

---

# 세종 전자사전 개발의 성과와 전망

이성현 · 서울대학교 불어불문학과 교수

---

## 1. 머리말

세종 전자사전은 문화체육관광부와 국립국어원이 순차적으로 주관한 21세기 세종 계획의 일환으로 추진된 현대 한국어 전자사전 개발 사업의 결과물이다. 각종 언어 정보를 처리하는 소프트웨어의 개발과 실행에 전자사전이 필수불가결한 핵심 부분이라는 인식에서 출발한 이 사업은 한국어 관련 언어 정보의 전산 처리에 필수적이면서도 보편적으로 활용될 수 있는 기반 전자사전의 구축을 목표로 1998년부터 2007년까지 10년 동안 추진되었다. 그 결과물인 세종 전자사전은 현대 한국어 어휘의 체계적인 분석·기술을 바탕으로 다양한 용도의 전산 처리에 실질적으로 활용될 수 있고 그 기술의 획기적 발전에도 공헌할 수 있는 대규모 범용 한국어 전자사전을 지향한다. 그런 까닭에 세종 전자사전 개발 사업은 한국어 어휘에 대한 사전적 기술 과정 외에도 언어학, 한국어학, 사전학, 전산학 등 모든 관련 분야의 기존 연구 성과에 대해 세밀하면서도 총체적인 비판적 검토, 이를 바탕으로 한 새로운 언어 분석 방법론 및 사전 구축 방법론 개발, 전자사전 구축과 검증을 위한 도구 개발, 전자사전 활용 방안 개발

등 다양한 과정들을 포괄하고, 이 과정들이 유기적으로 연계되도록 추진되었다.

다음에서는 이러한 배경과 과정을 통해 추진된 세종 전자사전 개발 사업의 성과와 의의를 정리하고 이와 함께 향후 과제 및 활용 방안 등에 대해 전망하도록 한다. 이를 위해 우선 세종 전자사전 개발 사업을 배경 및 목표, 기간별 사업 내용 등으로 나누어 개관하고, 그 성과를 세종 전자사전의 특성-기본 성격, 구성과 규모, 거시구조 및 미시구조, 전산어휘부로서의 기능- 과 관련하여 살필 것이다. 그런 다음, 세종 전자사전 개발의 의의를 정리하고 이와 함께 향후 과제에 대해서도 전망하도록 한다.

## 2. 세종 전자사전 개발 사업 개요

### 2.1. 사업의 배경 및 목표

세종 전자사전 개발 사업은 다음과 같은 배경에서 추진되었다. 정보화 시대를 맞아 정보 처리 산업에서 언어 전산 처리의 중요성이 크게 부각되었고, 그 결과 다양한 목적의 한국어 전산 처리에 활용될 수 있는 범용적인 한국어 언어 정보 데이터베이스 구축의 필요성이 강하게 제기되었다. 사실, 당시에는 한국어 자동 처리에 필요한 대규모 전자사전이 부재했고, 이로 인해 언어 처리 기술 개발과 이의 산업화, 상용화가 지체되고 있었다. 이러한 상황 인식이 공유, 확산됨에 따라 이 문제를 해결할 범용적인 한국어 언어 정보 데이터베이스 구축 필요성에 대한 인식도 그만큼 증대되고 있었다. 세종 전자사전 개발은 바로 이러한 요구에 부응하기 위해 추진되었는데, 그 목표는 다음과 같은 성격을 모두 갖는 전자사전의 구축이었다.

- ㉠ 한국어 자동 처리에 보편적으로 사용될 수 있는 전자사전

- ㉠ 다양한 전산 처리에 필수적이고 실용적인 전자사전
- ㉡ 현대 한국어 어휘 전반에 대한 종합적이고 방대한 정보를 담은 전자사전

## 2.2. 단계별 사업 내용

세종 전자사전 개발 사업은 총 10년(1998~2007) 동안 다음과 같이 4 단계로 나뉘어 추진되었다.

- ㉠ 1단계 사업: 1998년 - 2000년
- ㉡ 2단계 사업: 2001년 - 2003년
- ㉢ 3단계 사업: 2004년 - 2006년
- ㉣ 최종 단계 사업: 2007년

1단계 사업은 기초 연구 단계로서 한국어 전산 처리에 필요한 어휘 정보의 수집과 분석, 세종 전자사전의 모형 구축, 그리고 명사나 동사 등 주요 범주의 어휘들에 대한 기술 작업이 중점적으로 수행되었다.

2단계 사업은 다양한 분야에서 실제 활용될 응용 전자사전 개발을 위한 중규모 사전의 구축에 중점을 두었다. 체언사전과 용언사전의 어휘 항목 수를 대폭 확장하고, 한국어 전산 처리에 즉각 요구되는 필수 정보들에 집중하여 사전을 구축하는 한편, 조사, 어미, 관형사, 부사 등 어휘 규모는 상대적으로 작지만 전산 처리에 필수적인 범주들의 사전 구축을 완료하였다. 아울러 세종 전자사전의 실용성을 극대화하기 위한 방안들이 다각도로 연구되었고, 그 검증과 평가를 위한 도구들이 개발되었다.

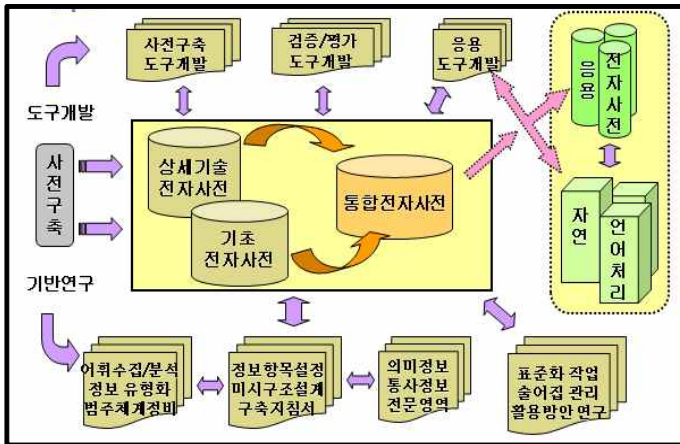
3단계 사업에서는 연어 사전, 관용 표현 사전, 특수어 사전, 복합 명사구 사전 등 2단계 사업 중 세종 전자사전의 실용성 제고를 위해 개발을 시작한 사전들의 구축을 완료하는 한편, 세종 전자사전의 핵심을 이루는

체인사전과 용언사전의 구축을 완료하였다. 이 기간 동안의 체인사전 및 용언사전 구축 작업은 신규 항목들의 기술 작업과 중규모 사전에 수록된 어휘 항목들의 정보를 확장하여 당초 미시구조에 따라 기술을 완료하는 정보 확장 작업이 병행 추진되었다. 이와 함께 이미 구축된 하위사전들에 대한 총체적인 정보 확장 작업과 보완·정제 작업도 수행되었다.

최종 단계 사업은 3단계 사업에서 실질적인 구축이 완료된 체인사전 및 용언사전에 대한 보완·정제 작업과 세종 전자사전의 하위 사전들 간의 통일성 및 연계성을 제고하는 작업에 중점이 두어졌다. 아울러 세종 전자사전의 검증과 활용을 위한 다각적인 연구와 도구 개발도 수행되었다.

### 2.3. 세종 전자사전 개발 사업의 내용 및 개발 과정

세종 전자사전 개발 사업은 크게 세 분야로 나눌 수 있는데, 전자사전 구축 분야와 이를 위한 기반 연구 분야, 그리고 전자사전의 구축, 검증 및 평가, 활용을 위한 도구 개발 분야가 그것이다. 세종 전자사전은 이 세 분야의 작업들이 아래에서처럼 유기적인 관계를 맺으며 수행된 결과이다.



〈표1: 세종 전자사전 개발 과정〉

### 3. 세종 전자사전의 특성

#### 3.1. 기본 성격

세종 전자사전의 특성은 다음 네 가지 기본 성격으로 요약될 수 있다.

- ㉠ 현대 한국어 사용자가 구사하는 어휘 지식을 총체적으로 표상한 사전
- ㉡ 범용적 대규모 현대 한국어 전산 어휘부
- ㉢ 1990년~2000년대 어휘에 대한 이론적·기술적 연구 성과를 통합한 이론 절충적 사전
- ㉣ 인문학과 첨단 공학기술이 융합된 학제적·통합적 연구의 결과물

세종 전자사전은 현대 한국어 성인 사용자가 한국어로 된 텍스트의 산출과 이해에 필요한 어휘 지식과 그 활용에 관한 지식을 총체적으로 담고 있는 현대 한국어 어휘 사전이다. 세종 전자사전에 기술된 어휘(군) 정보들은 제한된 수의 분석틀에 맞춘 기계적인 기술은 지양하는 대신 엄밀한 언어학적 분석에 근거하여 한국어의 작용 원리를 상세히 반영하고 있다. 이는 말뭉치, 기간 인쇄 사전 및 어휘 연구 자료 등 각종 풍부한 어휘 자료를 최대한 활용하되, 최신의 언어학과 언어 공학의 연구 성과들을 적극 수용하여 기존 자료들을 철저히 재분석하여 정밀하게 어휘 정보를 기술한 결과이다.

또한, 특정한 유형이나 영역의 기계 처리 작업에 국한되지 않고, 정보 검색, 텍스트의 분석과 산출, 자동번역 등 다양한 자연어 처리 분야에서 사용될 수 있음은 물론이고 다국어 사전 구축, 인쇄 사전 구축 또는 한국어에 대한 순수 언어학적 연구나 한국어 교육 등에 두루 활용되며, 더 나아가서는 향후의 진보된 인공지능 개발 환경에도 유연하게 적용될 수 있

다는 점에서 범용적 성격을 갖는다. 그런 까닭에 표제어의 수나 표제어별 정보 유형의 다양성과 정밀성 등 규모 면에서 대규모 사전이다.

한편, 내용 면에서는 1990년부터 사업완료 시점인 2007년까지 이루어진 어휘에 대한 이론적·기술적 연구 성과가 통합적으로 반영되고 여기에 첨단 컴퓨터 공학 기술이 융합된 학제적 연구의 결과물이라는 성격을 갖는다. 세종 전자사전은 사전 정보의 확장성을 극대화하고 특수 목적 전자사전 구축의 기반을 제공할 목적으로 어휘부 중심의 언어기술 모형을 적극 활용하였다. M. Gross의 어휘-문법 이론(Lexicon-Grammar), I. Mel'čuk의 의미-텍스트 이론(MTT), G. Gross의 대상부류 이론(Object Classes)이 적극 수용되었고, 문형 및 논항의 구조와 의미역 기술을 위해 생성문법이, 어휘 의미 관계의 기술을 위해서는 Cruse의 어휘 의미 관련 이론이 적극 수용되었다. 이와 함께 유형론적 관점에서의 한국어 어휘 부류의 위상과 특성을 적극적으로 고려한 이론 절충적인 사전 정보를 수록하고 있다.

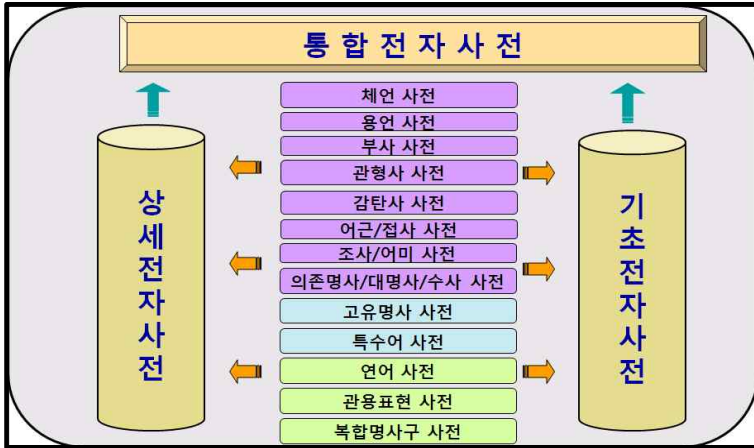
세종 전자사전의 이러한 특성을 구성과 규모, 거시구조와 미시구조, 그리고 기능 면으로 나누어 살펴보면 다음과 같다.

## 3.2. 구성과 규모

### 3.2.1. 중층적 복합 구성

세종 전자사전이 구성 면에서 갖는 특성은 한국어 전산 처리의 활용성과 효용성의 극대화를 위해 하위 모듈들이 통합된 형태로 구성되었다는 점이다. 즉, 세종 전자사전은 정보 항목의 유형 및 수, 그리고 표제어의 범주에 따라 구별되는 하위 사전들이 중층적·복합적으로 통합된 구성을 갖는다. 세종 전자사전은 우선 기본적인 문법 범주와 형태 정보만이 부착된 대규모 기초 전자사전과 상세한 어휘 정보가 부착된 상세 전자사전으로 구분되고, 이 두 사전은 각각 다시 표제어(항)의 문법 범주(18개 범주)

에 따라 13개의 하위 사전들로 나뉜다. 이렇게 별도로 나뉘어 구축된 하위 모듈들이 최종적으로 통합된 통합 전자사전이 세종 전자사전이 된다.



〈표2: 세종 전자사전의 구성〉

### 3.2.2. 대규모 어휘 수록

세종 전자사전은 상기한 바와 같이 중층적 구조를 갖는데, 이러한 위계적 구성을 통해 기술된 사전 항목의 규모는 총 60만 항목을 넘어선다. 다음은 하위 사전별 기술 항목의 규모이다.

구분	상세 전자사전	기초 전자사전	구분	상세 전자사전	기초 전자사전
체언	35,854	136,000	조사	559	563
용언	29,492	50,003	어미	1,507	1,554
의존명사	150	145	어근	2,511	1,607
대명사	200	191	접사	142	142
수사	182	182	연어	9,000	9,024
분류사	760	734	관용표현	5,052	5,055
부사	4,527	11,755	복합명사구	5,802	15,369
관형사	370	370	고유명사	24,278	168,123
감탄사	349	349	특수어	35,131	60,000
<b>통합 전자사전</b>			소 계	<b>155,866</b>	<b>461,166</b>
			총 계	<b>617,032</b>	

〈표3: 세종 전자사전의 기술 항목 규모〉

### 3.3. 거시구조

거시구조 측면에서 볼 때 세종 전자사전은 다양한 층위의 표제항을 수록하고 이에 대한 정밀한 정보를 제공하고 있는 것이 특징이다.

#### 3.3.1. 다단어 표현 사전 구축

세종 전자사전에서는 문장을 인식하거나 생성할 때 사전의 활용성을 극대화하기 위하여 연어 사전, 관용 표현 사전, 복합 명사구 사전 등 구절 단위 이상의 다단어 표현(multiword expression)을 표제항으로 하는 사전들을 별도로 구성하여, 단어나 형태소 층위의 표제항에서와 마찬가지로 형태, 의미, 통사 등 다양한 정보를 정밀하게 기술하였다.

#### 3.3.2. 부사 사전 구축

세종 전자사전에서는 또한 부사를 표제항으로 하는 부사 사전을 별도로 구축하여 어휘 의미 관계 정보 및 결합 정보, 결합 어휘와의 호응 관계 등에 대한 상세하고 풍부한 정보를 수록하였다. 이는 자연어 처리 과정에서 중의성 해소나 문장 생성의 정확도를 높이기 위함이었다.

#### 3.3.3. 특수어 사전 구축

세종 전자사전에서는 특수어 사전을 별도로 구축하여 외래어/외국어, 약어, 숫자 표기, 시사어, 고빈도 전문어 등에 대한 정보를 상세히 기술한다. 예컨대, ‘4·19혁명’, ‘사일구’, ‘4월 혁명’ 등 숫자를 포함하는 표현이나 ‘노사모’, ‘노찾사’, ‘전노협’, ‘민노총’ 등의 약어 표현 등이 특수어 사전에 수록, 기술된다. 이는 미등록어 처리 시의 효율성을 높일 목적에서였다.



### 3.4. 미시구조

세종 전자사전의 미시구조 설계에는, 거시구조의 경우와 마찬가지로, 한국어 전산 처리에의 활용을 목표로 하는 전산 어휘부로서의 기능이 최우선적으로 고려되었다. 즉 각 하위사전이 기술 대상으로 삼는 문법 범주에 속하는 어휘(군)들의 형태, 통사, 의미, 화용 속성과 전문어 영역에서의 쓰임들을 면밀히 검토하여 사전에 수록해야 할 정보의 유형을 분석한 후, 이를 토대로 정보 항목과 기술 내용이 설정되었다.

이렇게 설정된 정보 항목의 수는 하위 사전별로 16~53개에 이르는데, 이들 정보 항목은 표상하는 정보의 내용에 따라 유형별로 묶이고 동일 유형에 속하는 정보 항목들이 정보 구획을 구성하게끔 설계되었다. 그 결과 세종 전자사전의 어휘 항목은 세부 정보 항목들을 포함하는 여러 하위 구획으로 구성되는 위계적인 모듈의 형태를 갖는다. 즉 다양한 정보 항목이 체계적으로 위계화·조직화되어 있는 것이 세종 전자사전 미시구조의 가장 큰 특징이다.

#### 3.4.1. 상세 전자사전의 미시구조

예컨대 체언 상세 전자사전의 미시구조는 통사·의미 속성이 다른 술어명사와 논항명사(비술어 명사)의 구별 기술이 가능하도록 설계되었다. 술어명사 관련 정보 항목으로는 문형(<<frame>>), 명사구 최대 구조(<<max\_n\_str>>), 선택제약(<<sel\_rst>>), 기능동사(<<n\_v type="npred\_\_vsup">>) 등이 설정되었고, 비술어 명사 관련 정보 항목으로는 관형절 제약(<<s\_n>>), 단위표현(<<cl>>) 등이 설정되었다. 이 외에도 명사 범주의 특성을 반영한 조사 결합 제약 정보(<<prt>>)나 부사적 용법 가능 여부에 관한 정보(<<av>>), 그리고 명사를 포함한 연어 및 관용 표현 존재 여부에 관한 정보(<<idm\_grp>>) 등이 기술될 수 있는 다양하고 상세한 정보 항목들이 있다.

<superEntry>	최상위 표제형 구획	<domain>	전문 영역
<orth>	표기 형태	<sem_class>	의미 부류
<entry>	표제형 구획	<lr>	어휘 의미 관계 구획
<mnt_grp>	관리 정보 구획	<syn>	동의어
<see>	참조 표제형	<ant>	반의어
<morph_grp>	형태 정보 구획	<syn_grp>	통사 정보 구획
<var>	변이형	<comb_aj>	형용사 결합 정보
<str>	내부 구조	<comb_n>	명사 결합 정보
<org>	원어	<comb_v>	동사 결합 정보 구획
<hom>	동형어	<form>	동사 결합 형태
<der>	파생어 형성	<frame>	문형
<comp>	합성어 형성	<max_n_str>	명사구 최대 구조
<idm_grp>	숙어 구획	<sel_rst>	선택 제약
<idm>	숙어	<cl>	단위 표현
<sense>	센스 구획	<prt>	조사 제약
<sem_grp>	의미 정보 구획	<av>	부사적 용법
<eg>	용례	<s_n>	관형절 제약
<trans>	영어 대역어		

〈표4: 체언 상세 전자사전의 미시구조〉

또한, 이들 정보 항목들은 위의 표에서 볼 수 있듯이 정보의 유형별로 정보 구획을 구성하여 체계적으로 표상되고 있다. 이는 다른 하위 사전들에서도 마찬가지이다. 이렇게 동일한 방법으로 미시구조가 설계된 상세 전자사전의 범주별 하위 사전의 정보 항목 수는 다음과 같다.

범주별 하위사전	어휘정보항목 수	범주별 하위사전	어휘정보항목 수
체언 사전	35	어근/접사 사전	19
용언 사전	53	연어 사전	42
의·대·수 사전	41	관용표현 사전	44
부사 사전	39	고유명사 사전	8
관형사 사전	31	특수어 사전	16
감탄사 사전	22	복합명사구 사전	16
조사/어미 사전	33		

〈표5: 세종 전자사전의 하위 사전별 정보 항목 수〉

### 3.4.2. 기초 전자사전의 미시구조

한편, 기초 전자사전의 미시구조는 문법 범주 및 형태와 관련한 기본적인 정보 항목들만으로 구성되고 범주별 하위 사전 모두에 공통으로 사

용되는 단일 구조로 설계되었다. 단, 문법 범주별 속성을 반영한 정보 항목들을 설정하여 해당 어휘(군)을 기술할 때에만 사용하도록 한 것이 특징이다. 예컨대, 다음 표의 <infl>은 용언 표제항 기술을 위해 설정된 정보항목이고, <npTem>과 <subTem>은 복합명사구 표제항, <colC>와 <colCOrg>는 연어 표제항 기술을 위해 설정된 정보 항목들이다.

<superEntry>	최상위 표제항 구획
<entry>	표제항 구획
<orth>	표기 형태 정보
<dic>	품사 정보
<hom>	동형어 유무 정보
<class>	하위 부류 정보
<str>	내부 구조 정보
<org>	원어 정보
<infl>	용언 활용 정보
<npTem>	복합명사구 형판 정보
<subTem>	복합명사구 하위 형판 정보
<constant>	연어/관용표현 고정 요소 정보
<colC>	연어변 정보
<conOrg>	고정요소 원어 정보
<colCOrg>	연어변 원어 정보
<dom>	전문 영역 정보

〈표6: 기초 전자사전의 미시구조〉

### 3.5. 전산 어휘부로서의 특성

세종 전자사전의 전산 어휘부로서의 특성으로는 무엇보다도 사전 구성 및 어휘 정보의 위계적 모듈화와 형식화를 통해서 한국어 전산 처리 시에 필요한 정보를 즉각적으로 추출하고 사용할 수 있도록 구성되었다는 점을 들 수가 있다.

#### 3.5.1. XML 방식의 정보 표상

세종 전자사전은 전산 처리에 즉각적으로 활용될 수 있도록 사전 정보가 구조화·전산화되어 있을 뿐 아니라 이의 표상에서도 표준적이고 형식적인 방법이 채택되었다. 즉 XML 방식을 도입하여 전산학 분야의 데

이터 처리 및 표상 방법을 따른다. XML 방식이 정한 문법에 따라 개별 어휘 항목과 세부 정보 항목들이 특정 표지를 통해 구분되고 일정한 원칙에 따라 구체적인 정보들이 표상된다. 따라서 세종 전자사전과 세부 정보들은 현 상태 그대로 전산 프로그램에 활용될 수 있으며, 다른 유형의 데이터베이스 형태로도 쉽게 변환될 수 있다.

### 3.5.2. 확장성 및 연동성

세종 전자사전은 한국어 정보 처리에 손쉽게 활용될 수 있도록 확장성, 유연성, 연동성을 갖춘 기반 전자사전이다. 모듈적 구성을 가짐으로 해서 특정 모듈에 대해 언제라도 부분적인 보완 및 개선 작업이 가능한 확장성을 갖추고 있다. 따라서 향후 개발될 인공지능 또는 여러 새로운 전산기술에 입각한 언어 처리 시스템에도 손쉽게 적용, 활용될 수 있는 기술적 유연성을 갖는다. 또한, 세종 전자사전은 그 속에 표상되는 각종 정보들을 통해 고유명사, 전문용어, 방언, 북한어, 구어, 음성 사전 등의 위성 사전과의 연계 및 연동성을 갖는 하나의 핵심 모듈로 기능할 수가 있다. 이러한 연동성과 앞서 언급한 확장성, 기술적인 유연성이 맞물려 해당 분야에서 직접적으로 활용될 수 있는 전자 어휘부로서의 기능이 한층 강화된다.

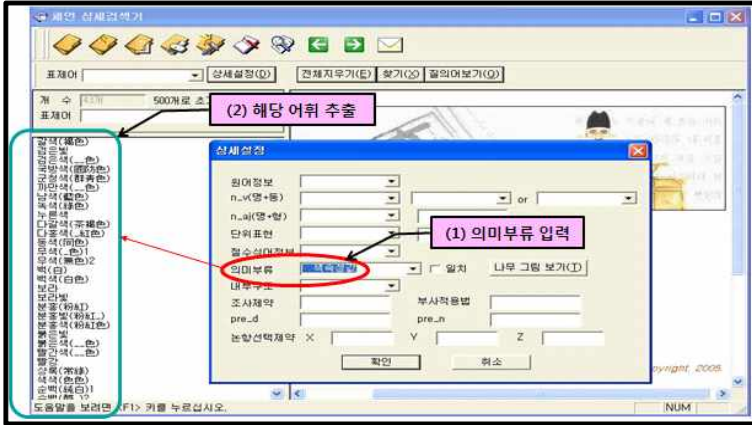
### 3.5.3. 정보 검색 및 추출의 용이성

세종 전자사전은 상기 특성들로 인해 필요한 정보를 검색하고 추출하기가 용이하다. 즉 필요한 정보를 유형별로 검색하거나 추출(단순검색)할 수 있을 뿐 아니라, 여러 유형의 정보를 묶어서 검색하고 추출(복합검색)할 수가 있다. 이는 세종 전자사전의 검증·평가를 위해 개발된 검색기를 통한 정보 추출의 예를 통해서 확인할 수 있다.

#### ㉠ 단순 검색

아래의 <표7>은 체언사전 표제항의 의미 기술에 사용되는 의미 부류

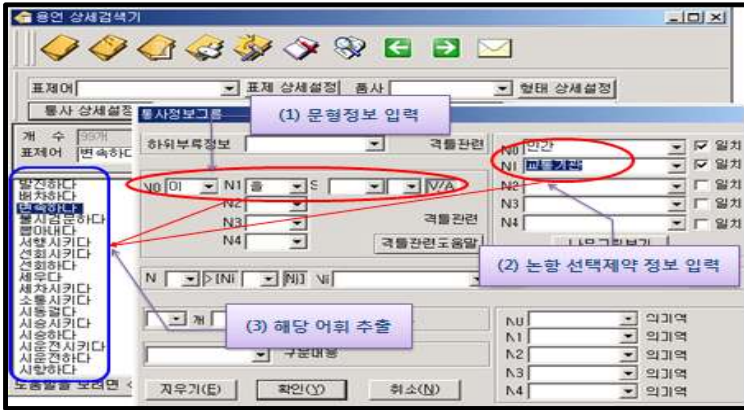
정보 항목을 이용하여 명사 어휘를 검색하는 과정과 결과를 보여주는 것으로서 단순 검색의 한 예이다. 의미 부류를 이용하여 의미 유형별 어휘 검색을 하기 위해 의미 부류 중의 하나인 <색속성값>을 입력하고 상세설정 창의 '확인'과 검색기 툴바의 '찾기'를 누르면, '갈색', '검은색' 등 <색속성값> 의미 유형에 속하는 43개의 어휘들이 검색된다.



<표7: 단순 검색의 예>

### ㉠ 복합 검색

다음의 <표8>은 복합 검색의 한 예로 용언 상세 검색기를 이용해 <교통기관> 부류를 목적으로 취하는 동사 어휘들을 검색하는 과정과 그 결과를 보여주는 것이다. 우선, 검색기의 상세 설정에서 문형 정보로 두 개의 논항을 취하는 동사들이 검색 대상임을 명기한다. 그런 다음 각 논항의 의미 유형을 세종 의미 부류 체계의 부류 명으로 입력하는데, 여기서 는 주어 논항으로 <인간>부류가, 목적어 논항으로는 <교통기관>부류가 조건으로 설정되었다. 이렇게 설정된 조건을 '확인'하고 '찾기'를 누르면, 이 조건을 충족시키는 동사 총 99개가 추출된다. 표에서 드러나듯이, 대부분이 <교통기관>에 속하는 명사들과 특징적으로 결합하는 동사들이다.



〈표8: 복합검색의 예〉

이상의 예는 세종 전자사전이 정보 항목별 정보의 추출, 그리고 특정 정보 항목들의 정보들을 조합해서 추출하는 것 모두가 용이하도록 구축되어 있음을 보여주는 것이다. 이는 결국 세종 전자사전이 한국어 전산 처리 시에 필요한 정보들을 즉각적으로 제공할 수 있는 전산 어휘부로서의 기능을 효과적으로 수행할 수 있음을 보여주는 것이라고 하겠다.

## 4. 세종 전자사전 개발의 의의 및 전망

세종 전자사전은 앞에서 살펴본 특성으로 인해 여러 측면에서 그 개발의 의의를 갖는다.

### 4.1. 문화적 의의

국가 지원 사업으로 성사되어 10년간 추진된 세종 전자사전 개발은 획기적이고 의미 있는 어문정책 사례로서 큰 의의를 갖는다. 국가 지원의 장기 계획이었다는 점, 기초·기반 구축을 위한 사업으로 학제적 연구로

추진되었다는 점, 그리고 그 결과물이 연구자 및 기업에 공공재로 개방되었다는 점에서 특히 그러하다. 또한, 언어학 연구와 전산학 연구 간의 긴밀한 협업을 통해 인문학적 지식과 첨단 정보 처리 기술이 융합된 문화재를 산출했다는 점에서 역사적으로 문화적으로 큰 의의를 갖는다.

## 4.2. 학술적 의의

세종 전자사전의 개발은 언어학과 컴퓨터 공학의 긴밀한 협력을 통해 수행된 결과이므로 우선 학제간 연구의 성공 사례로서도 의의가 크다고 하겠다. 뿐만 아니라, 해당 학문 분야 각각에서도 여러 가지 의의를 찾을 수 있다.

### 4.2.1. 사전학 및 국어학 분야

세종 전자사전은 NLP지원을 위한 전산 어휘부의 성격이 우선이긴 하지만 한국어 연구를 위한 전산 자료체의 성격도 가진다고 할 수 있다. 그 구축 과정에서 언어 유형론적 시각에서 한국어 어휘의 특성을 분석하여 지금까지 명시적으로 기술되어 본 적이 없는 한국어 어휘들의 복잡하고 다양한 언어적 속성들을 명시적·형식적으로 상세히 기술했다는 점은 사전학과 국어학 분야의 큰 성과라고 할 수 있다. 또한, 이 과정에서 어휘부 중심 연구의 이론 및 최근 성과들을 정리하고 종합하여 적극 활용한 점도 언어학 분야에서 세종 전자사전 개발이 갖는 의의라고 할 것이다.

### 4.2.2. 전산학 분야

세종 전자사전의 개발은 NLP 연구의 기반을 구축했다는 점에서 전산학 분야에서도 큰 의의를 갖는다. 사전·어휘부 등 언어 지식 기반의 NLP 시스템을 지원하는 대규모 범용 데이터베이스인 세종 전자사전이 구축됨으로써, 무엇보다도 현대 한국어 어휘 전반에 관한 효율적인 정보

구축 및 관리를 가능하게 되었다. 또한, 진화하는 NLP기술 및 인공지능 기술에 적응할 수 있는 다양한 목적의 응용 전자사전과 다국어 전자사전 등 특수 목적의 (전자)사전 구축을 위한 기반도 조성하였다. 이를 통해 궁극적으로는 기계 번역, 정보 검색, 문서 자동 요약, 지능형 질의 응답 시스템 등 한국어와 관련한 다양한 언어 처리 응용 시스템의 개발이나 고도의 인공지능 개발에 필요한 온톨로지 구축의 기반을 제공했다고 할 것이다. 세종 전자사전이 이렇게 한국어 텍스트 분석 도구의 개발을 지원하고 그 기술의 발전에 기여할 것이라는 점이 전산학 분야에서 갖는 개발 의의라고 하겠다.

### 4.3. 산업적 의의

세종 전자사전 개발은 산업적 측면에서도 의의를 갖는다. 언어 자동 처리를 통해서 인터넷을 기반으로 하는 신산업 육성의 발판을 마련하고, 미래 정보 처리 기술 개발의 핵심 기반을 제공하였다는 점이 그것이다. 또한, 정보화의 기반을 구축함으로써, 국내 정보 처리 시장의 활성화와 기술의 대외 종속 방지, 지식 정보를 활용한 국가 생산성 향상에 기여할 것으로 기대된다.

## 5. 맺음말

한국어에 대한 다양한 전산 처리에 활용할 목적으로 개발된 세종 전자사전은 대규모 범용 현대 한국어 전자사전을 지향하고, 이에 부합하는 규모와 구성, 기능을 갖는 전산 어휘부이다. 이러한 성과는 무엇보다도 언어학 및 언어 공학, 컴퓨터 공학 등 관련 분야의 다양한 기존 연구 성과들을 총체적이면서도 치밀하게 검토·분석하는 한편, 한국어의 작용 원리를 면밀하게 관찰·분석하여 사전을 구축하고자 한 노력의 결과이다.



특히, 현재의 기술 환경을 반영하는데 그치지 않고, 미래의 기술 환경을 예견하여 유연성과 확장성, 연동성을 갖추도록 한 것은 큰 의미가 있다고 하겠다. 이로 인해, 세종 전자사전 개발이 한국어 전산 처리 분야의 요구에 답하는 수준에 머무르고 이 분야의 기술 혁신과 발전을 촉진할 기반 조성에까지 이를 수 있었기 때문이다.

물론, 세종 전자사전의 완성도를 높이기 위해서는 아직도 보완해야 할 사항들이 많이 있다. 상세기술 어휘(군)의 확충, 빈도 정보와 음성 정보 등 수록 정보의 확충, 기술된 어휘(군)들에 대한 형식 및 내용의 보완·정제 작업, 항목별 하위 사전들 간의 통일성과 연계성을 제고하기 위한 표준화 작업, 한국어 말뭉치와의 연동 시스템 구축, 한국어의 특성이 반영된 치침과 규약이 만들어지도록 국제 표준화 기구(ISO)의 활동에 지속적이고 적극적으로 참여하는 노력 등은 세종 전자사전 개발 사업의 공식적 종료와는 별도로 반드시, 지속적으로 수행되어야 할 과제라고 할 것이다.

그러나, 세종 전자사전의 개발 목적이 한국어 전산 처리에의 활용에 있는 만큼 개발된 전자사전을 한국어 전산 처리에 활용하는 다각적인 방안을 마련하는 것이 방금 지적한 과제들을 풀어나가는 것 못지않게 중요한 과제라고 할 것이다. 이러한 노력이 부단히 경주되고 그 결과들이 하나둘 축적될 때, 세종 전자사전은 그 유연성과 확장성에 힘입어 더욱 더 개선·개량될 것이고 그만큼 세종 전자사전 개발의 의의도 증대될 것이다. 이는 사전 텍스트란 늘 다시 쓰여질 수밖에 없는 태생적으로 미완의 텍스트라는 점에 비추어 볼 때 더욱 그러하다고 하겠다.

## 참고 문헌

- 목정수 · 임유중, “전자사전의 구축 및 활용: 세종전자사전을 중심으로”, 「2005 제6회 국어정보화 아카데미 강의자료」, 국어정보화 아카데미 조직위원회. 2005.
- 박만규 · 이선웅 외, “21세기 세종 계획 관용 표현 전자사전 구축에 대하여”, 「한글 및 한국어 정보처리」, 제13회 한글 및 한국어 정보 처리 학술대회, 한국정보과학회/한국인지과학회. 2001
- 이성현, “세종 전자사전의 전산적 활용 방안”, 「2006 제7회 국어정보화 아카데미 강의자료」, 국어정보화 아카데미 조직위원회. 2006.
- 이용훈 · 이종혁, “국내외 전자사전 개발의 현황과 전망”, 「한국사전학」, 제8호, 한국사전학회. 2006.
- 홍재성, “21세기 세종 계획 전자사전 개발 연구보고서”, 문화체육관광부/국립국어원. 2000-2007
- 홍재성, “21세기 세종 계획 전자사전 개발 결과물 CD”, 문화체육관광부/국립국어원. 2000-2007
- 홍재성 · 이성현, “세종 전자사전: 전산 어휘부로서의 특성과 의의”, 「한글 및 한국어 정보처리」, 제19회 한글 및 한국어 정보 처리 학술대회, 한국정보과학회/한국인지과학회. 2007.
- 홍재성 · 이성현, “전산 어휘부로서의 세종 전자사전 - 그 특징과 구축 의의”, 제9회 인문학 열린 광장, 서울대학교 인문대학. 2008.
- Courtois, B. & Silberztein, M.(1990), Dictionnaires électroniques du français, *Langue française 87*, Paris, Larousse..
- Corréar, M.-H.(2002), *Lexicography and Natural Language Processing* : A Festschrift in Honor of B.T.S. Atkins, EURALEX.
- Gross, G.(1992a), Formes d'un dictionnaire électronique, *L'environnement traductionnel*, Silley, Presses de

l'Université du Québec.

Gross, M.(1975), *Méthodes en syntaxe*, Hermann.

Hong, C.-S. & Lee, S.-H.(2003), Representation of Lexico-Syntactic Information for the Description of Predicate Nouns in the Sejong Electronic Dictionary, *Proceedings of ICKL-TU Berlin International Conference on Korean/Corpus Linguistics*, ICKL-TU Berlin.

Hong, C.-S. & Pak, M.(2007), Developing a large scale computational lexical database of contemporary Korean : SELK, *Lux Coreana 2*, Paris : Han-Seine

Lee, S. H. & Hong, C.-S.(2008), Le dictionnaire électronique du coréen contemporain : le Dic Sejong - ses caractéristiques et son intérêt, *Lexicographie et Informatique : Bilan et perspectives*, ATILF/Nancy-Université, Nancy.

Mel'čuk, I.-A. et al.(1984, 1988, 1992), *Dictionnaire explicatif et combinatoire du français contemporain I, II, III*, Les presses de l'université de Montréal.

Mel'čuk, I.-A. et al. (1995), *Introduction à la lexicologie explicative et combinatoire*,