

1. 서론

오늘날 국제사회는 양질의 통신망의 확산으로 신속하고 정확한 정보를 쉽게 얻을 수 있게 되었다. 국제간에 유통되는 정보의 양이 급격하게 증가함에 따라, 이들 사이의 언어 장벽은 커다란 장애가 되어 왔다. 따라서 계산기가 실용화된 이후 번역작업을 계산기로 하고자 하는 연구가 꾸준히 계속되어 왔다. 하지만 기계번역의 대상이 되는 자연언어는 자의적(恣意的)이며 무한하고 개방적인 기호체계이기 때문에[조규빈93] 그 처리가 쉽지 않다.

기계번역에서 생성기는 번역 결과의 문장이 원래 문장의 의미를 가능한한 정확하게 표현하면 서도 자연스러운 문장이 되도록 해 주는 중요한 부 시스템이다[최운천91]. 일반적으로 기계번역의 품질은 생성된 문장의 다양성과 자연스러움으로 평가받기 쉽다. 따라서 기계번역에서 언어 생성부는 기계번역의 성능을 결정짓는 매우 중요한 역할을 하게 된다. 하지만 최근까지 기계번역 분야에서 한국어 생성기에 대한 본격적인 연구는 별로 없는 형편이다. 지금까지의 한국어 생성기에 대한 연구는 주로, 전체 기계번역 시스템의 일부로서 시스템 의존적인 방식으로 다루어 지거나, 한국어 생성에서 매우 중요한 역할을 하는 용언의 활용과 용언어절의 결합순서, 음운규칙의 적용순서 등의 용언어절 생성알고리즘에 대한 구체적인 연구가 부족한 실정이었다.

이 논문은 한.일 기계번역 시스템과 일.한 기계번역 시스템의 언어 생성부에 대한 연구의 일부로서 수행되었다. 따라서 이 한국어 생성기는 특정한 언어 생성기에 종속적이 되지 않도록 하기 위해 조사와 어미, 어순 등 특정 언어의 문법적인 정보들을 모두 별도의 파일(file)에 간단한 표(table) 형식으로 만들어 뒀으로써 다른 언어의 생성부로 이식이 쉽도록 하였다.

실제로 이 시스템은 그 기본적인 틀은 그대로 둔 채로 일본어 생성기로의 이식을 고려하고 있다. 일반적으로 한국어 생성부는 그 입력으로서, 원문 해석결과로부터 일련의 구조 변환과정을 거친 한국어 의존구조를 받게 된다. 이 논문에서 설계. 구현한 한국어 생성기는 입력이 되는 한국어 의존구조를 속성구조로 표현하여 다른 일반적인 기계번역시스템의 한국어 생성부로 사용하기 쉽도록 하였다. 용언어절의 다양하고 자연스러운 생성을 위해 어미를 체계적으로 분류하고, 어미결정에 필요한 인자들을 정리함으로써 구체적인 생성 알고리즘을 제시한다.

2장에서는 기계번역 시스템의 언어 생성부를 살펴본다. 현재 실용적인 기계번역 시스템의 대부분을 차지하고 있는 변환방식 기계번역 시스템에 대해서 고찰한다.

3장에서는 한국어의 형태적, 문법적 특징들을 알아본다. 구문생성을 위한 특성으로서 성분배열 순서인 어순에 관련된 내용을 다루며, 형태소 생성을 위한 여러가지 특성들도 정리한다.

4장에서는 이 논문에서 제안한 알고리즘을 기반으로 실제 한국어 문장 생성기를 설계한다. 전체 한국어 문장 생성기의 구조를 먼저 살펴본 후 시스템에서 쓰일 속성 구조의 실제 자료구조를 설계한다. 그 다음 구문생성 알고리즘과 형태소 생성 알고리즘을 제안하며 이때 어미 부분의 생성에 대해 체계적으로 정리한다. 다른 시스템으로의 이식성을 높이기 위해 사전은 어떻게 작성하는 것이 좋은지에 대해 논한다.

5장에서는 실제 한국어 문장 생성기의 구현에 대한 설명과 전체 프로그램의 구조를 알아본 후 이 생성기의 한계에 대해 분명히 정리함으로써 이후 시스템 확장에 도움을 주고

6장에서 이 논문의 결론을 맺는다.

2. 변환방식 기계번역 시스템의 구성

변환방식의 기계번역 시스템은 원문을 사전, 문법, 지식을 이용하여 원문의 의존구조로 바꾸는 해석과정, 이를 다시 역문의 의존구조로 변환하는 변환과정, 그리고 이 역문의 의존구조에서 사전, 문법, 지식등을 이용해 번역문을 생성하는 생성과정으로 이루어진다.

다음 그림 2.1은 일반적인 기계번역 시스템의 구성이다.



그림 2.1 변환방식 기계번역 시스템의 구성

2.1. 해석부

해석부는 크게 형태소 해석부와 구문. 의미 해석부 둘로 나눌 수 있다.

2.1.1. 형태소 해석

형태소란 일정한 음에 일정한 뜻이 결합되어 있는 가장 작은 단위이다. 뜻을 갖는 최소의 단위로 정의되며, 어소(語素)라고도 한다[조규빈93]. 형태소 해석이란 입력문장을 형태소 단위로 구분하고 사전에서 어휘항목에 대한 정보를 꺼내어 형태소의 해석을 수행하는 것을 말한다[임명현91].

한국어는 형태상으로 볼 때, 교착어에 속하는 언어로서[조규빈93] 특히 용언의 활용과 음운의 변화가 심해서 형태소 해석에 어려움이 있으며, 이러한 특징으로 인해 형태소 해석 부분은 전체 해석부에서 특히 중요한 비중을 차지하게 된다. 일반적인 한국어 해석 시스템에서 형태소를 규칙적용의 기본단위로 하는 경우가 많다. 이 방식은 정확한 해석 결과를 담보하지만 실제로 해석부에서 소요되는 시간의 대부분이 규칙을 적용하는데 걸리는 시간임을 생각하면, 규칙적용의 횟수가 지나치게 많은 이러한 방식의 해석은 계산기 공학적 측면에서 보면 비효율적이다. 규칙적용의 기본단위를 형태소로 하지 않고 어절로 한다면, 즉 어절내의 형태소들은 먼저 결합을 시키고 이들 어절을 대상으로 규칙을 적용한다면, 해석 효율이 크게 좋아진다. 하지만 이러한 어절단위의 해석 시스템도 몇가지 문제점을 가지게 된다. 이러한 문제점을 해결하기 위해서는 용언의 비종결어미(전성어미, 연결어미)와 조사가 붙은 어절의 경우에 이 어절을 먼저 서로 분리한 후, 각 어절에 대해 어절단위해석을 수행해야 한다.

2.1.2. 구문. 의미 해석

구문. 의미 해석 부분은 어절간의 관계를 문법 규칙에 의해 해석하는 부분이다. 현재의 자연 언어 처리는 1980년대 이후 가장 주목받아 온 단일화 문법(Unification-based Grammar)에 의해 주로 이루어 지고 있는데, 구문. 의미 해석에 쓰이는 문법 규칙도 대부분 이러한 단일화 문법을 근간으로 하고 있다. 하지만 기존의 단일화 연산이 지나치게 포괄적인 연산만을 정의함

으로써 실제 기계번역 시스템의 구문. 의미해석부분으로 사용하기에는 비효율적인 면이 많았다. 따라서 최근에는 PATR-II등에서와 같이 단일화 연산에다 선언적인 주석(declarative annotation)을 덧붙이는 방식을 사용한다. 그러나 실제 실용적인 기계번역 시스템의 제작에는 번역의 질 못지 않게 번역의 속도가 중요한 역할을 하게 된다. 따라서 단일화 연산을 더욱 세분화하고 이런 세분화된 연산을 기반으로 절차적인 주석(procedural annotation)을 부가해 줌으로써 효율적인 단일화 연산을 수행하는 방법이 논의되고 있다.[우요섭91]

2.2. 변환부

변환방식의 기계번역 시스템에서 언어 특질이 서로 크게 다른 두 언어를 번역하려면 원문의 의존구조를 역문의 의존구조로 적절히 변환시켜 주는 변환부 또한 중요한 역할을 한다.

2.3. 생성부

자연언어 처리에 대한 연구가 아직은 해석 부분에 집중되고 있기 때문에 생성부에 대한 연구는 별로 없는 형편이다. 또한 생성부는 해석부와는 달리 변환부와 더불어서 시스템에 종속적으로 설계되는 경우가 많으며 따라서 생성부를 별개의 시스템으로 설계하는 예는 별로 없었다. 생성부는 크게 원문 의존구조에서 변환부를 거친 후의 역문 의존구조를 입력으로 하여 역문의 실제 구문을 생성하고 각 어절에 대해 형태소 생성을 하게 된다.

2.3.1. 구문 생성

생성부의 입력이 되는 역문의 의존구조에서 실제로 문장을 만들어 내기 위해서는 구문 생성부가 필요하다. 구문 생성 규칙을 이용하여 구문을 생성하며 번역문의 어순도 이 부분에서 결정한다.

2.3.2. 형태소 생성

구문구조에서 각 단어의 품사와 어순을 결정한 다음 형태소 생성에서는 단, 복수 처리, 시제, 상(相) 정보등에 의한 어미 변화, 용언의 활용 처리 등을 수행하여 번역문을 생성하게 된다.

3. 생성기 관련 연구

3.1 ATOM

ATOM(jApanese To kOrean translation Machine)은 한양대학교 전자통신공학과 대학원 인공지능연구실에서 개발한 일한 기계번역 시스템이다. ATOM은 한국어와 일본어가 그 언어 구조가 비슷함에 주목하여 완전한 구문, 의미 해석을 유보하는 대신 시스템의 전체적인 효율에 중점을 둔 직접번역 방식의 일한 기계번역 시스템이다.[강석훈91] ATOM의 한국어 생성부는 별도의 생성 시스템으로 설계된 것이 아니라 일한 기계번역 시스템 전체와 유기적으로 연결되어 있어 별개의 한국어 생성 시스템을 분리해 내기 어렵다. 또한 한국어 생성기에서 다루어야 할 한국어 용언의 활용 문제라든가 하나의 어절 내에 형태소들의 결합관계에 대한 고찰이 부족하다.

3.2 MATE-EK

MATE-EK(MAchine Translation Environment - English to Korean)는 한국 과학기술원 전산학과 컴퓨터 시스템 연구실에서 연구되고 있는 영한 기계번역 시스템이다. 이 시스템의 한국어 생성부는 생성기의 입력이 되는 한국어 의존 구조를 정의하여 입력형태를 명확히 함으로써 생성기를 독립적인 시스템으로 구성하였으며[최운천91] 한국어 생성에 필요한 한국어의 특성과 문법을 비교적 자세하게 다루고 있다. 하지만 실제 시스템의 설계 면에서 보면 자세히 기술한 한국어 문법 정보들을 어떤 식으로 시스템에 적용하는지에 대한 부분과 형태소 생성 부분에서 용언의 어미 생성 부분에 대한 알고리즘이 제시되어 있지 않다.

3.3 속성구조를 이용한 한국어 생성기

이 논문에서 설계, 구현한 한국어 생성기는 속성 구조(feature structure)를 이용한 정보 기반 형태를 입력으로 설계하여 그 입력 양식이 명확하다. 따라서 다른 기계번역 시스템의 한국어 생성기로의 이식이 쉽다. 교착어인 한국어는 어미의 종류가 다양하며, 그 활용과 활용 과정에서 음운의 변이가 특히 심하다. 이러한 한국어의 문법적 특징을 잘 반영하기 위해 용언 어절 안에서 형태소 결합 순서와 구문 정보와 생성 형태소 사이의 관계를 간결하면서도 체계적으로 정리하였다. 이때 특정 언어 의존적인 정보인 어말어미 정보, 선어말 어미 정보, 조사 정보 등을 별도의 파일(file)에 표(table)형식으로 구성함으로써 같은 언어구조를 가지는 다른 언어의 생성기로도 이용할 수 있게 하였다.

최근 자동통역전화에 대한 연구가 활발하게 진행되고 있는데, 기존의 기계번역 시스템은 주로 과학기술용 서적 등의 문장체 말을 그 대상으로 하고 있기 때문에 생성기가 다양한 양식의 어미를 다루지 않았지만 인간의 말을 주 대상으로 하고 있는 이러한 대화체 기계번역 시스템의 한국어 생성부에서는 상대 높임의 정도, 시제 등에 따라 같은 뜻이더라도 형태가 조금씩 다른 다양한 어미가 쓰인다. 다양하고 자연스러운 한국어를 생성하기 위해 한국어 어미 부분의 형태소 생성에 특별히 중점을 두었으며, 이들 부분의 생성에 대한 간결하고 명확한 알고리즘을 제시하였다.

4. 생성기의 입장에서 본 한국어의 특성

한국어는 형태상 교착어에 속한다. 교착어는 문법형태의 기능이 단일하여 그 기능과 형태 사이에 1 : 1의 대응 관계가 성립하는 언어이다. 예를 들어 ‘가시었다’에서 ‘시’는 높임, ‘었’은 시제, ‘다’는 서법(Mood) 기능을 가지는 것처럼 하나의 형태에 하나의 기능이 있다는 말이다. 또한 첨가어라고도 하는데 이는 의미를 나타내는 실질 형태소에 조사, 어미 따위의 어법적 관계를 나타내는 형식 형태소가 붙어 문법기능을 함을 말한다[조규빈93]. 실질 형태소와 형식 형태소가 붙을 때, 특히 용언의 어간과 어미가 결합할 때 매개모음의 삽입이나, 불규칙 활용, 또는 규칙적인 음운 현상 등이 특히 심해서 형태소 해석이나 생성에 어려움이 있다.

문장을 구성하는 방식은 영어가 ‘주어 + 동사 + 목적어’의 구성을 갖는데 대해 한국어는

‘주어 + 목적어 + 동사’의 구성을 갖는 중심어 후행언어로서[서영훈90][성기철92] 수식어가 수식받는 말 앞에 놓인다. 주어의 생략이 잦고 조사가 발달하여 어절의 문법기능이 명확하기 때문에 문장의 어순이 비교적 자유롭다[남기삼93][조규빈93].

4.1 한국어 구문 생성에 필요한 특성

한국어 구문생성에 필요한 특성은 주로 어순과 관련된 것들이다. 어순은 단순히 단어의 순서가 아니라 문장 성분의 배열 순서이다[성기철92][김승렬90]. 구문 생성의 단위는 통사적인 단위인 단어가 아니라 문장의 성분이다. 다음은 구문생성에 필요한 한국어의 특징을 정리한 것이다.

① 국어 어순의 가장 기본적인 원리는 중심어(head) 후치이다[성기철92]

모든 종류의 수식 성분은 수식을 받는 성분, 즉 중심어 성분 앞에 와야 한다. 한가지 예외가 본용언과 보조용언의 관계이다. 본용언과 보조용언은 그 이름이 말해주듯이 앞에 오는 본용언의 의미상 중심이 되어 중심어 후행의 원칙에 어긋난다. 하지만 생성기의 입장에서는 보조용언을 별개의 성분으로 구별하지 않고, 일종의 특수한 어미의 형태로 처리한다. 이러한 보조용언은 하나의 속성으로 대표되어 용언 어절내의 형태소 결합에서 처리한다.

② 서술어를 제외한 주성분의 위치가 자유롭다.

서술어를 제외한 성분은 자유로운 자리바꿈이 허용되는데 이는 문장안에서 문법기능을 나타내는 조사가 매우 발달하여 문장의 위치에 상관없이 문법관계가 명확하기 때문이다[김승렬90]. 주성분의 위치가 자유롭다하더라도문장 유형에 따라, 일정한 격들의 자연스러운 배열 순서가 존재한다. 따라서 생성과정을 수행할 때는 해석과정과는 달리 한국어의 이러한 특징은 크게 문제가 되지 않는다.

③ 부사어의 위치가 자유롭다

문장의 주요 구성 성분 중 부사어만큼 어순이 자유로운 성분은 없다[성기철92]. 시간부사어(오늘, 어제,...)나 처소부사어(학교에서, 집에서,...)등은 자리바꿈에 별 문제가 없다.

ㄱ. 어제 나는 학교에 갔다.

ㄴ. 나는 어제 학교에 갔다.

ㄷ. 나는 학교에 어제 갔다.

하지만 모든 부사어가 자유로운 어순을 취하는 것은 아니다. 부사에 따라서는 어순에 상당한 제약이 수반되는 것도 있다. ‘조금’, ‘매우’ 따위의 부사어는 다음 예에서처럼 문장의 제일 앞에 오는 것은 불가능해 보인다.

ㄱ. *조금 친구는 바둑을 좋아한다.

ㄴ. 친구는 바둑을 조금 좋아한다.

ㄷ. 친구는 조금 바둑을 좋아한다.

생성기의 입장에서 보면 부사어의 어순은 다음 몇가지 예외를 빼면 서술어 바로 앞에 두는 것이 무난하다.

첫째, 부사어들이 여러개 겹쳐 있을 때는 시간부사가 처소부사보다 앞에 오는 것이 자연스럽다.

ㄱ. 나는 어제 그곳에 가지 않았다.

ㄴ. 나는 그곳에 어제 가지 않았다. (‘어제’라는 시제의 강조 의미)

둘째, 부정부사 ‘아니’, ‘못’은 중심어 바로 앞에만 온다.

ㄱ. 나는 담배를 안 핀다.

ㄴ. *나는 안 담배를 핀다.

ㄷ. *안 나는 담배를 핀다.

셋째, ‘마치’와 같은 부사는 대상이 되는 부사어 바로 앞에 온다

ㄱ. 저 남자는 마치 여자와 같다

ㄴ. ?마치 저 남자는 여자와 같다

ㄷ. *저 남자는 여자와 마치 같다

부사어의 어순 결정 원칙에서 문제가 되는 것은 다음과 같은 예이다.

ㄱ. 그들은 모두 술을 마셨다.

ㄴ. 그들은 술을 모두 마셨다.

이처럼 생성기의 입력이 되는 한국어 의존구조에서 부사가 어떤 것을 수식하는지를 명확히 해 주지 않으면 위와 같이 부사의 위치에 따라 문장의 뜻이 달라지는 것을 처리할 수 없다.

③ 관형사 사이의 어순 비교적 고정되어 있다[성기철92]

지시관형사 - 의문관형사 - 성상관형사 - 수관형사 - 명사상당어구

ㄱ. 그 새 두 건물

ㄴ. 어떤 세 물건

ㄷ. 그 어떤 것

4.2 한국어 형태소 생성에 필요한 특성

한국어 형태소 생성에 필요한 특징들은 주로 용언의 활용과 각 성분들 사이의 결합에서 나타나는 음운 현상에 대한 것들이다.

① 불규칙 활용이 일어나기도 한다.

용언어간과 어미의 결합에서 특정한 용언들은 불규칙적으로 활용한다. 불규칙 활용의 종류는 ㅅ, ㅈ, ㅊ, ㄹ, 으, 여, 거라, 너라, ㅎ 불규칙이 있다.

② 결합때 나타나는 음운현상이 복잡하다.

어간과 어미 결합 때, 모음조화에 따라 어미의 모양이 달라지기도 하고, 결합에서 특정한 조

건이 만족하는 모든 용언에 대해 ‘으’가 삽입/탈락되기도 하며 ‘ㄴ’이 탈락되는 현상이 나타나기도 한다. 이러한 현상은 불규칙 활용과는 달리 특정한 음운끼리 결합할 때 나타나는 규칙적인 음운 현상이기 때문에 자동적인 처리가 가능하다.

③ 같은 기능이라도 다른 형태의 조사가 있다.

‘이/가’, ‘을/를’, ‘은/는’ 따위의 조사는 앞음절에 따라서 조사의 형태가 달라진다. 주격조사의 경우, 앞음절이 받침이 있다면 ‘이’가 쓰이고 없다면 ‘가’가 쓰인다.

④ 조사끼리의 결합양상이 다양하다.

조사가 단독으로만 쓰이는 것이 아니라 여러개의 조사가 복합적으로 결합하여 쓰이는 경우도 흔하다. 격조사끼리의 결합은 처소 부사격 조사와 일반격조사가 서로 결합하여 쓰이며, 격조사와 보조사의 결합은 비교적 자유로운 편이다. 또한 보조사끼리의 결합은 꽤 제한되어 있긴 하지만 상당히 많은 수의 보조사가 서로 결합가능한 것으로 보인다.

ㄱ. 격조사 + 격조사

거기에서의 싸움은 치열했지.

ㄴ. 격조사 + 보조사

로마에서는 로마의 법을 따라야 한다.

ㄷ. 보조사 + 보조사 [+ 보조사 ……]

지금부터라도 잘 해 봅시다.

5. 한국어 생성기의 설계

5.1 한국어 생성기의 구조

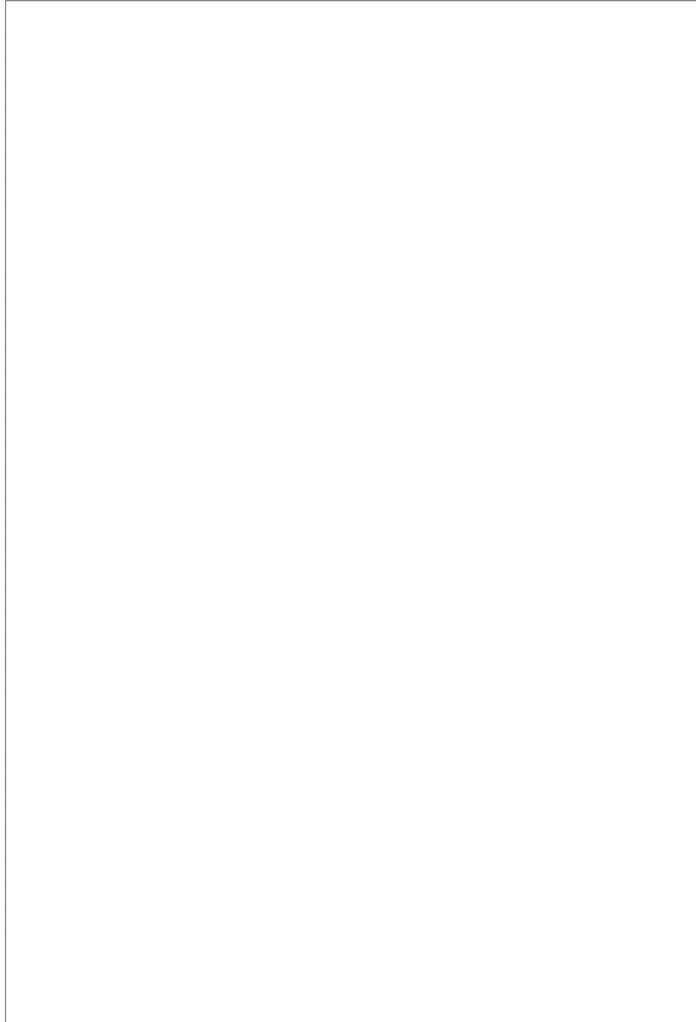


그림 5.1 속성구조를 이용한 한국어 생성기의 구조

기계번역 시스템에서 생성기는 번역하기 위해 주어진 문장의 의존구조를 입력으로 하여 번역문을 만들어 내는 부분이다. 생성기는 일반적으로 크게 구문생성부와 형태소 생성부로 나눈다.

① 속성값의 구조

속성값의 구조	들어갈 값
속성값사이의 관계	AND, OR, NONE
속성값(들)	속성값 문자열
다른 곳의 속성값의 포인터	속성값 문자열의 포인터

표 5.1 속성값의 구조

속성구조의 기본단위의 [속성-속성값]에서 속성값의 구조는 위의 표 5.1과 같이 설계하였다. 하나의 속성에는 하나의 속성값이 있는 것이 단일화 알고리즘의 원칙이나 이러한 원칙은 생성은 물론 해석에서도 부적절하다. 따라서 하나의 속성에 여러개의 속성값이 연접(Conjunction : AND) 또는 이접(Disjunction : OR)하는 형태의 속성구조가 도입되어야 한다[우요섭91]. 간단한 예를 든다면 ‘예약하다’라는 동사의 하위범주화에서 동작주(Agent)가 될 수 있는 것은 사람 또는 법인, 단체 등인데 이러한 ‘사람, 법인, 단체’의 속성값은 [사람 OR 법인 OR 단체]의 이접 구조를 가짐으로써 여러개의 속성구조로 되어 있는 것을 하나의 속성구조로 통합시켜, 효율을 높일 수 있다는 것이다. 그리고 속성값의 포인터는 다른 곳의 속성값을 공유하기 위해 필요하다. 다음 그림 5.3에서는 관형어 ‘쓰다’의 목적어가 그 수식성분인 ‘책’이 됨을 보여준다.

● 관재락이 쓴 책이

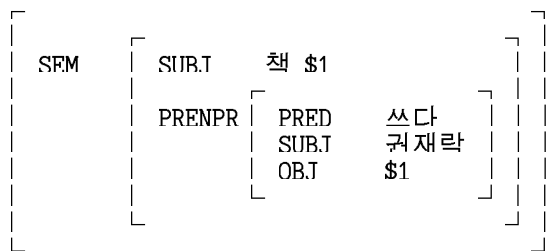


그림 5.3 속성값 포인터의 예

② 마디 구조

마디구조는 여러개의 [속성-속성값] 쌍이 하나의 마디에 존재하는 형식이다.

이러한 속성구조를 다루기 위한 자료 구조는 다음과 같다. [속성-속성값] 쌍을 갖는 마디구조 (node structure)가 있어야 하며 또한 그 속성값으로서 자기자신과 똑같은 마디구조를 가질 수도 있는 구조가 되어야 한다.그림 5.4는 마디구조의 개념도이다.

5.2.2 속성구조의 정의

이상과 같은 방법으로 설계한 마디구조를 이용하여, 한국어 의존구조를 표현할 속성구조를 정의한다. 이 한국어 생성기에서 사용하는 속성구조는기본적으로 다섯가지의 속성으로 구성되며, 일반 형태를 정의하면 다음과 같다. 다음은 이 생성기에서 사용하는 속성구조의 일반형태이다.

POS	문자열		
CAT	{ 품사 : 명사, 대명사, 수사, 조사, }		
HEAD	SYN	EWFORM	{ ① 종결형 : 평서, 의문, 감탄, 명령 ... }
			{ ② 비종결형 - 전성형 : 관형사형 ... }
			{ ③ 비종결형 - 연결형 : 단순나열 ... }
		SHONOR	{ +, - }
		TENSE	{ 과거, 과과거, 현재, ... }
		OHONOR	{ 상대높임 : 아주, 두루, 예사낮춤/높임 }
		
		
	SEM	NNFORM	{ 보통, 고유-사람, 고유-땅, ... }
		SUBJ	REF < ... >
		OBJ	REF < ... >
		PRED	< ... >
		
		
VAL	SELF	동사유형, 또는 명사의 의미소성	
SUBCAT	HEADD	중심어쪽의 속성구조 포인터	
	XXXXX	중심어 아닌 쪽의 속성구조 포인터	

다음은 ‘특선으로 뽑혔다’라는 간단한 문장의 속성구조 예이다.

㉠ ‘특선으로 뽑혔다’에 해당하는 속성구조의 예

```
( (POS      특선으로 뽑혔다)
  (CAT      VERBP)
  (HEAD     (SYN      (EWFORM  DECLM)
                      (TENSE   PAST)
                      (PASSIVE  +)
                )
    (SEM     (PRED     뽑)
              (QUAL    (REF      특선))
            )
  (VAL      (SELF     KVP11A))
  (SUBCAT   (HEAD     (POS      뽑혔다)
                      (CAT      VERBP)
                      (HEAD     (SYN      (EWFORM  DECLM)
                                          (TENSE   PAST)
                                          (PASSIVE  +)
                                    )
                (SEM     (PRED     뽑))
              )
    )
  (QUAL     (POS      특선으로)
            (CAT      NOUNP)
            (HEAD     (SYN      (NNFORM  NORMLN))
                      (SEM     (QUAL    (REF      특선)))
                    )
    )
  )
)
```

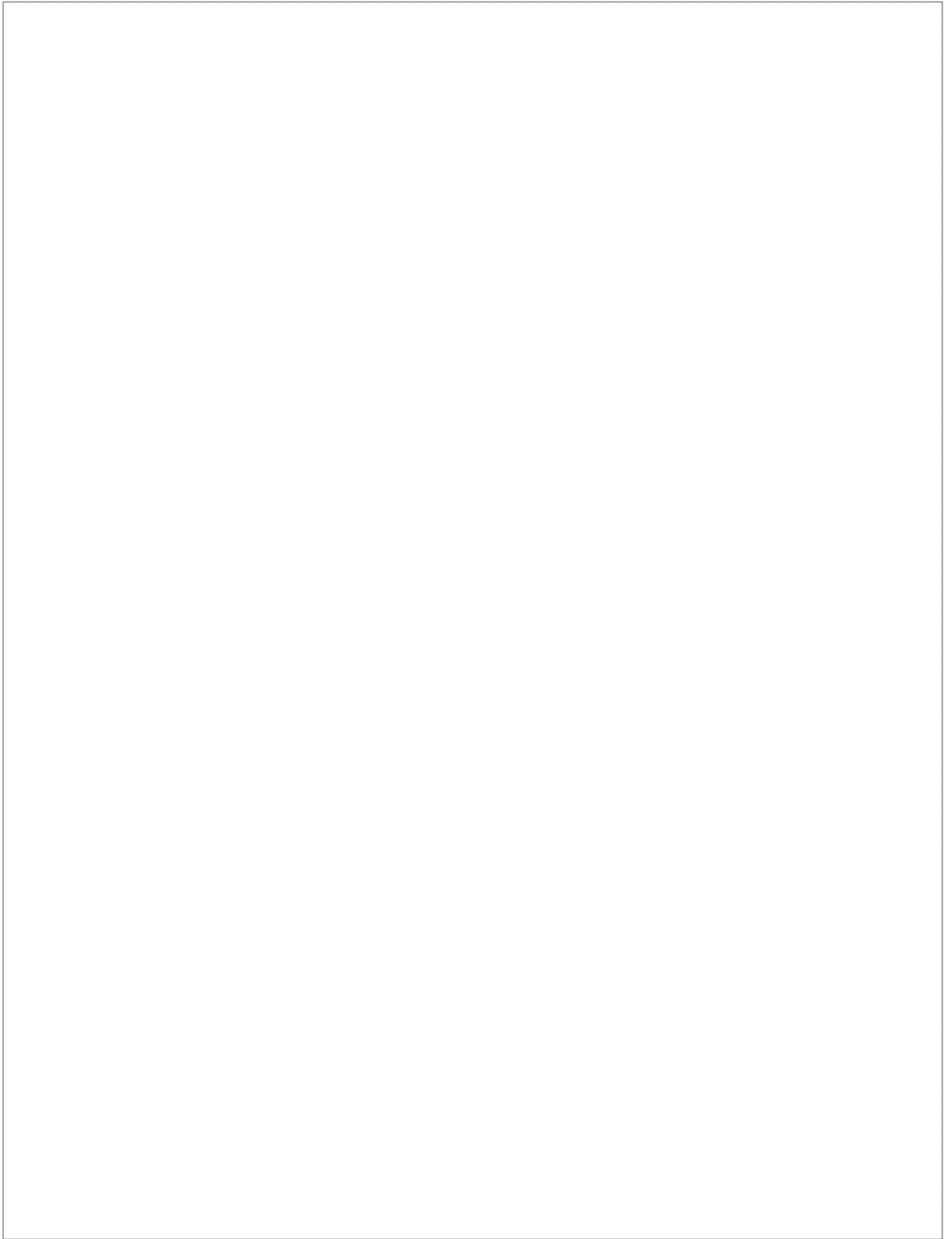



그림 5.4 마디구조의 개념도

5.3. 구문 생성

한국어 의존구조를 입력으로 하는 구문 생성부에서는, 먼저 의존구조에 나타난 문장의 의미구조를 구하고 그 의미구조에 나타난 문장성분들을 한국어의 구문구조에 맞도록 늘어 놓는다. 이렇게 늘어 놓은 문장성분들의 배열 순서를 어순이라 한다. 한국어 문장 생성에서 어순을 결정하기 위해서는 다음의 두 단계를 거친다.

① 동사유형에 따른 문장 성분의 순서화

한국어 동사를 성분배열에 따라 크게 열 여덟유형으로 나누고 이들 각각을 다시 세분하여 서른 여덟개의 유형으로 나눈다. 동사 유형의 분류는 [신효필90]의 분류를 따랐다. 이 정보를 기초로 성분배열간에 1차적인 순서화를 마친다.

② 중심어 후치 원칙에 따른 순서화

한국어 성분배열 순서의 대원칙인 중심어 후치 원칙에 따라 기타 속성들의 일반순서를 매긴다. 수식하는 성분은 수식받는 성분 앞에 오며 중심어인 문장의 서술어는 언제나 가장 뒤에 위치한다.

다음의 두 단계에 의해 다음의 문장에서 성분배열의 순서를 구해본다.

● 철수는 영희에게 순이가 내일 미국에서 온다고 말했다.

말하다 (SUBJ, STLOC, QUAD) 동사유형 KVP15A

오다 (SUBJ, SPLOC) 동사유형 KVP3A

밑줄 친 인용절 내부의 성분 배열 순서를 결정하는 단계는 다음과 같다. 우선 오다의 동사유형은 KVP3A[신효필90]로서 술어 이외의 성분 배열 순서가 주격, 장소부사격인 동사이다. 따라서 [순이 ▷ 미국 ▷ 오]로 순서를 정할 수 있다. 그 다음, 동사유형만으로는 부사의 위치를 결정할 수 없으므로 중심어 후치원칙에 따라 매긴 일반 성분 배열 순서를 이용한다. ‘시간부사

는 서술어 바로 앞에 두되, 장소부사격 보다는 앞에 두는 것이 자연스럽다'는 원칙에 따라 [순이 ▷ 내일 ▷ 미국 ▷ 오]로 순서화 시킬 수 있다. 주절의 성분 배열 순서도 같은 과정을 거쳐 그 순서를 정할 수 있지만, SUBJ로 대표되는 표층격이 존재하지 않는 경우라면 이런 성분이 문장내에 하나만이 존재할 경우에 한해서, 명사의 의미소성값과 동사의 하위범주화 정보에 의해 그 순서를 결정할 수 있다.

5.4. 형태소 생성

형태소 생성부에서는 구문 구조에서 결정된 각 단어의 품사와 어순에 따라, 나머지 형태소를 생성하게 된다. 형태소 생성부의 입력이 되는 것은 구문 생성부의 출력으로서 한국어 어순에 맞도록 배열된 기본적인 어절이 담겨진 리스트와 문장 전체에 대한 의존구조이다. 한국어는 의미를 나타내는 실질형태소에 조사나 어미같은 어법적 관계를 나타내는 형식 형태소가 붙음으로써 문법기능을 나타내는 첨가어이며, 하나의 어법관계에 하나의 형태소가 대응하는 교착어이다[조규빈93]. 때문에, 한국어 형태소를 생성하는 과정은 기본 형태소에 문법기능을 하는 형태소, 즉 조사나 어미 따위를 붙여나가는 과정이다. 그러므로 한국어 생성기에서 어절내의 형태소 결합 관계 특성을 체계적으로 파악하는 것이 무엇보다도 중요하다. 그런데 한 어절 내에서 형태소 사이의 결합에는 엄격한 순서가 존재한다[이미선93]. 생성기에서 다루는 어절의 개념은 해석기와는 조금 다르다. 생성기에서 다루는 어절의 개념은 선행 형태소나 후행 형태소를 필요로 하지 않는 발화 단위로서의 어절이다[김종미92]. 예를 들어, '먹는 척하다'는 실제 두개의 어절이긴 하지만 의미상으로 '척하다'가 '먹다'의 원래의 의미에 어떤 '시늉을 한다'는 부가적인 의미를 첨가하는 형태이다. 따라서 이런 것들을 하나의 단위로 생각해서 이를 발화단위의 어절로 정의한다. 실제로 생성부에서는 이와 같은 의존명사 구문이나 보조용언 구문을 모두 하나의 생성단위로서 생성을 수행한다.

접미사는 일반적으로 사전 처리하는 하는 것이 합리적이나, 사동화와 피동화에 사용되는 용언 접미사와 '-들, -쯤, -짜리' 등의 접미사는 예외이다. 용언 접미사의 경우, 용언 사전에 사. 피동화의 유형을 기술해 주어 서, 문장의 사동화, 피동화에 이용하고, '-들, -쯤, -짜리' 등의 접미사는 취하는 말의 수효가 많아 사전에 올리지 않는 것이 일반적이다. 이와 같이 사용범위가 넓은 '-들, -쯤, -짜리' 따위의 접미사를 특수접미사로 정의한다. 따라서 사동사. 피동사와 특

수접미사는 형태소 생성과정에서 일종의 전처리 과정으로 수행되어야 한다.

다음 그림 5.5는 국어의 발화단위 어절 내에서 형태소 사이의 결합순서를 보인 것이다. [김종미92]의 어절 구성도를 일부 수정하고 몇가지 단계를 추가했다.



그림 5.5 한국어 발화 단위 어절 구성도

5.4.1. 조사 생성

조사는 자립성을 갖지 못하고 자립형태소(주로 체언)에 붙어서 다른 말과의 문법적인 관계를 나타내거나 뜻을 더해주는 단어를 말한다[조규빈93]. 조사에는 다른 말과의 문법적인 관계를 나타내는 격조사와 여러격에 두루 쓰이면서 일정한 의미를 더해주는 보조사가 있다. 조사를 생성하기 위해서 필요한 요소는 체언의 표층격 정보와 체언말음에 따라 같은 격이라도 다른 형태의 조사가 있는지에 대한 정보, 이 두가지가 필요하다. ‘이/가, 은/는, 을/를’같이 같은 문법관계라도 체언말음에 따라 다른 형태를 갖는 조사가 있기 때문에 적절한 조사를 결정하기 위해서는 조사 생성 사전에 받침있음, 받침없음, 받침무관 어떤 경우에 해당하는 조사인지를 명시해 주어야 한다. ‘(으)ㄹ랑은’, ‘(으)ㄴ들’ 따위의 일부 보조사는 체언 말음에 따라 매개모음 ‘으’가

삽입되는데 이와 같은 규칙적인 음운현상은 자동적 처리가 가능하므로, 사전에 이를 기술해 줄 필요는 없다.

이밖에 조사가 복수개 결합하여 쓰이는 경우가 있는데, 이들 중 숫적으로 가장 많은 비율을 차지하는 것은 부사격 조사와 보조사의 결합 형태이다. 조사 결합 양식은 크게 네가지로 구분된다.

① 격조사 + 격조사

격조사 끼리의 결합은 『처소부사격 + 일반격 조사』의 형태가 있다[남기심93]고 하나 실제로 거부감 없이 쓰일 수 있는 것은 『처소부사격 + 주격』과 『처소부사격 + 목적격』 정도인 것으로 보인다.

② 격조사 + 보조사

격조사 중에는 부사격 조사만이 보조사와의 결합이 가능하며 보조사 중에서 ‘씩, 곧, 마는’ 등의 일부 보조사는 예외이다. 또한 뒤에 오는 보조사는 보조사끼리 결합이 가능한 형태라면 그 결합형도 대부분 가능하다. 부사격조사에서 같은 의미로 분류된 ‘에게서, 한테서 ……」 따위를 하나의 대표 형태만 취하는 방법으로 조사를 해보면, 이들 부사격조사와 보조사와의 결합 형태의 복합조사 갯수는

부사격조사 + 보조사 의 형태가 약 250개

부사격조사 + 보조사 + 보조사 의 형태가 약 260개

부사격조사 + 보조사 + 보조사 + 보조사 가 100여개 정도 이다.

③ 보조사 + 격조사

격조사의 후행을 허락하는 보조사는 ‘만’, ‘마다’, ‘부터’, ‘까지’ ‘조차’, ‘마저’등이고[남기심93] 이들은 일반격조사 앞에서 광범위하게 쓰인다.

④ 보조사 + 보조사

보조사 끼리의 결합은 상당히 제한되어 있는 듯하다. 반면, 특별한 특징 없이 서로 어울려 쓰

이는 경우가 많기 때문에 모든 보조사 끼리의 결합을 만들어 보아야 한다. 보조사는 [조규빈93]에서 분류한 것 중 종결보조사를 제외한 21개의 보조사를 대상으로 하였다. 이들끼리의 결합형을 모두 조사하여 이들 중 별 거부감 없이 쓰이는 것들만을 모았다. 표 5.2는 두개의 보조사가 결합한 것들을 정리한 것이다.

다음 표 5.3은 세개의 보조사가 결합한 형태이다.

결합 형태	예 문
부터만은	지금부터만은 그러지 말자.
부터만이라도	지금부터만이라도 잘 해 봅시다.
씩만은	두개씩만은 줄 수 없어. 세개라면 모를까.
씩만이라도	한개씩만이라도 주세요.
씩밖에는	한개씩밖에는 줄 수가 없군요.
까지만은	그때까지만은 참을 수 있어.
까지만이라도	그때까지만이라도 참아 봅시다.
뿐만은	그뿐만이 아니야.
뿐만이라도	가진 건 이것 뿐이라도 나는 행복해.

표 5.3 보조사 세개 사이의 결합

다음은 실제 조사 생성용 사전의 일부이다.

[조사]

```
#=====
# 속성      받침      조사
#=====
SUBJ      FINALX      가
SUBJ      FINALO      이
OBJ       FINALX      를
OBJ       FINALO      을
< 중간 선택 >
TOO       FINALZ      도
FOLLOW    FINALZ      마저
< 중간 선택 >
FROMTOPI  FINALZ      부터는
FROMNSEL  FINALZ      부터라도
#-----
```

결합 형태	예 문
만은	너만은 나를 지켜줘.
만이라도	당신만이라도 여기 남아주세요.
만도	이 짐승만도 못한 자식.
부터는	지금부터는 잘 할 수 있다.
부터만	지금부터만 잘 하면 된다.
부터라도	지금부터라도 잘 합시다.
부터야말로	지금부터야말로 잘 해야 한다.
밖에는	너밖에는 없구나.
씩은	한사람이 두개씩은 먹어야지.
씩만	한사람이 두개씩만 먹자.
씩이든지	두개씩이든지 세개씩이든지 상관없다
씩이라도	한개씩이라도 줘.
씩인들	두개씩인들 못 주겠니.
씩이나	세개씩이나 주다니.
씩은커녕	세개씩은커녕 두개도 못 받았다.
씩밖에	한개씩밖에 줄 수 없다.
씩도	세개씩도 못 준단 말이나?
까지는	두개까지는 줄 수 있어.
까지만	세개까지만 줘.
까지라도	그럼, 두개까지라도 줘라.
까지인들	내일까지인들 무슨 상관이라?
까지는커녕	내일까지는커녕 모래까지도 힘들겠어요.
까지밖에	내일까지밖에 시간이 없어.
까질랑은	그때까질랑은 해보지요.
까지도	그때까지도 안 된다면?
조차도	너조차도 그런 말을 하다니.
마저도	날씨마저도 좋지 않군.

표 5.2 보조사 두개 사이의 결합

5.4.2. 어미 생성

활용어의 어간에 붙어서 다른 말과의 관계를 나타내는 의존 형태소인 어미(語尾)는 크게 선어말어미와 어말어미로 나눌 수 있다[조규빈93]. 선어말어미는 실질형태소인 어간과 형식형태소인 어미 사이에 끼여, 높임. 공손. 시제 등을 나타내는 어미를 말한다. 어말어미는 어간이나 선어말어미의 뒤에 오는 어미를 말하는 것으로 크게 종결어미와 비종결어미로 구분되며, 비종결어미는 다시 연결어미와 전성어미로 가를 수 있다.

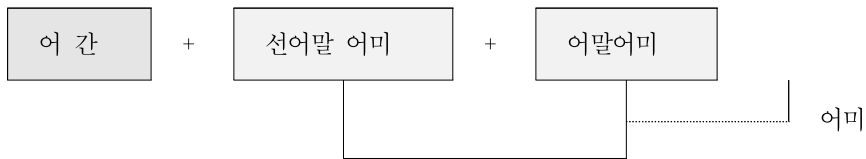


그림 5.6 용언 어절의 기본 구성

5.4.2.1. 어미 생성 알고리즘

생성 전처리 과정으로서 피동이나 사동 접미사를 먼저 어간에 붙여준다. 피동이나 사동의 뜻을 가지는 것은 접미사뿐만 아니라 보조용언도 있는데 어느 것이 더 적절한 지는 용언에 따라 다르다. 따라서 어떤 형식의 사동문 또는 피동문이 적절한 지는 용언의 생성 이전에 기술해 주어야 한다. 다음의 그림 5.7은 어미를 생성하는 알고리즘이다.

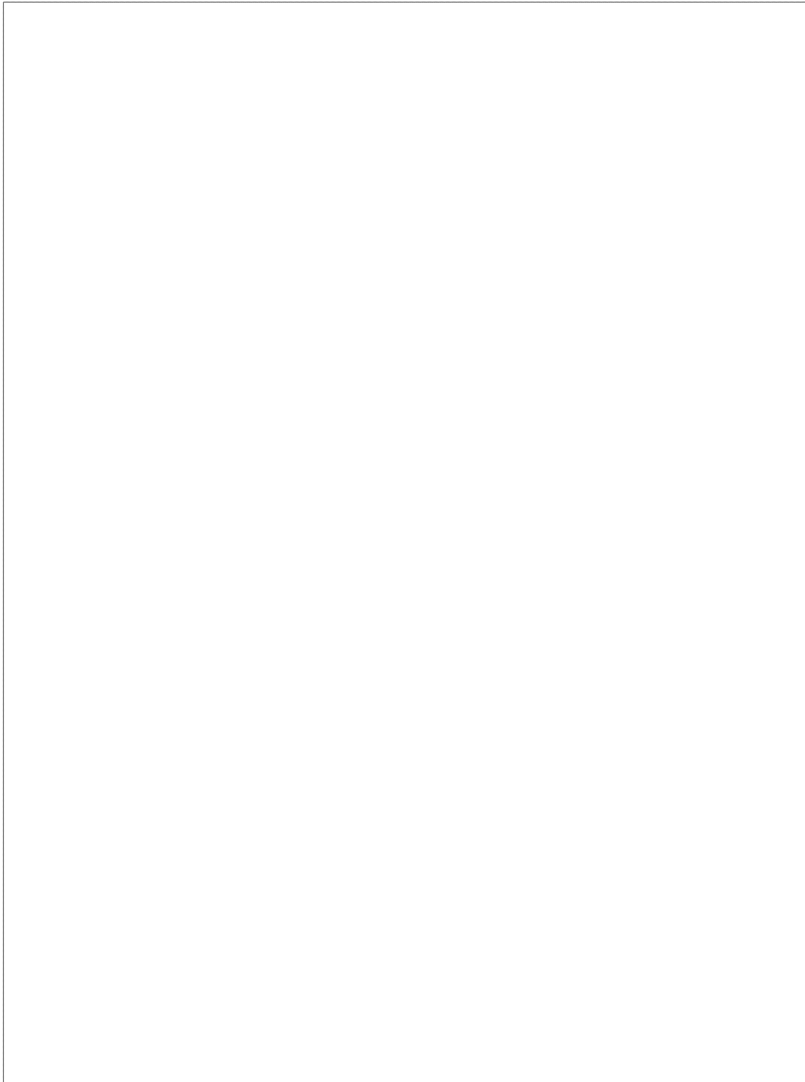


그림 5.7 어미 생성 알고리즘

5.4.2.1.1. 용언의 활용 처리

용언에 어미가 결합하면서 일정한 문법적인 관계가 표시된다. 이 때 용언의 끝이 여러가지로 바뀌는 현상이 일어나는데 이를 활용(活用 : inflection)이라 한다[조규빈93]. 특정한 용언에 대해 일정한 어미들이 결합할 때 어간 또는 어미가 음운규칙으로 설명할 수 없는 방식으로 달라지는 것을 불규칙 활용이라 한다. [강승석92]은 규칙적인 모음축약이나 일정한 조건을 만족하면 언제나 일어나는 ‘으’탈락, ‘ㄷ’탈락 따위의 음운현상도 ‘형태소의 모양이 바뀌면 모두 불규칙’이라는 원칙 하에 위와 같은 규칙적인 음운현상을 모두 불규칙으로 정의하고 있는데, 이러한 규칙적 음운현상들은 계산기 상에서도 자동 처리가 가능한 것을 생각하면 전산적 입장에서나 문법적 입장에서나 바람직한 정의는 아니다. 규칙적 음운현상을 자동적으로 처리하는 방법을 살펴보자. 어미 중에서 ‘-아서, -어서’처럼 뜻은 같으나 ‘아/어’의 이형태가 존재하는 어미들은 사전에 양성 모음으로 시작하는 ‘-아서’만을 등록한다. 이렇게 등록하고 나서, 그림 5.7의 어미 생성 알고리즘에 따라, 적절한 순서에 의해 음운현상들을 고려해 주면 같은 의미의 어미를 사전에 중복 등록할 필요가 없다.

어간모음		어미모음	축 약 형
ㅏ	+	ㅑ	ㅓ
ㅓ	+	ㅕ	ㅗ
ㅗ	+	ㅛ	ㅜ
ㅛ	+	ㅝ	ㅡ
ㅜ	+	ㅠ	ㅔ

표 5.4 모음 축약의 양상

자동적인 음운 현상으로서 모음축약이 있다. 어간말음이 모음이고 어미가 모음으로 시작한다면 다음 조건에 따라 모음축약이 일어난다. 표 5.4는 모음 축약의 양상을 정리한 것이다. 위의 모음 축약 표에서 ‘ㅗ’와 ‘ㅡ’의 축약형 ‘ㅔ’는 국어문법에서 정리한 것이 아니라 이 논문에서 제안한 알고리즘에 의해 어미를 생성할 때 필요한 축약현상이다.

다음의 표 5.5는 실제 문법에서 다루는 규칙적 음운변이 규칙과는 약간 차이가 있다.

갈 래	조 건
으 삽입	어간 말음이 ‘ㄷ’ 이외이고 어미가 ‘ㄴ, ㄹ, ㅂ, ㅇ, 시, ㅁ’으로 시작
으 탈락	어간 말음이 ‘ㄴ’이고 어미가 ‘아’로 시작
ㄹ 탈락	어간 말음이 ‘ㄷ’이고 어미가 ‘ㄴ, ㅂ, ㅅ, ㅇ’로 시작
어 변화	어미 시작이 ‘아’이고 어간 말음이 중성, 음성 모음이면

표 5.5 규칙적 음운 현상

이와 같이 규칙적 음운 첨가, 또는 탈락 현상이나 모음의 축약 현상등은 전산적 입장에서 약간의 수정만 하면 자동적 처리가 가능하다.

불규칙한 활용을 하는 용언들에 대한 활용 처리는 다음과 같은 방법으로 한다. 불규칙 활용이라는 것이 특정 불규칙 용언에 대해서는 규칙적인 활용 양상을 하기 때문에 불규칙 유형을 사전에 기술하면, 어미 생성 알고리즘 내에서 자동적 처리가 가능하다. 다음의 표 5.6은 용언들의 불규칙 갈래를 이 논문에서 제안하는 알고리즘에 따라 재정의 한 것이다.

※ 아/으 : ‘아’ 또는 매개모음 ‘으’로 시작하는 어미

※ ㅂ 불규칙 예외 : ㅂ 불규칙 용언 중에서 단음절 어간은 ‘돕다, 곱다’만 ‘오’삽입, 그 이외에는 앞음절 모음의 성(性)에 상관없이 무조건 ‘우’삽입이다.

갈 래	조 건	작 용	예
ㅅ 불규칙	아/으	어간말음 ㅅ 탈락	있다, 져다
ㄷ 불규칙	아/으	어간말음 ㄷ → ㄹ	긴다, 걷다(步) ..
ㅂ 불규칙	아/으	어간말음 ㅂ 탈락 어간 끝글자 모음이 양성이면 오 삼입 음성이면 우 삼입	깎다, 눅다, 줍다 돕다, 춥다, 곱다 덥다....
ㄹ 불규칙	아/으	어간말음 ㄹ 탈락 어간끝에서 두번째 글자받침에 ㄹ 첨가	흐르다, 가르다, 오르다, ...
우 불규칙	아/으	어간말음 ㅜ 탈락	푸다 (하나 뿐)
여 불규칙	아/으	어미가 ㄹ 로 변함	‘-하다’동사
러 불규칙	아/으	어미 아 → 러	이르다, 누르다
ㅎ 불규칙	어미초성이 자음	어간말음 ㅎ 탈락	까맣다,노랗다,..
	어미초성이 모음	어간말음 ㅎ 탈락 어미의 아 → 애	
거라 불규칙	아/으	아라 → 거라	
너라 불규칙	아/으	아라 → 너라	

표 5.6 불규칙 활용의 종류

용언의 활용에서 음운현상의 처리 순서는 다음과 같다.

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

5.4.2.2. 사동。피동 접미사 생성

동사를 사동화 또는 피동화 하는 방법은 크게 나누어서 접미사에 의한 방법과 보조용언에 의한 방법, 두가지가 있는데, 피동화의 경우에는 다른 단어로 대체해야 하는 경우도 생긴다. 접미

사에 의한 것과 다른 음운으로 대체되는 것은 어미 생성 전처리 단계라 할 수 있는 사동. 피동화 과정에서 생성작업이 수행되며, 보조용언에 의한 것은 다음 단계인 보조용언 생성부에서 처리하게 된다. 피동 어미의 종류에는 크게 다섯가지 유형이 있으며 이를 정리하면 다음 표 5.7과 같다 [최 윤 천 1 9 9 1] [조 규 빈 9 3] .

구 분	어미/피동형	보 기
어미 삽입	-이	쌓다, 놓다...
	-히	먹다, 잡다, 접다, 건다(收)...
	-리	날다, 휩쓸다, 졸다, 뚫다...
	-기	찢다,
보조용언 이 용	-아 지다	파랗다, 굵다....
'하다'가 다른 어간으로 바뀜 (수혜, 충고)	-되다	공부하다, 연구하다,...
	-받다	사랑하다, 존경하다, ..
	-당하다	조롱하다, 문책하다, ..
	-듣다	꾸중하다, ..
	-먹다	욕하다, ..
대응하는 동사가 있는 경우	맞다 ↔	때리다
	지다 ↔	이기다
	받다 ↔	주다
	배우다 ↔	가르치다
	받다 ↔	던지다
	사다 ↔	팔다

표 5.7 피동어미 처리의 유형

다음은 표 5.8은 사동 어미의 유형이다.

구 분	어미/사동형	보 기
어미 삽입	-이	녹다, 높다, 먹다, ...
	-히	입다, 앓다, 밝다, ...
	-리	들다, 날다, 불다, ...
	-기	안다, 남다, ...
	-우	지다, 비다, ...
	-구	돌다, ...
	-추	맞다, 늦다, ...
보조용언 이 용	-게 하다	‘-하다’동사, ...

표 5.8 사동 어미의 유형

5.4.2.3. 보조용언 구문 생성

자립성을 가지고 실질적인 의미를 나타내며 단독으로 서술능력을 가지는 동사와 형용사를 본용언이라 하고, 겉모습은 일반적인 동사, 형용사와 다름이 없지만 자립성이 없거나 약하여 본용언에 기대어 그 말의 뜻을 도와주는 동사, 형용사를 보조용언이라 한다. 이러한 보조용언은 양태적(樣態的) 의미만 더할 뿐, 주어에 대한 서술기능은 없다.

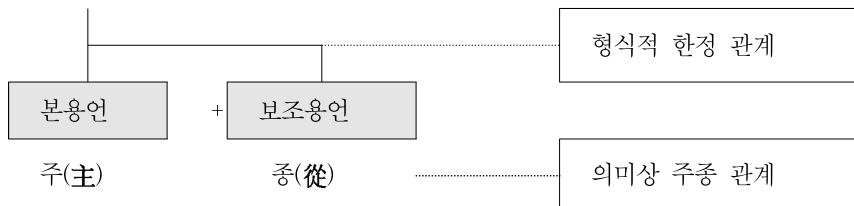


그림 5.8 보조용언과 본용언과의 관계

보조용언을 생성할 때는 보조용언이 하나의 속성에 대응하므로 이 속성 정보를 이용하여 보조용언 형태소를 본용언에 결합시켜 주면 되므로, 실제 생성 처리는 간단한 편이다.

① 보조동사

갈 래		형태, 종류	보 기
피동		(-아/-어) 지다 (-게) 되다	글씨가 잘 써 진다. 그도 성공하게 되었다.
사동		(-게) 하다 (-게) 만들다	누구를 가게 하느냐? 그를 행복하게 만들다.
진행		(-아/-어) 가다 (-아/-어) 오다 (-고) 있다 (-고) 계시다	그 일은 잘 되어 간다 날이 점점 어두워 온다 지금 원고를 쓰고 있다 선생님께서 원고를 쓰고 계신다
종결		(-아/-어) 내다 (-아/-어) 버리다 (-고) 나다 (-고야) 말다	어려움을 견디어 낸다. 도망가 버렸다. 공부를 하고 나서 놀아라. 해내고야 말겠다.
봉사		(-아/-어) 주다 (-아/-어) 드리다	장난감을 사 주었다. 선물을 사 드리다.
보유		(-아/-어) 두다 (-아/-어) 놓다 (-아/-어) 가지다	책을 책꽂이에 꽂아 두어라. 그대로 책상위에 놓아라. 그 책을 집어 가지고 오너라.
강제		(-아/-어) 대다	먹어 대다.
부 정	금지 불능 의지	(-지) 말다 (-지) 못하다 (-지) 아니하다	너무 멀리 가지 말아라. 넌 가지 못해. 나는 가지 않아.
시행		(-아/-어) 보다	이거 한번 먹어 봐.
집작		(-아/-어) 보이다	멋있어 보인다.
당위		(-어야) 한다	결혼은 해야 한다.
시인		(-기는) 하다	네가 좋기는 하다.

표 5.9 보조동사의 갈래

② 보조형용사

보조형용사에는 ‘희망, 부정, 상태, 추측, 시인’ 따위의 종류가 있는데, 정리하면 다음의 표 5.10과 같다

갈 래	형태, 종류	보 기
희망	(-고) 싶다	그가 보고 싶군요.
부정	(-지) 않다 (-지) 못하다.	그다지 답지 않군요. 마음이 넓지 못하군.
상태	(-아/어 있다) (-아/어 계시다)	매일 누워 있네. 늘 앉아 계시곤 했지요.
추측	(는가, ㄴ가, 나) 보다	정말로 그가 떠나는가 보다.
시인	(-기는) 하다	예쁘기는 하군.

표 5.10 보조형용사의 갈래

보조용언을 생성할 때 가장 주의해야 할 점은 본용언과 결합시킬 때 품사의 변화에 버금가는 일련의 변환 작용이 일어난다는 것이다.

ㄱ. 그가 보고 싶군요.

ㄴ. 정말 예뻐졌군.

‘보고 싶다’는 ‘보다’라는 동사에 희망을 나타내는 ‘-고 싶다’라는 보조 형용사가 붙은 경우인데, 이때 이 용언구의 품사는 동사이므로 이 후 생성과정에서 어미 결합 때 동사로서 활용하게 된다. 하지만 실제로는 ‘싶다’라는 보조형용사에 어미들이 결합하면서 그 다음 과정에서는 품사 정보는 동사이나 형용사처럼 활용하게 된다. 따라서 품사 속성 이외에 품사특성이란 속성

이러한 보조용언들이 이중, 삼중으로 나타나는 경우도 있다.

– 33 –

5.4.2.4. 선어말 어미 생성

어미(語尾) 구조체는 단어의 끝에 쓰이는 어말 어미(final ending)와 중간에 쓰이는 선어말 어미(prefinal ending)가 있다. 어말어미는 그 다음에 다른 형태소가 오지 않는 폐쇄 형태소(closing morpheme)인데 반하여, 선어말어미는 다음에 반드시 어말어미를 요구하는 개방 형태소(nonclosing morpheme)이다[고영근89][조규빈93][강승식91]. 이러한 선어말 어미는 다시, 결합가능한 어말어미가 제한적이어서 그 관계가 밀착되어 있는 교착적(膠着的) 선어말 어미와 용법 상의 별다른 제약이 따르지 않는 분리적(分離的) 선어말 어미로 나눌 수 있다. 교착적 선어말 어미에는 ‘-ㅁ-, -느-, -더-, -리-, -니-’ 따위가 있는데 이들은 실제로 어말어미와의 결합형태가 극히 제한적이므로 이를 어말어미의 일부로 생각해서 어말어미와의 통합형으로서 어말어미 부분에서 다룬다. 높임, 시제 선어말 어미같이 별다른 용법 상의 제약이 없이 광범위하게 쓰이는 것들을 분리적 선어말 어미라 한다. 따라서 실제로 선어말어미 생성부에서 생성하게 되는 선어말 어미는 이러한 분리적 선어말 어미이고, 일반적으로 한국어 해석이나 생성에서 선어말 어미라 함은 이 분리적 선어말 어미를 말하는 것이다.

① 높임 선어말 어미

술어(Predicate)의 주체가 되는 대상을 높이는 주체 높임법으로 ‘-시-’가 대표적 형태이다.

② 시제 선어말 어미

행위가 이루어진 때를 표시하는 선어말어미로서 현재시제의 ‘-ㄴ-/-는’과 과거시제로 ‘-었-/-았-’, 미래 시제로서 ‘-겠-’이 있다.[조규빈93] 과거시제를 세분해 보면 ‘-었-’이 중첩된 과거-과거 시제[장석진93]가 있다. 이는 과거의 경험을 나타내는 시제이다. 미래시제 선어말 어미 ‘-겠-’은 미래 시제라기 보다는 말하는 이 또는 듣는 이의 의지나 의도, 추측 따위를 나타내는 형태소이고, 실제로 단순한 미래 시제에는 보통 현재형이나 ‘어간+ㄴ 것이다’의 형태가 쓰인다.

ㄱ. 나 내일 떠나.
 ㄴ. 나 내일 떠나겠어.
 ㄷ. 너 좋았겠다.

미래 시제
 의지
 추측

[서정수90]에 의하면, 한국어에는 명시적으로 미래를 나타내는 시제 형태소는 없다고 한다. 이 한국어 생성기에서는 일반적으로 단순한 미래 시제의 선어말 어미라고 정의된 ‘-겠-’을 의지. 의도. 추측의 형태소로 정의하고 선어말 어미로 제정의한다. 단순 미래의 경우에는 선어말 어미를 생성하지 않는다.

③ 공손 선어말 어미

공손을 뜻하는 선어말 어미로서 ‘-옵-’이 대표적인 형태이나 현대 국어에서는 거의 쓰이지 않아 이 논문에서는 다루지 않는다.

둘 이상의 선어말 어미가 어울릴 때는 『높임-시제-의지』의 순서이다.

[강승식91]은 53개의 선어말어미 결합형을 선어말어미 구조체라 하여 제시하였지만, 이 논문에서 제시한 생성 알고리즘에 의하면 나타날 수 있는 선어말어미 구조체는 다음의

시, 시었, 시겠, 시었겠, 시었었겠
었, 었었, 었겠, 었었겠,
겠,

10개로서 충분하며 나머지의 다른 형태들은 ‘아 변화’나 ‘으’모음삽입, 또는 모음축약 따위의 음운 현상으로 규칙적인 현상이므로 별도로 취급할 필요는 없다.

5.4.2.5. 어말 어미 생성

어말어미는 문장을 종결시키는 종결어미와 그렇지 못한 비종결어미로 나뉜다. 비종결어미는 다시 명사형이나 관형사형으로의 전성이 일어나는 전성어미와 문장이 새로운 문장으로 대등적 또는 종속적으로 이어지는 연결어미로 나뉜다. 이러한 어말어미의 생성에는 여러가지 정보가 복합적으로 필요하며 이들 정보를 이용해 하나의 적절한 어말어미를 결정하게 된다.

5.4.2.5.1. 종결 어미 생성

종결어미가 우리말에서 차지하는 비중은 다른 문법 요소에 비해 아주 크다[한길91]. 종결어미는 발화에서 다양한 기능을 하고 있지만, 다른 한국어 생성부에서는 상대적으로 소홀히 다루지는 부분이다. 문어체의 종결어미는 상대적으로 제한되어 있지만, 대화체에서는 다양한 형태의 종결어미가 쓰이고 있어서 이에 대한 체계적인 연구 없이는 자연스러운 한국어 문장을 생성해 낼 수 없다.

종결어미에는 문장을 종결시킨다는 기능적인 특수성 때문에 말하는 이와 듣는 이 사이에 성립하는 사회적 등분인 존비법과 기능적 관계인 문체법이 성립한다. 또한 종결어미에는 말하는 이와 문장의 내용사이에 성립하는 서법(Mood)의 범주가 체계적으로 파악된다.

종결어미를 생성하기 위해 필요한 중요한 정보는 상대 높임, 시제, 서법이다. 여기에 품사적 성질, 즉 품사의 변형 처리를 않기 위해 도입했던 속성이 하나 더 추가된다. 종결어미의 양상은 어간부 특성이 형용사성인지, 지정사성인지, 그 이외의 것인지에 따라 달라지기 때문이다.

① 서법(Mood)

구 분	개 념	보 기
평서법	어떤 사실을 판단하여 말하는 형식	나는 너를 <u>사랑한다</u> .
의문법	문거나 의문을 품는 형식	너는 나를 <u>사랑하니</u> ?
감탄법	어떤 사실을 느끼어 말하는 형식	정말 <u>아름답구나</u> !
명령법	남에게 어떤 행동을 시키는 형식	얘들아, 조용히 <u>공부해라</u> .
허락법	남에게 어떤 행동을 허가하는 형식	그만 나가 <u>놀으렴</u> .
약속법	남에게 어떤 행동을 할 것을 약속하는 형식	그날은 내가 꼭 <u>가리다</u> .
청유법	행동을 같이 하자고 권유하는 형식	야, 밥 <u>먹자</u> .

표 5.11 법(Mood)의 구분

② 상대 높임

상대 높임법은 말하는 이가 특정한 종결어미를 씌으로써 말 듣는 이를 높이거나 낮추어 말하는 방법을 말한다.[남기심93] 그 높임의 정도에 따라 아주 낮춤, 두루 낮춤, 예사 낮춤, 예사 높임, 두루 높임, 아주 높임의 여섯가지 단계로 구분할 수 있다. 각각의 단계는 ‘해라, 해, 하게, 하오, 해요, 합니다’체라 한다.

대구분	소구분	소구분	종결어미	보 기
낮춤	아주 낮춤	해라	ㄴ 다,는다	밥을 <u>먹는다</u> .
	두루 낮춤	해	아	밥을 <u>먹어</u> .
	예사 낮춤	하게	네	밥을 <u>먹네</u> .
높임	예사 높임	하오	오	밥을 <u>먹으오</u> .
	두루 높임	해요	아요	밥을 <u>먹어요</u> .
	아주 높임	합니다	습니다	밥을 <u>먹습니다</u> .

표 5.12 상대높임에 따른 현재형 종결어미의 예

③ 시제

시제 정보는 이미, 선어말 어미 처리부분에서 형태소화했기 때문에 종결어미에서는 다음과 같은 부가적인 처리에 이용된다.

예를 들어 형용사성도 지정사성도 아니고 시제는 ‘현재’이며 ‘평서법’에 ‘아주 낮춤’의 성질을 갖는 종결어미는 ‘-는다’와 ‘-ㄴ다’ 두가지가 있다. 어간의 받침유무에 따라 같은 의미에 다른 형태소 두개가 대응하게 된다. 이와 같은 현상을 처리하기 위해서는 어간과 어미의 결합 때 몇가지 부가적인 작업을 수행해야 하지만 시제가 과거라면 이미 용언의 어간부에는 선어말 어미 처리부에서 생성된 형태소 ‘-었-’이 결합되어 있기 때문에 종결어미의 형태도 한가지 밖에 존재하지 않게 된다. 왜냐하면 과거 시제의 경우는 어떤 용언이라 하더라도 그 용언의 끝글자가 ‘었/았’ 또는 ‘였’이기 때문이다.

④ 어간부의 특성

표 5.13에서 보는 것처럼 품사 또는 어간부 특성(형용사성, 지정사성)에 따라 종결어미가 달라진다. 또한 특정의 서법과는 어울리지 않는 품사 특성이 있기 때문에, 어미 결정의 한 인자로써 사용하면 비정상적인 형태소 결합을 체계적으로 막아준다.

구 분	평서	의문	감탄	명령	허락	약속	청유
동 사	ㄴ다 는다	느냐	는구나	아라	렴	마	자
형용사성	다	냐	구나	。	。	。	。
지정사성	다	냐	로구나	。	。	。	。

표 5.13 현재시제 종결어미의 예

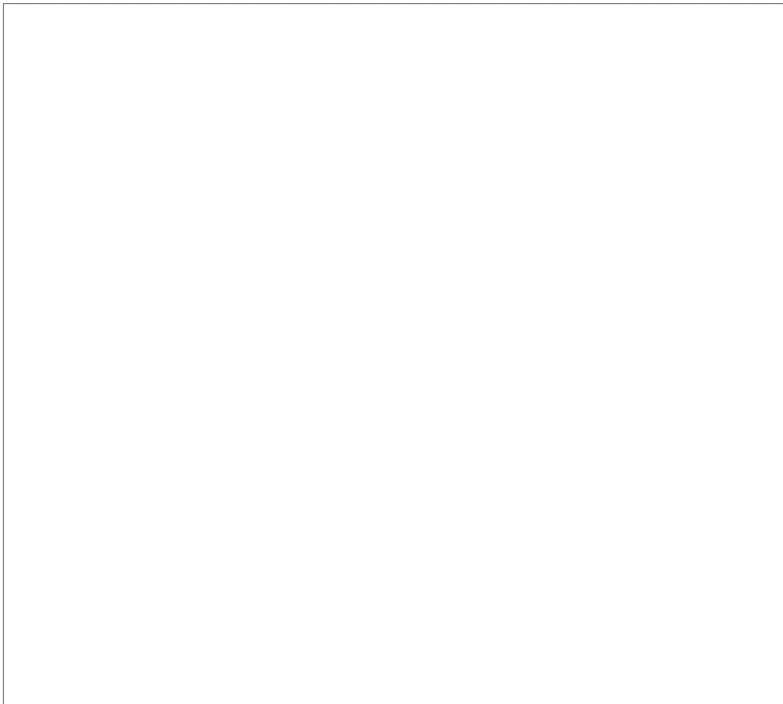


그림 5.9 종결어미를 결정하는 요소

앞의 그림 5.9는 종결어미는 결정하는 요소를 표현한 것이다.

다음은 종결어미 사전의 일부이다.

[종결어미]

#	품사속성	시제	법	상대높임	어미변경	종결어미
OTHERS	PRESNT	DECLM	LOWER1	1		는다
OTHERS	PRESNT	DECLM	LOWER2	1		아
OTHERS	PRESNT	DECLM	LOWER3	0		네
OTHERS	PRESNT	DECLM	HIGHR1	1		오
OTHERS	PRESNT	DECLM	HIGHR2	1		아요
OTHERS	PRESNT	DECLM	HIGHR3	1		습니다

< 중간 생략 >

ADJNAT	PRESNT	DECLM	LOWER1	0		다
ADJNAT	PRESNT	DECLM	LOWER2	1		아
ADJNAT	PRESNT	DECLM	LOWER3	1		네
ADJNAT	PRESNT	DECLM	HIGHR1	1		오
ADJNAT	PRESNT	DECLM	HIGHR2	1		아요
ADJNAT	PRESNT	DECLM	HIGHR3	1		습니다

< 중간 생략 >

COPULA	PRESNT	DECLM	LOWER2	0		야
COPULA	PRESNT	DECLM	LOWER3	0		ㄴ 세
COPULA	PRESNT	DECLM	HIGHR1	0		오
COPULA	PRESNT	DECLM	HIGHR2	0		요
COPULA	PRESNT	DECLM	HIGHR3	0		ㅂ니다
COPULA	OTHERS	EXCLM	LOWER1	0		구나
COPULA	OTHERS	EXCLM	LOWER2	0		군
COPULA	OTHERS	EXCLM	LOWER3	0		구면
COPULA	OTHERS	EXCLM	HIGHR1	0		구려
COPULA	OTHERS	EXCLM	HIGHR2	0		균요
COPULA	OTHERS	EXCLM	HIGHR3	0		습니다그러

#-----

5.4.2.5.2. 비종결 어미 생성

비종결어미는 종결어미와는 그 특징이 크게 다르다. 종결어미에서 나타나던 서법이나 존비법 따위가 비종결어미에서는 나타나지 않는다. 따라서 종결어미에 비해 상대적으로 단순한 모양을 갖는다. 비종결어미는 크게 전성어미와 연결어미로 나뉜다.

①-1 관형사형 전성어미

앞절의 주체에 대하여 서술어의 기능은 그대로 유지한 채로, 뒤따르는 체언에 대해서 관형어 구실을 하게 하는 어미를 말한다.

활용어	현재	과거	미래	회상
동사	-는	-ㄴ	-ㄹ	-던
형용사	-ㄴ	-던	-ㄹ	。
지정사	-ㄴ	-던	-ㄹ	。

표 5.14 시제와 품사에 따른 관형사형 어미

①-2 명사형 전성어미

용언을 임시로 명사와 같은 자격을 갖도록 하는 어미로서 ‘-ㅁ’과 ‘-기’가 있다. ‘-ㅁ’은 동작을 관념적으로 나타내는 것이라 하여 정법어미(定法語尾)라 하고 ‘-기’는 동작의 진행을 뜻한다고 하여 미정법어미(未定法語尾)라 하는데, 이 논문의 생성기에서는 아직 이러한 구별은 하지 않고 있다. 하지만 실제 구어체에서는 ‘-ㅁ’보다는 ‘-는 것’이란 어구가 더 잘 쓰이므로 명사형 전성어미는 ‘-는 것’이라는 명사형 상당어구를 생성하는 것으로 설계하였다.

② 연결어미

의 미	형 태	의 미	형 태
단순나열	-고	가정구속	-거든
순차나열	-고, -고서	가정양보과거	-던들
결과나열	-서, -아/어서	가정양보현재	-ㄴ 들
동시나열	-며, -면서	가정양보방임	-더라도
이어생김	-자(마자)	결과	-도록, -게끔
이유	-니까, -아/어서	대립	-나마
원인	-느라(고)	의도	-려고, -고자
원인방임	-거니와	경과설명	-는데
도입	-건대	의심과거	-ㄴ 지
양보	-어도, -든(지)	의심현재	-는지
양보방임	-아/어도, -나	의심회상	-던지
인정	-거니	의심추측	-ㄴ 지
추측인정	-려니	중단	-다가
추측방임	-려니와	그같음	-다시피, 듯이
가정설명	-면	방임	-ㄴ 망정
더함	-ㄴ 뿐더러	심해짐	-ㄴ 수록
목적	-러	필요	-어야만

표 5.15 연결어미

비종결 어미를 생성하기 위해 필요한 인자는 비종결인 어미의 속성과 시제 어미 결합 정보이다. 시제 어미로 분류된 ‘-었-, -겠-, -더-’와의 결합 제약을 사전에 기술해 주어야 하는 이유

는 비종결어미를 이루는 부분의 용언의 시제에 대한 정보가 속성구조로 전달되는 경우를 상상해 볼 수 있기 때문이다. 이를 사전에 기술해 두지 않는다면, 불필요한 시제 형태소들의 결합으로 틀린 형태소가 생성되기 때문이다. 일부 어미들은 어미자체가 가지는 의미 때문에 시제 형태소와 결합하지 못하는 경우가 있다. 예를 들어 사건의 순차적인 나열을 뜻하는 연결어미 ‘-고서’는 과거 시제 형태소와 결합하지 못한다. 따라서, 용언이 실제로 과거시제이더라도, 시제 선택어말어미를 결합하지 못하도록 제어해야 한다. 다음은 비종결어미의 실제 사전 일부이다.

[비종결어미]

#	속성	었	겠	더	어미변경	비종결어미
SMPLAR	1	1	0	0		고
ORDRAR	0	0	0	0		고서
RSLTAR	0	0	0	1		아서
SMLTAR	1	1	0	1		며
IMMDET	0	0	0	0		자
< 중간 생략 >						
ACKNWL	1	1	0	0		거니
GUSACK	1	1	0	1		려니
< 중간 생략 >						
NNINTF	1	0	0	1		ㄴ 망정
ADDDET	1	0	0	1		ㄴ 뿐더러
< 중간 생략 >						
PRENPA	0	0	0	1		ㄴ
PRENFU	0	0	0	1		ㄴ
GENCON	1	1	0	1		+

5.5. 사전

속성구조를 이용한 한국어 생성기의 사전은 일반사전으로 용언사전과 체언사전, 어미사전으로 종결어미 사전, 비종결어미 사전, 그밖에 보조용언 사전, 조사 사전으로 이루어져 있다. 기존의 한국어 생성시스템들이 Lisp이나 Prolog 등의 특정 프로그래밍언어로 사전을 기술하여 범용성과 유지. 보수가 어려운 반면, 이 논문에서 사용한 사전들은 모두 일반적인 형식의 표로 구성하

여 확장성과 유지. 보수성이 뛰어나다. 한국어에서 문법기능을 하는 이러한 요소들이 별도의 파일에 표 형식으로 저장. 관리되므로 언어 구조가 같은 경우에 한해서는 약간의 수정만으로 다른 언어의 생성 시스템으로 구성할 수 있다.

6. 한국어 생성기의 구현과 성능 평가

이 논문에서 설계. 구현한 속성구조를 이용한 한국어 생성기는 IBM-PC AT 386기종에서 실험하였으며 프로그래밍 언어로는 C 언어를 사용하였다. ANSI-C를 기준으로 작성하여 SUN Workstation 등의 UNIX 기반의 시스템에서도 수정없이 동작함을 확인하였다. UNIX기반의 시스템에서의 동작을 보장하기 위해 한글 코드는 KS-5601완성형 코드를 사용하였다. 완성형 코드로는 형태소 처리에 필수적인 자소 처리를 할 수 없기 때문에 조합형으로의 실시간 코드 변환 프로그램을 작성하여 내부적으로는 상용조합형 코드를 사용했다. 하지만 사전 시스템이나 구현 프로그램 소스에서 사용된 모든 한글코드는 완성형이다. 처리의 효율을 높이기 위해 사전등을 조합형 코드로 저장해 놓는다거나 프로그램 상에서 음운 현상 비교를 처리를 위한 부분에서 조합형 코드를 사용하는 것은, 현재 완성형 한글 코드만이 지원되는 UNIX기반의 시스템에서 프로그램의 유지, 보수를 어렵게 하는 한 요인이 된다.

생성기는 각 기능별로 다른 파일로 구성함으로써 프로그램의 유지, 보수와 재사용을 쉽게 하였다.

```

((POS    철수가 그런 풍경화가 전람회에서 특선으로 뽑혔다)
 (CAT    VERBP)
 (HEAD   (SYN    (EWFOM DECLM)
                  (TENSE  PAST)
                  (PASSIV  +)
            )
          (SEM    (PRED    뽑)
                  (QUAL    (REF    특선))
                  (SPLOC   (REF    전람회))
                  (SUBJ    (PREN   (PRED   그리)
                                   (SUBJ   (REF    철수)))
            )
          )
        )
      )
    )
  )
  (VAL.   (SELF   KVP5A))
  (SUBCAT (HEAD   (POS    전람회에서 특선으로 뽑혔다)
                  (CAT    VERBP)
                  (HEAD   (SYN    (EWFOM DECLM)
                                   (TENSE  PAST)
                                   (PASSIV  +)
                             )
                        (SEM    (PRED    뽑)
                                (QUAL    (REF    특선))
                                (SPLOC   (REF    전람회))
                        )
                  )
                )
            (SUBCAT (HEAD   (POS    특선으로 뽑혔다)
                            (CAT    VERBP)
                            (HEAD   (SYN    (EWFOM DECLM)
                                             (TENSE  PAST)
                                             (PASSIV  +)
                                           )
                                (SEM    (PRED    뽑)
                                        (QUAL    (REF    특선))
                                )
                            )
                        (SUBCAT (HEAD   (POS    뽑혔다)
                                        (CAT    VERBP)
                                        (HEAD   (SYN    (EWFOM DECLM)
                                                         (TENSE  PAST)
                                                         (PASSIV  +)
                                                       )
                                                (SEM    (PRED    뽑)
                                                        )
                                        )
                                    (QUAL    (REF    특선으로)
                                                (CAT    NOINP)
                                                (HEAD   (SYN    (NNFORM NORM.N))
                                                         (SEM    (QUAL    (REF    특선)))
                                                )
                                    )
                                )
                            )
                        (SPLOC (POS    전람회에서)
                            (CAT    NOINP)
                            (HEAD   (SYN    (NNFORM NORM.N))
                                    (SEM    (SPLOC (REF    전람회)))
                            )
                        )
                    )
                )
            )
        )
    )
  )

```

< 다음 장에 계속 >

7. 결론

변환방식의 기계번역 시스템의 한국어 생성부로서 속성구조를 이용한 한국어 생성기를 설계하였다. 생성기의 입력이 되는 의존구조를 속성구조를 이용하여 표현함으로써 생성부의 독립성을 높여, 다른 일반적인 변환방식 기계번역 시스템으로의 이식이 쉽도록 설계하였다. 조사, 어미 따위의 한국어 생성정보들을 별도의 파일에 표형식으로 저장. 관리함으로써 같은 한국어와 비슷한 언어구조를 가지는 일본어같은 언어의 생성기로의 변환도 쉽도록 하였다.

이 한국어 생성기는 일반적인 자동번역 시스템의 한국어 생성부로서 이용이 가능하며, 자동통역 전화 시스템등의 대화체 한국어 생성부로 사용할 수 있도록 다양한 종결어미를 처리할 수 있도록 중점을 두었다. 그간의 한국어 정보 처리 분야에서 연구가 부진했던 한국어 생성부에 대한 연구를 수행하였으며, 체계적이고 간결한 생성 알고리즘과 이를 위한 사전 구성방법을 제안하였다.

앞으로, 구문생성부에서 성분배열 순서를 위한 원칙들이 좀 더 검토되어야 할 것이며, 형태소 생성부에서는 준비관계를 종결어미부분에서만 처리해 주었는데, 역어선정이나 조사 선정에서도 이러한 정보를 이용하는 방법을 연구해야 할 것이며, 병렬 복합문을 처리하기 위해서는 이를 표현하기 위한 속성구조에 대한 고찰도 필요할 것이다.

참고문헌

- [권재일85] 권재일, 국어의 복합문 구성 연구, 집문당, 1985
- [성기철92] 성기철, “국어 어순 연구”, *한글* 218, pp.101-138 1992.12.
- [우요섭87] 우요섭, “영한번역을 위한 GPSG Parser의 설계 및 구현에 관한 연구”, 한양대학교 대학원 석사학위논문, 1987
- [우요섭91] 우요섭, “데이터베이스 자연언어 인터페이스”,
한양대학교 대학원 박사학위논문, 1991.12
- [강석훈91] 강석훈, “일한 번역 시스템(ATOM)에서의 한국어 생성 모델”
한양대학교 대학원 석사학위논문, 1991
- [조규빈93] 조규빈, 하이라이트 고교문법, 지학사, 1993
- [최운천91] 최운천, “변환방식의 기계번역을 위한 한국어 생성기의 설계 및 구현”, 한국과학기술원 석사학위논문, 1991
- [임명현91] 임명현, “단일화 문법에 의한 한국어 문장 해석기”,
한양대학교 대학원 석사학위논문, 1991
- [남기심93] 남기심. 고영근, 표준국어문법론, 탑출판사, 1993
- [김승렬90] 김승렬, 국어어순연구, 한신문화사, 1990
- [서태룡92] 서태룡, 국어활용어미의 형태와 의미, 국어학회, 1992
- [한 길91] 한 길, 국어종결어미연구, 강원대학교 출판부, 1991
- [권혁철92] 권혁철, “단일화 기반 의존 문법을 이용한 한국어 분석기”,
한국정보과학회 논문지 92, 9 Vol.19, No.5 Sep. 1992
- [이미선93] 이미선, “정형화된 한국어 형태소 분석기의 설계”,
제1회 ETRI 음성, 언어및 음향정보처리 워크샵 논문집, 1993
- [김종미92] 김종미, “국어의 형태소 결합순서”,
한국음향학회 논문집, 1992
- [강승식91] 강승식, 김영택, “한국어 형태소 해석기에서 선어말 어미의 분석모형”, *한국정보과학회 논문지*, Vol.18 No.5, 1991

- [강승식92] 강승식, 김영택, “한국어 형태소 분석기에서 불규칙 용언의 분석모형”, *한국정보과학회 논문지* Vol.19 No.2 Mar 1992
- [신호필90] 신호필, HPSG를 기초로 한 한국어 동사의 하위범주화, 언어학연구 제7호, 서울대학교 대학원 언어학과, 1990
- [고영근89] 고영근, 국어 형태론 연구, 서울대학교 출판부, 1989
- [서정수90] 서정수, 국어 문법의 연구, I, II, 한국문화사, 1990
- [서영훈90] 서영훈, 김영택, “활성차트를 이용한 중심어 후행언어의 파싱” *한국정보과학회 논문지* 90. 1. Vol. 17. No. 1 Jan, 1990
- [최재혁93] 최재혁, 이상재, “양방향 최장일치법을 이용한 한국어 형태소 분석기“, *한국정보과학회 봄 학술발표논문집* Vol.20 No.1
- [水谷靜85] 水谷靜夫, 외. 文法と意味 I, 朝倉書店, 1985
- [牧野武89] 牧野武則, 機械翻譯, オ-ム社, 1989
- [野村浩88] 野村浩郷, 機械翻譯, 共立出版株式會社, 1988
- [Gazdar89] Gerald Gazdar, Chris Mellish, Natural Language Processing In Lisp. Addison-Wesley, 1989
- [Shieber86] Stuart M. Shieber, An Introduction to Unification-Based Approaches To Grammar, 1986
- [Paris91] Cecile L. Paris, Natural Language Generation in Artificial Intelligence and Computational Linguistics.

Abstract

A Design and Implementation of Korean Generator Using Feature Structure

This thesis describes the design and implementation of a Korean generator using feature structure.

This Korean generator is independent of a specific machine translation system in transfer method. The Korean dependency structure, which is the input of the Korean generator, is described using feature structure. A grammatical information for the Korean generator is saved and controlled in the separate file system, so this generator system has a good extensibility.

The Korean is an agglutinative language, in which the post-endings are severely inflected. This generator uses a simple and efficient algorithm for post-ending generation. This Korean generator is able to generate various types of sentence. And morphological phenomena are well reflected for a natural and various type of Korean language.