

論文94-31B-12-11

일한 번역시스템을 위한 한국어 생성기 설계

(A Design of Korean Generator for Japanese to Korean Translation System)

姜 哲 燾*, 崔 炳 旭*

(Seok Hoon Kang and Byung Uk Choi)

要約

본 논문에서는 일한 기계번역 시스템을 대상으로 한, 한국어 생성기를 제안한다. 기존의 일한 번역을 위한 한국어 생성기는 문장 전체의 구조에 의해 한국어를 생성하거나 각 형태소의 직접번역에 의한 것이었으나, 본 논문에서는 어절을 중심으로 한국어를 생성한다. 일본어 해석에 의해 생성된 어절과 한국어의 어절 사이에는 어순을 비롯하여 유사점이 발견되지만 한국어 생성의 입장에서 본다면 어절 내에서는 많은 차이점이 존재한다. 일본어 해석시 부가되는 속성들은 한국어의 그것과는 다른 순서를 가지므로 이를 고려한 한국어 생성기가 필요하다. 이를 위해 주석적인 한국어 생성 알고리즘을 제안하며 조사와 용언, 그리고 조동사의 번역에 이를 적용하였다. 또한 일본어 해석정보에서 추출이 불가능한 한국어 음운현상도 아울러 고려하여, 그 결과 절차적인 생성 방식에서는 고려하기 힘들었던 일본어 용언의 다양한 활용현상을 충분히 번역할 수 있었다.

Abstract

This paper proposes the Korean generator which is used in a Japanese to Korean translation system. The established Korean generator for Japanese to Korean translation uses the entire sentence structure to generate Korean, or directly translates each morpheme. But in this paper, the generator generates Korean, based on a short clause. Japanese and Korean have similarities in short clauses, including the word order. But in the Korean generation, they also have many differences within short clauses. The analyzed features in Japanese and Korean are in the wrong order. Therefore the Korean generator which takes it into consideration is needed. We propose an annotational algorithm for the Korean generation and apply it to the translation of postpositions, declinable words and auxiliary verbs. We also consider Korean phonological phenomenon that can't be extracted from the analyzed Japanese informations. In the end, we can translate the Japanese conjugation that can't be easily dealt in the procedural method.

1. 서론

* 正會員, 漢陽大學校 電子通信工學科
(Dept. of Elec. Communication Eng.,
Hanyang Univ.)
接受日字: 1994年 2月 3日

한일, 일한 기계번역은 한일 양국의 지역적 문화적 특성으로 많은 연구의 대상이 되어왔다. 특히 양국의

언어는 전산언어학적인 측면에서 보았을 때 많은 유사점을 가지고 있어서 시스템의 구현에 관한 연구가 활발히 진행되어 왔다. 그러나 일한 기계번역 시스템의 구현에 대한 연구는 대부분 일본어의 문법적, 의미적인 해석에 중점을 둔 것이 사실이며 입력된 일본어에 대응하는 한국어 생성에 관한 연구는 상대적으로 부족한 실정이다.^[16, 19, 20] 그러나 한국어는 일본어보다 활용의 범위가 넓고 음운현상이 더욱 다양하여 한국어의 형태소 생성에도 많은 난점이 있다.^[1, 2, 21, 22, 25]

일본어와 한국어는 어절간의 어순이 거의 일치하며, 한국어로 번역될 문장이 일본어의 어순을 따른다고 하더라도 거의 어색하지 않기 때문에, 구분적인 정보에 의해 문장의 주중에 따라 번역을 수행하는 것보다는 각 어절 내의 정확한 번역에 중점을 두는 것이 중요하다고 생각된다. 일한 번역을 위한 한국어 생성에 관한 기존의 연구에서는 절차적인 방법에 의한 생성이^[26] 주류를 이루었으나 최근 주석적 방법에 의한 연구가 활발히 진행되고 있는 실정이다. [26]과 같이 테이블에 의한 방식은 번역의 품질은 높일 수 있는 반면 정확한 테이블의 작성이 난해하며 이의 시스템 적용시 처리 효율에 문제가 생길 수 있을 것으로 보인다.

본 논문에서는 형태소 해석 결과 생성된 일본어 어절을 중심으로 한국어 어절을 생성하며 이와 같은 관점에서 일한 기계번역 시스템의 한국어 생성시 대두될 수 있는 중요한 점을 다음과 같은 5가지로 분류하였다.^[3, 4, 5, 6, 7, 8, 9, 10]

(1) 조사류의 번역

일본어는 격조사, 보조사 등 다양한 형태의 조사를 가지고 있으며 정도의 차이는 있으나 한국어의 경우도 마찬가지이다. 그러나 전산언어학적인 측면에서 살펴본다면 한국어의 조사와 일본어 조사는 전후접 단어와 결합 양상이 다르게 나타나므로 이의 문법적 대응이 중요하다.

(2) 동사류의 번역

일본어 동사는 기본형에 활용하는 부속어가 추가하여 문법적, 의미적인 뜻을 달리한다. 특히 활용의 형태가 아주 다양하고, 복합적으로 결합되어 사용된다. 이에 대응하는 한국어도 동사의 어간에 어미가 활용하는 형태로 이루어지므로 유사한 형태를 가지지만 대응하는 말의 문법적 기능이 상이한 경우가 많고, 특히 일본어의 부속하는 활용형과 한국어 어미의 어순이 틀린 경우가 많기 때문에 이에 대한 처리가 필요하다.

(3) 형용사류의 번역

일본어의 형용사는 동사와 마찬가지로 활용을

한다. 특히 일본어 형용사에 대응하는 한국어 형용사는 한국어 동사와 문법적 기능에 유사성이 많아서, 일한 기계번역에서는 일본어의 형용사와 동사를 같은 범주에 포함시켜 용언의 처리로 대처하려는 경향도 없지 않으나 일본어 문장의 해석시 나타나는 문법적 정보가 상이하고 활용의 형태가 서로 다르므로 이에 대한 구분이 필요하다.

(4) 조동사의 번역

일본어는 조동사라는 품사를 가지지만 한국어에서는 일본어 조동사의 기능을 용언의 어미와 서술격조사 등이 대신한다. 따라서 용언과 별도로 활용하는 일본어 조동사를 일본어 용언에 대응하는 한국어에 어미의 형태로 변환시킬 필요가 있다.

(5) 명사의 의미분류

조사와 용언의 번역에 근간이 되는 것은 이에 관련되는 명사의 의미와 기능적 분류이다. 대부분의 어절 단위의 번역시 일차적인 번역어는 전후접하는 중심어의 의미에 따른다. 그러나 이 명사의미의 분류 방법은 본 논문의 대상인 일한 번역을 위한 한국어 생성의 범위를 벗어나는 것이므로 상세히 다루지는 않기로 한다.

이 밖에도 각 품사별로 번역어를 생성하는데 문제점이 있을 수 있겠으나 위의 4가지에 중점을 두어 일한 번역을 위한 한국어 생성에 관해서 기술한다. 그 밖의 경우는 정도의 차이는 있으나 대부분 대표 뜻에 의한 일대일 대응으로 형태소별 번역이 가능하다. 본 논문에서 대상으로 하는 시스템의 입력은 일본어 문장의 형태소 해석결과^[23]이며, 이 해석결과에는 어절로 분리되어 있는 일본어 문장에 문법적인 정보가 부가되어 있다. 아울러 적합한 한국어의 생성을 위해 일한 번역의 입장에서 전술한 각 경우에 대한 한국어 문법을 파악하였고, 해석된 일본어 정보에 의해 적절한 한국어가 생성되도록 한국어 생성 알고리즘을 설계한다.

II. 한국어 생성

일본어문의 해석 결과가 포함하고 있는 문법적인 정보는 한국어의 그것과는 다르게 나타날 수 있다. 예를 들어 용언에 후접하는 일본어 조사는 한국어로 번역될 때 용언의 어미로 생성되어야 하며 일본어 조동사의 경우도 대부분 이와 같은 경우이다. 특히 용언의 경우 하나의 자립어에 여러 개의 부속어가 추가되어 표현되며 이 때 한국어로의 번역순서는 일본어의 해석순서와 일치하지 않는다. 또한 일본어 동사와

형용사는 문법적 규칙이 상이하지만 한국어로 번역될 때는 유사한 특징을 가지게 된다. 그러므로 이들을 고려한 한국어 생성기가 번역의 품질을 높이는데 중요하다 고 할 수 있다.

본 논문에서는 한국어의 생성방식을 주석적인 방법으로 설계한다. 이것은 Flag와 속성정보를 이용하는 것으로 하나의 중심어에 해석된 모든 속성을 병렬로 부가시키는 방법이다.¹¹⁾ 즉 한국어 "먹었습니다"라는 말은 "먹다"라는 중심어에 '과거'와 '존중'의 속성이 부가된 형태로 정의할 수 있는 것이다. 이 방식으로 한국어와 상이한 표현 양식을 가지는 일본어를 같은 의미를 가지는 적합한 한국어로 번역할 수 있으며, 언어표현에 따라 번역 시스템의 구성이 달라지는 것을 방지할 수 있다. 기존의 절차적인 방법의 경우, 다양한 표현을 한국어로 번역하기에는 시스템 설계의 입장에서 많은 난점이 있었으나, 본 논문의 방식은 부가적인 속성을 이용하므로 시스템 구성과는 무관하게 근접한 번역어를 얻을 수 있으며 품사의 천이까지 일부 처리할 수 있다. 본 논문에서 제안하는 방법에 의한 실제 번역 방식은 이후의 각 번역 방식에서 자세히 논하기로 한다.

1. 조사의 번역

일본어에서 조사는 여러 가지의 기능을 가지며 존재 가능한 위치도 한국어의 조사와 다소 다르다. 본 논문에서는 한국어 생성의 입장에서 일본어 조사를 두 종류로 분류하였다. 이 분류는 사전적인 정보가 아니라 형태소 해석 후 부가되는 정보에 의한 것으로, 첫 번째는 체언에 부속하는 형태를 가지는 것으로 격조사 등이 이에 해당하고 한국어의 조사로 번역할 수 있다. 두 번째는 용언 등 그 밖의 자립어에 부속하는 것으로 접속조사 등이 있으며 이들은 한국어로 번역할 때 조사가 아닌 어미의 기능을 수행하여야 한다. 이 구분은 한국어로의 번역 시, 대상이 되는 한국어로 조사를 선택할 지 아니면 어미를 선택할 지를 결정하는 기본이 된다. 본 논문에서 제안하는 각각에 대한 번역 방식은 다음과 같다.

1) 단독 조사의 번역

조사가 하나의 어절에 하나만 사용되었을 경우는 문법적 기능에 의해서만 번역을 수행할 수 없으므로 어절간의 의미정보에 의해 용례를 구분해야 한다. 조사가 포함된 어절의 중심어가 체언이라면 조사는 어절에 속성부여는 하지 않으며 문법적 기능만 수행하는 것으로 해석한다. 만약 포함된 어절의 중심어가 용언이라면 어절에 속성을 부여하며 문법적 기능도 아울러 수행하는 것으로 한다.

何かありますか. (nanika arimasuka) (1)
무엇인가 있습니까

예문 (1)에서 첫 번째 "か"는 체언 "何"에 접속한 부조사로, 한국어로 번역될 때 체언에 부속하여 "(인)가"로 된다. ("何か"를 하나의 연어로 볼 수도 있다.) "か"는 체언에 접속한 부조사이므로 체언 "何"에 의미적 속성은 부여하지 않는 것으로 형태소 해석한다. 그러나 문장의 끝에 나타나는 두 번째 "か"는 용언의 활용형 "あります"에 접속한 종조사로, 한국어로는 "-까"로 번역되어야 한다. 이것은 어간에 부가되는 어미이며, "ある"라는 중심어에 '의문'이라는 속성을 부여하고 동시에 문의 종결의 기능을 수행한다. 이와 같이 일본어에서 문법적 기능은 조사이지만 한국어에서는 조사와 어미 등으로 구별되므로 각 조사마다 사전에 이를 분류하여 등록할 필요가 있다. 위의 경우를 문법적 기능에 의한 한국어 생성에 관한 문제라면 일한 기계번역 중 조사의 번역에서 가장 문제가 되는 것은 의미에 따른 번역어의 결정이다. 다음은 일본어 조사 "と"의 경우에 대한 번역의 경우이다.

友と遊ぶ. (tomoto asobu) (2)
친구와 놀다.

이 경우는 체언에 부속하는 격조사로 어떤 사항에 대한 공존이나 명사를 병렬하여 말하는 것으로 한국어로는 "와/과"로 번역된다.

社長となる. (syachoto naru) (3)
사장이 되다.

동작의 변화나 결과를 표시하기 위해 격조사 "と"가 사용되었을 때는 "으로/로, 이/가"로 번역된다. 이 때는 대부분 후접하는 형태소로 "なる"의 활용형이 위치한다.

いいと思う. (iito omou) (4)
좋다고 생각하다.

禁煙と書く. (kinento kaku) (5)
금연이라고 쓰다.

(4)와 (5)번 예문은 내용을 가리키거나 내용의 전달을 위한 격조사 "と"의 사용례로 "라고/고, 하고" 등으로 번역하여야 한다.

이와 같이 일본어 조사는 전후접하는 형태소와 그들의 의미에 따라 해석이 다양하게 이루어지며 각 해

석에 맞게 일본어 해석 정보를 분류하는 것이 중요하다. 일례로 일본어 조사 “と”에 대한 본 논문의 번역 방식을 표 1에 나타내었다. 조사의 경우는 번역의 형태가 아주 다양하게 나타나므로 각각의 조사에 대해 전후접 어절의 속성정보와 문법정보를 이용한 의미 분류가 필요하다.

표 1. 조사 “と”의 번역

Table. 1. A translation of the postposition “と”.

문법기능	전립형태소 분류	후립형태소 분류	번역어	예문	기타
격조사	궤속성을 가지는 제언 용언	부정의 속성을 가지는 용언	도	三時間と眠らなかつた。	
격조사	동사·형용사·조동사 활용형(거의 종지형)	용언, 상대성 명사	(이)라고	いとおもう。	
격조사	제언	용언, 상대성 명사	(이)라고	頼むと書く。	
격조사	부사어	(일반 접속정보)	(무해석)	「ここに」と笑う。	
격조사	제언	“なる”활용형	이/가	社長となる。	
격조사	용언의 종지형, 제언	(일반 접속정보)	와/과	友と遊ぶ。	참고. 2
접조사	용언의 종지형	(일반 접속정보)·시제 면		春になると花がさく。	
접조사	용언의 종지형	(일반 접속정보)·시제 면	(이)가	顔を見るとなまだした。	참고3

*참고 1. “と”가 두 어절 이상 연속하여 나열되고 마지막의 “と”가 격조사에 전접하면 이 “と”는 번역하지 않는다.

*참고 2. 용언의 종지형 + “と”가 “와/과”로 번역 되려면 “と”를 포함한 어절이 두개이상 연속하여 나열되어야 한다.

*참고 3. 용언의 종지형 + “と”가 “(하)자”의 경우로 번역되기 위해서는 “と” 이전의 어절의 속성에 부가된 시제가 현재이고 “と” 이후의 어절의 속성에 부가된 시제가 과거이어야 한다.

2) 조사의 결합

한국어와 마찬가지로 일본어 조사도 서로 결합하여 사용될 수 있다. 이것은 기계번역시 많은 문제점을 야기시킬 수 있으므로 이들의 구분이 필요하다. 본 논문에서는 두 가지로 구분하여 번역을 수행하였다. 첫 번째는 형태소 해석의 차이이다. 예를 들어 일본어 조사 “でも”는 계조사 또는 접조사로 하나의 조사 기능을 수행할 때도 있지만, 때로는 “で” + “も”의 형태로 다양한 접속관계를 규정지을 수도 있다. 이 경우는 사전에 “で”, “も”, “でも”의 3가지 경우를 모두 사전에 등록시키고 어절 내에 조사의 기능으로 “でも”가 포함되어면 이는 더 이상 분리하지 않고, 분리되었을 때의 의미 구분까지 “でも”가 수행하도록 한다.²³ 이를 위해 형태소 해석시 조사의 결합규칙을 제한하여 두 개의 조사가 각각 등록되는 해석목의 생성을 억제하였다. 그러나 “にだけ”와 같이 “に” + “だけ”로 구분되어 해석은 가능하지만 하나의 조사로 사전에 등록되어 있지 않는 경우는 앞절과 같이 각각 조사 처리과정을 수

행하여야 한다. 두번째는 한국어 조사의 결합에 관한 문제이다. 일본어 부조사는 격조사의 앞에도 위치할 수 있고 뒤에도 위치할 수 있다. 즉 “にだけ”도 “だけ”도 모두 “한테만”이라는 한국어로 번역할 수 있다. 그러나 “만한테”라는 말은 사용하지 않으므로 조사의 재조정이 필요하다. 한국어 조사의 경우 생성의 입장에서 보면 그 결합형태가 “격조사 + 격조사”, “격조사 + 보조사”, “보조사 + 격조사”, “보조사 + 보조사”의 4가지 경우로 나눌 수 있다. 그러나 실제로 거부감 없이 사용하는 격조사끼리의 결합은 “처소부사격 + 주격”, “처소부사격 + 목적격” 밖에 없는 등 각각의 결합에 대해서는 상당한 제약이 따른다. 본 논문에서는 일본어의 조사의 결합순서에 관계없이 2개 이상의 조사의 결합으로 해석된 경우 그림 1과 같은 별도의 조사 생성 순서를 이용하여 적절한 한국어 조사를 생성하도록 하였다. 이에 따라 전술한 조사의 경우 행위귀착(그림 1의 402번)의 한 유형으로 “만한테”라는 한국어는 생성하지 않는다.

번호	문법식	부조사	격조사	부조사	격조사
#define SPT10P1 380	○ 처소 + 에	○ 에서	○ 에서	○	○
#define THT10P1 381	○ 시간 + 에	○	○	○	○
#define ST10P1 382	○ 행위귀착 + 에	○	○	○	○
#define SRC10P1 383	○ 출발점 + 에	○	○	○	○
#define GOV10P1 384	○ 직할점 + (으)로	○	○	○	○
#define SPT10NE 400	○ 처소 + 에	○	○	○	○
#define THT10NE 401	○ 시간 + 에	○	○	○	○
#define ST10NE 402	○ 행위귀착 + 에	○	○	○	○
#define SRC10NE 403	○ 출발점 + 에	○	○	○	○

그림 1. 한국어 조사의 결합의 일례

Fig. 1. An example of the combination of the postposition.

2. 용언의 번역

본 논문에서 용언은 크게 동사와 형용사로 나누어 번역하지만 형용사의 경우 한국어로 번역될 때 “-다”로 종결하는 등 한국어 생성시는 큰 어법상의 차이가 나타나지 않으므로 본 논문에서 구분하는 것은 생성시 필요한 것이라기 보다는 해석시 문법적 정보 추출을 위한 구분이라고 할 수 있다.

1) 동사의 번역

일본어 동사는 한국어 동사와 마찬가지로 다양하게 활용한다. 그리고 ㅅ변형과 ㄴ변형 등 몇 개의 유형을 제외하면 대부분 규칙적인 활용이 이루어지며 문법적 접속이 명확하다. 그러나 이에 해당하는 한국어는 일본어와 활용의 순서와 범위에 차이가 있으므로 질차적으로는 완전한 번역을 수행할 수 없다. 그림 2와 같이 (a)의 일본어 해석 결과와 이의 번역결과에 해당하는 (b)의 한국어 해석 결과는 의미정보의 순서가 완전히 다르므로 (a)의 문법 결합 순서로 (b)의

문장을 생성해 내는 것은 난해한 일이다. 즉, 일본어 해석정보의 순서에 의해서 한국어를 생성하는데는 상당한 무리가 따른다.

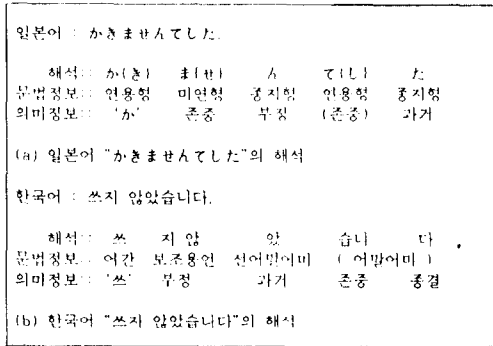


그림 2. 한국어와 일본어의 의미해석 결과의 비교
Fig. 2. A comparison of the semantic representation between Korean and Japanese.

그러므로 일본어 해석순서에 관계없이 적절한 한국어를 생성하기 위해서는 주석적인 한국어 생성방법이 필요하다. 본 논문에서 설계하는 주석적 생성방법은 일본어 해석에서 발생하는 속성정보를 한국어의 어말어미와 어미 등의 정보로 변환하고 이를 하나의 어간에 전부 부가시키는 방식이다.^{[19][20]} 그림 2의 (a)의 일본어의 해석정보는 그림 3과 같은 구조로 해석되며 활용정보는 FLAG의 형태로 부가시켰다. 이 부가된 정보는 별도의 한국어 생성 Network을 통해 적절한 한국어 용언으로 번역한다.

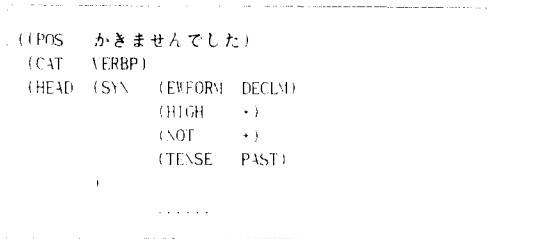


그림 3. 일본어 입력문 해석의 일례
Fig. 3. An example of the analysis of Japanese input.

한국어 용언의 어절은 그림 4와 같은 기본 구조를 가진다. 일반적으로 어미는 선어말 어미와 어말어미

로 나눌 수 있다. 선어말어미는 실질 형태소인 어간과 형식 형태소인 어미 사이에 위치하여 높임, 공손, 시제 등을 나타내는 어미를 말한다. 어말어미는 어간이나 선어말어미 뒤에 오는 어미를 말하는 것으로 크게 종결어미와 비종결어미로 구분된다.^[24]

한국어 용언의 어미를 생성하기 위해서는 전처리 과정으로 피동이나 사동 접미사를 먼저 어간에 붙여야 한다. 피동이나 사동의 뜻을 가지는 것은 접미사뿐만 아니라 보조용언도 있는데 어느 것이 더 적절한지는 사용되는 용언에 따라 다르다. 그러므로 어떤 형식의 사동문 또는 피동문이 적절한지는 한국어 생성사전의 용언부에 기술해 주도록 하였다.

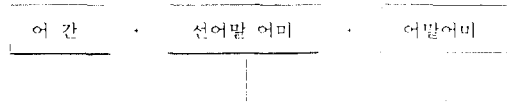


그림 4. 용언어절의 기본구성
Fig. 4. The basic configuration of declinable words.

한국어 용언의 구체적인 생성 알고리즘은 그림 5와 같다.

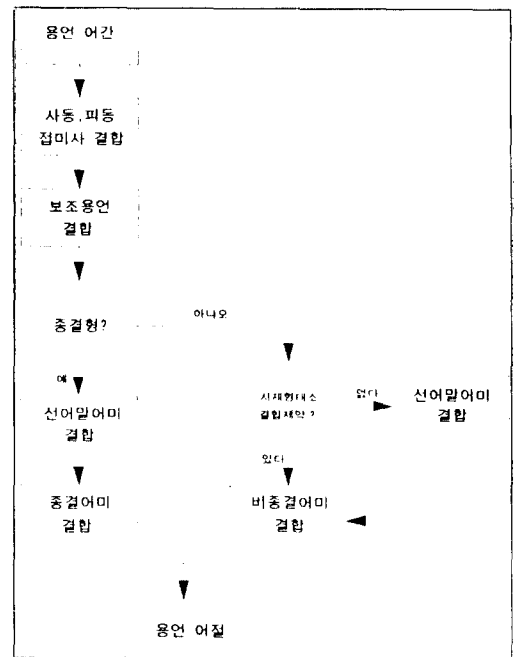


그림 5. 용언 생성 알고리즘
Fig. 5. An algorithm for generating the declinable word.

하나의 어간에 일본어 해석에서 추출한 여러 개의 정보가 부가되어 하나의 어절¹⁾을 생성하게 한다. 선어말 어미는 높임, 시제, 공손의 3가지 유형이 있으며 공손의 경우는 "음"이 대표적인 경우이나 현대 국어에서는 거의 사용하고 있지 않으므로 본 논문에서는 다루지 않기로 한다. 또한 미래 시제 선어말어미인 "겠"은 미래 시제라기 보다는 듣는 이의 의지나 의도, 추측을 나타내는 말이므로 본 논문에서는 '의지'의 경우로 재정의하여 사용한다. 일반적으로 미래를 나타내기 위해서는 "어간 + 르 것이다"가 사용된다. 둘 이상의 선어말어미가 사용되면 그 사용순서는 "높임-시제-의지"로 정의하였다. 선어말어미와 어미에 사용되는 속성정보가 여러 개라고 하더라도 그 사이의 어순과 최종 생성되는 어절은 명확하다. 그러나 보조용언의 경우 종류가 복잡하고 그 사이의 순서를 정하기가 어렵다. 이 경우는 해석된 속성정보의 순서에 따르도록 한다. [22]에 의하면 대상 문헌의 분석결과 사용된 보조용언 중 홀보조용언은 전체 보조용언의 93.45%였고 두결보조 용언은 6.53%. 그리고 세결보조용언은 현대 한국어에서 거의 사용되지 않는 것으로 나타났다. 따라서 두 개 이상의 보조용언이 결합하여 사용되는 경우는 극히 드물며, 해석의 순서를 따라 번역하여도 예문 (6)과 같이 비교적 적절한 번역문이 생성되었다.

かきたくない。(kakitakunai) (6)
쓰고 싶지 않다.

예문 (6)은 보조용언으로 '희망'의 보조형용사와 '부정'의 보조형용사가 사용된 문장으로 희망-부정의 결합이 생성된다. 이 경우 해석의 순서대로 "쓰고 싶지 않다"라고 번역하는 것은 올바른 것이다.

어미는 여러 개의 속성이 부가되어 하나의 번역어를 결정하도록 하였다. 어미는 한국어 문법에서 비종결어미와 종결어미로 구분하는데 본 논문에서 정의하여 사용하는 종결어미에 대한 요소를 그림 6에 보인다. 시제의 경우 이미 선어말어미에서 처리되었기 때문에 종결어미에서는 부가적인 처리만을 고려한다.

그림 6에서 낮춤의 경우는 높임의 경우와 아래 노드의 요소가 동일하며, 과거·미래는 현재 노드 중에서 평서·의문·감탄의 3개 요소를 가진다. 시스템에서 사용하는 종결어미의 일부를 그림 7에 보인다.

1) 일본어 해석에서 추출한 하나의 어절정보로 한국어를 생성하면 보조용언 등이 삽입되므로 최종 생성된 한국어 용언은 하나의 어절이 아닐 수 있지만 본 논문에서는 해석정보가 하나의 어절정보를 가지고 있으므로 이를 용언의 상 하나의 어절로 간주하기로 한다.

시제와 법, 높임의 정도에 따라 별도의 종결어미를 생성하는 것을 보이고 있다. 이 때 각각의 속성구조는 용언의 경우와 그렇지 않은 경우로 크게 나누어 구성하였다.^{[22]에}

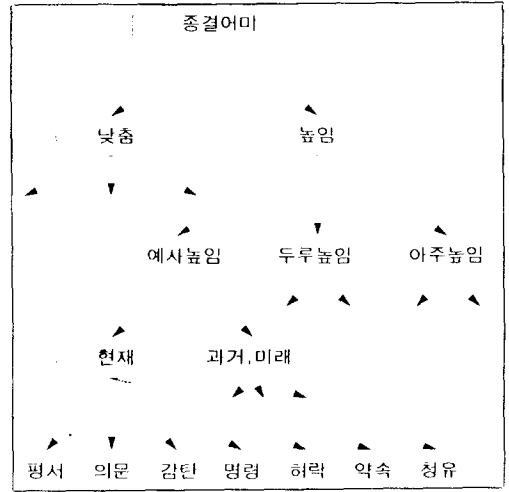


그림 6. 종결어미를 결정하는 요소
Fig. 6. Components to determine the final ending.

```

#-----
# 품사 속성 시제 법 미래/높임 어미변경 종결어미
#-----
< 평서 >
ADJNAT PRESNT DECLM LOWER1 0 단
ADJNAT PRESNT DECLM LOWER2 1 어
ADJNAT PRESNT DECLM LOWER3 1 네
ADJNAT PRESNT DECLM HIGHR1 1 오
ADJNAT PRESNT DECLM HIGHR2 1 아요
ADJNAT PRESNT DECLM HIGHR3 1 습니다
,
COPULA PRESNT DECLM LOWER2 0 어
COPULA PRESNT DECLM LOWER3 0 고요
COPULA PRESNT DECLM HIGHR1 0 요
#-----
< 평약 >

```

그림 7. 종결어미의 일례
Fig. 7. An example of the final ending.

2) 형용사의 번역

일본어의 형용사로 해석된 정보가 한국어로 번역될 때는 동사의 경우와 거의 일치하지만 제약은 더욱 심해진다. 그러나 이는 일본어 해석이 문법적으로 한국어와 거의 동일한 제약을 가지고 수행되기 때문에 생성이 난해해지는 문제점은 아니다. 예를 들면 진행동의 보조동사를 사용하지 못하지만 이는 일본어의 경우에도 마찬가지이기 때문에 크게 문제점이 되지

않는다. 다만 일본어 형용사의 문법규칙은 동사의 그것과는 다르기 때문에 속성정보를 문장에서 추출하는 방법은 차이가 있다. 그러므로 해석시에는 형용사와 동사가 전혀 다른 절차를 가지게 되며 한국어의 생성시에는 유사한 절차를 가질 수 있다. 본 논문에서는 동사와는 다른 해석과정을 거친 형용사의 번역을 동사와 결합하여 용언의 생성으로 한다. 이를 위해 동사와 형용사에 부가되는 속성을 같은 형식으로 정의하였다. 예를 들어 “美しいです”는 형용사 “美しい”에 ‘존중’의 속성이 부가된 형태로 해석하여 “아름답습니다”로 번역할 수 있다.

3. 조동사의 번역

일본어 조동사의 경우 이에 해당하는 한국어 품사가 별도로 존재하지 않으므로 번역시 어미의 속성으로 부가시켜야 한다. 예를 들어 “かきません”의 경우는 “かく”에 조동사 “ます”의 부정형이 결합된 형태이며 한국어로는 “쓰다”에 ‘존중’과 ‘부정’의 속성이 부가되어 “쓰지 않습니다”로 번역한다. 이처럼 일본어 조동사는 용언의 속성으로 부가되기도 하지만, “です”, “らしい”류의 조동사는 체언에 직접 접속할 수 있으므로 별도의 처리가 필요하다. “です”의 경우 전접 형태소가 명사이면 “이다”라는 서술격 조사를 삽입하고 ‘존중’의 속성 및 활용에 따라 속성을 부가하며, 만약 명사가 상대성 명사라면 ‘이다’대신 ‘하다’라는 보조어간을 삽입한다. 즉 “決定です”의 경우 “결정이다”라고 번역하지 않고 “결정하다”라고 번역하게 된다. 또한 전접의 단어가 형용사의 종지형이면 ‘존중’ 속성을 부가한 뒤 이후의 속성을 계속 부가한다. 이외의 경우는 부가되는 속성에 따라 번역한다. 일본어 문장 “學生でした”는 “學生”이라는 명사에 “です”의 과거 활용형이 부가된 형태이므로 ‘이다’라는 서술격 조사에 ‘존중’의 속성이 부가되고 이후에 ‘과거’의 속성이 계속 부가되어 “학생이었습니다”라고 번역한다.

Ⅲ. 한국어 음운현상

해석된 일본어 정보와는 무관하게 번역된 명사와 조사의 결합이나 용언의 어간에 부가된 속성의 결합에서 한국어에는 특별한 음운현상이 발생한다. 본 논문에서 제안한 한국어 생성 시스템에는 다음과 같은 것을 대상으로 음운 현상을 고려하였다.

체언과 조사간의 대표적인 음운현상으로는 받침의 유무에 의한 조사의 변화가 있다. 격조사 “이”는 전접하는 형태소에 종성이 존재하는 경우이고, 종성이 없으면

“가”로 번역해야 한다. 많은 조사들이 전접 형태소의 종성 유무에 따라 두 가지의 변화를 가진다. 그러나 특히 종성이 “ㄹ”인 경우는 좀 더 다양한 조사의 변화가 발생하는데, 이 경우 종성이 있어도 이후에는 “로”라는 조사가 결합된다. 일반적으로 조사 “로”는 전접하는 형태소에 종성이 없는 경우에 해당하는 것이다.

용언에 어미가 결합하면서도, 혹은 어미끼리의 결합에도 규칙적인, 혹은 불규칙적인 음운현상이 발생한다. 규칙적인 음운현상으로는 “으 삽입”, “으 탈락”, “ㄹ 탈락”, “어 변화” 등이 있으며 이들은 전산 언어학적인 입장에서는 처리가 용이한 것이다. 시스템에서 고려하는 불규칙 활용의 종류의 일부는 표 2와 같다. 이들 불규칙 용언은 이것이 발생하는 용언에서만 규칙적으로 일어나므로 사전에 이의 유형을 등록하여 생성시 고려하도록 한다.

표 2. 불규칙 활용의 종류

Table 2. The type of irregular conjugation.

갈래	조건	작용	예
ㅅ 불규칙	아/으	어간말음 ㅅ 탈락	있다, 적다, ...
ㄷ 불규칙	아/으	어간말음 ㄷ → ㄹ	간다, 건다(夢)...
ㅂ 불규칙	아/으	어간 말음 ㅂ 모음이 양성이면	오 삽입
		음성이면	우 삽입
ㄹ 불규칙	아/으	어간말음 ㄹ 탈락	호르다, 가르다, 오르다, ...
		어간말에서 두번째 글자받침에 ㄹ 첨가	푸다 (하나 뿐)
우 불규칙	아/으	어간말음 ㅍ 탈락	푸다 (하나 뿐)
여 불규칙	아/으	어미가 ㄹ로 변화	‘하다’ 동사
리 불규칙	아/으	어미 아 → 리	이리다, 누리다
		어미초성이 자음	어간말음 ㅎ 탈락
ㅎ 불규칙	어미초성이 모음	어간말음 ㅎ 탈락	까말다, 노말다...
거라 불규칙	아/으	아라 → 거라	
나라 불규칙	아/으	아라 → 나라	

용언의 활용현상은 여러가지가 동시에 발생할 수 있으며, 본 논문에서 고려하는 용언의 활용에서의 음운현상의 처리순서는 다음과 같다.

- ① 으 삽입 ② 으 탈락 ③ ㄹ 탈락 ④ 어 변화
- ⑤ 불규칙처리 ⑥ 모음 축약 ⑦ 자소취성

Ⅳ. 번역 결과와 시스템 구현

본 논문에서 제안한 한국어 생성 시스템은 [23]의 형태소 해석시스템에서 일본어 문장을 처리한 결과를 전송 받아 번역된 한국어어를 생성한다. 한국어 생성 시스템의 표면적인 입력은 한국어 어절을 기준으로

띄어쓰기된 일본어 문장이며, 구체적으로는 그림 3과 같은 해석결과가 스택에 저장된 형태로 입력된다. 해석결과를 어절을 단위로 중심어와 부가되는 속성으로 구분된 정보를 가지며 이 속성정보를 중심으로 전절까지 제안한 방식으로 한국어를 생성한다. 한국어와 일본어는 중심어가 문장의 말미에 위치하는 등 어순이 비슷하다는 것을 가정하여 구문의 재구성은 하지 않았으며 어절을 단위로 적절한 한국어를 생성하였다. 번역된 일례를 표 3에 보인다.

표 3. 번역결과의 일례

Table 3. An example of the translated output.

No	입력문	번역문	기타
1	美しいと思います	아름답다고 생각합니다.	조사"と", 조동사"ます" 나이의 편이결어 체언결어 아니고 중심어가 형용사이므로 "않다"로 번역 평가 명사이므로 보조어간 "감"삽입
2	強くはない。	강하지는 않다	속성으로 "중중" 부가 "은" 탈락 음운현상
3	花らしいです。	꽃같습니다	다중의 속성부가 병행해석 "동사조사"의 기본 번역 예 ("から" 앞의 예)
4	暑:いた	뜨다	일본어 "동사조사"이후에 계속해서 해석되는 경우(からの 번역 예)
5	かきませんでしたか。	쓰지 않았습니까	
6	構成するから	구성하기 때문에	
7	構成するからである	구성하기 때문이다	

표 3의 5번 예문은 본 논문에서 제안한 한국어생성 알고리즘의 전형적인 일례로 절차적인 방식으로 번역한다면 중심어가 되는 "かく"에서 중조사 "か"까지 일괄적으로 번역을 수행하기 난해한 경우이다. 6번과 7번의 "構成する."의 경우는 명사+동사로 표층 해석되며 "構成"이라는 명사가 상태성 명사이므로 "する"는 이전 명사 "構成"의 번역어에 "하"이라는 동사형 어간을 추가시킨다. 그리고 "する" 자체는 기본형만 가지기 때문에 속성 구조에 '현재+평서'를 추가시키며 이의 결합된 형태 "구성하다"를 결과로 출력하게 된다. "構成するから"의 경우는 위의 경우에 일본어 접조사 "から"가 접속하여 일본어 "명사 + 동사 + 조사"의 형태를 한국어 "(어간 + 명사형 어미 = 명사) + (명사 + 조사)"로 변환시키는 것이다. 실제의 경우, (어간 + 명사형 어미)는 (명사 + 동사)가 동사의 속성 구조 때문에 변환된 것으로 생성 시스템 내에서는 동사의 형태를 유지하게 되며(이는 이후의 번역을 원활하게 하기 위함이다.), 결국 "構成する" 자체가 "구성하다"라는 하나의 동사를 이루는 것과 같은 결과를 가진다. "から"의 경우는 번역된 결과가 '명사 + 조사'로 분리되지만 이는 이후의 접속에 의한 번역을 위한 것이며 "때문에"라는 연어의 형식으로 시스템 내에서 처리된다. 본 한국어 생성 시스템은 Workstation 상에서 C 언어로 구현하였으며 [23]의 형태소 해석기와 함께 일한 기계번역을 수행한다.

V. 결론

일본어와 한국어는 구문적으로 많은 유사성을 가지고 있지만 한국어를 중심으로 할 때 하나의 어절 내에서는 많은 의미적 차이를 가지고 있다. 일한 기계번역기를 설계하기 위해서는 유사성이 많은 구문적 정보의 처리에 중점을 두기보다는 문법적, 해석적 차이점이 많은 어절내 번역에 중점을 두어야 한다. 또한 일본어 해석시에 추출할 수 없는 한국어 음운정보도 별도로 규정하여야 하며, 다양하고 적절한 한국어를 생성하기 위해 주석적인 번역방식이 필요하다.

본 논문에서는 주석적 방식에 의한 어절단위의 일한 기계번역 시스템을 위한 한국어 생성기를 설계하였다. 명사구의 번역은 체언에 부속하는 조사류를 의미정보에 의해 세분하여 생성하고, 용언의 번역을 위해 주석적인 한국어 생성방식을 설계하였다. 이를 위해 한국어 용언 생성 알고리즘을 제시하였으며 생성기의 입장에서 용언 내의 어간과 어미 등의 결합 순서를 규정하였다. 조동사는 용언에 접속할 경우 속성 정보로 어간에 부가되게 하였으며 체언에 접속할 경우 적절한 보조어간 또는 서술격 조사를 삽입하고 속성정보를 부가하게 하여 적절한 한국어를 생성할 수 있었다. 절차적인 한국어 생성 방식이 확장성에 문제점을 가지고 있으며 입력문이 복잡할 경우 적절한 한국어의 생성에 많은 난점을 가진 것에 비해, 본 논문에서 설계한 주석적인 방식은 속성의 추가로 세분된 한국어를 생성할 수 있으므로 확장이 용이하고, 특히 대상이 될 한국어 어절을 중심으로 일본어 해석 후 한국어를 생성하므로 생성의 범위가 명확하여 적절한 결과를 얻을 수 있었다. 아울러 본 논문에서 제안한 한국어 생성기는 언어 의존적인 처리가 아니라 해석된 속성에 의한 것이므로, 별도의 해석기에 의한다면 약간의 확장으로 영한번역 등 다른 언어의 번역에도 사용할 수 있을 것이다.

參考文獻

- [1] 日本人工知能學會, 人工知能ハンドブック, オーム社, 1990
- [2] 博松明, "自動翻譯電話のための音聲處理と言語處理", 電子情報通信學會論文誌 D-II Vol.75, No.10, 1992
- [3] 吉村賢治, 日高達, 吉田淳, "文節數最小法を用いたべた昔き日本語文の形態素解析", 情報處理學會論文誌, Vol.24, No.1, 1983
- [4] 吉村賢治, 武内美津乃, 津田健藏, 首藤公昭, "未

- 登録語を含む日本語文の形態素解析”, 情報處理學會論文誌, Vol.30, No.3, 1989
- [5] 平井誠, 北橋忠安, “格の強度と述語の構文および意味屬性を用いた格構造の變換生成について”, 情報處理學會論文誌, Vol.28, No.3, 1987
- [6] 奥雅博, “日本文解析における述語相當の慣用的表現の扱い”, 情報處理學會論文誌, Vol.31, No.12, 1990
- [7] 平井章博, 梶博行, 芦尺實, “機械翻譯向け前編集のための日本語係り受け構造の曖昧性検出方式”, 情報處理學會論文誌, Vol.31, No.10, 1990
- [8] Masahiro Oku, “A Method for Analyzing Japanese Idioms”.
- [9] 元吉文男, 大場健司, etc, “未定義語を含む文の多段階構文解析法”, 電子情報通信學會論文誌 D-II Vol.72, No.10, 1989
- [10] 池田尚志, “語法規則方式による日本語文の構文意味解析”, 情報處理學會論文誌, Vol.26, No.6, 1985
- [11] Gosse Bouma, etc, “A Flexible graph-unification formalism and its application to natural-language processing”, *Journal of Research and Development*, 1988.
- [12] N.Maruyama, M.Morohashi, etc, “A Japanese sentence analyzer”, *Journal of Research and Development*, 1988.
- [13] Shalom Lappin, Michael McCord, “Anaphora Resolution in Slot Grammar”, *Computational Linguistics* Vol.16, No.4, 1990.
- [14] Paola Velardi, Maria Teresa Pienza, “How to Encode Semantic Knowledge”, *Computational Linguistics* Vol.17, No.2, 1991.
- [15] Daniel D.K.Sleator, Davy Temperley, “Parsing English with a Link Grammar”.
- [16] 강석훈, 우요섭, 김한우, 최명욱, “다의어 처리를 고려한 일한번역 시스템의 구현에 관한 연구”, *대한전자공학회 추계학술발표대회*, 1990
- [17] Seok-hoon Kang, Han-woo Kim, Byung-uk Choi, “On-Line Morphological Analyzer for Japanese to Korean Translation”, *JTC*, 1992.
- [18] Young-kil Kim, Seok-hoon Kang, Han-woo Kim, Byung-uk Choi, “Plan Recognition for Q/A System Using Natural Language Processing”, *ICEIC*, 1993.
- [19] 한양대, 과학기술원, “대화체 기계번역에 관한 연구”, 한국통신 장기기초연구과제 최종보고서, 1992
- [20] 한양대, “자동통역전화를 위한 요소기술의 개발에 관한 연구”, 전자통신연구소 최종보고서, 1992
- [21] 佐治圭三 外, 日本語學の理解
- [22] 김승렬, 국어어순연구, 한신출판사
- [23] 강석훈, 최명욱, “일한번역 시스템을 위한 일본어 형태소 해석시스템 설계”, 대한전자공학회 논문 심사중, 1993
- [24] 하일라이트 국어문법
- [25] 남기심, 고영근, 표준국어 문법론
- [26] 金泰錫, 浦昭二, “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*, Vol.33, No.12, 1992

 著者紹介



姜哲堦(正會員)

1966年 8月 31日生. 1989年 2月
한양대학교 전자통신공학과 졸업
(공학사). 1991年 8月 한양대학교
대학원 전자통신공학과 졸업(공학
석사). 1992年 2月 ~ 현재 한양
대학교 대학원 전자통신공학과 박

사과정 재학중

崔炳旭(正會員) 第31卷 B編 第3號 參照

현재 한양대학교 전자통신공학과
교수