

# 한국어의 형태론적 모호성 유형 및 해결 방안

강 승 식  
한성대학교 정보전산학부  
Email: kang@ice.hansung.ac.kr

Classification and Disambiguation of Morphological Ambiguity of the Korean Language

Kang, Seung-Shik  
School of Information and Computer Engineering, Hansung University

## 요 약

한국어는 실질형태소와 형식형태소가 결합되는 교착어라는 특성 때문에 품사 모호성을 비롯한 여러 가지 유형의 형태론적 모호성이 발생한다. 형태론적 모호성 해결의 관점에서 형태론적 모호성을 한국어의 특성에 따라 어근 유형 모호성, 형태소 분리 모호성, 형태소 길이 모호성, 불규칙 용언의 원형 복원 모호성, '아/에/이' 탈락 모호성 등으로 분류한다. 이 때 임의의 두 분석 결과에서 발생하는 모호성이 특정 유형에만 속하도록 모호성 유형들을 서로 독립적으로 정의한다. 또한 품사 모호성을 계층적 품사 분류 체계에 따라 1~3차적 품사 모호성으로 구분하고 국어사전에서 발견되는 품사 모호성을 분석한다. 이를 기반으로 형태론적 모호성의 유형을 단어 내에서 해결 가능한 것과 그렇지 않은 것으로 구분하여, 단어 내에서 해결 가능한 모호성을 해결하는 방법을 제안한다.

## 1. 서 론

한국어 형태소 분석 결과는 교착어라는 특성 때문에 영어와 같은 굴절어의 분석 결과와 비교할 때 그 유형 및 모호성 유형이 매우 다양하다[1,2]. 따라서 영어의 형태론적 모호성이 품사 모호성으로 규정되는데 비해, 한국어의 경우는 품사 모호성뿐만 아니라 형태소 분리 위치, 단어의 유형, 형태소의 원형 복원, 탈락 및 축약 현상 등을 처리하는 과정에서 여러 가지 모호성 유형들이 발견된다[3,4].

지금까지 한국어의 형태론적 모호성 문제는 태깅된 말뭉치를 구축하기 위한 태깅 시스템을 구현할 필요성에 의해 연구되었고, 한국어 형태소 태깅 시스템은 대부분 통계적 기법을 기반으로 한 방법론을 사용하고 있다[5,6,7,8]. 따라서 형태론적 모호성이 일어나는 원인을 파악하고 유형을 분류하여 분석적 접근 방법에 대한 연구는 미흡하였다. 최근에는 통계적 기법의 제약을 극복하기 위해 규칙을 혼합하는 방법론을 도입하고 있기 때문에 모호성의 원인을 명확하게 규명하고 이를 체계적으로 분류할 필요가 있다[9,10,20].

본 논문에서는 한국어에서 형태론적 모호성이 발생하는 원인에 따라 모호성 유형들을 세분화하고 모호성 해결 방법을 제안한다. 모호성 유형은 형태소 수준에서 해결이 가능한 것을 중심으로 한국어의 단어 구성 및 품사 분류 특성, 조음소의 삽입이나 탈락 현상 등 형태소 분석 결과 및 분석 과정에서 발견되는 특성들을 기준으로 분류한다.

## 2. 형태론적 모호성의 유형 분류

임희석(1993)은 형태소 분석을 위하여 모호성이 발생하는 단어들의 유형을 형태소의 길이와 불규칙/축약 현상에 따라 4가지로 분류하고 있다. 이는 형태소 분석기가 가능한 모든 분석 결과를 생성하도록 하기 위한 목적으로 사용되었다. 김충원(1994)은 말뭉치 구축시에 발견된 모호성 유형들을 형태 모호성, 품사 모호성, 의미 모호성으로 구분하고 의미정보를 이용한 모호성 해결 방법을 제시한 바 있다.

이러한 기존의 연구를 기반으로 형태론적 모호성 해결의 관점에서 모호성 유형을 세분화하기 위한 형태론적 모호성 및 모호성 유형의 분류 기준은 다음과 같다.

### 형태론적 모호성이 있는 단어

임의의 단어에 대해 가능한 모든 형태소 분석 결과를 생성하는 형태소 분석기  $M$ 이 단어  $w$ 에 대한 분석 결과로  $r_1, r_2, \dots, r_n$ 을 생성할 때  $n \geq 2$ 이면  $w$ 는 형태론적 모호성이 있는 단어이다.

### 형태론적 모호성의 유형 분류 기준

단어  $w$ 에 대한 임의의 분석 결과  $r_i$ 와  $r_j$ 의 형태론적 차이점을 형태론적 특성에 따라 유형별로 분류한다. 이 모호성 유형들을 정의할 때 단어  $w$ 에 대한 임의의 두 분석 결과  $r_i$ 와  $r_j$ 의 모호성 유형이 유일하게 결정되도록 한다.

위 모호성 유형 분류 기준은 형태론적 모호성이 발생하는 원인 및 모호성 주체를 두 분석 결과의 형태론적 차이점에 초점을 맞추고, 모호성 해결의 편의를 위하여 한 단어에 대한 임의의 분석 결과가 한 가지 유형에 속하도록 모호성 유형을 정의하기 위한 것이다.

한국어 형태소 분석 결과는 크게 체언, 용언, 독립언 분석으로 분류되고, 각 분석 결과는 다시 형태소 분리 위치와 어근의 품사에 따라 모호성이 증폭되는 이중적 구조를 가지고 있다. 따라서 모호성의 유형은 체언-용언-독립언 모호성이라는 '1차적 모호성'과 체언/용언/독립언 내에서 모호성이 발생하는 '2차적 모호성'으로 구분된다.

1차적 모호성과 2차적 모호성, 혹은 2차적 모호성 내에서 두 가지 이상의 유형에 동시에 속하는 모호성이 발생하면 모호성 해결 문제가 어려워지므로 이를 허용하지 않은 것이다. 이러한 모호성 분류 원칙에 따라 임의의 두 분석 결과에 대한 형태론적 모호성을 유형별로 정의하면 다음과 같다.

## 2.1 어근 유형 모호성

임의의 두 형태소 분석 결과에 대해 실질형태소의 품사가 체언과 용언, 혹은 체언과 독립언 등과 같이 대분류 차원에서 다른 경우에 단어 유형이 달라지므로 이를 어근 유형 모호성(stem-type ambiguity)이라고 정의한다. 이 유형은 체언-용언 모호성, 체언-독립언 모호성, 용언-독립언 모호성으로 세분화되며, 그 예는 아래와 같다.

'자는' → '자'체언 + '는'어미, '자'용언 + '는'어미  
 '보고서' → '보고서'체언, '보'용언 + '고서'어미  
 '걸음마' → '걸음마'체언, '걸음마'독립언

## 2.2 형태소 분리 모호성

아래 예와 같이 명사형 어미, 접미사, 보조용언, 선어말 어미 '시'의 분리에 의하여 발생하는 모호성으로 형태소 분리 여부에 따라 모호성이 결정되기 때문에 이를 형태소 분리 모호성(morpheme segmentation ambiguity)이라고 정의한다.

'녹음이' → '녹음'용언 + '이'조사, '녹'용언+'음'어미+'이'조사  
 '그들은' → '그들'체언 + '은'조사, '그'체언+'들'접미사+'은'조사  
 '가시는' → '가시'용언 + '는'어미, '가'용언+'시'선어미+'는'어미  
 '보관함' → '보관함'체언, '보관'체언+'하'접미사+'다'어미  
 '사랑하다' → '사랑하'용언 + '다'어미,  
 '사랑'용언+'하'접미사+'다'어미  
 '예뻐지다' → '예뻐지'용언 + '다'어미,  
 '예쁘'용언+'어'어미+'지'접미사+'다'어미

## 2.3 형태소 길이 모호성

형태소 길이 모호성(morpheme-length ambiguity)은 어근의 품사가 동일한 품사범주<sup>1)</sup>에 속하면서 형태소 분리 위치에 따라 형태소, 특히 어근의 길이가 달라지는 모호성이다.

이 유형은 아래 예의 '가는'과 같이 문법형태소의 길이는 같더라도 실질형태소의 길이가 달라지는 것을 포함한다. 그러나 형태소의 길이가 다르다 할지라도 다른 모호성 유형에 포함되는 것은 제외한다.

'가는' → '가'용언 + '는'어미, '가늘'용언 + 'ㄴ'어미  
 '자라는' → '자라'용언 + '는'어미, '자'용언 + '라는'어미  
 '경기도' → '경기도'체언, '경기'체언 + '도'조사

## 2.4 불규칙 용언의 원형 복원 모호성

실질형태소가 분리되는 위치는 같으나 불규칙 용언의 원형 복원에 의하여 발생하는 모호성을 불규칙 용언의 원형 복원 모호성(irregular stem ambiguity)이라 한다.

'가는' → '가'용언 + '는'어미, '갈'용언 + '는'어미

## 2.5 '아/에/이' 탈락 모호성

이 유형은 탈락된 '아/에/이'를 형태소 분석 과정에서 복원할 때 발생하기 때문에 이를 '아/에/이' 탈락 모호성(a/e/i-drop ambiguity)이라고 한다. 이 유형에 속하는 분석 결과의 예는 아래와 같다.

'사과라고' → '사과'체언 + '라고'조사,  
 '사과'체언 + '이'서술격조사 + '라고'어미  
 '경찰서는' → '경찰서'체언 + '는'조사,  
 '경찰'체언 + '에서'는'조사

## 2.6 품사 모호성

넓은 의미의 품사 모호성은 형태소 분석 결과에서 한 형태소에 대한 품사가 두 개 이상이 가능한 경우이다. 그런데 이 모호성은 그 적용 범위가 너무 넓기 때문에 좁은 의미의 품사 모호성으로 두 분석 결과의 형태소 개수 및 각각의 형태소가 동일하지만 단지 품사만 다른 것을 품사 모호성(part of speech ambiguity)이라고 정의한다.

'곳' → 명사 또는 의존명사  
 '가다' → 자동사 또는 보조용언  
 '는' → 조사 또는 어미

좁은 의미의 품사 모호성은 하나의 형태소에 대한 품사만 다르고 나머지 형태소에 대한 품사는 모두 같은 경우와 두 개 이상의 형태소에 대한 품사가 다른 경우가 있다. 따라서 '객관적'명사/관형사, '잘못'명사/부사 등과 같이 대분류 품사 체계에서 발생하는 품사 모호성이나 '자라는'('자라'체언+'는'조사, '자라'용언+'는'어미)과 같이 두 개 이상의 형태소에서 품사 모호성이

1) 여기서 품사 범주는 대분류 체계에 따라 실질형태소의 품사를 체언/용언/독립언으로만 분류한 것을 의미한다.

발생하는 것은 ‘단어 유형 모호성’에 속하므로 제외한다.

### 3. 품사 유형과 품사 모호성

두 가지 이상의 품사가 가능한 형태소는 품사 모호성뿐만 아니라 대부분의 모호성들을 유발하는 가장 큰 요인이 되고 있다. 따라서 형태소 분석 사전에 수록된 형태소들의 품사 유형 및 특성을 분석함으로써 그 결과를 모호성 해결에 활용할 수 있다.

형태소의 품사 유형은 한국어의 품사 분류 체계가 계층적으로 정의되는 특성 때문에 대분류 체계에서 발생하는 것과 중분류 체계 혹은 소분류 체계에서 발생하는 것으로 구분된다. 대분류 체계는 어근의 품사 유형을 체언, 용언, 독립언 세 가지로만 분류한 것으로 이 범주 내에서 발생하는 품사 모호성을 ‘1차적 품사 모호성’이라 한다.

중분류 체계는 체언을 명사/대명사/수사/의존명사, 용언을 자동사/타동사/형용사/보조용언, 독립언을 부사/관형사/감탄사로 세분화한 것으로 이러한 세분화로 인하여 발생하는 품사 모호성을 ‘2차적 품사 모호성’이라 한다. 또한 명사를 보통명사와 고유명사 등으로 세분하는 소분류 체계에서는 ‘3차적 품사 모호성’이 발생하지만 이러한 품사 모호성 문제는 형태소 수준에서 처리하기 어려우므로 제외하기로 한다.

표 1. 형태소의 품사 유형

체언-용언 (297개)	명사-타동사	87
	명사-자동사	32
	명사-자동사타동사	46
	명사-형용사	26
	명사-형용사타동사	12
기타(명대-자타 등)	94	
체언-용언-독립언		35
체언-독립언 (488개)	명사-부사	332
	명사-관형사	28
	의존명사-관형사	29
	명사의존명사-관형사	11
	명사-감탄사	33
기타(명대-부관 등)	55	
용언-독립언		28
총		848

대분류 체계에서 발생하는 1차적 품사 모호성을 분석하기 위하여 6만 단어 수준의 국어 사전에서 발견되는 품사 모호성의 유형을 조사하였는데 각 유형에 속하는 어휘수는 표 1과 같다. 표 1에서 ‘-적’으로 끝나는 어휘의 명사-관형사 모호성 493개와 명사-부사 모호성이 있는 ‘가급적’, 명사-부사-관형사 모호성이 있는 ‘비교적’은 제외하였다. 그 이유는 관점에 따라 ‘-적’을 접미사로 간주할 수도 있기 때문이다.

중분류 체계에서 발생하는 품사 모호성은 체언이 명사/대명사/의존명사/수사, 용언이 자동사/타동사/형용사/보조용언, 독립언이 부사/관형사/감탄사 중에서 두 가지 이상의 품사로

사용되는 어휘에서 발생한다. 이 때 중분류 체계에서 발생하는 품사 모호성은 더 이상 세부적으로 분류하지 않는다.

### 4. 모호성 해결 방법

모호성 해결 방법은 한 단어의 분석 결과 내에서 해결되는 것과 좌우 문맥의 분석 결과를 참조하는 방법으로 구분된다. 한 단어의 분석 결과만으로 해결하는 방법은 각 분석 결과의 형태론적 특성에 따라 선호도가 달라지는 특성을 이용하여 각 분석 결과에 대해 절대적인 가중치 부여 방법(absolute weighting)과 두 개의 분석결과를 비교하여 가중치를 부여하는 상대적 가중치 부여 방법(relative weighting)을 이용한다.

#### 4.1 절대적인 가중치 부여 방법

절대적 가중치는 임의의 형태소 분석 결과에 대해 형태소의 길이, 단어 유형, 형태론적 특성 등 그 분석 결과 자체가 가지는 특성을 가중치 형태로 반영하기 위한 것이다. 아래와 같은 절대적 가중치 부여 원칙에 따라 가중치를 부여한다.

##### [원칙 1] 형태소 길이 비례의 원칙

형태소 분석 결과에 대한 가중치는 어근과 문법형태소 등 분리된 형태소의 길이에 비례하여 가중치를 증감한다.

‘형태소 길이 비례의 원칙’은 분리된 형태소의 길이가 모호성 해결을 위한 요소로서 작용할 수 있음을 의미한다. 이를 적용한 예를 들면, 분석 결과의 모든 형태소가 사전에 수록된 것은 어근의 길이가 길수록 가중치를 높이고, 어근이 미등록어인 경우에는 길이가 짧을수록 가중치를 높이는 방법이다. 즉, 분석 성공인 결과는 어근의 길이가 긴 것을 선호하고 분석 실패한 결과는 문법형태소의 길이가 긴 것을 선호한다.

##### [원칙 2] 형태소 개수 비례의 원칙

형태소 분석 결과에 대한 가중치는 분석 결과를 구성하고 있는 형태소의 개수에 비례하여 가중치를 증감한다.

일반적으로 형태소 분석 결과를 구성하고 있는 형태소의 수가 많을수록 가중치를 높이는 방법이다. 그러나 분석 성공인 결과는 형태소의 개수가 적을수록 가중치를 높이는 것이 더 유리하다.

##### [원칙 3] 원형 보존의 원칙

축약이나 탈락에 의한 형태소의 변이체가 포함된 분석 결과의 가중치를 낮춘다.

‘형태소 개수 비례의 원칙’은 형태소의 개수가 많은 결과를 우선적으로 선택하게 되는데, ‘아/어’나 ‘이’ 등 탈락 또는 축약 현상에 의하여 형태소 개수가 많아진 것은 선호도가 낮아지는 것이 일반적이다. 즉, 탈락이나 축약 현상보다는 그렇지

않은 분석 결과가 선호된다. 따라서 탈락이나 축약 현상에 의하여 형태소가 복원된 것은 가중치를 낮춰야 한다.

[원칙 4] 통계적 선호도의 원칙

모호성이 있는 단어들에 대해 통계적으로 선택되는 빈도가 가장 높은 분석 결과와 두 번째 높은 결과와의 편차 정보를 이용하여 가중치를 증감한다.

4.2 통계적 빈도수를 이용하는 방법

사용 빈도수가 높은 단어(어절)들에 대해서만 모호성이 해결되더라도 실제 문서에서는 대부분의 모호성이 해결될 것이라 예상된다. 그런데 형태론적 모호성이 일어나는 어절의 정확한 개수가 얼마나 되는지, 현실적으로 통계적 빈도수를 계산할 수 있는지를 판단하기가 쉽지 않다. 한국어에서 형태론적 모호성이 발생하는 단어수(어절수)는 예상보다 훨씬 적을 것으로 판단되며, 그 근거는 다음과 같다.

표 1을 기준으로 어근의 형태소가 동일한 모호성이 발생하는 어절수를 대략 계산하면 다음과 같다.

- 체언-독립언 모호성 어절수 : 523개
- 길이가 같은 체언-용언 모호성 어절수 : 332개 × x

1차적 품사 모호성에 따라 길이가 같은 체언-용언 모호성이 발생하는 어절수는 체언과 용언이 동시에 가능한 어근수와 그 뒤에 올 수 있는 조사/어미의 평균 개수 x를 곱한 것이다. 즉, x는 조사/어미로 가능한 문법형태소의 수보다 작은 값이다. 그러나 체언과 용언이 모두 가능하더라도 '먹은('먹' 체언/용언 + '은'조사/어미)과 '먹는('먹' 용언 + '는'조사/어미)의 예에서 알 수 있듯이 모호성이 발생하지 않는 경우가 많기 때문에 그 값은 더 작아진다. 실제로 조사-어미 모호성을 갖는 형태소는 20여개지만, 그 중에는 사용 빈도가 낮은 것들이 포함되어 있으므로 x는 5~10일 것으로 예측된다.

- 길이가 다른 체언-용언 모호성 어절수
- 2차적 품사 모호성 어절수 :  
(모호성 어근수) × (문법형태소의 수)

길이가 다른 체언-용언 모호성이 일어나는 어절들은 실제로 그 개수가 많지 않으나, 2차적 품사 모호성이 일어나는 어절들은 그 수가 매우 많다. 그러나 2차적 품사 모호성 어절들은 문법형태소가 어근이 모호성을 유발하므로 통계적 선호도를 계산할 때는 어근 단위로 빈도수를 측정해도 된다.

'통계적 선호도의 원칙'을 적용하기 위해 계산된 모호성 어절수는 '2차적 품사 모호성 어절수'를 빈도가 높은 1만 어휘 정도로 제한할 경우 약 2만 어절이다<sup>2)</sup>. 따라서 이에 대한 빈도수 통계를 수집함으로써 형태론적 모호성을 해결하는데

2) 2.5절 '사과라고'의 예와 같이 모든 명사에 대해 발생하는 모호성 어절은 제외하였다.

많은 도움이 될 것으로 예상된다.

4.3 상대적인 가중치 부여 방법

한 단어의 분석 결과들 중에서 임의의 두 개를 비교했을 때 형태론적 차이점에 의해 발생하는 모호성의 유형에 따라 상대적인 가중치를 부여할 수 있다.

(1) 어근 유형 모호성의 해결

• 체언-독립언 모호성

이 유형은 조사/어미가 결합되지 않은 단일어 유형에서 발생한다. 그런데 일반적으로 체언-부사, 체언-관형사 유형에서는 부사/관형사로 사용되는 빈도가 매우 높고, 체언-감탄사 유형에서는 체언 빈도가 매우 높다. 따라서 빈도가 높은 분석 결과에 가중치를 높게 부여한다. 즉, 독립언이 부사 혹은 관형사이면 부사/관형사의 가중치를 높이고, 감탄사이면 가중치를 낮춘다.

• 명사-관형사 모호성

품사 모호성 중에서 많은 부분이 '-적'으로 끝나는 명사-관형사 문제이다. 이 경우에 '-적'으로 끝난 어근 뒤에 조사가 결합된 경우에는 명사로, 그렇지 않은 경우에는 관형사로 간주한다.

• 체언-용언 모호성

이 유형은 비교적 자주 발생하고 있으나 형태론적 특성에 의해 가중치를 부여하기가 어렵다. 따라서 '통계적 선호도'나 좌우 문맥을 고려하는 방법, 또는 구문분석을 이용한 해결 방법 등이 모색되어야 한다.

(2) 단어 유형에 따른 가중치 부여

단어 유형이 복잡할수록 가중치를 높이고 단순할수록 가중치를 낮춘다. 예를 들어, 접미사가 분리되지 않은 유형보다 접미사가 분리된 유형에 대하여 가중치를 더 높게 한다. 그러나 미등록어가 포함된 것은 '명사 + 조사', '명사 + 서술격 조사 + 어미' 유형에 대한 가중치를 가장 높게 부여한다.

(3) 문법형태소의 빈도/길이/구어체 정보

두 개의 분석 결과가 모두 문법형태소가 분리된 경우에 문법형태소의 길이가 긴 것, 빈도가 높은 것에 대하여 가중치를 높인다. 길이가 짧고 빈도가 낮은 문법형태소, 구어체 문법형태소는 가중치를 낮춘다.

(4) '아/에/이' 탈락 모호성

'아/어'나 '에', '이'가 탈락된 것을 복원한 분석결과는 그렇지 않은 유형에 비해 가중치를 낮춘다. 형태소 분석 결과 중에서 탈락 현상이 일어난 유형은 일반적으로 그렇지 않은 것에 비해 오분석일 가능성이 높기 때문이다.

그러나 2.5절 '사과라고'의 예와 같이 체언과 문법형태소가 동일하고 단지 서술격 조사 '이'만 차이가 나는 것은 서술

격 조사가 복원된 분석 결과에 높은 가중치를 부여한다.

#### (5) 미등록어 추정 결과

미등록어의 어근은 일반적으로 명사일 가능성이 매우 높다. 따라서 용언으로 분석된 것보다는 체언으로 분석된 결과에 대하여 가중치를 높게 부여한다.

### 5. 결 론

한국어에서 발생하는 형태론적 모호성의 유형들을 모호성 해결의 관점에서 체계적으로 분석하고 모호성 해결 방안을 제시하였다. 모호성의 유형을 1차적 모호성과 2차적 모호성으로 규정하여 교착어에서 광범위하게 발생하는 1차적 모호성을 중심으로 그 유형들을 정리하였다. 또한 각 유형들을 기반으로 모호성을 해결하는 방안을 제시하였다.

이 기능이 구현되면 형태소 분석기가 분석 결과를 하나만 제시해 주게 되어 응용 시스템의 요구를 다양하게 충족시킬 수 있을 것이다. 특히, 구문분석시에 가중치가 높은 분석 결과를 중심으로 파싱할 수 있기 때문에 파서의 효율을 높이는 데 많은 기여를 할 것으로 예상된다.

#### 참고문헌

- [1] 임권묵, 김병희, 송만석, “형태 중의성 해결을 위한 말마디 사전 설계”, 한국정보과학회 봄 학술발표 논문집, 제20권 1호, pp.789-792, 1993.
- [2] 안미정, 옥철영, “형태소적 중의성 해소를 통한 한국어 복수동사의 통합”, 한국 인지과학회 춘계 학술발표 논문집, pp.18-31, 1995.
- [3] 임희석, 이호, 임해창, “형태소 분석 단계에서 발생하는 어절의 중의성 분석 방안”, 한국정보과학회 봄 학술발표 논문집, 제20권 1호, pp.773-776, 1993.
- [4] 김충원, 임권묵, 송만석, “의미정보를 이용한 형태소 중의성 해결”, 한국정보과학회 가을 학술발표 논문집, 제21권 2호, pp.649-652, 1994.
- [5] 임해창, 임희석, 윤보현, “자연어처리를 위한 품사 태깅 시스템의 고찰”, 한국정보과학회지, 14권 7호, pp.36-57, 1996.
- [6] 이하규, “어말-어두 공기 정보를 이용한 한국어 어휘 중의성 해소”, 정보과학회 논문지(B), 24권 1호, pp.82-89, 1997.
- [7] 김재훈, 임철수, 서정연, “은닉 마르코프 모델을 이용한 효율적인 한국어 품사의 태깅”, 정보과학회 논문지, 22권 1호, pp.136-146, 1995.
- [8] 신중호, 한영석, 박영찬, 최기선, “어절구조를 반영한 은닉 마르코프 모델을 이용한 한국어 품사 태깅”, 제6회 한글 및 한국어 정보처리 학술발표 논문집, pp.389-394, 1994.
- [9] 신상현, 이근배, 이종혁, “통계와 규칙에 기반한 2단계 한국어 품사 태깅 시스템”, 한국정보과학회 논문지(B), 24권 2호, pp.318-321, 1997.
- [10] 임희석, 김진동, 임해창, “변형 규칙 기반 한국어 품사 태거의 개선”, 제8회 한글 및 한국어 정보처리 학술발표 논문집, pp.216-221, 1996.
- [11] 신상현, 이근배, 이종혁, “TAKTAG: 통계와 규칙에 기반한 2단계 학습을 통한 품사 모호성 해결”, 제7회 한글 및 한국어 정보처리 학술발표 논문집, pp.169-174, 1995.
- [12] 신상현, 이근배, 홍남희, 이종혁, “확률과 규칙을 사용한 품사 태깅”, 제6회 한글 및 한국어 정보처리 학술발표 논문집, pp.318-321, 1994.
- [13] 이하규, 김영택, “통계정보에 기반을 둔 한국어 어휘중의성 해소”, 한국통신학회지, 19권 2호, pp.265-275, 1994.
- [14] 김재훈, 장병규, 김길창, 서정연, “형태소의 모호성 축소를 위한 포섭조건의 자동 추론”, 제7회 한글 및 한국어 정보처리 학술발표 논문집, pp.175-180, 1995.
- [15] 김재훈, 조정미, 김창현, 서정연, 김길창, “퍼지망을 이용한 한국어 품사 태깅”, 제5회 한글 및 한국어 정보처리 학술발표 논문집, pp.593-603, 1993.
- [16] 이운재, 최기선, 김길창, “한국어 문서 태깅 시스템”, 한국정보과학회 봄 학술발표 논문집, 제20권 1호, pp.805-808, 1993.
- [17] 이상주, 임희석, 임해창, “은닉 마르코프 모델을 이용한 두단계 한국어 품사 태깅”, 제6회 한글 및 한국어 정보처리 학술발표 논문집, pp.305-312, 1994.
- [18] 임희석, 김진동, 임해창, “변형 규칙 기반 한국어 품사 태거의 개선”, 제8회 한글 및 한국어 정보처리 학술발표 논문집, pp.216-221, 1996.
- [19] 박혜준, 윤준태, 송만석, “말뭉치 꼬리달기 시스템 구현”, 한국정보과학회 봄 학술발표 논문집, 제21권 1호, pp.829-832, 1994.
- [20] 임희석, 언어지식과 통계정보를 이용한 한국어 품사 태깅 모델, 고려대학교 컴퓨터과학과 박사학위 논문, 1997.