음절 특성을 이용한 한국어 불규칙 활용 어절의 형태소 분석 방법

강 승 식
컴퓨터 신기술 공동 연구소
Tel: 02-880-7302, E-mail: nlpkang@krsnucc1.bitnet

Analysis of Korean Irregular Verbs Using Syllable Characteristics

Kang, Seung-Shik
Research Institute of Advanced Computer Technology

요 약

한국어 형태소 분석 시스템은 형태소를 분리하거나 원형을 복원하는 후보 생성 과정에서 많은 후보를 생성하고 이에 대한 사전의 검색이 요구되는 부담이 있다. 특히, 불규칙 활용 어절을 분석하려면 불규칙 활용 어절뿐만 아니라 체언 어절이나 불규칙 활용이 일어나지 않은 모든 어절에 대해서도 불규칙 어절일 가능성을 검사하고, 원형을 복원하기 위해 원형의 후보들을 역으로 추정한 후에, 각 후보에 대해 사전을 검색하는 과정을 거치게 된다. 이 때 불규칙 활용 가능성으로 인한 후보들의 과다한 생성은 사전 검색 횟수의 증가를 유발하여 시스템의 성능을 저하시키는 요인이 되어 왔다. 본 논문에서는 한글의 음절 특성을 이용하여 불규칙 활용이 일어난 후보 어절의 수를 줄임으로써 사전의 검색 횟수를 적게 하고 형태소 분석 시스템의 성능을 향상시키는 방법을 제안한다.

I. 서 론

형태소 분석(morphological analysis)은 기계 번역(machine translation)이나 자연어 이해 시스템(natural language understanding system)을 구현하기 위하여 구문 분석(syntactic analysis)과 의미 분석(semantic analysis)의 전 단계로 시작되었다. 형태소 분석의 문제는 자연어 처리에서 가장 기본적으로 해결되어야만 하는 문제이기 때문에 대부분의 자연어 처리 시스템에 필수적으로 요구될 뿐만 아니라 그 자체만으로도 여러 가지 응용 분야에 이용될

수가 있다[Jens86, Nagao86]. 특히, 현대 사회에서는 컴퓨터를 사용하여 많은 양의 자료를 관리하고 이를 유익하게 이용하고자 하는 정보 검색(information retrieval)에 대한 수요가 증가하고 있으며, 자연어 인터페이스와 맞춤법 검사 등에 대한 요구가 늘어남에 따라 형태소 분석의 중요성이 더욱 강조되고 있다. 영어의 경우에는 이미 형태소 분석 문제가 해결되어 이를 기반으로 하는 기계 번역, 맞춤법 검사, 정보 검색 시스템 등에 유용하게 사용되고 있다[Heid82, Macd82].

한국어에 대해서도 많은 연구가 진행되어 한국어의 특성에 맞는 여러 가지 방법론이 제시되어 왔다[김성용87, 김덕봉90, 박종만90, 이은철92, Zhan90]. 최근에는 형태소 분석 성공률 99% 이상의 실용화 수준에 이르렀으나, 알고리즘의 효율이라든지 복합 명사 및 사전 미등록어 처리, 모호성 제거 문제, 단어의 사용 빈도에 따른 사전의 구성 등 부분적으로 개선의 여지가 남아 있다[강승식93]. 형태소 분석 알고리즘의 효율은 입력 단어(어절)로부터 분석 후보를 생성하는데 필요한 비교 연산(comparison)의 수와 사전 검색 횟수에 의하여 결정된다. 비교 연산은 형태소의 분리와 불규칙 어절의 원형 복원 과정에서 주로 발생하고 사전 검색 횟수는 분석 후보의 수에 비례한다. 문법 형태소의 분리에 관해서는 음절 특성을 이용한 효율적인 알고리즘이 제시되었다[Kang92].

용언의 불규칙 현상 처리 방법에 대해서는 현재 형태론적 변형 규칙에 의한 것과 사전으로 처리하는 두 가지 방법이 사용되고 있다. Zhang(90)은 기계 학습에 의한 Two-level 규칙의 습득에 의한 처리 방법을 제안하였고, 강승식(92)은 한국어의 형태론적 변형 규칙은 그 수가 많지 않으므로 규칙 기반 시스템 대신 프로시쥬어에 의한 형태론적 변형 규칙을 기술하고 있다. 형태론적 변형 규칙에 의하여 불규칙 어절을 분석하는 경우에는 어떤 방법론에 따르더라도 불규칙 어절로부터 원형을 복원하는 과정에서 과생성(over generation) 문제가 발생한다. 접속 정보를 이용하는 형태소 분석에서는 형태 변이가 일어난 어간을 모두 사전에 기술하고 접속 정보에 의하여 어미와의 결합을 제약하는 방법으로 처리하기 때문에 원형 복원으로 인한 과생성 문제가 발생하지 않으나 사전을 기술하고 관리하기가 어렵다.

본 논문에서는 불규칙 어절을 형태론적 변형 규칙으로 처리할 경우에 형태소 분석 시스템의 성능 저하 요인으로 크게 작용하고 있는 분석 후보의 과다한 생성을 막기 위하여 한글의 음절 특성을 이용한 불규칙 어절의 분석 방법을 제안한다.

Ⅱ. 형태소 분석 개요

입력된 단어(한국어의 경우에는 어절)에 대한 형태소 분석은 분석의 중간 단계로 가능한 모든 분석 후보를 생성하는 후보 생성(candidate generation) 단계와 분석 후보들로부터 옳은 것을 선택하는 후보 선택(candidate selection) 단계로 이루어진다. 분석 후보의 생성은 여러 개의 형태소로 이루어진 단어(어절)로부터 각 형태소를 분리(morpheme identification)하는 과정과 활용이나 곡용, 축약, 탈락 현상 등으로 인하여 형태론적 변형(morphological alternation)이 일어난 형태소로부터 원형을 복원하는 과정에서 발생한다. 후보 선택에서는

분석 후보들에 대하여 사전 검색을 거쳐 옳은 후보를 선택하는 과정으로 기술된다(그림 1). 그림 1의 결합 제약 단계는 복합어(compound word)나 한국어에서 여러 개의 형태소로 이루어진 단어(어절)를 분석할 때 이웃하는 형태소 간의 결합 제약 조건을 검사하는 과정이다[Pach92].

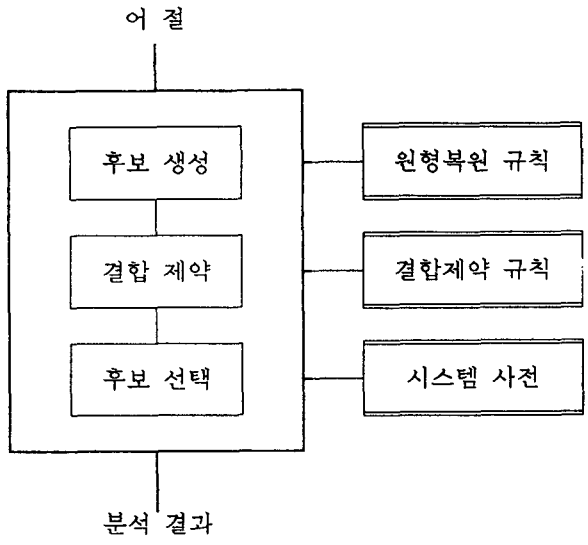

그림 1. 일반적인 형태소 분석 과정

형태소 분석 시스템의 구조는 언어의 특성에 따라 달라질 수 있다. 중국어의 경우에는 형태론적 변형이 일어나지 않는 대신 단어를 구분하는 문자(word delimeter)가 없다. 따라서 중국어의 형태소 분석은 문장을 입력 단위로 하여 단어를 분리(word segmentation) 과정으로 기술된다[Spro90, Chen92]. 그러나 일반적으로 형태론적 변형 현상을 처리하는 문제가 가장 크게 대두되었기 때문에 전산언어학에서 형태론은 형태론적 변형을 중심으로 발전되어 왔다[Kosk83, Russ86, Cahi90].

한국어의 형태소 분석은 형태소의 분리와 결합 제약, 그리고 인식이라는 문제가 복합되어 있어서 어떤 문제에 비중을 두었는가 하는 기준에 따라 표 1과 같이 분류된다. 표 1에서 같은 부류의 방법론들은 서로 비교 분석이 가능하나 다른 부류에 속하는 것들은 형태소 분석 문제를 접근하는 관점이나 알고리즘을 기술하는 초기 상태가 다르기 때문에 비교-분석의 의미가 없다.

최장일치법과 최단일치법, 그리고 Tabular 파싱법은 어절에서 분리될 수 있는 가능한 모든 형태소들이 생성되었다고 가정하고 이로부터 어절을 이루는 연속된 형태소들의 조합을 찾는 방법론이다. 이에 비해 Head-Tail 구분법과 음절 단위 분석법은 형태소를 어떻게 분리할 것인가 하는 문제를 중심으로 알고리즘을 기술하고 있고, Two-level 형태론과 음절

기반 형태론은 불규칙 어절의 원형을 복원하는 방법에 관한 것이다.

표 1. 한국어 형태소 분석 기법의 분류

분 류 기 준	한 국 어 형 태 소 분 석 기 법
형태소 분석 방향	좌-우 분석법, 우-좌 분석법 양방향 분석법, 역방향 분석법
형태소 인식 방법	최장일치법, 최단일치법, Tabular 파싱법
문법형태소 기술 방법	기본 단위 기술 방법, 통합형 기술 방법
형태소 분리 방법	Head-Tail 구분법, 음절 단위 분석법
형태소 결합조건 기술 방법	접속 정보표를 이용하는 방법, 결합 제약 규칙에 의한 방법
불규칙 어절의 원형 복원법	Two-level 형태론, 음절 기반 형태론, 기계 학습에 의한 방법
입력 단위에 따른 분류	어절 단위 분석법, 문장 단위 분석법

Ⅲ. 불규칙 어절의 형태소 분석

국어학에서 용언의 활용은 규칙 활용과 불규칙 활용, 자동적 교체 현상으로 구분하고 있다[남기심86]. 그러나 전산언어학의 관점에서는 불규칙 활용과 자동적 교체 현상이 모두 어간이나 어미의 형태론적 변형으로 나타나기 때문에 분석 후보 생성시에는 따로 구분할 필요가 없다. 따라서 본 논문에서 한국어의 불규칙 현상의 분류 기준은 강승식(92)에 정의된 축약이나 탈락을 포함한 넓은 의미의 분류 기준에 따른다. 다만 사전에 불규칙 정보가 기술되어야 하는 불규칙 용언에 대해서는 후보 선택 과정에서 사전에 기술된 불규칙 정보와 후보 생성시에 얻은 정보가 일치하는지를 검사하는 과정이 필요하다.

3.1 용언의 음절 특성

용언 어절을 분석할 때 미리 어간일 가능성이 없는 후보를 제거할 수 있다면 분석 후보의 수가 줄어들게 되고 사전을 검색하는 부담을 덜 수가 있다. 예를 들어, 'ㅂ' 불규칙의 원형을 복원할 때도 '싸웠다'를 '쌉다'+'-었-'+'-다'라는 후보를 생성하게 된다. 그런데 불규칙 '쌉'은 'ㅂ' 불규칙 용언의 끝음절로 사용될 수 없기 때문에 후보에서 제외할 수 있다. 마찬가지로 '살'을 '살다'+'-을'로 분리하는 후보 생성 과정은 '후보를'을 '후

보를다'+'-을'로, '분석할'을 '분석할다'+'-을'로 분리하는 후보를 생성하는 과정과 동일하다. 이와 같은 과생성은 'ㄹ' 탈락 어절의 원형을 복원하기 위하여 'ㄹ'로 끝나는 모든 어절을 'ㄹ' 탈락 후보로 생성하는 잘못으로 인하여 발생한다. 만일 용언의 끝음절이 '를'이나 '할'인 것이 존재하지 않는다는 사실을 알면 이러한 과생성을 막을 수 있다. 즉, 용언의 끝음절로 사용되는 음절 특성 집합이 구성되면 쉽게 후보를 걸러낼 수가 있다([강승식93]의 부록 참조).

불규칙 용언의 경우에도 'ㄷ' 불규칙 용언은 37개인데 종성이 'ㄷ'인 한글 음절 399(=19×21)개 중에서 단지 10개만 'ㄷ' 불규칙의 끝음절로 사용되고, 'ㅂ' 불규칙은 446단어에서 46개의 음절만이 끝음절로 사용된다([강승식93]의 4장 참조). 특히, '우' 불규칙 용언은 '푸다'라는 동사 하나밖에 없기 때문에 원형을 복원할 때 '-ㅓ', '-ㅓ서', '-ㅆ다', '-ㅆ고'로 끝나는 모든 어절을 '우' 불규칙으로 추정할 때 '퍼', '퍼서', '펐다', '펐고'와 같이 어간이 한 음절이고 복원된 어간의 원형이 '푸'인 것으로 제한하면 후보 어절의 과생성을 막을 수 있다.

불규칙 유형 T에 속하는 용언의 끝음절에 대한 음절 특성 집합을 V_T라 하고 V_T에 대한 음절 특성 함수를 C_T라 하면 C_T는 다음과 같이 기술된다.

$$C_T(x) = \begin{cases} 1, & \text{if } x \in V_T \\ 0, & \text{otherwise} \end{cases}$$

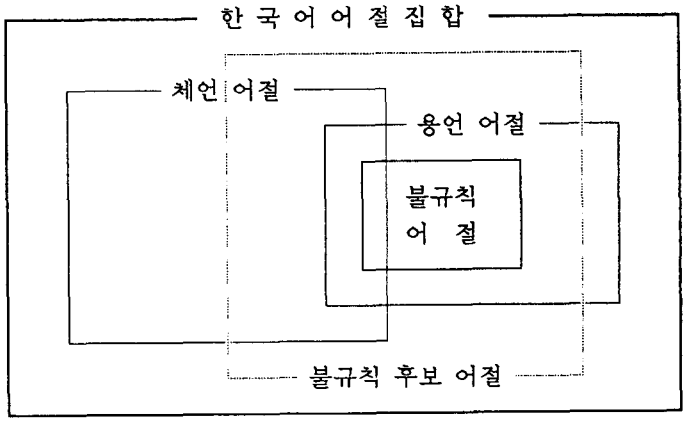

그림 2. 불규칙 추정 어절의 포함 관계

3.2 불규칙 어절의 후보 생성

일반적으로 입력된 단어(어절)가 불규칙 어절인지 아닌지는 형태소 분석을 한 후에만 알 수가 있기 때문에 불규칙 어절에 대한 분석 후보 어절을 생성하는 대상은 모든 단어(어

절)이다. 즉, 분석하고자 하는 한국어 어절의 집합(원소의 중복이 허용되는 집합)을 W라 할 때 불규칙 후보 어절의 대상이 되는 어절의 수는 n(W)이다. 그런데 'ㄴ'이나 'ㄹ'로 끝나는 모든 어절에 대해 'ㅎ' 불규칙과 'ㄹ' 탈락 후보로 생성되는 것과 같이 불규칙 후보 어절 중에 실제로 불규칙 어절로 분석되지 않는 후보가 많이 발생하고, 생성된 후보 어절의 수만큼 사전 검색이 요구된다.

체언 어절 집합을 N, 용언 어절 집합을 V, 불규칙 어절 집합을 I, 그외의 어절 집합을 R이라 하면, 어절 집합간의 포함관계는 다음과 같다(그림 2).

$$R = W - (N \cup V), \quad V \supset I$$

그림 2에서 체언 어절과 용언 어절의 교집합은 한 어절이 체언 혹은 용언으로 분석 가능한 형태론적 모호성(어휘 모호성과 품사 모호성)이 일어나는 경우이다. 용언 어절에 대한 불규칙 어절의 비율을 α, 용언 어절과 용언이 아닌 어절의 비율을 β, 후보 생성시에 용언 어절에 대해서 불규칙 어절로 추정 가능한 확률을 γ, 용언 어절이 아닌 어절을 불규칙 어절로 추정 가능한 확률을 δ라 하면, 불규칙 어절로 추정하는 어절의 집합을 I_c라 할 때 $n(I_c)$는 다음과 같이 계산된다.

$$\alpha = n(V)/n(I), \quad \beta = n(W-V)/n(V)$$
$$n(I_c) = \gamma \cdot n(V) + \delta \cdot n(W-V) = \alpha\gamma \cdot n(I) + \alpha\beta\delta \cdot n(I)$$
$$= \alpha \cdot (\gamma + \beta\delta) \cdot n(I)$$

$\alpha, \beta, \gamma, \delta$의 정확한 값은 알 수 없으나 대략 α, β를 5~10, γ를 0.1~0.5, δ를 0.05~0.3라 가정할 때 $n(I_c)$은 $1.75n(I) \sim 35n(I)$, 즉, 불규칙 어절을 분석하기 위하여 실제 불규칙 어절의 수보다 적게는 2배에서 많게는 30여배나 되는 후보들을 생성하게 된다. 그런데 위에서 $n(I_c)$를 계산할 때 한 어절에 대해 두 개 이상의 불규칙 후보를 생성하는 경우에 1로 계산했기 때문에 $n(I_c)$의 값은 더 크게 된다. 따라서 불규칙 후보 어절에 대하여 사전을 검색하는 횟수는 위에서 계산한 $n(I_c)$의 값보다 더 크다.

3.3 불규칙 후보 어절의 여과

불규칙 용언을 분석할 때 발생하는 과생성의 부담을 줄이기 위해 불규칙 용언의 끝음절 특성을 이용하여 아래와 같은 가정이 성립한다고 할 때 불규칙 후보 어절일 가능성이 없는 후보들을 미리 배제할 수 있다.

> 가 정
> 형태소 M의 끝음절이 불규칙 유형 T에 속하는 용언의 끝음절로 사용되지 않으면, M은 T-불규칙 용언으로 분석되지 않는다.

음절의 연속으로 이루어진 어절 $S_1S_2...S_n$에 대해 $S_1S_2...S_i$이 어간, $S_{i+1}S_{i+2}...S_n$이 어미로 분리되고, 불규칙 추정에 의하여 음절 S_i를 S_i'로 어간의 원형이 복원되었다고 하면, 임의의 불규칙 유형 T로 추정된 어간 후보의 끝음절 S_i'가 T-불규칙 용언의 끝음절로 사용될 확률을 σ라 할 때, 불규칙 용언의 끝음절에 대한 음절 특성 함수를 이용하여 불규칙 후보 어절의 수가 $n(I_c)$에서 $\sigma \cdot n(I_c)$로 줄게 된다.

불규칙 용언의 끝음절 특성을 이용하여 한국어의 불규칙 어절을 분석하는 과정은 그림 3과 같다. 형태소 분리와 원형 복원은 기존의 방법과 같으나 네 번째 단계인 용언의 끝음절 특성을 이용한 불규칙 후보 어절 여과 부분이 추가되어 불규칙 어절일 가능성이 없는 후보를 제외한 것이다.

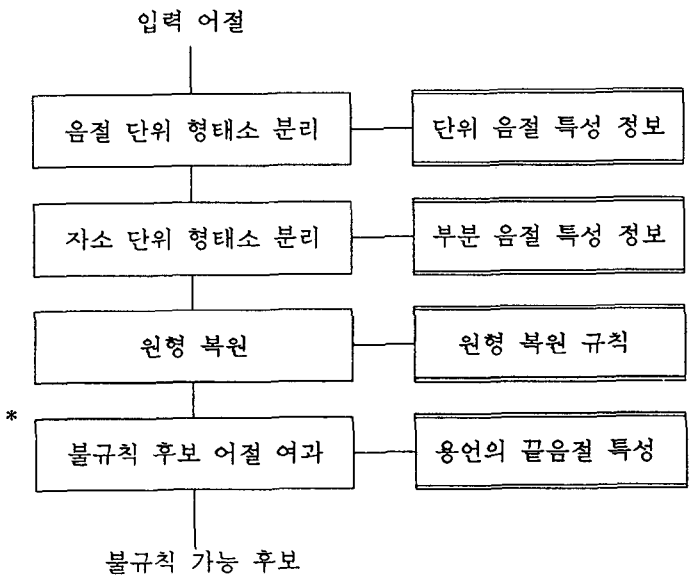

그림 3. 불규칙 어절의 분석 후보 생성 과정

용언의 음절 특성을 이용한 불규칙 후보 어절의 여과 과정을 알고리즘으로 기술하면 그림 4와 같다.

```
algorithm irr_word_filtering(cstem, T)
syllable cstem[]; /* stem candidate */
int T;     /* type of irregular verb */
begin
    n = index of the last syllable of stem candidate;
    if (CT(cstem[n]) == 1) return OK;
    else return FAIL;
end
```

그림 4. 불규칙 후보 어절의 여과 알고리즘

3.4 어간과 어미의 결합 제약

어간과 어미 사이의 결합 제약 검사는 형태소를 분리할 때와 불규칙 어절의 원형 복원, 그리고 어간을 사전에서 검색한 후에 필요하다. 형태소 분리시에는 모음조화 현상을 검사하고 어미의 유형에 관한 정보를 표시한다. 어미의 유형에는 품사(동사, 형용사)를 제약하는 것과 서술격 조사 '-이'에만 결합하는 것, 형태 변이가 일어난 것과 그렇지 않은 것 등이 있다. 불규칙 어간의 원형을 복원할 때는 불규칙 현상이 일어나기 위해 요구되는 어미에 대한 제약이 있으며, 사전 검색 후에는 어미에 의한 품사 제약, 어간과 어미의 불규칙 유형 제약을 검사한다. 불규칙 어간과 어미의 결합에서는 결합 가능 조건과 결합 불가 조건을 모두 검사해야 한다.

불규칙 어절의 추정 대상이 되는 어미는 형태 변이가 일어난 모든 어미와 그렇지 않은 어미의 일부이다. 어간은 규칙 어간과 형태 변이가 일어난 어간, 그리고 형태 변이가 일어나지 않은 불규칙 용언의 어간으로 분류된다. 규칙 어간은 어미의 원형하고만 결합하고, 불규칙 어간은 일부 어미에 대해서 형태 변이를 요구하므로 형태 변이가 요구되는 어미의 원형과 불규칙 어간이 결합되지 않도록 하여야만 '낫으니'를 '낫다'+'-으니'로 분석하는 것과 같은 엉뚱한 결과를 막을 수 있다.

결합 제약 조건을 검사하기 위해서는 형태소를 분리할 때 어미의 형태론적 변이 여부와 불규칙 어절을 위해 분리된 어미에 대해서는 요구되는 불규칙 유형에 관한 정보를 전달하여 사전 검색 후에 어간과 어미의 결합 여부를 검사할 수 있도록 하여야 한다. 불규칙 어절의 원형 복원 과정에서는 사전에 불규칙 정보가 저장되어야 하는, 형태 변이가 일어난 어간에 대해서만 불규칙 유형을 기술한다. 어간과 어미에 대한 형태론적 정보를 기반으로 하여 어간과 어미의 결합 조건 검사 알고리즘을 기술하면 그림 5와 같다. 이외에도 불구동사와 '있다', '없다' 처럼 어미 활용에 제약이 있는 단어에 대한 결합 조건을 검사해야 한다.

```
algorithm coocur(cstem, ceomi)
syllable cstem[], ceomi[]; /* stem and eomi candidate */
begin
    if (!vowel_harmony(cstem, ceomi) or
        !irr_restriction(cstem, ceomi) or
        !pos_restriction(cstem, ceomi)) return FAIL;
    else return OK;
end
```

그림 5. 어간과 어미의 결합 조건 검사 알고리즘

Ⅳ. 결 론

형태론적 변형 현상은 여러 가지 방법으로 처리할 수 있지만 모두 분석 후보의 과다한 생성이 효율을 저하시키는 요인으로 작용되었다. 한국어 형태소 분석시에 발생하는 분석 후보의 과다한 생성을 막고 시스템의 성능을 향상시키기 위하여 음절 특성을 이용함으로써 사전을 검색하기 전에 분석 후보의 수를 줄이는 불규칙 어절의 분석 방법을 제안하였다. 한국어의 경우에는 한글의 우수한 특징이라 할 수 있는 음절 단위 표기법과 표음 문자라는 음절 특성을 이용하여 분석 후보의 수를 줄임으로써 사전 검색의 부담을 감소시키고 보다 효율적인 형태소 분석 시스템을 개발할 수 있었다.

참 고 문 헌

[강승식92] 강승식, 김영택, "한국어 형태소 분석기에서 불규칙 용언의 분석 모형", 한국정보과학회 논문지, 19권, 2호, pp.151-164, 1992.

[강승식93] 강승식, 음절 정보와 복수어 단위 정보를 이용한 한국어 형태소 분석, 서울대학교 컴퓨터공학과 박사학위 논문, 1993년 2월.

[김성용87] 김성용, 최기선, 김길창, "Tabular Parsing 방법과 접속 정보를 이용한 한국어 형태소 분석기", 한국정보과학회 인공지능연구회 춘계 인공지능 학술발표회 논문집, pp.133-147, 1987.

[김덕봉90] 김덕봉, 최기선, 강재우, "한국어 형태소 처리와 사전 - 접속정보를 이용한 한글 철자 및 띄어쓰기 검사기 -", 어학연구, 26권, 1호, pp.87-113, 1990.

[남기심86] 남기심, 고영근, 표준 국어 문법론, 탑출판사, 1986.

[박종만90] 박종만, 효율적인 한국어 형태소 분석기 및 철자 검사 교정기의 구현, 서울대학교 공학석사 학위논문, 1990.

[이은철92] 이은철, 이종혁, "계층적 기호 접속정보를 이용한 한국어 형태소 분석기의 구현", 제4회 한글 및 한국어 정보처리 학술발표 논문집, pp.95-104, 1992.

[Cahi90] L.J. Cahill, "Syllable-based Morphology," Proceedings of the 13th Internatioanl Conference on Computational Linguistics, Vol.3, pp.48-53, 1990.

[Chen92] K.J. Chen and S.H. Liu, "Word Identification for Mandarin Chinese Sentences," Proceedings of the 14th Internatioanl Conference on Computational Linguistics, Vol.1, pp.101-107, 1992.

[Heid82] G.E. Heidorn, K. Jensen, L.A. Miller, R.J. Byrd, and M.S. Chodorow, "The EPISTLE Text-Critiquing System," IBM System Journal, Vol.21, No.3, pp.305-326, 1982.

[Jens86] K. Jensen, G. Heidorn, and S. Richardson, "PLNLP, PEG and CRITIQUE: Three Contributions to Computing in Humanities," Research Report RC-11841, IBM Thomas J. Watson Research Center, Yorktown Heights, New York, 1986.

[Kang92] S.S. Kang, "A Statistical Approach to Syllable-based Morphological Analysis," Proceedings of the International Conference on Computer Processing of Chinese and

Oriental Languages, 1992.

[Kosk83] K. Koskenniemi, "Two-level Model for Morphological Analysis," Proceedings of the 8th International Joint Conference on Artificial Intelligence, pp.683-685, 1983.

[Macd82] N.H. Macdonald, L.T. Frase, P. Gingrich, and S.A. Keenan, "The WRITER'S WORKBENCH: Computer aids for text analysis," IEEE Transactions on Communication, COMM-30, pp.105-110, 1982.

[Nagao86] M. Nagao, J.I. Tsujii and J.I. Nakamura, "Machine Translation from Japanese into English," Proceedings of IEEE, Vol.74, No.7, pp.993- 1012, 1986.

[Pach92] T. Pachunke, O. Mertineit, K. Wothke and R. Schmidt, "Broad Coverage Automatic Morphological Segmentation of German Words," Proceedings of the 14th Conference on Computational Linguistics, pp.1219-1222, 1992.

[Russ86] G.J. Russel, G.D. Richie, S.G. Pulman and A.W. Black, "A Dictionary and Morphological Analyzer for English," Proceedings of the 11th International Conference on Computational Linguistics, pp.277-279, 1986.

[Spro90] R. Sproat and C. Shih, "A Statistical Method for Finding Word Boundaries in Chinese Text," Proceedings of International Conference on Computer Processing of Chinese and Oriental Languages, Vol.4, No.4, 1990.

[Zhan90] 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, Vol.2, pp.431-436, 1990.