

# 다층 형태론과 한국어 형태소 분석 모델

강 승 식  
한성대학교 전산통계학과

Multi-level Morphology and Morphological Analysis Model for Korean

Kang, Seung-Shik

Dept. of Computer Science and Statistics, Hansung University

## 요 약

형태소 분석은 단위 형태소를 분리한 후에 변형이 일어난 형태소의 원형을 복원하고, 분리된 단위 형태소들로부터 단어 형성 규칙에 맞는 연속된 형태소들을 구하는 과정이다. 이러한 일련의 분석 과정은 독립적인 특성이 강하면서 각 모듈이 서로 밀접하게 연관되어 있으므로 Two-level 모델에서는 형태론적 변형뿐만 아니라 형태소 분리 문제를 통합 규칙으로 처리하고 있다. 그러나 한국어에 Two-level 모델을 적용해 보면 형태소 분리와 형태론적 변형이 복합되어 있어서 교착어의 특성과 관계되는 단어 유형을 분석할 때 비효율적인 요소가 발견된다. 따라서 본 논문에서는 교착어인 한국어의 형태소 분석시에 발생하는 문제점들을 해결하는데 적합한 방법론으로 다층 형태론(multi-level morphology)과 다단계 모델(multi-level model)을 제안한다.

## 1. 서 론

자연 언어의 분석 과정은 분석 후보를 생성하는 과정(candidate generation)과 적합한 후보를 선택하는 과정(candidate selection)으로 기술된다. 이와 같이 형태소 분석 과정을 분석 후보 생성과 후보 선택의 과정이라 할 때 후보 생성 과정은 다시 형태소 분리(morpheme isolation)와 형태론적 변형(morphological alternation)으로 구분된다. 이러한 후보 생성과 후보 선택이라는 패러다임에 따라 한국어 형태소 분석은 (1) 단어에 포함되어 있는 단위 형태소(unit morpheme)들을 분리하는 과정과 (2) 변형이 일어난 형태소의 원형을 복원하는 과정, 그리고 (3) 단어 형성 규칙에 따라 단어를 구성하는 적합한 단위 형태소들의 조합(sequence of morphemes)을 구하는 과정으로 이루어진다[Kang94].

한국어 형태소 분석론은 주로 위의 3 가지 과정에 대해 별개의 방법론으로 제시되어 왔다. 단위 형태소를 분리하는 방법으로는 변형이 일어나지 않는 어간부(Head)와 변형이 일어나는 어미부(Tail)로 분리하는 Head-tail 구분법, 한국어의 음절 특성에 따라 문

법형태소의 경계와 형태론적 변형 음절을 효율적으로 인식하는 음절 단위 분석법이 제시되었다[최형석84, 강승식93]. 형태소의 원형을 복원하는 방법으로는 규칙 학습에 의해 형태론적 변형 규칙을 인식하는 방법과 변형이 일어난 어간을 사전에 수록하여 접속 정보를 이용하는 방법이 있다[Zhang90, Kwon91]. 단어 형성 규칙에 따라 단위 형태소들의 조합을 구하는 방법으로는 최장-최단 일치법과 Tabular 파싱법이 제시되었다[김덕봉90, 김성용87, 이은철92]. 이외에도 형태소간의 분리-결합 조건을 검사하기 위하여 계층적 접속 정보를 이용하는 방법과 밑몽치에 기반한 형태소 분석법, 중의성이 있는 단어를 분석하는 방법, overlap 정보를 이용한 양방향 최장일치법 등이 있다[김병희93, 김재한93, 최재혁94, 임희석94].

언어 독립적인 방법론으로 제시된 Two-level 모델은 여러 언어에 대해서 실험해 본 결과 매우 효율적이라고 알려져 있으며 따라서 Two-level 모델을 한국어에 대해 적용하는 방법에 대해서도 연구되었다[Kim94, Kwon94]. Two-level 모델에서는 개별 언어의

언어 현상들을 일반화시킴으로써 언어 종속적인 현상들을 처리할 때 backtracking overhead나 과분석 오류를 유발하기 때문에 개별 언어의 형태소 분석은 그 언어의 특성을 충분히 반영할 수 있어야 한다. 본 논문에서는 기존의 방법론을 한국어에 적용할 때 발생하는 문제점을 제시하고 이를 해결하기 위하여 현재까지 제시된 형태소 분석 방법론을 토대로 하여 한국어에 가장 적합한 형태소 분석 모델을 제안하고자 한다.

## 2. Two-level 모델과 한국어 형태소 분석

### 2.1 Two-level 모델

변형 문법에서는 형태론적 변형의 발생 유형과 이론적인 규명을 위하여 여러 계층(level)을 거쳐 일어나는 형태소의 변형 과정을 각 계층마다 개별적으로 기술한다. 따라서 이를 컴퓨터로 처리하려면 규칙의 적용 순서를 결정하기 어려운 문제가 발생한다. Two-level 모델은 이러한 문제를 해결하기 위하여 중간 계층(intermediate level)을 인정하지 않고 변형 규칙 대신에 표면층(surface level)으로부터 어휘층(lexical level)으로 직접 대응하는 합성규칙(Two-level 규칙)에 의하여 처리하는 전산언어학적 방법론이다. 그런데 Two-level 모델은 규칙을 기반으로 하고 있어서 컴퓨터로 구현시에 효율을 향상시키기 위하여 문자열 일치(string match)에 의하여 기술되는 Two-level 규칙을 유한 자동 기계(FST: Finite State Transducer)로 변환하여 bottom-up parallel 방식으로 입력 단어를 적용한다.

이 모델에 의하여 형태소 분석을 처리하는 알고리즘을 간략하게 기술하면 다음과 같다. 입력 단어와 트라이(trie) 구조로 된 사전을 한 문자씩 비교하면서 사전을 검색한다. 이 때 규칙의 조건부(LHS)와 일치되는 단어의 부분문자열이 발견되면 규칙의 실행부(RHS)에 기술된 문자열로 대치한 후 사전을 비교함으로써 backtracking에 의하여 형태소 분석을 수행한다[Kosk83].

형태소 분석의 가장 중요한 문제는 형태소 분리와 형태론적 변형의 처리이다. Two-level 모델은 굴절어에서 발생하는 형태론적 변형과 형태소 분리를 동시에 Two-level 규칙으로 처리하고 있다. 예를 들어, 영어 동사의 3인칭 단수형 'studies'를 분석하여 ('study' + '-s')라는 결과를 얻기 위하여 [ies\$ --> \_y+s\$]와 같이 string match를 기반으로 하여 기술되는 규칙을 사용한다. 따라서 굴절 현상의 유형이 다양한 언어일수록 규칙의 수가 늘어난다. 대부분의 언어에서는 현실적으로 규칙의 수가 수십 개에서 백 여개에 불과하므로 규칙에 의하여 처리하더라도 크게 문제가 되지 않는다[Bart86, Kosk88, Kart92].

Two-level 규칙은 문자열 일치를 기반으로 하는 제약이 있기 때문에 Two-level 모델을 확장하는 방안이 제시되기도 하였다[Bear88, Cahi90]. 그러나 언어 종속적인 형태론적 현상 중 형태소 분리 문제는 모든 분리 가능한 문자 위치마다 Two-level 규칙을 적용할 때 많은 backtracking 현상이 일어나므로 이를 감소해야 하는 경우가 발생한다.

### 2.2 한국어 형태소 분석

한국어 용언의 불규칙 활용은 형태론적 변형 규칙으로 기술되므로 어간의 원형을 복원할 때 Two-level 모델이 바로 적용될 수 있다. 하지만 한국어는 형태소 분리와 형태론적 변형이 복합되어 있어서 Two-level 모델을 적용했을 때 많은 overhead가 발생한다. Two-level 모델을 한국어에 적용할 때 발생하는 문제점은 다음과 같다.

첫째, Two-level 규칙은 문자열 일치를 기반으로 기술되는 제약이 있으므로 규칙의 조건부(LHS)와 실행부(RHS)가 모두 문자열로 기술된다. 이러한 제약으로 인하여 형태소 분리와 원형 복원 과정이 반복될 때 중간 분석 결과(intermediate result)를 다음 형태소의 분리-결합 제약으로 사용하기가 어렵다. 예를 들어, '가'를 '가-+'-'아'로 분석하기 위하여 중성이 없는 음절 중 중성이 ' / '인 것은 항상 '아'의 원형을 복원하여 후보를 생성해야 하는 부담이 있다. 이를 방지하기 위해서는 규칙들을 제약 조건에 따라 몇 개의 그룹으로 분류하여 현재까지의 중간 분석 결과에 따라 적용가능한 규칙의 수를 제한하여 과분석 오류를 생성하는 규칙이 잘못 적용되지 않도록 해야 한다.

Two-level 모델에서는 중간 계층을 인정하지 않기 때문에 여러 부류의 규칙이 있을 때 입력 단어와 분석 과정에 따라 특정 유형의 규칙이 적용되도록 하거나 특정 유형의 규칙은 적용 대상에서 제외되도록 제한하기가 어렵다. 그런데 한국어는 형태소 분리와 관련된 규칙, 불규칙 원형을 복원하는 규칙, 준말 처리에 관련되는 규칙 등 유형 분류가 명확하게 이루어지므로 분석 과정에 따라 적용되는 규칙의 유형을 제한할 필요가 있다. 한국어에서는 분석 결과의 수가 여러 개일 때 적용된 규칙의 가중치에 따라 분석 결과에 대한 가중치를 부여하거나 상대적으로 가능성이 매우 낮은 분석 결과를 무시할 수 있다면 형태소 분석의 질(quality)이 좋아지게 된다.

둘째, 단어 형성 규칙과 관련된 형태소의 분리 문제이다. Two-level 모델에서는 하나의 형태소와 그 활용(inflexion)이라는 관점에서 형태소는 Trie 구조 사전으로 기술되고 활용은 규칙으로 기술된다. 그런데 두 개 이상의 형태소가 하나의 단어를 구성할 때 각 형태소를 인식하고 형태소간의 결합 제약을 처리하기

위해서는 제어구조가 복잡해지므로 원래의 효율을 기대하기가 어려워진다.

불규칙 현상이 발생하는 각 경우에 대하여 Two-level 규칙이 기술된다. 또한 선어말어미 '-았/었-'과 결합되는 모든 어미에 대해서도 각각 규칙을 기술하여야 하므로 규칙의 수가 많아지게 된다. 예를 들어, 'ㅂ'-불규칙 현상이 발생하는 '아/어'로 시작하는 모든 어미에 대해 다음과 같은 Two-level 규칙이 기술되어야 한다.

셋째, 띄어쓰기 현상과 관련된 문제점이다. '해보다'와 같이 보조용언은 붙여쓰기가 허용되기 때문에 이러한 유형에 해당되는 모든 보조용언은 ['\_ ||보다\$ --> \_ | +어+보+다\$']와 같이 어미가 결합된 형태로 규칙을 기술하여야 한다. 이러한 경우에는 '해보-'를 어휘형태소 사전에 수록하는 방법으로 처리할 수도 있다. 그러나 각 형태소의 원형을 분석하지 못하게 되고, 접미사(명사 + '하-') 유형을 처리하기 어려우며, '같이'로부터 '같이+어+보+다'를 분석하지 못하는 등 여러 가지 파생되는 문제가 발생한다. 또한 맞춤법 규정에는 어긋나지만 '-고싶다, -게되다, -ㄴ 것-'과 같이 일반적으로 붙여 쓰기 쉬운 단어를 분석해야 할 필요가 발생한 경우도 있다.

넷째, 복합 명사와 미등락어의 처리 문제이다. 한국어는 복합 명사의 붙여쓰기가 허용되는 관계로 일반적으로 붙여쓴 복합명사는 어휘형태소 사전에 수록되지 않는다. 따라서 붙여쓴 복합 명사나 사전 미등락어는 추정에 의하여 처리해야 하는데 왼쪽에서 오른쪽으로 형태소 분석을 행하는 Two-level 모델에서는 문법형태소의 분리가 선행되지 않은 상황에서 미등락어를 추정하기가 어렵다. 물론 Two-level 모델을 오른쪽에서 왼쪽으로 처리하도록 수정할 수 있지만 그렇게 할 경우에는 trie 구조로 된 사전의 장점이라 할 수 있는 사전 탐색이 저하된다.

다섯째, 한국어에서는 형태소 분석할 때 분석 가능한 후보의 수가 많다는 점이다. Two-level 모델은 문자열 일치로 기술되는 Two-level 규칙의 조건에 만족하는 모든 분석 후보들을 backtracking에 의해 어휘형태소 사전과 비교하면서 처리하기 때문에 한국어의 경우 '아/어'가 생략된 어미를 복원하거나 '으' 탈락이나 'ㄹ' 탈락 단어의 원형 복원이 요구되는 현상들을 처리하기 위하여 backtracking이 매우 빈번하게 발생하게 된다.

.이와 같이 Two-level 모델을 한국어에 적용하면 한국어의 형태론적 특성과 관련된 여러 가지 문제들을 Two-level 모델이라는 틀에 맞춰야 하므로 Two-level 모델의 가장 큰 장점이라고 할 수 있는 효율성을 기대하기가 어렵다. 따라서 효율적인 한국어 형태소 분석 모델은 단어 형성 규칙에 따른 형태소

분리와 형태소간의 결합 제약, 띄어쓰기 등 여러 가지 특성들을 고려해서 적합한 알고리즘과 제어구조를 갖추어야 한다.

### 3. 다층 형태론(multi-level morphology)

Two-level 모델의 특징은 Two-level 형태론에 따라 "표층형 + 규칙<sub>i</sub> = 어휘형"으로 입력스트링(표층형)이 규칙<sub>i</sub>(Two-level 규칙)에 의해 어휘형으로 직접 mapping이 된다는 것과 규칙<sub>i</sub>의 적용 여부가 단지 문자열 일치에 의하여 기술된다는 점이다. 이렇나 특징은 단어의 활용 자체를 처리하기에 유리하지만 형태소 분석이 여러 단계를 거쳐서 이루어지는 경우에는 규칙을 어떻게 기술할 것인지의 문제가 발생한다. 예를 들어, ㅂ-불규칙 현상을 Two-level 규칙으로 기술한 예는 다음과 같다.

-운\$ --> -ㅂ + ㄴ \$  
 -우니\$ --> -ㅂ + 으니\$  
 -워서\$ --> -ㅂ + 어서\$  
 -워서는\$ --> -ㅂ + 어서는\$  
 -워지고\$ --> -ㅂ + 어+지+고\$  
 -워져서\$ --> -ㅂ + 어+지+어서\$  
 -워졌다\$ --> -ㅂ + 어+지+었+다\$

.....

위 규칙은 ㅂ-불규칙 현상이 일어나는 모든 어미와 선어말어미에 대하여 기술한 것으로 규칙의 수가 매우 많아지는데 이를 통합하여 기술할 수도 있다.

-우- --> -ㅂ -  
 -워- --> -ㅂ + 어-

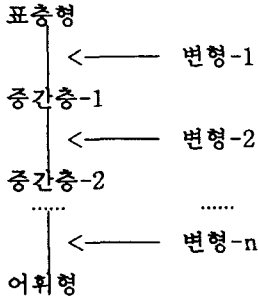
규칙-1'으로 통합했을 때, 즉 어미의 일부만 규칙에 기술된 경우에는 규칙-1'이 적용된 후에 어미 형태소 사전에서 확인하는 과정을 거쳐야 하고, '-워졌다'의 경우에는 계속해서 '-졌다'에 대한 분석을 해야 한다. 이는

"표층형 + 규칙<sub>i</sub> = 어휘형"

이라는 Two-level 형태론의 틀을 벗어나는 것으로 ㅂ-불규칙에서와 같이 형태소 분리와 형태론적 변형이 복합된 경우에는 Two-level 형태론이 적합하지 않다.

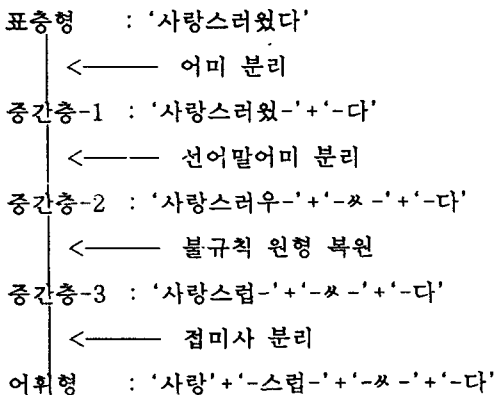
다층형태론은 Two-level 형태론에서 표층형으로부터 바로 어휘층을 이끌어내는 과정을 수정하여 중간 계층으로 문법 형태소가 분리되거나 원형을 복원한 형태를 저장하기 위하여 다수의 중간층(intermediate level)을 두는 것이다(그림 1). 다층 형태론은 변형 생성 문법에서 중간 계층을 두고 변형

규칙에 따라 표층형으로부터 어휘형을 생성하는 과정과 Two-level 형태론에서 형태론적 변형을 두 개의 계층으로 제한하여 처리하는 과정을 복합하여 형태소 분석에 적합하게 재구성한 것이다.



(그림 1) 다층 형태론

변형 생성 문법에서는 개별화될 수 있는 언어 현상을 최소 단위의 변형 규칙으로 기술함으로써 변형 규칙들 사이에 적용 순서를 결정하기 어려운 문제가 있다. 그런데 다층 형태론에서는 변형 과정을 반드시 변형 규칙으로 기술할 필요가 없다. 즉, 사전을 이용하여 형태소를 분리할 수도 있고, 형태소 분리와 변형을 하나의 단계로 기술하는 알고리즘에 의한 변형을 허용하기도 하며, 부적합한 변형이 일어나지 않도록 변형 금지 제약을 가할 수도 있다. 변형 과정을 기술하는 제약을 없애는 대신에 중간 계층 사이의 변형은 단어 형성 규칙에 따라 일정한 순서를 갖고 각 다층마다 필요한 규칙을 적용하거나 형태소 분리를 행한다.

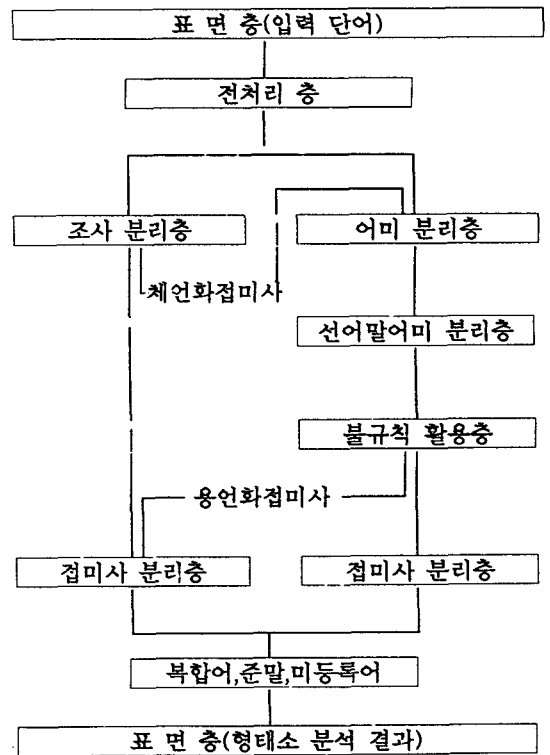


(그림 2) '사랑스러웠다'의 변형 과정

형태론에 따라 형태소의 원형을 복원한다. 즉, 다층 형태론은 Two-level 형태론을 강제로 적용할 때 발생하는 비효율성을 방지하기 위하여 변형 생성 문법에서 취하는 중간층을 도입한 것이다. 다층 형태론에 따라 '사랑스러웠다'를 분석하는 예는 (그림 2)와 같다.

#### 4. 한국어 형태소 분석

한국어 형태소 분석은 어휘 형성 규칙에 따라 문법형태소 분리와 불규칙 활용, 접미사 분리 등 분석 과정이 여러 단계로 명확하게 구분된다. 각 분석 단계에서 적용되는 분석 규칙(계층-i로부터 계층-j로의 전이 과정)은 어휘 형성 규칙에 따라 정유한자등(DFA)으로 구성된다. (그림 3)은 한국어 형태소 분석을 다단계 분석 모델에 따라 각 계층간의 분석 과정을 그림으로 나타낸 것이다.



(그림 3) 한국어 형태소 분석 단계

변형 과정에서 형태론적 변형(morphological alternation)이 일어나는 단계는 Two-level

각 계층간의 전이 과정에 대하여 간략하게 설명하면 다음과 같다.

#### ① 전처리

형태소 분석의 전처리 단계는 입력 문서에 나타난 단어들을 분석하기 위한 준비 과정이다. 전처리에서 필요한 기능은 먼저 입력 문자열을 분리하여 형태소 분석 단위의 단어를 추출하는 일이다. 즉, 입력 문자열로부터 문장 부호, 숫자, 영문자와 같이 형태소 분석 대상이 되지 않는 문자들을 분리하거나 제거한다. 단어를 추출할 때 문장부호는 별개의 단어로 독립시키고 불필요한 문자(제어문자, 특수문자, Tag 등)는 공백으로 대체한다. 입력 문자열에서 단어를 인식한 후에는 한글 코드 변환이 이루어진다. 형태소 분석기가 처리할 수 있는 코드는 자음과 모음을 분리할 수 있는 조합형 코드이므로 입력 단어의 코드를 형태소 분석기의 내부 코드로 변환한다. 현 형태소분석기는 내부 코드로 상용 조합형 한글 코드(2-바이트 KSSM 조합형)를 사용하므로 입력문서가 다른 코드로 작성된 경우에는 KSSM 코드로 변환하는 모듈이 추가되어야 한다.

#### ② 문법형태소 분리

형태소 분리 단계에서는 단어를 구성하고 있는 각 형태소들을 분리한다. 한국어의 단어 형성 규칙(word-formation rule)에 따라 문법형태소인 조사와 어미를 분리하고 어미가 분리된 경우에는 선어말어미를 분리한다. 이 때 어휘형태소와 문법형태소간의 결합관계를 검사하여야 한다. 어미가 분리되어 용언으로 추정되는 단어에 대하여 불규칙 활용이 일어났을 가능성이 있는 것은 불규칙 원형복원 루틴에 의해 어휘형태소의 원형을 추정한다.

#### ③ 체언 분석

체언 분석은 조사가 분리된 분석후보에 대하여 체언 분석을 행하는 과정이다. 형태소 분리 결과 조사가 분리된 단어는 체언으로 추정되므로 어휘형태소 사전에서 이를 확인하여 체언 분석을 시도한다. 어근부가 체언이 아닌 경우에는 접미사를 분리하고 다시 체언 분석을 시도하며, 명사형 전성어미가 발견된 단어는 용언 분석 단계에서 처리한다.

#### ④ 용언 분석

용언 분석은 어미가 분리된 분석후보에 대하여 체언 분석을 행하는 과정이다. 형태소 분리 결과 어미가 분리된 단어는 용언으로 추정되므로 어휘형태소 사전에서 이를 확인하여 용언 분석을 시도한다. 어근부가 용언이 아닌 경우에는 용언화 접미사를 분리하여 '-하다/되다/시키다'와 같은 접미사가 발견되면 체언 분석을 시도한다. 서술격 조사 '이'가 발견되거나 서술격 조사가 생략된 경우에도 체언 분석 과정에서 처리한다. 본 용언과 보조 용언이 결합된 유형은 보조 용언을 먼저 분석하고 난 후에 본 용언을 분석한다.

#### ⑤ 단일형태소 분석

형태소 분리 단계에서 조사나 어미가 발견되지 않은 단어는 단일형태소로 이루어진 단어라고 추정할 수 있다. 이러한 유형의 단어는 어휘형태소를 사전에서 체언이나 부사어, 감탄사, 관형사인지 확인하고 접미사가 분리되는 경우에는 접미사를 분리한 후에 어휘형태소 사전을 확인한다.

#### ⑥ 복합어 추정

조사가 분리되어 체언으로 추정된 어근부가 체언으로 분석되지 않은 단어는 어근부가 복합명사로 인식될 수 있는지 확인하여야 한다. 복합명사는 4음절, 5음절, 6음절, 7음절 어근부에 대해서만 명사들의 조합이 가능한지를 검사하여 명사와 명사의 조합으로 구성된 것은 복합명사로 분석한다.

#### ⑦ 준말처리

준말은 단어의 일부가 축약되거나 생략되어 일반적인 단어형성규칙에 의해 형태소를 분리할 수가 없다. 따라서 준말의 유형에 따라 준말 유형이 적용 범위가 넓은 것은 본디말 규칙으로 처리하고 적용의 범위가 좁은 것은 분석 결과를 기본적 사전에 수록하여 처리한다.

#### ⑧ 미등록어 추정

위 단계까지 분석되지 않은 단어들은 오류어거나 사전에 수록되지 않은 형태소를 포함하고 있는 단어이다. 고유명사나 전문용어, 신조어 등과 같이 사전에 수록되지 않은 형태소들은 형태소 분리 과정에서 얻은 결과와 단어형성 규칙을 바탕으로 하여 미등록어를 추정한다.

#### ⑨ 후처리

후처리 단계는 형태소 분석기의 응용 분야에 맞게 출력 양식이 다를 수 있으므로 형태소 분석결과를 사용할 수 있도록 분석결과를 재구성하여 출력하는 과정이다. 또한 영문자나 숫자 등 한글이외의 문자열에 대하여 분석이 필요하면 이 모듈을 연결하여 결과를 출력한다. 그 이외에도 모호성 해결(ambiguity resolution)과 같이 여러 가지 분석결과가 생성된 경우에 옳지 않은 것을 제거하거나 옳은 결과를 선택하는 모듈이나 분석실패한 단어를 처리하는 모듈이 완성되면 추가할 수 있도록 한다.

### 5. 결론

한국어 형태소 분석의 근본적인 문제는 형태소 분리와 관련되어 나타난다. 형태소 분석을 하기 위해서 분리되어야 할 단위형태소의 수가 많을 뿐 아니라 형태소 분리와 형태론적 변형 현상이 복합됨으로써 형태소 분리 과정에서 backtracking이 빈번하게 일어난다. 또한 복합명사와 보조용언의 붙여쓰기가 허용되는 등 여러 가지 단어 유형을 처리해야 하는 부담이 있다. 따라서 한국어 형태소 분석시에 미분석 오류와 과분석 오류, 그리고 분석 알고리즘의 비효율성 문제가 발생한다.

본 논문에서는 Two-level 모델을 한국어에 적용할 때 발생하는 문제점을 제시하고 이를 극복하기 위한 방법으로 다층 형태론과 다단계 모델을 제안하였다. 다층 형태론은 단어의 표층형으로부터 어휘형을 생성하기 위하여 언어의 특성에 따라 중간 계층을 정의하고 계층간의 변형을 형태소 분리 또는 형태론적 변형 규칙으로 처리한다. 즉, Two-level 모델에서 형태소 분리로 인하여 발생하는 backtracking 문제를 해결하기 위하여 중간 계층을 도입한 것이다.

다단계 모델은 다층 형태론을 기반으로 형태소 분석기를 구현하기 위한 방법론이다. 이 모델에서는 형태소 분리, 형태론적 변형 등을 처리할 때 불필요한 분석 후보의 수를 최소화하기 위해 언어의 특성과 단어 유형에 따라 중간 계층을 두고 계층과 계층간의 변형을 규칙과 사전에 의하여 처리한다. 이 모델에 따라 형태소 분석기를 구현할 때 형태론적 변형을 처리하는 계층간의 변형은 Two-level 모델을 적용하는 방법도 가능하며, 문제의 특성에 따라 규칙의 적용이 제한되는 경우에 계층간의 변형 과정에서 유용한 정보를 사용할 수 있으므로 매우 효율적이다.

#### 참 고 문 헌

[강승식93] 강승식, 음절 정보와 복수어 단위 정보를 이용한 한국어 형태소 분석, 서울대학교 컴퓨터공학과 박사학위 논문, 1993년 2월.  
 [김덕봉90] 김덕봉, 최기선, 강재우, "한국어 형태소 처리와 사전 - 접속정보를 이용한 한글 철자 및 띄어쓰기 검사기 -", 어학연구, 26권, 1호, pp.87-113, 1990.  
 [김병희93] 김병희, 임근목, 송만석, "형태소 접속 특성과 인접 말마디 정보를 이용한 형태소 분석기", 제5회 한글 및 한국어 정보처리 학술발표 논문집, pp.395-404, 1993.  
 [김성용87] 김성용, 최기선, 김길창, "Tabular Parsing 방법과 접속 정보를 이용한 한국어 형태소 분석기", 한국정보과학회 인공지능연구회 춘계 인공지능 학술발표회: 논문집, pp.133-147, 1987.  
 [김재한93] 김재한, 안미정, 옥철영, "활용 형태소에 기반한 한국어 형태소 분석기", 한국정보과학회 가을 학술발표 논문집, pp.1139-1142, 1993.  
 [이은철92] 이은철, 이종혁, "계층적 기호 접속정보를 이용한 한국어 형태소 분석기의 구현", 제4회 한글 및 한국어 정보처리 학술발표 논문집, pp.95-104, 1992.  
 [최형석84] 최형석, 이주근, "자연어 어절 처리 알고리즘", 한국정보과학회 추계 학술발표회 논문집, 11권 2호, 1984.

[Bart86] E. Barton, "Computational Complexity in Two-Level Morphology," in 24th Annual Meeting of the Association for Computational Linguistics, 1986.  
 [Bear88] J. Bear, "Morphology and Two-level Rules and Negative Rule Features," Proceedings of the 12th International Conference on Computational Linguistics, pp.28-31, 1988.  
 [Cahi90] L.J. Cahill, "Syllable-based Morphology," Proceedings of the 13th International Conference on Computational Linguistics, vol.3, pp.48-53, 1990.  
 [Kang94] S.S. Kang, Y.T. Kim, "Syllable-based model for the Korean Morphology," Proceedings of the 15-th International Conference on Computational Linguistics (COLING-94), Vol.1, pp.221-226, 1994.  
 [Kart92] L. karttunen, R. M. Kaplan, and A. Zaenen "Two-Level Morphology with Composition," Proceedings of the 15th Conference on Computational Linguistics, pp.141-148, 1992.  
 [Kim94] D.B. Kim, S.J. Lee, K.S. Choi, G.C. Kim, "A Two-level Morphological Analysis of Korean," Proceedings of the 15-th International Conference on Computational Linguistics (COLING-94), Vol.1, pp.535-539, 1994.  
 [Kosk83] K. Koskenniemi, "Two-level Model for Morphological Analysis," Proc. of the 8th International Joint Conference on Artificial Intelligence, pp.683-685, 1983.  
 [Kosk88] K. Koskenniemi, "Complexity, Two-Level Morphology and Finnish," Proceedings of the 12th Conference on Computational Linguistics, pp.335-339, 1988.  
 [Kwon91] H.C. Kwon, Y.S. Chae, and G.O. Jeong, "A Dictionary-based Morphological Analysis," Proceedings of Natural Language Processing Pacific Rim Symposium, pp.87-91, 1991.  
 [Kwon94] H.C. Kwon, L. Karttunen, "Incremental Construction of a Lexical Transducer for Korean," Proceedings of the 15-th International Conference on Computational Linguistics (COLING-94), Vol.2, pp.1262-1266, 1994.  
 [Zhang90] B.T. Zhang and Y.T. Kim, "Morphological Analysis and Synthesis by Automated Discovery and Acquisition of Linguistic Rules," Proceedings of the 13th International Conference on Computational Linguistics, pp.431-436, 1990.