는 것을 나타낸다.

IV. 실험결과

1. 한.일 번역

(KJMT)  
KORBAN=>말 가운데에서, 사람이나 물건의 움직임, 상태의 변화, 존재 등을 나타내는 것을 동사라고 말한다. 동사는 어간과 어미가 있고, 어간은 전혀 변하지 않는다. 어미는 쓰이는 경우에 따라 여러가지로 모양이 변한다. 따라서, 어미가 변하는 것을 활용이라 말하며, 그 방법은 다섯 종류가 있다. \$

JAPANESE=>KOTOB A NAKA DE , HITO YA MOMO NO UKOKI , ZYOUTAI NO HENKA , SONZAI NADO O ARAWASU MONO O DOUSI TO IU . DOUSI WA GOKAN TO GOBI GA ARI , GAKAN WA ZENZEN K.WARA NAI . GOBI WA TSUKAWARERU BAAI NI YOTTE IROIRO NI KATACHI GA KAWARU . SITAGATTE , GOBI NO KAWARU KOTO O KATSUYOU TO II , SOMO SIKATA WA GO SYURUI GA ARU .

2. 일.한 번역

참고 문헌

(JKMT)

JAPANESE=>KOTOBANO NAKADE, HITOYA  
MONONO UKOKI, Z'OUTAINO HENKA,  
SONZAINADOO ARAWASU MONOO DOUSITO  
IU. DOUSIWA GOKANTO GOBIGA ARI,  
GAKANWA ZENZEN KAWARA NAI. GOBIWA  
TSUKAWARBRU BAAINI YOTTE IROIRONI  
KATACHIGA KAWARU. SITAGATTE,  
GOBINO KAWARU KOTOO KATSUYOUTO II,  
SONO SIKATA WA GO SYURUIGA ARU.\$

KOREAN=>말 가운데에서, 사람이나 물건의 움직임을, 상태의 변화, 존재등을 나타내는 것을 동사라고 말한다. 동사는 어간과 어미가 있고, 어간은 전혀 변하지 않는다. 어미는 쓰이는 경우에 따라 여러가지로 모양이 변한다. 따라서, 어미가 변하는 것을 활용이라 말하며, 그 방법은 다섯 종류가 있다.

V. 결론

본 논문에서는 한국어와 일본어에 대한 양방향 번역 시스템에 대하여 논했다. 두 언어에서 문법 형식이 유사하다는 점에 주목하고, 구절 구조를 구문적으로 같은 level로 작성하여 동일한 program으로서 양방향을 동시 번역한다. 이 때 구조적 차이가 나는 부분은 subroutine에서 처리한다. 사건은 종래의 단방향 번역과는 달리 양방향 입력과 수식이 가능하도록 구성하고 두 언어 사이에서 의미 정보에 의하여 control된다. 따라서 단방향 시스템으로 두 언어를 번역하는 system에 비하여 H/W와 S/W의 cost가 반으로 감소된다.

1. S.Takamatsu, F.Nishida, "English-Japanese Translation Based on Verb Pattern and Their Case-structure," 전자통신학회논문집, vol. J64-D, No.9, pp. 815-822, 1981.
2. W.A.Woods, "Transition Network Grammer for Natural Language Analysis," CACM vol.13, No.10, pp. 591-606, 1970.
3. 이주근, 한성국, "한국어 구문 분석 System(III)", 전자공학회 논문집, vol.3, No.1, 1980.
4. 이주근, 한성국, "한국어의 Machine Translation을 위한 구문 구조 분석," 전자공학회지, vol.18, No.5, pp. 15-21, 5, 1981.
5. 이주근, 한성국, "자연어 의미의 Schema 표현", 전자공학회 논문집, vol.7, No.5, pp. 322-327, 7. 1984.
6. 이주근, 윤민기, "자연어의 형태소 분석", 전자공학회 논문집, vol.7, No.1, pp.320-327, 7. 1984.
7. 이주근, 이수현, "한일 기계 번역 시스템(I)", 전자공학회 논문집, vol.8, No.1, pp.202-204, 6. 1985.
8. 이주근, 한광복, "한일 기계 번역 시스템(II)", 전자공학회 논문집, vol.8, pp. 205-207, 6. 1985.
9. J.K.Lee, K.R.Han, "Korean-Japanese Machine Translation system(III)", Seoul International Symposium on Electronics Engeering, pp. 134-137, Nov. 1985.
10. 이주근,한광복, "의미 frame에 의한 기계 번역", 전자공학회 논문집, vol.9, No.1, 6.1986.