

고성능 한국어 형태소 분석을 위한 어미 분류

은종진 박선영
LG 전자기술원
{zzeun, choktak}@LG-Elite.com

A Classification of Endings for an Efficient Morphological Analysis of Korean

Zong-Zin Eun Sun-Young Park
LG Electronics Institute of Technology

요 약

한국어 형태소 분석에서 가장 중요한 부분 중 하나가 바로 용언구(서술어)를 분석하는 것이다. 형태소 분석 뿐만 아니라 구문 분석, 의미 분석 단계에서도 정확한 용언구 분석은 매우 중요한 작업 중의 하나이다. 또한, 용언구에는 [체언+지정사+어미] 패턴도 포함되므로, 정보 검색기의 핵심 모듈인 명사 추출기(색인기)의 성능에도 용언구의 분석은 높은 비중을 차지한다. 본 논문에서는 용언구 분석의 정확성을 높이고, 견고하면서 속도도 향상시킬 수 있는 방법으로 새로운 어미 분류를 제안하고자 한다.

1. 서론

지금까지 한국어 형태소 분석기에 대해서 많은 연구들이 이루어져 왔으며[1,2,6,7,8,10], 자연어 처리를 연구하는 학교나 기업에서는 대부분 자체 개발한 형태소 분석기를 가지고 있을 정도로 필수적인 요소로 자리잡고 있다. 또한, 정보 검색기, 음성 합성/인식기, 기계 번역, 맞춤법 교정기 등 많은 분야에서 활용되고 있다.

그러나, 형태소 분석기를 개발하는 사람들 대부분은 전산학을 전공하였기 때문에, 국어학적 지식을 기반으로 한 형태소 분석 연구가 부족한 상태이다. 특히, 용언과 어미가 결합할 때 용언과 어미에 많은 형태 변화가 있지만, 용언의 원형을 잘 분석해 내는 반면, 어미에 대해서는 원형을 밝혀 내는데 상대적으로 소홀하고, 단순히 용언으로부터 분리해 내는 데에만 치중했다. 따라서, 어미의 정확한 의미를 필요로 하는 응용 프로그램 등에서 좋은 성능을 발휘하지 못하는 원인이 되었다.

이에 본 논문은 어미의 형태 변이에 대한 분석을 통해 어미의 원형을 찾아내고, 용언과 어미가 결합할 때 나타나는 형태 변이의 특성에 따라 어미를 분류함으로써, 기존의 형태소 분석기들에 비해 메모리를 적게 쓰면서도 속도와 성능을 향상시킬 수 있음을 보이고자 한다.

2 장에서는 어미를 분류하기에 앞서 대표 어미와 변형 어미에 대한 뜻을 정의하고, 어미 분류를 활용하지 않은 기존의 형태소 분석기들의 오분석 예들을 보여준다. 3 장에서는 본 논문에서 제안하는 새로운 어미 분류에 대해 기술한다. 4 장에서는 본 논문에서 새로이 제안한 어미 분류를 형태소 분석기에서 어떻게 활용하는지를 보이고자 한다.

2. 관련 연구

이번 장에서는 어미 분류의 기준이 된 대표 어미와 대표 어미로부터 파생된 변형 어미의 뜻을 정의하고, 기존의 형태소 분석기들

의 문제점들을 살펴 보겠다.

2.1 대표 어미와 변형 어미

한국어에서 어미는 동일한 의미의 어미더라도 어미 앞에 오는 형태소에 따라 다양한 형태로 변형된다. 어미 앞에 올 수 있는 형태소로는 동사, 형용사, 선어말 어미¹, 지정사²가 있는데, 그 형태소들의 특성에 따라 어미는 다양한 형태로 변화하며, 앞에 오는 형태소의 형태가 바뀌기도 한다.

다음은 어미 ‘는다’가 여러 종류의 형태소들과 결합할 때의 예이다.

- (예1) a. 먹는다 (는다)
- b. 나온다 (는다)
- (예2) a. 작다 (다)
- b. 예쁘다 (다)
- (예3) a. 먹었다 (다)
- b. 먹으신다 (는다)
- (예4) a. 책이다 (다)
- b. 선수이다 (다)

(예 1)은 동사와 어미 ‘는다’와의 결합을 나타내며, (예 2)는 형용사와 어미 ‘는다’의 결합을 나타내고, (예 3)는 선어말 어미와의 결합, (예 4)은 지정사 ‘이’와 어미 ‘는다’가 결합한 예를 보여 준다. 위 예문에서 볼 수 있듯이 하나의 어미 ‘는다’는 앞에 오는 형태소에 따라 여러 가지 형태의 어미로 나타난다. 이러한 경우, ‘는다’를 대표 어미라 정의하고, 예문에 나온 각각의 변한 형태의 어미(‘는다’, ‘는다’, ‘다’)를 변형 어미라 정의 하겠다. 즉, 대표 어미는 어미의 원형이라 할 수 있겠다.

형태소 분석에서 대표 어미를 찾아 내는 것은 어미의 정확한 의미를 파악하는데 중요한 열쇠가 되지만, 지금까지의 형태소 분석기들은 어미 원형을 찾는데 소홀했다. 다음 절에서는 지금까지의 형태소 분석 방법들의 오분석 예들을 살펴 보도록 하겠다.

2.2 기존의 형태소 분석 방법들

¹ 여기서 말하는 선어말 어미는 ‘으시/였/겠’ 중의 하나로 제한한다.

² 체언과 어미가 결합할 때 필요한 ‘이’를 말한다.

형태소 분석의 방법들을 여러 가지 측면에서 분류할 수 있겠지만, 본 논문에서는 용언구를 분석하는 방법에 따라 크게 ‘음소 단위의 분석’과 ‘음절 단위의 분석’으로 구분하였다. 각각의 분석 방법들에 대해 간단히 소개하고, 예문들을 통해 문제점을 하나씩 짚어 보도록 하겠다.

음소 단위의 분석

어미를 용언 등으로부터 분리해 내는 방법 중에 가장 쉽게 생각할 수 있는 것은 음절을 음소 단위로 쪼개는 것이다. 이 방법은 많은 형태소 분석기들이 채택해 온 방법인데[2,6,7,8,10], 여기에는 여러 가지 문제점이 있다. 다음의 예를 보자.

- (예1) 먹는다
- ⇨ 먹 + 느 + ㄴ + 다
- ⇨ 먹 + 는다

(예 1)에서 정답은 ‘먹+는다’지만 이를 미리 알 수 없기 때문에, 일단 ‘느’를 음소 단위로 쪼개어 ‘느+ㄴ’으로 만들고, ‘먹+는다’와 ‘먹느+ㄴ다’ 중 어느 것이 정답인지 판별해야 하므로 복잡도가 증가한다. 즉, 모든 음절에 대해 중성이 ‘ㄴ/ㄹ/ㅁ/ㅂ’이면 음소 단위로 쪼개야 하고, 또한 자소 단위로 분석이 이루어지므로 분석의 복잡도도 증가하고, 속도도 훨씬 느려진다.

음절 단위의 분석

이를 해결하는 방법으로 ‘음절 단위의 분석’을 시도한 연구가 있다[1]. 이는 입력문을 음소 단위로 쪼개지 않고, 음절 단위로 형태소 분석을 시도한다. 따라서 분석 복잡도가 음소 단위에 비해 훨씬 줄어든다.

그러나, 이 방법에서도 어미의 원형을 제대로 찾아 주지 못함으로써 어미의 의미에 대한 모호성을 갖고 있는 상태로 분석 결과를 낸다. 다음의 예를 보자.

- (예2) a. 나간
- ⇨ 나가 + ㄴ (과거)
- b. 나간다
- ⇨ 나가 + ㄴ다 (현재)
- c. 예쁜
- ⇨ 예쁘 + ㄴ (현재)

위의 예문의 분석 결과를 보면 어미는 모두 같은 ‘ㄴ’으로 시작하지만, 현재형인지 과거형인지를 밝히지 못한 채 결과를 낸다. 뿐만 아니라, 기존의 형태소 분석 방법들에서는 변형 어미들이 모두 어미 사전에 등록되어 있어야 하므로 사전의 크기가 커질 뿐만 아니라, 사전 검색 시간 또한 증가하기 때문에 전체적인 형태소 분석 효율이 떨어지게 된다. 즉, 메모리와 속도 모두 나빠지는 결과를 가져 온다.

또 다른 문제점으로 잘못된 용언구를 정상적인 문장으로 오분석하는 것이다. 즉, 다음의 예를 보자.

- (예 3) a. 예쁘다
 ⇨ 예쁘 + ㄴ다
 b. 나가는다
 => 나가 + 는다

위의 예는 분명히 잘못된 입력이지만, 기존의 형태소 분석 방법으로는 잘못된 입력임을 판별하기 어렵다.

이처럼 기존의 형태소 분석 연구들은 여러 가지 문제점을 갖고 있음을 알 수 있다. 따라서, 본 논문에서는 이러한 여러 가지 문제점들을 해결하고, 속도 또한 증가할 수 있는 방법으로 새로운 어미 분류를 제안하고자 한다.

3. 어미 분류

일반적으로 전산언어학에서 어미 분류는 어미가 나타나는 순서에 따라 선어말 어미와 어말 어미로 구분하는 정도가 거의 전부이다[1,4]. 선어말 어미는 주로 ‘으시/였/졌’만 해당되며, 국어 문법에서 선어말 어미로 보는 것들 대부분은 전산언어학에서는 편의상 어말 어미로 간주한다[1]. 어말 어미는 더 이상 분류하지 않고, 앞장에서 보았던 예에서처럼 하나의 대표 어미로부터 파생된 모든 변형 어미들을 전혀 다른 어미인 것처럼 사전에 각각 등록하여 처리한다.

국어 문법에서는 어미를 의미적 관점에서 분류하기도 한다. 즉, 어미가 의문형인지, 명령형인지 등에 따라 분류하는 것이다. 하지만, 이러한 것이 전산언어학에서 사용되고 있지는 않다.

본 논문에서는 용언과 어미가 결합할 때 나타나는 용언과 어미의 형태 변화를 바탕으로, 어미를 새로이 분류하여 전산언어학에

서 유용하게 활용할 수 있도록 하고자 한다.

어미를 분류하기에 앞서 먼저 어미가 나타나는 용언구 패턴을 살펴 보면 다음과 같다.

[어미의 출현 패턴]

- (1) 용언 + 어미
- (2) 용언 + 선어말어미 + 어미
- (3) 체언 + 지정사 + 어미

어미가 나타나는 패턴을 보면, 어미 앞에 올 수 있는 형태소로는 동사, 형용사, 선어말 어미, 지정사임을 알 수 있다. 하나의 대표 어미는 어미 앞에 올 수 있는 형태소들의 특성, 즉 품사, 용언 불규칙 현상, 마지막 음절의 종성 여부(받침이 있는가 없는가) 등에 따라 여러 가지 형태의 변형 어미를 갖는 것을 알 수가 있다. 따라서, 먼저 대표 어미와 그의 변형 어미를 파악하는 것이 무엇보다 중요하다.

본 논문은 어미들을 분석한 결과로 얻은 대표 어미를 앞에 나오는 형태소들과 결합할 때의 형태 변이 특성에 따라 다음의 4 가지로 분류하였다.

[어미 분류]

- (1) ‘으’ 매개 어미
- (2) ‘는’ 매개 어미
- (3) ‘어’ 매개 어미
- (4) 규칙 어미

각각의 어미 종류에 대해서 하나씩 살펴볼도록 하자.

3.1 ‘으’ 매개 어미

‘으’ 매개 어미는 어미 앞에 오는 형태소의 종성 여부에 따라 어미의 형태가 바뀌는 어미를 말한다. 즉, 앞에 오는 형태소의 마지막 음절이 유종성이면 ‘으’가 탈락되지 않고, 무종성이면 ‘으’가 탈락된다. 단 ‘ㄹ’ 불규칙 용언과 ‘ㅅ’ 불규칙 용언은 예외이다. 이에 대해서는 뒤에서 다시 설명하겠다.

‘으’ 매개 어미는 ‘으/은/을/읍/읍/습’으로 시작하는 어미들로 구성된다. 여기서, ‘습’으로 시작하는 어미는 유종성 뒤에서는 ‘습’으로 시작하고, 무종성 뒤에서는 ‘스’가 탈락된 ‘ㅂ’이 되므로 정확히 표현하자면 ‘스’ 매개 어미이지만, 현상이 ‘으’ 매개 어미와 유사

하므로 ‘으’ 매개 어미로 분류하였다.

대표어미		을
동사	유종성	을 (먹을)
	무종성	리 (나갈) 리* (도울)
형용사	유종성	을 (좋을)
	무종성	리 (예쁠) 리* (하얗)
선어말어미	유종성	을 (먹었을)
	무종성	리 (먹으실)
지정사		리 (책임)

표 3-1 ‘으’ 매개 어미의 예

위의 표에서도 알 수 있듯이 ‘으’ 매개 어미는 어미 앞에 오는 형태소의 품사 혹은 불규칙 용언인가 하는 특성들과는 상관없이 없고, 오직 마지막 음절에 종성이 있는가에 따라 ‘으’가 존재하기도 하고, 탈락되기도 한다. 불규칙 용언에서도 마찬가지로 적용되는데 불규칙 용언인 경우는 원형에서 발생하는 것이 아니라, 불규칙 현상에 의해 바뀐 형태에서 적용된다. 다음의 예를 보자

(예1) ‘ㅂ’ 불규칙 용언

- 돕 + 을
- ⇒ 도우 + 을 (용언 불규칙 현상)
- ⇒ 도울 (‘으’ 탈락)

(예2) ‘ㅎ’ 불규칙 용언

- 하얗 + 을
- ⇒ 하야 + 을 (용언 불규칙 현상)
- ⇒ 하얗 (‘으’ 탈락)

위의 예를 보면 불규칙 원형(‘돕’, ‘하얗’)이 아닌 불규칙 현상에 의해 바뀐 형태(‘도우’, ‘하야’)의 종성 여부에 따라 ‘으’ 탈락이 결정됨을 알 수 있다. 즉, 바뀐 불규칙 용언이 무종성이므로 ‘을’에서 ‘으’가 탈락된 형태와 결합한 것이다.

‘으’ 매개 어미의 예외

그러나, 일부 불규칙 용언에서 ‘으’ 매개 어미와 결합할 때 예외 현상을 나타낸다. 다음의 예를 보자.

(예3) ‘ㄹ’ 불규칙 용언

- 알 + 으면
- ⇒ 알 + 면 (‘으’ 탈락)

(예4) ‘ㅅ’ 불규칙 용언

- 짓 + 으면
- ⇒ 지 + 으면 (용언 불규칙 현상)

(예3)에서 ‘ㄹ’ 불규칙 용언 ‘알’은 무종성임에도 불구하고, ‘으’ 매개 어미 ‘으면’의 ‘으’가 탈락됨을 볼 수 있다. (예4)는 ‘ㅅ’ 불규칙 용언 ‘짓다’가 ‘ㅅ’ 불규칙 현상에 의해 무종성(‘지’)이 되었는데도 어미 ‘으면’의 ‘으’는 탈락되지 않았다.

‘으’ 매개 어미에 속하는 어미들의 대표 어미들은 다음과 같다.

[‘으’ 매개 어미의 예]

- (a) 으며, 으니, 으나, 으라고, 으시³⁾, ...
- (b) 은, 은지, 은테, 은가, ...
- (c) 을, 을수록, ...
- (d) 음, 음에도, ...
- (e) 읍시다, ...
- (f) 습니까, 습니다, ...

3.2 ‘어’ 매개 어미

‘어’ 매개 어미는 모음조화 규칙에 따라 어미 앞에 오는 형태소가 양성이면 ‘아’, 음성이면 ‘어’로 시작하는 어미들을 말한다. ‘어’ 매개 어미는 또한 양성/음성에 관계없이 어미 앞에 오는 형태소의 마지막 음절이 무종성이면 ‘어’와 축약이 일어난다. 불규칙 용언과 결합할 때는 용언 불규칙 현상을 일으키기도 한다. 다음의 표를 보자.

대표어미		어도
동사	유종성	어도 (먹어도) 아도 (잡아도)
	무종성	도 (나가도) 도* (나와도) 도* (도와도) 여도 (하여도) 러도 (이르러도)
형용사	유종성	아도 (좋아도) 어도 (적어도)
	무종성	도* (예뻐도) 도* (빨라도)
선어말어미	유종성	어도 (먹었어도)
	무종성	어도 (먹으시어도) 도* (먹으셔도)

³⁾ ‘으시’는 선어말 어미이지만, 어미 분류로 보면 ‘으’ 매개 어미에 해당한다.

지정사	어도 (책이어도)
-----	-----------

표 3-2 ‘어’ 매개 어미의 예

위의 표에서 별표(*)가 있는 경우는 ‘어’ 매개 어미가 무중성 형태소와 결합하여 축약이 일어난 경우와, 불규칙 용언과 결합하여 용언 불규칙 현상이 일어난 경우들이다.

‘어’ 매개 어미의 대표 어미들은 다음과 같다.

[‘어’ 매개 어미의 예]

- (a) 어, 어서, 어라, 어야, ...
- (b) 었

위의 예에서 ‘었’은 선어말 어미이지만, 현상은 ‘어’ 매개 어미와 같으므로, ‘어’ 매개 어미로 분류한다.

3.3 ‘는’ 매개 어미

‘는’ 매개 어미는 동사와 결합할 때 ‘는’ 혹은 ‘ㄴ’이 삽입되고, 형용사와 결합할 때는 ‘는’ 혹은 ‘ㄴ’이 탈락되는 어미들을 말한다. 어미 앞에 오는 형태소가 동사인 경우 ‘는’ 혹은 ‘ㄴ’이 삽입된다고 했는데, 이는 동사의 종성 여부에 따라 달라진다. 즉, 동사의 마지막 음절이 유중성이면 ‘는’이 삽입되고, 무중성이면 ‘ㄴ’이 된다.

대표어미	는다
동사	유중성 는다 (먹는다)
	무중성 ㄴ다 (나간다)
형용사	유중성 다 (좋다)
	무중성 다 (예쁘다)
선어말어미	유중성 다 (먹었다)
	무중성 ㄴ다 (먹으신다)
지정사	다 (책이다)

표 3-3 ‘는’ 매개 어미의 예

동사나 형용사 다음에 선어말 어미가 오는 경우는 선어말 어미마다 다른 현상을 보임을 알 수가 있다. 즉, 아래의 (예 1)에서처럼 선어말 어미 ‘었/겠’이 오는 경우는 항상 ‘는’ 혹은 ‘ㄴ’이 탈락되고, (예 2)에서처럼 선어말 어미 ‘으시’의 경우는 선어말 어미가 없을 때와 마찬가지로 선어말 어미 앞에 오는 용언의 품사에 따라 동사이면 ‘ㄴ’이 되고, 형용사이면 ‘ㄴ’이 탈락된다.

- (예 1) a. 먹었다
- b. 예뻤다
- (예 2) a. 먹으신다
- b. 예쁘시다

선어말 어미 ‘으시’다음에 ‘ㄴ’이 되는 것은 ‘으시’가 무중성으로 끝나기 때문이다.

‘는’ 매개 어미의 예는 다음과 같다.

[‘는’ 매개 어미의 예]

- (a) 다, 다기에, 다네, 단다, ...
- (b) 구나, 구려, ...

3.4 규칙 어미

규칙 어미는 어미 앞에 오는 형태소에 상관없이 결합할 때 항상 똑같은 형태로 나타나는 어미를 말한다.

대표 어미	거나
동사	유중성 거나 (먹거나)
	무중성 거나 (나가거나)
형용사	유중성 거나 (좋거나)
	무중성 거나 (예쁘거나)
선어말어미	유중성 거나 (먹었거나)
	무중성 거나 (먹으시거나)
지정사	거나 (책이거나)

표 3-4 규칙 어미의 예

위의 표에서 대표 어미는 ‘거나’이며 어미 앞에 오는 형태소에 상관없이 ‘거나’의 형태가 유지된다. 규칙 어미는 선행하는 형태소에 관계없이 형태가 변하지 않으므로 대표 어미만 존재하고, 변형 어미는 없다.

다음은 규칙 어미에 해당하는 어미의 예들이다. 지금까지의 어미 분류에 해당하지 않는 모든 어미는 규칙 어미라 할 수 있겠다.

[규칙 어미의 예]

고, 기, 게, 지, 도록, 든, 라는, 던, 나이다, 건, 더라도, 되, 자는, 냐, 다면, 네, 구나, 니⁴, ...

‘으’ 매개 어미의 ‘으니’와는 다르다. 규칙 어미 ‘니’는 의문형 종결 어미이고, ‘으’ 매개 어미의 ‘으니’는 까닭을 나타내는 연결 어미이다.

4. 어미 분류의 활용

이번 장에서는 새로운 어미 분류를 용언구 분석에 어떻게 활용하며, 어떠한 장점들을 가지는지를 보이고자 한다.

용언은 어미와 결합할 때 용언의 불규칙 현상이나, 어미와의 축약 현상 등에 의해 많은 형태 변이가 일어나는데, 새로운 어미 분류 정보를 이용하면 용언의 변형된 표층 형태만으로 다음에 올 어미 종류를 알아낼 수 있다. 몇가지 예문들을 살펴 보자.

불규칙 용언의 분석

(예1) 도우면 => 돕 + 으면

위의 예에 나온 ‘도우면’은 ‘되’ 불규칙 용언 ‘돕다’가 ‘으’ 매개 어미 ‘으면’과 결합한 형태이다. ‘도우면’을 분석하기 위한 첫번째 단계는 먼저 용언 부분인 ‘도우’를 찾는 것이다. 검색된 ‘도우’는 ‘되’ 불규칙 용언 ‘돕다’가 ‘으’로 시작하는 어미와 결합할 때 변화한 형태이므로, 어미 ‘으’를 복원하여 ‘으’ 매개 어미 사전을 검색함으로써 분석된다. 즉, 전체 어미 사전을 검색할 필요없이 ‘으’ 매개 어미 사전만 검색하면 된다. 또한, 어미 ‘으면’과 ‘면’을 모두 어미 사전에 등록할 필요가 없으므로, 어미 사전 크기도 줄고, 검색 대상이 한정되므로, 검색 속도도 증가한다.

무중성 용언의 분석

(예2) a. 예쁘면 => 예쁘 + 으면
b. 예쁘다 => 예쁘 + 는다
c. 예쁘지만 => 예쁘 + 지만

무중성 용언의 경우도 마찬가지로 먼저 용언 부분(‘예쁘’)을 찾는다. ‘예쁘’는 무중성 형용사 ‘예쁘다’의 원형이므로 결합 가능한 어미 종류는 다음과 같다.

- [‘예쁘’와 결합 가능한 어미 종류]
(1) ‘으’ 매개 어미 (‘으’ 탈락 형태)
(2) ‘는’ 매개 어미 (‘는’ 탈락 형태)
(3) 규칙 어미

(예2)의 a는 형용사 ‘예쁘다’와 ‘으’ 매개 어미가 결합한 경우이다. ‘예쁘다’가 무중성

이므로 ‘으’ 매개 어미에서 ‘으’가 탈락된 형태로 결합되고, 따라서, ‘으’를 복원하여 ‘으’ 매개 어미 사전을 검색하여야 한다.

(예2)의 b는 ‘는’ 매개 어미와 결합한 경우인데 ‘예쁘다’가 형용사이므로 ‘는’ 매개 어미에서 ‘는’이 탈락된 형태로 결합되고, 따라서, ‘는’을 복원하여 ‘는’ 매개 어미 사전을 검색한다. (예3)의 c는 규칙 어미와의 결합이므로 어미 복원없이 바로 규칙 어미 사전을 검색하면 된다.

시제 판별

아래의 예문은 새로운 어미 분류를 활용하였을 때 어미 ‘ㄴ’의 대표 어미가 ‘는’인지 ‘은’인지를 판별하여 줌으로써 용언의 시제를 판별할 수 있는데 도움이 됨을 보인다.

(예3) a. 나간 => 나가+은
b. 나간다 => 나가+는다

(예3)의 a, b 모두 첫번째 단계는 용언 부분 ‘나간’을 찾는다. ‘나간’은 무중성 용언 ‘나간다’에 ‘ㄴ’이 결합한 형태이므로 다음과 같은 어미 종류와 결합할 때만 가능하다.

- [‘나간’의 형태가 나올 수 있는 어미 종류]
(1) ‘으’ 매개 어미(‘은’으로 시작하는 어미)
(2) ‘는’ 매개 어미

‘나간’ 다음에 올 수 있는 어미는 위의 두가지 가능성이 있으므로, 각각에 대해 어미 사전을 검색해 본다. 먼저 (1)의 경우인지 보기 위해 ‘나간’의 형태로부터 ‘은’을 복원하여 ‘으’ 매개 어미 사전 중 ‘은’으로 시작하는 사전을 검색하면 되고, (2)의 경우는 ‘나간’으로부터 ‘는’을 복원하여 ‘는’ 매개 어미 사전을 검색하면 된다.

지금까지의 예에서 알 수 있듯이 새로운 어미 분류를 활용하여 용언을 분석하면 어미의 분류가 결정되고, 따라서 어미의 시제가 파악된다. 이는 구문 분석, 의미 분석 등에서 매우 중요한 정보로 활용될 수 있다.

5. 결론

한국어에서 용언과 어미들은 결합하는 과정에서 많은 형태 변이가 일어난다. 하지만, 용언과 어미는 모두 원형이 존재하며, 결합하는 과정에서 일정한 규칙으로 형태가 변

함을 알 수 있다.

본 논문에서는 어미의 원형을 대표 어미라 정의하고, 이 대표 어미를 대상으로 용언과 결합할 때의 형태 변이 특성을 기준으로 어미를 분류하였다. 본 논문에서 제안한 어미 분류를 용언구 분석에 활용하면 다음과 같은 여러 가지 장점을 얻을 수 있다.

첫째, 용언과 어미의 변환 형태들을 기준으로 어미를 분류하였기 때문에, 형태가 변화된 용언 부분만으로 뒤에 올 수 있는 어미 분류가 정해지기 때문에 검색할 어미 사전의 엔트리수가 훨씬 작아져 속도가 빨라지며, 잘못 입력된 예문(‘예쁘다’, ‘나가는다’, ...)을 오분석하는 것을 방지할 수 있어, 맞춤법 교정기 등의 응용 프로그램에서 유용하게 활용될 수 있다.

둘째, 대표 어미들로 어미 사전을 구성하고 변형 어미들은 사전에 등록할 필요가 없으므로, 메모리를 적게 사용하는 효과가 있다. 이는 PDA 등과 같이 작은 메모리를 요구하는 환경에서는 필수적인 것이다.

셋째, 어미의 원형을 분석 결과로 뉘뉘으로써 동일한 의미이면 하나의 어미 대표 어미 형태로 나오게 되어, 구문 분석이나 의미 분석 단계에서 대표 어미만을 기준으로 분석할 수 있기 때문에 의미의 모호성도 줄고, 효율적이다.

참고문헌

- [1] 강승식, “음절 정보와 복수어 단위 정보를 이용한 한국어 형태소 분석”, 서울대학교 컴퓨터공학과 박사학위 논문, 1993.
- [2] 권오욱, 정유진, 김미영, 류동원, 이문기, 이종혁, “음절단위 CYK 알고리즘에 기반한 형태소 분석기 및 품사태거”, 제 1 회 형태소 분석기 및 품사태거 평가 워크숍, 1999
- [3] 김진수, “국어 접속조사와 어미 연구”, 탑출판사, 1987
- [4] 남기심, 고영근 “표준 국어문법론”, 탑출판사, 1995
- [5] 서정수 “국어 문법”, 한양대학교 출판원, 1996
- [6] 송연정, 이근용, 이용석, “최장일치를 이용한 구문 분석용 형태소 분석기”, 제 1 회 형태소 분석기 및 품사태거 평가 워크숍, 1999
- [7] 심준혁, 김준석, 차정원, 이근배, “통계와 규칙을 이용한 강인한 품사 태거”, 제 1

회 형태소 분석기 및 품사태거 평가 워크숍, 1999

- [8] 안동연, “좌우접속정보를 이용한 명사추출기”, 제 1 회 형태소 분석기 및 품사태거 평가 워크숍, 1999
- [9] 이상태, “국어접속 어미 연구”, 계명대학교 국어국문학과 박사학위 논문, 1988
- [10] 이운재, 김선배, 김길연, 최기선, “모듈화된 형태소 분석기의 구현”, 제 1 회 형태소 분석기 및 품사태거 평가 워크숍, 1999
- [11] 이희자, 이종희 “(사전식) 텍스트 분석적 국어 어미의 연구”, 서울, 언어정보개발연구원, 한국문화사, 1999