

세종전자사전에서의 조사·어미 기술

김창섭 김진형
서울대학교 국어국문학과
(kcsro@snu.ac.kr, quark@hananet.net)

The Description of Korean particles and endings in the Sejong Electronic Dictionary

Chang-Seop Kim Jin-Hyeong Kim
Dept. of Korean language and Literature, Seoul National University

요약

조사어미사전은 한국어 연구 및 교육, 정보처리 분야에 두루 이용될 수 있는 범용적 전자사전을 지향하는 세종전자사전의 한 위성사전으로서, 한국어 조사와 어미에 대한 각종 언어 정보를 체계적이고 일관된 형식으로 표상하는 것을 목적으로 하고 있다. 그러한 목적을 달성하기 위해 본 연구 작업에서는 조사와 어미의 형태적 변이 양상과 조건을 상세히 밝히는 한편, 의미·통사적 특성과 제약에 관한 다양한 정보들을 가능한 한 풍부하게 제시하고 있다. 조사와 어미에 대하여 사전에 풍부하고 다양한 언어 정보를 표상하는 작업은 기존의 한국어 인쇄사전은 물론 전자사전에서도 본격적으로 시도되지 않았던 것으로, 본 사전에서 현재까지 기술하고 있는 다양한 정보들은 앞으로 한국어에 대한 순수 언어학적 연구만이 아니라 한국어 정보 처리 분야의 발전에 있어서도 기여하는 바가 적지 않을 것으로 기대한다.

1. 서언

본 연구는 정보 검색, 텍스트의 분석과 산출, 자동번역, 다국어 사전 구축, 인쇄 사전 구축 또는 한국어 교육이나 순수 언어학적 연구 등에 두루 활용될 수 있고, 더 나아가서는 향후의 진보된 인공지능 개발 환경에도 유연하게 적용될 수 있는 범용적인 전자사전의 구축을 목표로 한 21세기 세종계획 전자사전 구축 사업의 한 하위 과제이다. 조사어미사전은 21세기 세종계획 착수 연도인 1998년에는 계획 단계에 있다가 1999년부터 구축이 시작되었으며, 2001년 작업이 완료된다. 3개년에 걸친 조사어미 전자사전 구축 작업의 결과 얻어질 최종 결과물은 다음과 같다.

- 조사 및 어미 2,000 여 항목 기술
- 조사 및 어미 사전 기술 지침서 작성
- 조사어미사전 입력기 개발(통합사전분과로부터 지원 받음)

이 글에서는 올해로 3년째를 맞는 조사어미사전 구축 작업이 갖는 의의와 위상을 생각해 보고, 세종전자사전에서의 조사·어미 기술에 대하여 간단히 살펴보기로 한다.

2. 조사어미사전의 의의와 위상

일반적으로 조사는 문장에 참여하고 있는 명사구들의 격을 표시해 주거나(격조사) 담화에 관여하는 여러 대상들

사이의 의미·화용적 관계, 그 대상들에 대한 화자의 심리적 태도 등을 표시해 준다(보조사)고 할 수 있고, 어미는 서술어에 결합하여 문장 전체가 표현하고 있는 사건이나 상황의 시제나 상적 특성, 그리고 해당 사건, 상황에 대한 화자의 심리적 태도를 드러내는 한편, 문장과 문장의 논리적 연결을 담당하기도 하고 화자-청자 관계의 문법적 반영(경어법)이나 문장 유형의 결정에까지 관여하고 있다고 할 수 있다. 이렇게 조사와 어미는 한국어에서 통사, 의미, 화용의 층위에 걸쳐 매우 중요하고도 다양한 기능을 담당하고 있으므로, 이에 대한 체계적인 연구가 뒷받침되지 않고서는 한국어 문법 기술이나 이의 실제적 응용이 만족스러운 수준에 도달하기 어렵다.

한국어 정보 처리 수준을 점차 높여 가기 위해서는 지금까지 쌓여 온 언어학적 성과들이 폭넓게 활용될 필요가 있으나, 현재로서는 그러한 작업이 만족할 만하게 이루어지지 않고 있다. 특히 조사와 어미에 대한 것은 그 중요성에 비해 매우 제한된 정보만이 실제 언어 정보 처리 분야에 이용되고 있는 실정이다. 그것은 언어 정보의 전산 처리와 언어학적 연구 양쪽이 현 단계에서 가지고 있는 한계 때문이라고 할 수 있다. 따라서 언어학적 연구 성과를 전산적으로 이용할 수 있도록 체계적으로 정리하는 작업과 함께 공동 연구를 통해 실제적인 응용 범위를 확대해 갈 필요가 있다.

조사어미전자사전은 세종전자사전의 위성 사전으로서, 지금까지 조사와 어미에 대하여 이루어진 언어학적 연구 성과와 전산적 요구를 적절하게 조화시켜 기존의 인쇄사전이나 전자사전에 비해 풍부하고 체계적인 언어 정보를 일관성있는 형식으로 표상하는 것을 궁극적인 목적으로 하고 있다. 그러한 목적을 달성하기 위하여 본 연구에서는 기존 전자사전에서 담고 있던 기초적인 형태 정보뿐만 아니라 전자사전에서 표상 가능한 통사·의미 정보의 범위를 가능한 한 확장하려 하였다. 이에 따라 본 사전에서는 조사와 어미들의 다양한 형태적 변이 양상과 조건, 결합 양상, 문법적 기능과 의미, 선행어의 문법적 범주 및 선택 제약 등에 대한 정보들을 기술하고 있다. 이러한 시도는 한국어 연구 및 언어 정보 처리 분야가 지금까지 도달한 수준에서 한 걸음 더 나아가는 데 적극적으로 기여할 수 있으리라 믿는다.

3. 조사어미사전의 언어 정보 기술

3.1. 표제어의 선정과 분할

일차적인 사전의 기술 대상이 되는 항목은 우선 단일형인 조사와 어미이다. 그러나 표제어 선정에 있어서 둘 이상의 조사와 어미들로 이루어지는 수많은 공시적 결합체들의 등재 여부가 문제가 되었다. 언어학적인 고려만으로는 그들 결합체 중 대부분이 표제어로 올리기 어려운 것이지만, 전산적 처리를 고려하여 임시적인 조사끼리의 연속체, 어미끼리의 연속체, 어미와 조사의 연속체 등도 표제어로 삼았다.¹⁾ 이들을 표제화할 경우 기계에 의한 텍스트 분석에서 조사·어미가 복수로 결합한 어형을 만났을 때 그 각각을 분해하였다가 다시 조합하여 그 기능과 의미를 해석하는 식의 연산 과정을 단순화할 수 있다는 이점이 있다.

1) 본 사전에서는 조사끼리의 연속 통합체, 어미끼리의 연속 통합체, 어미와 조사의 연속 통합체 등을 '연속'이라는 이름으로 부르며, 품사 정보 역시 '연속'으로 제시한다. 어미의 경우 통사 구성이 축약되어 단일 어미와 같은 모습을 갖는 것도 표제화하기로 했는데, 앞의 경우와 차이는 있지만 단일 어미로 취급할 수 없으므로 '연속'이라는 이름으로 같이 취급한다. 이에 따라 다음의 항목들이 '연속'으로서 표제어가 될 수 있다.

- a. 조사끼리의 연속 (예) 예도, 예를, 예까지도...
- b. 선어말어미끼리의 연속 (예) 시었, 시었겠, 시었겠사옵...
- c. 어말어미와 어말어미의 연속 (예) 단다(다-고_하-ㄴ다)...
- d. 관례적으로 붙여쓰는 '보조적 연결어미+보조용언' 구성 (예) 어지('열어지다'의 '어_지')

한편 조사 '으로부터', 어미 '단다(환원불가능형)'와 같은 것은 '으로+부터, 다+ㄴ다'처럼 공시적 분석이 가능하기는 하지만 통사적으로 하나의 조사와 어미로 문법화된 것으로 본다. 따라서 이들을 '합성조사, 합성어미'로 부르고, 단일조사, 단일어미와 마찬가지로 취급한다.

그런데 이 경우 실제 텍스트에서 출현 가능한 모든 임시적 결합체 가운데 어떤 것들을 한정된 수의 표제어 목록에 포함시킬 것인가가 문제이다. 결국 그들 각각의 상대적인 빈도와 작업 참여자의 언어 직관, 기존 사전들의 표제어 목록 등에 의지하여 그 수를 한정할 수밖에 없다. 그리하여 기존 사전의 표제어 항목들을 검토하여 제한적으로 수용하는 한편 전산화된 텍스트에서 추출된 조사 어미의 목록을 참조하여 상세기술 대상 항목들을 선정하였다. 선정된 항목들은 변이형(이형식 포함)과 대표형의 관계를 포착하기 위해 동등하게 표제화하였고, 표준어형을 우선적으로 표제화하되 비표준어형을 제한적으로 수용하였다.

표제어의 선정과 관련한 또 하나의 어려운 문제는 선어말어미의 처리이다. 공시적으로 분석해 낼 수 있는 각각의 선어말어미는 모두가 표제어가 될 수 있어야 하지만 ‘-느/는/ㄴ-, -더-, -리-, -것-’ 등은 분리성이 약하고, ‘-(으)시-, -었-, -있었-, -겠-, -(으)옵-’ 등은 분리성이 강한 차이를 보인다.²⁾ 이에 따라 후자는 표제어로서의 지위를 인정할 수 있으나 전자는 그것이 곤란해진다. 본 사전에서는 후자를 따로 표제어로 세우지 않고, 그 뒤에 놓이는 어말어미와의 통합체 전체를 하나의 어말어미로 표제화하기로 하였다.

표제어의 분할 문제에 관하여는 다음의 원칙을 따르기로 하였다.

(1) 조사 표제어의 분할

- 가. 격조사와 접속조사 혹은 보조사는 별개 표제어로 분할한다.
- 나. 격조사 내에서도 격이 다르면 별개의 표제어로 분할한다.

(2) 어미 표제어의 분할

- 가. 어말어미가 연결 어미와 종결 어미의 두 용법을 가지면 별개 표제어로 분할한다.
- 나. 종결어미 안에서 문체법(평서, 감탄, 의문, 명령, 청유 등)에 따라서는 표제어를 분할하지 않는다.
- 다. 종결어미 안에서 직접문과 간접문 구성에 나뉘어 쓰이면 별개 표제어로 분할한다.
- 라. 종결어미 안에서 상대경어법(화제)에 차이를 보일 때는 별개 표제어로 분할한다.
- 마. 대등적 연결 어미, 종속적 연결 어미, 보조적 연결 어미는 각각 별개 표제어로 분할한다.

일반적으로 인쇄사전의 경우에는 국립국어연구원(1999)에서의 처리와 같이 조사나 어미의 하위 분류에 따라서는 별개의 표제어로 분할하지 않는다. 그러나 본 사전에서는 한 표제어 아래 표상되는 언어 정보가 위계적으로 깊어지지 않도록 하기 위하여 가능한 한 다의의 수를 줄이고 동형어의 수를 늘리는 쪽을 택하였다.

3.2. 의미 및 통사 정보의 기술

세종전자사전의 통합구조에서는 통사 정보 그룹(<synGrp>)과 의미 정보 그룹(<semGrp>)이 나뉘어 있다. 그러나 조사나 어미의 경우 그 어휘적 의미가 미약하거나 불분명하기 때문에 의미 정보 그룹 아래 표제어의 의미 구분, 선행어의 범주 및 의미 선택 제약, 문법적 제약, 용례 등이 제시된다.

조사와 어미의 의미 구분이나 문법적 특성 기술과 관련하여 발생하는 가장 큰 문제는 이들이 어휘적 요소가 아닌 문법적 요소라는 데에서 기인한다. 이들의 관계적 의미는 단독으로 파악될 수 없으며, 관련 체언 및 용언, 전후 문맥 등에 매우 밀접하게 관련되어 있다. 그러나 본 사전에서는 조사와 어미만이 기술 대상이므로 이들과 관련되는 체언이나 용언의 의미나 통사에 관한 정보는 제한될 필요가 있다. 가령, 조사에서 의미역 정보를 제공할 것인가 하는 문제에 대하여, 의미역은 선행 명사구와 서술어와의 의미적 관계에 의해 결정되는 것이지 특정 조사와 필연적인 관계에 있지는 않다고 간주하여 의미역 정보를 따로 표상하지 않는다. 그러나 조사의 의미를 구분하여 제시할 때에는 그 자체의 어휘적 의미에 국한하지 않고 선행 명사구와 서술어의 의미 관계에 따라 드러나는 의미들을 포착하여

2) 고영근(1993)의 용어에 따르면 전자는 문장 구성소, 후자는 문장 형성소에 해당한다.

기술하도록 하였다.

3.2.1. 의미의 구분과 정의

순수하게 기계를 위한 사전용을 구축한다면 의미의 구분이나 정의는 당장 필요한 정보가 아니다. 그러나 본 사전의 다양한 활용과 미래에의 응용을 위해서는 일반적인 인쇄사전 이상의 다양한 의미 구분과 정의가 이루어질 필요가 있다. 그러나 조사, 어미의 문법적 기능에 따른 의미 구분은 비교적 분명한 데 반해 전후 문맥이나 상황에 따라 드러나는 조사나 어미에 얽혀 나타나는 의미들의 구분은 모호한 경우가 많다. 가능한 한 객관적인 기준에 의해 의미를 구분하고, 명시적인 방식을 통해 의미를 정의할 필요가 있다. 다음은 조사, 어미 의미 구분의 예이다.

(3) '(으)로'의 경우(명사구와 서술어의 의미역 관계에 따른 의미 구분)

- 가. 그는 오늘 부산으로 갔다. (방향·지향점)
- 나. 계곡의 물이 바위 틈으로 흐른다. (경로)
- 다. 물이 얼음으로 되었다. (도착·접·방향)
- 라. 이 시간 이후로 나는 서울에 없을 것이다. (출발점)
- 마. 목수가 고목을 톱으로 자른다. (도구·수단)
- 바. 그는 나무로 책상을 만들었다. (재료)
- 사. 그는 잔걸음으로 걸었다. 가을엔 아침 저녁으로 기온차가 크다. (방식·양상)
- 아. 그는 주장으로 경기에 나섰다. 그녀는 그를 한 가족으로 여겼다. (자격·속성)
- 자. 여러 지방이 지진으로 큰 피해를 입었다. (원인·이유)

(4) '에서'의 경우(선행 명사구의 의미 특성에 따른 의미 구분)

- 가. 오늘 고급 식당에서 근사한 식사를 했다. (구체적 장소)
- 나. 이런 상황에서 무엇을 할 수 있는가? 경마에서 재미 좀 봤다며? (비유적 장소)

(5) '는'의 경우(보조사는 의미(화용) 기능에 따라 의미 구분을 함)

- 가. 서울은 우리 나라의 수도이다. (화제)
- 나. 철수는 오지 말고, 영희는 와라. (대조)
- 다. 흥, 그렇게 쉽게는 날 이길 수 없을걸? (강조)

(6) '는데'의 경우(문체법 상의 차이에 따른 의미 구분)

- 가. 내가 보기엔 이 옷이 더 예쁘는데. (반말, 평서형)
- 나. 네가 보기엔 어느 옷이 더 예쁘는데? (반말, 의문형)

(7) '더러'의 경우(형태·통사론적 제약의 차이에 의한 의미 구분)

- 가. 그 일이 끝나니 나는 배가 몹시 고프더러. (주어=1인칭)
- 나. 선생님께서 무척 네 칭찬을 하시더러. (주어=2인칭)

(8) '게'의 경우(연결어미의 경우 선행행절의 의미 관계에 따른 의미 구분)

- 가. 할아버지 쉬시게 다들 조용히 해라. (목적)
- 나. 아기가 착하고 건강하게 자라기를 바란다. (결과)
- 다. 일을 신속하게 마무리지어야 한다. (방식)

3.2.2. 통사 정보 기술

조사와 어미의 문법적 행동에 대한 연구가 좀더 진전된다면 사전에서 체계적으로 표상할 수 있는 통사 정보의 내용이 더욱 풍부해질 수 있을 것이다. 그러나 현재로서는 선행어와의 결합 유형, 선택 제약에 관한 정보, 조사와 어미의 다중 결합 시에 앞뒤에 놓일 수 있는 조사와 어미에 관한 정보, 그 밖의 제한적인 문법 제약 정도가 기술되고 있다.

3.2.2.1 선행 요소와의 결합 관계 유형 정보 <frame>

조사 및 어미가 어떤 범주의 선행 요소와 결합하는가를 밝혀 주는 정보항으로, 출현 가능한 결합 유형을 간략화된 범주 기호를 사용하여 제시한다.³⁾

(9) 가. 표제어가 보조사 '가'일 때: <frame>N-KIAdv-KIEf-K</frame>

나. 표제어가 종결어미 'ㄴ다'일 때: <frame>Vv(시)-E</frame>

3.2.2.2 선행 요소 선택제약 <selRst>

조사나 어미의 선행 요소가 특정한 부류로 제약될 때 그 선택 제약을 밝혀 주는 정보항으로, 그 예는 다음과 같다.

(10) 표제어가 부사격 조사 '에게'인 경우 : <selRst>N=유정</selRst>

3.2.2.3 선행 조사·어미 <frn>와 후행 조사·어미 <fol>

이 정보항은 조사, 어미 표제어에 직접 선행하거나 후행하는 조사, 어미 항목을 밝혀 주기 위해 마련되었다. 한국어의 조사와 어미는 선행어에 다중적으로 결합하는 경우가 많은데, 현재의 구조 분석론에 따르면 조사와 조사끼리, 혹은 어미와 어미끼리 연속 결합한 것이 아니라 선행어에 한 항목이 결합해 만들어진 구성체에 다시 다른 하나가 결합하였다는 식으로 이해한다. 그러나 전통적으로 조사와 어미의 다중 결합 순서나 그에 작용하는 제약 등에 학계의 관심이 쏠려 왔고, 전산 처리의 측면에서도 유용성을 지닐 수 있으리라는 기대 아래 표제어 앞뒤에 놓일 수 있는 항목들을 나열해 주도록 하였다.

(11) 가. 표제어가 '으로'일 때: <frn>만|만큼</frn>

<fol>가|까지|나|나머는|도|든|든지|라도|라야|만|밖에|야|와|요|의|니</fol>

나. 표제어가 '겠'일 때: <frn>으|시|았|았|었|었|습</frn>

<fol>다|더|라|지|습|습|니다|고|으니|으면...</fol>

3.2.2.4 동절 내의 문법적 제약 <ownS>과 후행절에 걸리는 문법적 제약 <folS>

이 정보항은 주로 어미 표제어가 자신이 위치한 절 내부나 후행절에 대하여 가지고 있는 통사적 제약을 표상하기 위해 마련되었다.

(12) 가. '관테'는 동절 내에 의문사가 와야 한다. <ownS>의문사</ownS> (누가 왔관테 ~)

나. '러'는 동절 내에 人性 주어와 와야 한다. <ownS>인성 주어</ownS> (우리 고기 잡으러 ~)

(13) '관테'는 후행절이 의문문이어야 한다. <folS>의문문</folS>

그러나 이 정보항의 기술 내용과 방식은 앞으로 많은 보완이 필요하다. 현재 이 정보항에 기술되는 내용은 기계를 이용하여 유의미한 정보를 뽑아 내기는 어려운 방식으로 제시되어 있다. 그것은 작업 참여자가 이 정보항에 기술할 내용을 체계화하는 데 충분한 시간을 투자하지 못한 탓이기도 하지만 모든 조사와 어미에 대한 이런 식의 문법적 제약이 아직 체계적으로 기술되어 있지 못한 탓이기도 하다. 이 정보항은 조사, 어미의 기술이 끝난 뒤 다시 검

3) 이때 선행 요소는 조사와 어미에 직접 결합하고 있는 단어 차원의 요소를 말한다. 조사와 어미는 둘 다 선행 요소로서 통사 구성을 취하지만 여기서는 그 통사 구성의 맨 오른쪽 요소만을 결합 관계식에 포함시켜 나타낸다.

토하여 일관되고 체계적인 표상 방식을 갖추도록 보완할 필요가 있다.

3.3 형태 정보의 기술

조사와 어미의 형태와 관련된 정보는 형태 정보 그룹에 기술된다. 이 정보 그룹에서는 표제어의 품사 및 하위 품사 분류, 이형태와 변이형 등의 정보들이 제시된다.

3.3.1. 품사(part-of-speech) 정보 <pos> 및 하위 품사 분류 <subpos>

본 사전에서는 기초 품사 단위를 조사, 어미, 연속으로 나누고 그 각각을 <pos>항에 J, E, ser로 제시한다. 그리고 조사, 어미 각각의 하위 품사 분류 체계는 임홍빈·송철의(1998)의 것을 따라 다음 표의 맨 오른쪽 약호를 <subpos>항에 제시한다.

기본품사	확장 품사		약호	
	저단위	고단위		
조사 j	격조사 jk	주격 조사	jks	
		보격 조사	jkc	
		관형격 조사	jkg	
		목적격 조사	jko	
		부사격 조사	jkb	
		호격 조사	jkv	
		인용격 조사	jkq	
	보조사		jx	
접속 조사		jc		
어미 e	선어말 어미		ep	
	어말 어미 ef	종결어미		eft
		연결 어미 efc	대등적 연결 어미	efcd
			종속적 연결 어미	efcs
			보조적 연결 어미	efcx
		전성 어미	명사형 전성 어미	efn
	관형사형 전성 어미		efd	
연속		ser		

3.3.2. 내부 구성 정보(constituents) <cnst>

한 표제어가 둘 이상의 표제어 요소의 결합으로 이루어졌을 경우 그 내부 구조를 밝혀 준다. 이 정보항은 단일 조사나 어미가 표제어일 경우보다는 '연속'형이 표제어일 경우에 주로 이용된다. 그러나 통시적으로 둘 이상의 요소가 하나로 굳어져 생겨난 항목일 경우, 공식적으로 분석이 가능한 선까지 이 정보항에 내부 구성을 밝혀 줄 수 있다.

(14) 어도: <cnst>어+도</cnst>

3.3.4. 이형태 및 변이형 <var>

조사와 어미는 선행 요소의 음운론적 환경이나 형태론적 환경에 따라 이형태를 갖는 경우가 많다. 이 정보항에는 이러한 이형태뿐 아니라 표제어의 다양한 변이형 관계까지 포착하여 기술하고 있다. 단, 이때의 변이형 개념은 표제어와 음운론적으로나 형태론적으로 엄밀하게 조건지어진 것을 말하는 것이 아니라 좀더 느슨하게 확장된 것이다. 아래에 본 사전에서 표상하고 있는 이형태 및 변이형의 종류와 예들을 제시한다.

(15) 가. 표제어의 음운론적 이형태

는: <var type=phn>은</var> 았: <var type=phn>있|쓰</var>

나. 표제어의 형태론적 이형태

았: <var type=mor>였1</var>

다. 선행어간의 품사에 따른 변이형

ㄴ다(간접문의 종결어미): <var type=stem>다|라</var>(쓴다고(동) : 착하다고(형) : 학생이라고(지))

라. 경어법적 변이형

에게: <var type=hon>께</var> 께: <var type=hon>에게</var>

마. 준말과 본디말

는(“철수는”): <var type=abb>ㄴ</var> ㄴ(“철순”): <var type=lng>는</var>

바. 강조형과 비강조형

라(“서라”): <var type=em_p>랏</var> 랏(“서랏”): <var type=em_n>라</var>

사. 구어형과 비구어형

느냐(“오느냐?”): <var type=spk_p>니</var> 니(“오니?”) : <var type=spk_n>느냐</var>

아. 표준어형과 비표준어형

기에: <var type=std_n>길래</var> 길래: <var type=std_p>기에</var>

자. 표제어의 기타 변이형

기에(당신이 누구기에”): <var type=etc>관데</var>

관데(당신이 누구관데”): <var type=etc>기에</var>

3.4. 기타 정보

의미 정보 그룹과 형태 정보 그룹 외에도 관용표현 정보 그룹(<idmGrp>)과 표제어 정보 그룹(<headGrp>), 관리 정보 그룹(<mntGrp>)을 두고 있다. 관용표현 정보 그룹에는 표제어가 참여하여 만들어지는 관용표현의 형식과 용례를 제시하도록 되어 있는데, 일반적인 관용표현은 어휘적인 요소들이 중심이 되기 때문에 이 정보 그룹에 기술할 만한 것은 거의 없다. 다만 2001년 들어 ‘-에 대하여, -에 관하여, -으로 인하여, -을 통하여...’ 등의 항목들을 조사로 처리할 수 있는 가능성이 제기되었고, 이를 조사어미사전에 수용하기로 하였는데, 이들을 관용표현의 하나로 간주할 수 있을 듯하다.

4. 조사어미사전의 한계 : 결론

본 사전이 기존의 인쇄사전이나 전자사전의 조사·어미 기술보다 한 단계 진보한 수준의 언어 정보 표상을 궁극적인 도달점으로 삼고 있으나 한국어의 조사와 어미와 관련된 언어 현상에는 학계의 연구가 아직 미진한 부분이 남아 있거나, 연구자의 관점에 따라 달리 해석될 수 있는 여지가 많다. 따라서 3년이라는 연구 기간 동안 언어학적으로 완벽한 사전 정보 표상에 도달하기 어려운 것이 사실이다. 따라서 본 사전은 기존의 연구 성과를 성급하게 뛰어넘으려 하기보다는 일차적으로 이미 이루어진 언어학적 연구 성과들을 재정리하여, 전산적 응용을 염두에 두고 체계화하는 것으로 일차적인 목표를 삼았다.

또한 본 사전은 기계가독형의 전자사전을 지향하고 있으므로 때로는 언어학적 기술의 엄밀성에 대한 요구와 전산 처리의 효율성에 대한 요구가 충돌을 일으키기도 하였다. 이 경우 세종전자사전의 전체 체계와 아울러 전자사전으로서의 특성을 우선적으로 고려할 수밖에 없었다. 그리고 아직까지는 조사어미사전에 표상되어 있는 언어학적 정

보들 전체가 전산적 처리에 곧바로 이용될 수 있는지는 미지수이다. 그 가운데에는 현재의 전산 처리 수준으로는 이용이 어렵지만 다음 단계에서 이용할 수 있으리라 예상하여 표상된 것도 있고, 현재까지의 사전적 기술 수준으로는 전산적 처리에 직접 활용할 만큼 적절한 체계성을 갖추기 어려워 추후 좀더 일관성있는 형식으로서의 보완이 필요한 것도 있다. 본 사전의 정보들이 다양하게 응용될 수 있으려면 기술된 정보의 검증과 보완뿐만 아니라 그것의 응용 방법론 역시 지속적으로 연구되어야 할 것이다.

참고문헌

- [1]고영근(1993) 우리말의 총체서술과 문법체계, 일지사.
- [2]국립국어연구원(1999) 표준국어대사전, 두산동아.
- [3]국립국어연구원(2000) <표준국어대사전> 편찬 지침 I · II.
- [4]김홍규 · 강범모 (2000) 한국어 형태소 및 어휘 사용 빈도의 분석1, 고려대학교 민족문화연구원.
- [5]연세대학교 언어정보개발연구원(1998), 연세한국어사전, 두산동아.
- [6]이희자 · 이종희 (1998) 텍스트 분석적 국어 조사의 연구, 한국문화사.
- [7]이희자 · 이종희 (1999) 텍스트 분석적 국어 어미의 연구, 한국문화사.
- [8]임홍빈 · 송철의(1998) 한국어 정보 처리를 위한 어절 분석 표지의 표준화 연구, 21세기 세종계획 기초언어자료기반 분과 세부연구과제 보고서.
- [9]채영숙 · 김재원 · 김민정 · 김지영 (1991) 한국어 조사 유형 사례집, PNUAI TR91-1.