

국립국어원 2007-01-12

발 간 등 록 번 호
11-1370252-000063-10

연구보고서 【2007년】

21세기 세종계획 전자사전 개발

연구책임자 : 홍 재 성 (서울대학교)

국립국어원

제 출 문

국립국어원장 귀하

21세기 세종계획 “전자사전 개발”에 관하여 귀 원와 체결한 연구 용역 계약에 의하여 연구 보고서를 작성하여 제출합니다.

2007년 12월 10일

총괄 연구 책임자 : 홍 재 성 (서울대학교)

■ 세부 과제명 : 핵심 전자사전 구축 및 통합지원·홍보

연구책임자 : 홍재성(서울대학교)

연구 원 : 이성현(서울대학교) 외 13명

연구보조원 : 임근석 외 25명

보 조 원 : 없음

■ 세부과제명 : 통합전자사전 구축

연구책임자 : 신호필(서울대학교)

연구 원 : 없음

연구보조원 : 이용훈 외 10명

보 조 원 : 없음

요 약 문

1. 사업명 :

한국어 전자사전 개발

2. 연구의 목적

- 본 연구는 현대 한국어 어휘의 체계적인 분석·기술에 근거하여, 한국어 관련 언어 정보의 자동 처리에 필수적이면서도, 보편적으로 활용될 수 있는 기반 전자사전을 구축하려는 작업이다.
- 전자사전이 각종 언어 정보를 처리하는 소프트웨어의 개발과 실행에 필수불가결한 핵심 부문임을 인식하여, 본 연구에서는 언어학/한국어학/사전학 및 전산학의 기존 연구성과 전반에 대한 비판적 검토를 바탕으로, 다양한 용도의 전산처리에 실질적으로 활용될 수 있고 그 기술의 획기적 발전에도 공헌할 수 있는 대규모 범용 한국어 전자사전을 개발하고자 한다.
- 범용 전자사전이란 특정한 유형이나 영역의 기계 처리 작업에 국한되지 않고, 정보 검색, 텍스트의 분석과 산출, 자동번역, 다국어 사전 구축, 인쇄 사전 구축 또는 한국어 순수연구나 교육 등에 두루 활용되며, 더 나아가서는 향후의 진보된 인공지능 개발환경에도 유연하게 적용될 수 있는 전자사전을 말한다.

3. 연구의 필요성

- 고도의 정보화 시대에 접어들어 선진 각국에서 이미 개발한, 또는 개발중인 대규모 전자사전 구축 작업들을 염두에 둘 때, 정밀하면서도 방대한 한국어 전자사전의 편찬 작업은 우리나라가 21세기를 주도할 정보, 통신 등 언어 정보 산업에서 우위를 점하기 위해서 갖추어야 할 기본 요건이다. 이는 또한 미국이나 일본 등 언어 정보 산업 선진국에 의한 문화 및 기술 종속 방지 차원에서도 시급히 요구되는 일이다.
- 언어 정보의 효율적인 자동처리를 위해서는 언어학과 전산학의 긴밀한 협력이 필수적으로 요구된다. 본 연구는 언어학적 관점에서 정밀하게 기술된 자료를 전산학적 관점의 연구와 체계적으로 접목시킴으로써 이제껏 미진했던 두 분야의 학제간 연구를 획기적으로 진척시키는 기회를 제공한다.
- 자연언어 기계처리 분야의 비약적 발전을 위해서는 특정 분야에만 제한적으로 응용될 수 있는 자동 처리 방식에서 탈피하여 언어 처리 전반에 걸쳐 폭넓게 활용될 수 있는 방대한 규모의 기본적 정보자료가 구축되어야 한다. 따라서 개별적이고 특수한 연구들과는 달리 본 연구에서 지향하는 범용적인 대규모 방대한 어휘 정보 베이스의 구축이 이루어져야 한다.

4. 연구내용 및 결과물

세종 전자사전 개발 최종단계 사업연도인 2007년의 주요 연구내용은 다음과 같다.

- 1) 핵심전자사전 보완·정제
 - ㉠ 체언 상세전자사전 보완·정제 : 35,100 항목
 - ㉡ 용언 상세전자사전 보완·정제 : 27,700 항목:
- 2) 핵심전자사전 구축지침서 보완·정제 : 체언사전, 용언사전
- 3) 세종 전자사전 개발과 활용을 위한 기반 연구
- 4) 통합 전자사전 구축
- 5) 세종 전자사전 입력 및 검색, 평가·검증, 활용을 위한 연구 및 도구 개발

이와 같은 연구내용은 각 하위분과에서 수행되는데, 1)-2)는 하위 전자사전 개발 분과에서, 3)은 하위 전자사전 개발 분과 및 통합지원·홍보 소분과에서, 4)는 통합 전자사전 구축 분과에서, 5)는 통합 전자사전 분과와 통합지원·홍보 소분과에서 수행되었다. 각 분과별 2007년 연구의 최종 결과물을 구체적으로 제시하면 아래와 같다.

○ 핵심전자사전 보완·정제 작업

구분		상세전자사전								
		계획 항목 수			실제 구축 항목 수					
		SuperEntry	Entry	Sense	SuperEntry	Entry	Sense			
보완 · 정제	체언			35,100	25,458		29,708		35,854	
	용언	형용사		27,700	4,398	19,578	4,530	21,060	6,244	
		동사			15,180		16,530	23,268	29,512	
처리항목					45,036		50,768		65,366	
합계				62,800						65,366

[참고 1] 체언 및 용언 상세전자사전 총 구축량 산출 단위 : Sense 단위

[참고 2] 실제 구축 항목 수는 보완·정제 과정에서 표제항 삭제 및 추가를 통해 계획 항목 수를 초과하게 되었다.

○ 전자사전 구축 지침서 보완

- ㉠ 체언사전 및 용언사전의 보완·정제 작업을 위해 구축지침서를 정비
- ㉡ 체언사전 및 용언사전의 보완·정제 작업 결과를 반영하여 구축지침서를 보완·정제

○ 전자사전 개발과 활용을 위한 기반 연구

- ㉠ 핵심 전자사전 보완·정제 작업 지원
 - 사전원고의 전산형식 검증, 평가 및 관리
 - 세종 전자사전의 각종 언어/어휘자료 및 목록 정비·관리
 - 분과 사전간 표준화 및 조정 작업을 통한 통합사전 구축 지원
- ㉡ 세종전자사전 표준미시구조/표준tag set 보완·정제
- ㉢ 세종전자사전 술어집 보완·정제
- ㉣ 세종전자사전 활용방안 연구 및 홍보자료 작성

○ 전자사전 구축도구 및 활용도구 개발 및 보완

- ㉠ XML형식의 통합입력기(통합에디터)의 수정·보완
- ㉡ 세종 전자사전의 상세 검색기 및 단순 검색기 보완·정제
- ㉢ 유니코드 형식의 웹 기반의 통합 관리시스템 보완·정제
- ㉣ 세종전자사전 검증 평가 및 활용을 위한 도구 개발

5. 연구결과의 활용

전자사전 및 언어정보 처리 소프트웨어 개발과 관련하여, 국외에서는, 대규모 프로젝트를 통한 지속적이고 일관성 있는 연구들이 활발하게 진행되고 있으며, 그 결과들의 활용도 이미 상당한 수준에 이른 상태이다. 반면 국내의 연구 현황은 아직 미미한 수준에 머무르고 있어 언어정보 선진국의 수준에 미치지 못하고 있다. 국내 최초로 장기간의 체계적인 연구 작업에 의한 대규모 한국어 전자사전 구축을 목표로 하는 세종전자사전 개발 사업은, 이미 중규모 전자사전의 개발과 배포를 통해 부분적으로 확인할 수 있었듯이, 다음과 같은 기대효과로 열악한 현 상황의 개선에 공헌할 수 있을 것이다.

- 본 연구는 한국어 정보 처리 분야에 필수적인 대규모의 어휘 정보의 전산화를 위한 본격적인 연구로서 아직도 미미한 수준에 있는 국내 관련분야의 연구 활성화를 촉진한다.
- 본 연구를 통해 구축될 한국어 전자사전은 범용적 어휘 데이터베이스의 성격을 가지므로, 이의 전산적 가공을 통해 각종 한국어 정보 처리 분야에서 사용될 수 있다. 따라서, 해당 분야의 국내 연구 및 기술 수준을 획기적으로 발전시킬 수 있는 계기를 제공한다.
- 본 연구는 언어 정보 데이터베이스를 공학적 실용성의 관점에서 뿐만 아니라 언어학적 이론과 방법론을 엄격하게 적용하여 구축하는 것으로서, 그 결과물로 제공되는 한국어 전자사전은 국어학·언어학 연구에도 요긴하게 활용될 수 있는 언어 자료를 제공할 수 있다. 이러한 관점에서 본 연구는 한국어 연구의 발전에도 크게 기여할 것이다.
- 전자사전의 개발은 언어학과 전산학의 진정한 학제간 협력 작업을 통해서 이루어질 때 바람직한 결과가 기대되는 성격의 작업이다. 본 연구 개발에는 다수의 국어학·언어학 분야의 전문 연구 인력과 전산학 분야의 전문 인력이 참여하여 긴밀한 상호 협력의 장

을 이루고 있어, 그간 한국어 자동처리 분야에서 문제점으로 나타났던 언어학적 분석/기술과 공학적 응용 사이의 불균형 문제를 해소시키고 있는 만큼 만족스러운 성과를 기대할 수 있을 것이다.

- 본 연구에서 목표하는 전자사전의 구축은 한국의 정보·통신산업 분야에서 미국, 유럽, 일본 등 기존 선진산업국들에의 기술 종속을 미연에 방지하고, 나아가서는 이들과의 경쟁에서 한국이 우위를 점할 수 있는 기반을 마련한다.

차례

I. 머리말	1
1. 세종 전자사전 개발의 목표와 의의	1
2. 세종전자사전의 특징	3
3. 3단계(2004-2006) 및 최종단계(2007) 사업 개요	5
II. 사업 개요	7
1. 목 적	7
2. 사업기간	7
3. 주요내용	8
4. 2007년 사업 목표 및 내용	11
III. 연구 추진 계획 및 결과	15
IV. 세부 추진 내역	17
1. 핵심사전 분과	17
1.1. 체언사전	17
1.2. 용언사전	39
1.3. 기획·운영 소분과	71
1.4. 통합지원·홍보 소분과	74
2. 통합전자사전 구축 분과	81
V. 결론 및 향후 전망	101
1. 문제점 및 개선방안	101
2. 전망 및 향후 계획	101
VI. 참고문헌	103

세부 차례

요 약 문	1
1. 사업명 :	1
2. 연구의 목적	1
3. 연구의 필요성	1
4. 연구내용 및 결과물	1
5. 연구결과의 활용	3
I. 머리말	1
1. 세종 전자사전 개발의 목표와 의의	1
1.1. 세종 전자사전 개발의 목표	1
1.2. 세종 전자사전 개발의 배경	1
1.3. 세종 전자사전 개발의 의의 또는 필요성	2
2. 세종전자사전의 특징	3
3. 3단계(2004-2006) 및 최종단계(2007) 사업 개요	5
II. 사업 개요	7
1. 목 적	7
2. 사업기간	7
3. 주요내용	8
4. 2007년 사업 목표 및 내용	11
4.1. 2007년도 연구의 목표	11
4.2. 2007년 사업 개요	12
4.2.1. 2007년 사업의 주요 내용 및 특징	12
4.2.2. 2007년도 하위분과별 연구 목표 및 내용	13
III. 연구 추진 계획 및 결과	15
IV. 세부 추진 내역	17

1. 핵심사전	17
1.1. 체언	17
1.1.1 2007년도 연구 계획 및 추진 결과	17
1.1.1.1 개요	17
1.1.1.2. 연구 내용 및 추진 결과	17
1.1.2. 체언 전자사전 연구 수행 내역	18
1.1.2.1. 체언 전자사전 보완·정제 작업	18
1.1.2.2. 체언 상세전자사전 미시구조 및 구축 지침 보완	18
1.1.2.2.1. 미시구조 및 구축 지침 보완 개요	18
1.1.2.2.2. 2007년도 체언 상세전자사전 미시 구조	22
1.1.2.2.3. 2007년도 체언 상세전자사전 구축 지침	23
1.1.2.3. 목록 관리 및 형식 검증	35
1.1.2.3.1. 체언 사전의 형식 검증	35
1.1.2.3.2. 체언 사전 superEntry, entry, sense 항목 수 검증	37
1.1.2.4. 사전 구축 관련 기반 연구	37
1.1.2.5. 회의	38
1.2. 용언	39
1.2.1. 2007년도 연구 계획 및 추진 결과	39
1.2.1.1. 개요	39
1.2.1.2. 연구 내용별 추진 결과	41
1.2.2. 용언 전자사전 연구 수행 내역	41
1.2.2.1. 용언 전자사전 보완·정제 작업	41
1.2.2.1.1. 의미역 보완·정제 작업	42
1.2.2.1.2. 용언의미부류 부착 작업	42
1.2.2.1.3. 선택제약 보완·정제 작업	42
1.2.2.2. 사전 구축 지침 및 보완	44
1.2.2.2.1. 지침 보완 개요	44
1.2.2.2.2. 2007년도 용언사전 미시 구조	45
1.2.2.2.3. 2007년도 용언사전 구축 지침	46
1.2.2.3. 목록 관리 및 형식 검증	65
1.2.2.3.1. 용언 사전의 형식 검증	65
1.2.2.3.2. 용언 사전 superEntry, entry, sense 항목 수 검증	67
1.2.2.4. 사전 구축 관련 기반 연구	68
1.2.2.4.1. 용언의미부류 관련 연구	68
1.2.2.4.2. 구문분석 관련 연구	69
1.2.2.5. 회의	70

1.3. 기획·운영 소분과	71
1.3.1. 2007년도 연구 계획 및 추진 결과	71
1.3.1.1. 개요	71
1.3.1.2. 연구 내용별 추진 결과	71
1.3.2. 기획·운영 소분과 연구 수행 내역	71
1.3.2.1. 세종전자사전 분과 전체 및 하위분과별 연구 기획·관리	71
1.3.2.2. 계획서 및 각종 보고서 작성	72
1.3.2.3. 워크숍, 세미나 및 각종 회의 기획과 진행	72
1.3.2.2.1. 전체회의 및 전문가 초청 세미나 개최	72
1.3.2.2.2. 분과사전 책임자, 연구원 회의 및 세미나	72
1.3.2.2.3. 핵심사전분과 회의 및 세미나	73
1.3.2.2.4. 분과 간 협의회 개최	73
1.3.2.4. 회의	73
1.4. 통합지원 홍보	74
1.4.1. 2007년도 연구 계획 및 추진 결과	74
1.4.1.1. 개요	74
1.4.1.2. 연구 내용별 추진 결과	74
1.4.2. 연구 수행 내역	75
1.4.2.1. 핵심사전분과 보완·정제작업 지원	75
1.4.2.1.1. 사전 데이터의 형식 검증, 평가 및 관리	75
1.4.2.1.2. 각종 어휘자료 수집 및 표제어 목록 관리	75
1.4.2.2. 세종전자사전의 표준미시구조 및 tag set 보완·정제	75
1.4.2.3. 세종전자사전의 표준 기술지침안 및 술어집 보완·정제	76
1.4.2.4. 세종전자사전의 활용방안 연구	76
1.4.2.4.1. 개요	76
1.4.2.4.2. 작성 지침	77
1.4.2.4.3. 작성 예시	78
1.4.2.5. 회의	80
2. 통합전자사전 구축 분과	81
2.1. 2007년도 연구 계획 및 추진 결과	81
2.1.1. 개요	81
2.1.2. 연구 내용별 추진 결과	82
2.2. 통합전자사전 연구 수행 내역	83
2.2.1. 통합전자사전 구축	83
2.2.1.1. 각 분과의 상세기술사전 결과물 통합 및 관리	84

2.2.1.2. 각 분과의 확장기초전자사전 결과물 통합 및 관리	86
2.2.2. 통합전자사전 관리	87
2.2.2.1. 웹 기반 통합 관리 시스템 수정 보완	87
2.2.3. 통합전자사전 검색	89
2.2.3.1. 단순검색기 및 상세검색기 수정 보완 및 사전 업데이트	89
2.2.3.2. 의대수 상세검색기 개발	90
2.2.4. 통합전자사전 검증 및 평가	95
2.2.4.1. 세종전자사전 기반 형태소 분석 프로그램 개발·보완	95
2.2.5. 사전 구축 분과 지원	96
2.2.5.1. 통합에디터 수정 및 보완	96
2.2.5.2. 외형 검증 도구 개발	98
2.2.6. 2007년도 중간 및 최종 결과물 CD 제작	99
2.2.6.1 2007년도 확장본 통합전자사전	99
2.2.6.2 기타 통합전자사전용 도구	99
2.2.7. 회의	100
V. 결론 및 향후 전망	101
1. 문제점 및 개선방안	101
2. 전망 및 차후 계획	101
VI. 참고문헌	103

I. 머리말

1. 세종 전자사전 개발의 목표와 의의

1.1. 세종 전자사전 개발의 목표

세종 전자사전 개발 사업은 다음과 같은 구체적인 목표를 지향한다.

- 현대 한국어 어휘의 체계적인 분석·기술에 근거하여, 언어 정보의 자동 처리에 가장 필요하고 보편적으로 사용될 수 있는 기반 전자사전을 구축한다.
- 전자사전이 각종 언어 정보를 처리하는 소프트웨어 개발과 실행에 필수불가결한 핵심 부문임을 인식하여, 특정한 유형이나 영역의 기계 처리 작업에 국한되지 않고, 정보 검색, 텍스트의 분석과 산출, 자동번역, 다국어 사전 구축, 인쇄 사전 구축 또는 한국어 교육이나 순수연구 등에 두루 활용되며, 더 나아가서는 향후의 진보된 인공지능 개발환경에도 유연하게 적용될 수 있는 대규모 범용 한국어 전자사전을 개발한다.
- 전자사전의 개발에 언어학과 전산학 분야의 연구 성과들을 적극 수용하고 연계시킴으로써, 학제 간 연구를 획기적으로 진작시키는 기회를 제공함과 동시에 향후 한국의 다양한 언어 정보 기반 산업의 도약적인 발전에 실질적으로 기여한다.

1.2. 세종 전자사전 개발의 배경

상기 목표 하에 추진되는 세종 전자사전 개발은 다음과 같은 상황 인식에서 출발한다.

- 지식의 정보화 및 산업화는 21세기에 들어선 현대 산업사회의 가장 핵심적인 과제들 중 하나이다. 지난 반세기 동안 경험한 전산학의 급속한 발전과 이에 힘입은 인터넷, 무선전화 등 통신 분야의 놀라운 성장을 볼 때, 금세기가 정보, 통신 산업에 의해 주도되리라는 데는 의문의 여지가 없다. 세계의 모든 선진 산업 국가들이 이미 오래전부터 지식정보화 및 산업화에 총력을 기울이고, 이에 근간이 되는 자연언어처리(NLP) 분야의 발전에 국가차원의 노력과 지원을 아끼지 않고 있는 것은 바로 이러한 이유에서이다.
- 정보화의 대상이 되는 지식체계는 분야를 막론하고 자연언어에 의해 표현된다. 따라서 이에 대한 정확한 인식이나 분석 및 산출을 보장하는 전산처리 기술의 확보가 지

식정보화 및 산업화의 관건이 된다. 그런데, 자연언어처리 분야의 발전은 자연언어의 특성과 논리에 근거한 체계적이고 망라적인 언어 데이터베이스의 구축이 없이는 기대하기 어렵다. 정보검색, 텍스트의 분석과 산출, 텍스트 자동요약, 자동번역, 다국어 사전 구축 등 언어의 각종 전산처리에는 자연언어에 관한 모든 정보를 담은 방대한 데이터베이스로서의 전산어휘부가 필수적이기 때문이다. 최근 각국이 앞다투어 전산 DB, 즉 전자사전 개발에 박차를 가하는 것은 바로 이와 같은 인식에서이다.

- 해외 선진국들에서는, 사실, 이미 오래 전부터 자국어의 기계 처리를 위한 대규모의 어휘 정보 구축과 전산화에 힘써왔다. 최근에는 그 성과들을 기반으로 이미 구축된 또는 구축될 어휘 정보 베이스가 재활용성, 범이론성, 범기능성을 확보할 수 있도록 이에 관한 연구들이 활발하게 이루어지고 있다. 기간 인쇄사전과 전자사전으로부터 다양한 어휘 정보들을 효과적으로 추출하고 통합하는 동시에, 이제까지 축적된 자료들로부터 유용한 언어 정보들을 추출하여 통합적으로 활용하는 방안에 대한 연구 등이 그것이라 하겠다. 또한, 기존의 자료들을 여러 응용 분야에 사용될 수 있도록 개편하고, 특정 언어이론이나 응용분야에 국한되지 않는 다양한 기능의 대규모 어휘 정보 베이스를 구축하는 활발한 시도들도 이같은 노력의 일환으로 지적될 수 있다.
- 국외의 이러한 동향에 비추어 보면, 국내의 자연언어처리 분야는 아직 그 기반이 취약한 상태이다. 이는 무엇보다도 한국어 전산 처리에 필수적인 대규모 한국어 전산어휘부의 구축이 아직까지 본격적으로 연구·개발되지 못한 데 기인하는 바가 크다. 사실, 국내에는 언어 정보의 전산화의 기초 자료로 활용될 수 있을 만큼 언어학적 기반한 정밀하고 한편 연구자나 개발자가 자유롭게 접근해서 활용할 수 있는 공공적 성격의 인쇄 사전이나 기계 가독형 사전이 개발되어 있지 못한 형편이다.
- 한국어 전산 처리를 위한 지금까지의 연구들도 대부분 한국어에 대한 부분적인 처리에 초점을 맞추고 있는 실정이다. 다시 말해, 국내에서 개발되었거나 개발 중인 전자사전의 대부분은 단순 검색 기능만을 갖추고 있으며 그 정보 내용도 기존의 인쇄 사전에 수록된 정보와 크게 다르지 않다. 뿐만 아니라, 본격적인 기계처리를 위한 사전인 경우에도 한정된 분야의 어휘나 기능에 국한되어 있고 정보 내용도 충분히 정밀하지 못한 경우가 대부분이다. 즉, 기존의 연구 성과들은 그것이 즉각적으로 활용될 만큼 충분히 범용적인지 호환성이 있는 것인지조차도 불확실한 상태라고 하겠다.

1.3. 세종 전자사전 개발의 의의 또는 필요성

상기한 상황 인식에서 출발한 세종 전자사전 개발 사업이 갖는 의의 또는 필요성은 다음과 같다.

- 효율적인 자연어 자동처리를 위해서는 특정 분야에만 응용될 수 있는 자동 처리 방

식에서 탈피하여 언어 전반에 걸쳐 폭넓게 활용할 수 있는 방대한 규모의 정보자료를 구축하는 것이 그 무엇보다도 우선하는 필수적 조건이 된다.

- 자연어의 효과적인 자동처리를 위해서는 언어학적 관점에서 체계적으로 기술된 자료를 전산학적 관점의 연구와 체계적으로 접목시키는 일이 필요하다. 그런데 이제까지의 언어공학 관련 연구에서는 언어학과 전산학 분야의 연구 성과가 유기적으로 연계되지 못했다고 할 수 있다. 사실, 후속 연구에 필수적이라 할 연구결과의 공개나, 연구자들 사이의 자료 공유조차도 잘 이루어지지 못했고 현재의 상황도 과거와 크게 다르지 않은 실정이다.
- 이러한 상황에 비추어 볼 때, 한국어의 자동처리 분야의 발전에 필수불가결한 정밀하면서도 방대한 전자사전의 개발은 시급한 과제라 하겠다. 이는 우리나라가 정보화 시대에 효과적으로 대처하여 미국이나 일본 등 선진국에의 기술 종속을 방지하고, 나아가 이들과의 경쟁에서 우위를 차지하기 위해서는 더욱 그러하다.

2. 세종전자사전의 특징

세종 전자사전은 범용적인 대규모 전자사전의 개발을 지향하고 있다. 이 때의 범용성이란 무엇보다도 공학적 활용의 범용성을 의미한다. 즉 다양하고도 구체적인 용도의 공학적 활용을 위한 응용 프로그램 및 응용 전자사전 개발에 두루 활용될 수 있다는 것이다. 바로 이 “범용성”이 세종 전자사전의 성격과 구성을 규정하여, 이에 수록되는 어휘 정보의 규모 및 깊이 등과 같은 사전의 내용과 정보 표상 등의 형식을 결정하는 핵심적인 요인으로 작용한다. 이런 점에서 세종 전자사전은 기간 인쇄사전 및 전자사전들과 구별되는 여러 특징들을 갖는다.

- 전산 어휘 자료체
세종 전자사전은 특정 분야에 국한되지 않는 이론 절충적인 전산 어휘자료체이다.
- 정밀하고 방대한 어휘 정보
세종 전자사전은 말뭉치, 기간 인쇄사전 및 어휘 연구자료 등 각종 풍부한 어휘자료를 최대한 활용하여 정밀한 어휘 정보를 수록한다.
- 철저한 기술
세종 전자사전은 최신의 언어학과 언어공학의 연구 성과들을 적극 반영함과 동시에 인쇄사전과 같은 기존의 자료를 철저하게 재분석하여 기술한다.
- 위계적인 사전 구성

세종 전자사전은 한국어 전산처리의 활용성과 효율성의 극대화를 위해 하위모듈이 통합된 형태로 구성된다. 기본적인 문법범주와 형태정보만이 부착된 대규모 (확장)기초전자사전과 상세한 어휘 정보가 부착된 상세전자사전으로 분리되어 개발된다. 각 전자사전은 다시 문법범주에 따라 별도의 사전으로 개발되고, 최종적으로 이들이 통합되어 전체 세종 전자사전을 구성하게 된다.

○ 어휘 정보 구성

모듈적 구성은 어휘 정보 표상에도 반영된다. 사전에 수록되는 어휘 정보들은 그 특성에 따라 여러 하위 유형으로 나뉜다. 세종 전자사전은 어휘 정보의 이러한 특성을 반영하여 정보 항목들을 유형에 따라 정보 구획으로 구분하여 표상한다. 이에 따라 세종 전자사전의 어휘 항목은 세부 정보항목들을 포함하는 여러 하위구획으로 구성되는 위계적인 모듈의 형태를 갖는다.

○ 기술적 유연성

세종 전자사전은 그 모듈적 구성으로 인해 특정 모듈에 대해 언제라도 부분적인 보완 및 개선 작업이 가능하다. 세종전자사전은 따라서 향후 개발될 인공지능 또는 여러 새로운 전산기술에 입각한 언어처리 시스템에도 손쉽게 적응, 활용될 수 있는 기술적 유연성을 갖는다.

○ 형식적인 표상

세종 전자사전은 전산처리에 즉각적으로 활용될 수 있도록 형식적인 표상방법을 채택하고 있다. 세종 전자사전은 XML 방식을 도입하여 전산학 분야의 데이터 처리 및 표상 방법을 따른다. 즉, 위의 방식이 정한 문법에 따라 개별 어휘 항목과 세부정보 항목은 특정 표지를 통해 구분되고, 일정한 원칙에 따라 구체적인 정보들이 표상된다. 따라서, 세종사전과 세부 정보들은 현 상태 그대로 전산 프로그램에 활용될 수 있으며, 다른 유형의 데이터베이스 형태로도 쉽게 변환될 수 있다.

○ 연동성 추구

세종 전자사전은 타분야 및 위성사전과의 연동성을 추구한다. 한국어 자동처리 지원이 기본 목표인 세종사전은 다각적인 언어공학적 타당성의 검증과 평가를 추구한다. 이는 앞서 언급한 기술적인 유연성과 맞물려 직접적으로 이 분야에 활용될 수 있는 전자사전의 개발을 가능하게 할 것이다. 또한, 세종사전은 하나의 핵심 모듈이 되어, 고유명사, 전문용어, 방언, 북한어, 구어, 음성 사전 등의 위성사전과의 연계 및 연동성을 추구하기 위해 필요한 정보들을 표상한다.

3. 3단계(2004-2006) 및 최종단계(2007) 사업 개요

- 3단계 사업(2004-2006)은 1-2단계의 성과를 기반으로 세종 전자사전 구축을 실질적으로 완성함을 목표로 한다. 이를 위해, 우선, 당초 계획된 규모의 범주별 신규 어휘 및 표현 기술 작업을 단계적으로 추진·완료하고, 아울러 중규모 전자사전 구축을 위해 축소·유보했던 정보 항목들에 대한 확장·기술을 완료하여 예정했던 어휘항목 규모와 정보 내용을 완벽히 갖춘 세종 전자사전의 구축을 완료한다. 또한, 이론적인 관점과 실용적인 관점 모두를 충족시킬 수 있는 전자사전에 대한 기초연구를 대폭 강화하여, 1-2단계에서 개발된 세종 전자사전의 미시구조와 정보내용 및 표기체계를 추가·보완함과 동시에 다양한 한국어 정보처리 작업에서 사용될 수 있는 응용 전자사전과 이를 구축하기 위한 인터페이스 구상·설계 등 기성 결과물들을 다각적이고 효과적으로 활용할 수 있는 방법들을 모색할 것이다.
- 최종단계인 2007년도 연구에서는 전자사전에 등재하기 위해 기술을 완료한 모든 어휘 및 표현들에 대해 최종 교열 및 보완 작업을 수행한다. 동시에, 세종 전자사전의 공개를 위한 이론적/기술적인 정비 작업을 치밀하고 광범위하게 수행한다. 또한, 1-2단계에서 마련된 세종 전자사전의 통합 관리 환경을 보완하여, 웹을 통한 보수 및 관리, 인터넷 서비스 등이 가능하도록 하여 궁극적으로는 한국어 연구 및 한국어 정보처리 도구 개발을 위한 언어정보 서비스 망 구축의 기반을 제공한다. 이같은 작업들을 통해 최종년도 사업이 완료되는 2007년 말에는 한국어 정보처리와 관련된 다양한 기술 영역에서 가장 보편적으로 그리고 용이하게 응용·활용될 수 있는, 명실상부한 대규모의 범용적 한국어 기반 전자사전이 구축될 수 있도록 한다.

이상 상술한 3단계 및 최종단계 연구의 목표 및 내용을 요약 정리하면 다음과 같다.

- **세종 전자사전 구축 완료 : 2007년도까지 완성**
 - 상세전자사전 :
 - 신규기술 : 48,800 항목 (총 누계 152,650 항목)
 - 정보확장 : 55,600 항목
 - 확장기초전자사전 : 280,800 항목 (총 누계 547,350 항목)
- **하위분과 전자사전 및 통합전자사전의 모형 보완과 개발 완료**
 - 전자사전 정보항목 확장 및 보완
 - 용언 의미 부류 기술
 - 의미역 기술·보완
 - 문형 기술 정제
 - 세종 전자사전 기반 parser 개선
 - 전자사전 정보구축 통합지침서 보완 및 완성

- 전자사전 통합을 위한 프레임 구축
 - 확장 기초전자사전 설계 및 구축
- 세종 전자사전의 내적 완결성과 외적 활용성 평가·검증 및 개선 : 2007년도
- 하위사전 내 기술정보들의 정확성 및 형식적 표상체계 검증·평가 및 보완
 - 하위사전 간의 유기적인 연결체계 평가·검증 및 보완·개선
 - 독립 기술된 하위사전 정보들 간의 상호 연결성 점검 및 제고
 - 통합전자사전 내의 기술정보들의 안정성 및 효율성 검증·평가 및 제고
- 실용적 전자사전 개발을 위한 기초 연구 및 기반 연구
- 세종 전자사전의 개발 및 응용을 위한 기초연구
 - 세종 전자사전의 의미정보 연구
 - 세종 전자사전의 통사정보 연구
 - 세종 전자사전의 전문영역 체계 및 표상방안 연구
 - 세종 전자사전의 응용성 검증 및 평가 수행
 - 응용 전자사전 개발을 위한 인터페이스 모형 연구
- 통합전자사전 관리 시스템 구축 및 검색기 보완 계획
- 데이터베이스의 다국어 지원 검색 - UNICODE 지원하도록 적절히 수정
 - XML 파일의 DB 입력기 개발
 - 세종 전자사전 정보의 통합 관리 시스템 구축
 - 세종 전자사전 상세검색기 및 단순검색기 개발
 - 상세 도움말 기능 등 부가적인 사용자 인터페이스 지원
- 통합전자사전 검증 계획
- 통합전자사전을 이용한 질의 응답 (Q&A) 시스템 구축
 - 중-한 기계 번역 시스템에서의 세종 전자사전 활용
 - 응용전자사전 개발을 위한 효율성 제고
 - 세종 전자사전 정보를 이용한 형태소 분석 및 구문 분석을 통한 검증
- 제2기 세종계획에 대비한 세종전자사전 확장 및 활용도 제고를 위한 기초 연구

Ⅱ. 사업 개요

1. 목 적

- 이 연구는 한국어 언어 정보의 자동 처리에 핵심적으로 필요하고 보편적으로 사용될 수 있는 기반 전자사전 개발을 목표로 한다. 특히, 다양한 용도의 전산처리에 실질적으로 활용될 수 있을 뿐 아니라, 그 기술을 획기적으로 발전시킬 수 있는 범용적인 대규모 한국어 전자사전을 구축하는 것이 이 연구의 주된 목표이다.
- 전자사전이 각종 언어 정보를 처리하는 소프트웨어의 개발과 실행에 필수불가결한 핵심 부문이라 할 때, 범용적 전자사전은 특정한 목적이나 특정한 영역의 기계 처리 작업에 국한되지 않고 다양한 용도로 활용될 수 있는 전자사전을 말한다. 즉, 정보 검색, 텍스트의 분석 및 산출, 자동번역, 다국어 사전 구축, 인쇄 사전 구축 그리고 한국어 교육이나 순수 언어학적 연구 등에 두루 활용되며, 더 나아가서는 향후의 진보된 인공지능 개발환경에도 유연하게 적용될 수 있는 전자사전을 의미한다.
- 언어학과 전산학의 긴밀한 연계 하에 추진되는 이 연구는 한국어 어휘를 대규모로 수집하고 그 언어적 행태와 속성을 최신 언어학 연구 성과에 기반을 둔 체계적이고 정밀한 방법론에 따라 분석·기술하는 작업으로 구성된다.
- 이렇게 개발되는 한국어 전자사전은 향후 한국의 다양한 언어 정보 기반 산업의 도약에 실질적이고 효과적으로 기여할 수 있는 필수적인 어휘정보 자료체가 될 것이다. 부수적으로 이렇게 전산화된 어휘정보 자료체는 한국어학/언어학 연구는 물론 심리학/인지과학 등의 영역에도 소중하게 활용될 수 있는 언어자료가 될 것이다.

2. 사업기간

- 총 사업기간 : 1998년 - 2007년(총 10년)
 - ▶ 1단계 사업기간 : 1998년 - 2000년
 - ▶ 2단계 사업기간 : 2001년 - 2003년
 - ▶ 3단계 사업기간 : 2004년 - 2006년
 - ▶ 최종 단계 사업기간 : 2007년

3. 주요내용

이 연구는 총 10년에 걸쳐 범용적인 대규모 한국어 전자사전의 개발을 목표로 이에 수반되는 다음과 같은 필수적인 작업들을 수행한다.

- 한국어 어휘 수집 및 분석
 - (1) 한국어 어휘 분석 및 정보 추출
 - (2) 한국어 어휘 정보의 유형 분석
- 전자사전의 설계
 - (1) 정보 항목 설정
 - (2) 전체 모형 설계
- 전자사전의 구축
 - (1) 정보 표상 방법 확정
 - (2) 전자사전의 정보 구축 지침 확정
 - (3) 사전 항목 기술
- 전자사전의 검증 및 평가
- 전자사전의 수정 및 보완

한국어 전자사전 개발은 무엇보다도 이에 수록될 한국어 어휘들을 수집하고 이들에 대한 정보들을 분석, 유형화하는 작업으로부터 시작된다. 이를 바탕으로 하여 전자사전 내 각 어휘항목에 표상될 정보 항목들을 확정하여 전자사전의 전체 모형을 설계·확정하는 것도 필수적인 작업이다. 이러한 작업들은 일련의 하위 작업들로 구성되는데 그 각각은 전자사전의 효용성을 극대화하기 위한 치밀하고도 체계적인 연구에 기반을 두고 추진된다.

예를 들어, 한국어 어휘의 수집은 단순히 기존의 인쇄 사전들의 표제어를 추출하여 기계적으로 재활용하기보다는 실용적인 측면에서 현대 한국어에서 사용 빈도와 중요도 등이 반영될 수 있도록 이에 대한 선행연구가 뒷받침해야 한다. 또한 어휘 정보의 분석·추출과 관련하여 기존의 인쇄 사전에 수록된 개별 어휘들에 대한 정보들이 다양한 측면에서 미흡하다는 점을 고려할 때, 이 작업은 한국어에 대한 최근의 언어학적 연구, 한국어 교육에서의 성과, 다양한 한국어 전산처리를 위한 어휘 정보의 분석 등을 토대로 하는 개별 어휘들에 대한 정밀하고도 체계적인 재분석을 필요로 한다.

아울러 전자사전의 정보 항목과 전체 모형은 대규모 범용 전자사전에 대한 연구가 아직 초기 단계에 있는 우리의 현실을 고려하여 이러한 작업들이 활발하게 진행되고 있는 선진국들에서의 연구 성과들을 적극적으로 활용하되, 본 연구에서 수행되는 어휘 정보에 대한 재분석 결과들을 토대로 새롭게 연구되어야 한다. 또한 정보 항목과 모형은 전자사전의 실질적 효능과 직결되는 부문이므로 기존의 한국어 전산처리 시스템에서의 어휘 정보의 이용과 데이터베이스 구축 등의 성과들이 충분히 검토·고려되어야 하며, 언어공학적 관점에서 우리 전자사전의 활용성에 대한 검증·평가 작업이 지속적으로 수행되고 그 결과가 다시 전자사전의 수정·보완에 활용되는 상시적 검증·보완 체계가 구축되어야 한다.

상기한 작업들은 바로 이러한 요구에 부응하기 위해 설정되었다. 하지만 이러한 작업들은 전자사전 개발에 수반되는 최소한 작업들이며, 이 외에도 보다 효율적으로 작업을 수행하고 완성도가 높은 전자사전의 개발을 위해 어휘 정보에 대한 다양한 유형의 자료 제작, 사전 구축에 활용되는 각종 도구 마련, 기성 결과물을 적극 활용하고 이를 검증·평가하기 위한 보조 도구 개발 등의 보조 작업들이 추가된다. 또한 각 하위 작업들은 다루어지는 내용과 그 중요도에 따라 보다 세분화된 세부 작업들을 통해 이루어지며, 각 세부작업을 연계/조정/통일하는 전체 통합의 작업도 수행된다.

우리의 연구는 여러 개의 하위 분과들을 두어 이러한 일련의 하위 작업과 세부 작업들을 분담하여 동시에 수행되도록 한다. 다만, 대규모 어휘를 대상으로 정밀한 어휘 정보를 체계적으로 수록하는 범용 전자사전의 개발이 일시에 이루어질 수 없고 작업의 성격에 따라 단계적으로 수행되어야 한다는 점을 염두에 두어 전체 작업을 3 단계로 구분하여 진행해왔는데, 이미 완료된 1-2단계 사업의 내용 및 성과는 다음과 같다.

• 1단계 연구(1998~2000)의 내용 및 성과

- ▶ 빈도 및 난이도에 의한 기본 어휘 수집
- ▶ 핵심 어휘에 대한 엄밀한 언어학적 분석과 기술
- ▶ 핵심 어휘를 수록하는 기본 전자사전 데이터베이스 구축
- ▶ 상세 정보사양을 명시한 전자사전 정보구축 지침서 작성
- ▶ 전자사전의 구축 및 관리·검색을 위한 도구 개발
- ▶ 기존 전자사전 통합을 위한 프레임 구축

• 2단계 연구(2001~2003)의 내용 및 성과

- 사전 구축 작업

- ▶ 중규모 전자사전 구축 완료 및 공개
 - ㉠ 다양한 분야에서 실질적으로 응용되는 실용 전자사전의 개발을 위한 중규모 전자사전 개발
 - ㉡ 중규모 전자사전 규모
 - 핵심 일반 어휘 상세기술 : 87,300 항목
 - 형태·범주 정보 등 1수준 정보 부착 기초 전자사전: 244,550 항목
- ▶ 중규모 전자사전 검색기 개발 및 공개
 - 중규모 전자사전의 효율적 검색·활용을 위한 검색 시스템
- ▶ 빈도 및 난이도에 의한 기본 어휘 정제
- ▶ 기본 전자사전의 데이터베이스 구축 및 확장
- ▶ 상세 정보사양을 명시한 전자사전 정보구축 지침서 완성
- ▶ 일부 분과사전 완성 :
 - 2001년: 의존명사사전, 대명사사전, 수사사전, 어근사전, 접사사전, 관형사사전, 조사사전, 어미사전, 감탄사사전,
 - 2002년 : 부사사전
 - 2003년 : 관용표현사전

- ▶ 전자사전의 구축 및 관리를 위한 다양한 도구 개발
- ▶ 전자사전 통합을 위한 프레임 구축
- ▶ 실용적 중규모 사전의 구축을 위한 기초 연구
- ▶ 사전정보의 위계화 및 활용성 분석 연구
- ▶ 시사어, 유행어, 외래어, 외래어 문자표기, 약어 등의 특수어사전 개발
- ▶ 통합 전자사전의 형태·어휘·구문정보 평가 및 검증
- ▶ 통합 전자사전의 응용 방안 탐구
- ▶ 공학적 면에서 전자사전의 응용성 평가·분석
- ▶ 범용DB와 응용 전자사전 개발을 위한 인터페이스 구축

- 3단계 사업 준비 : 전자사전 관련 기초연구를 대폭 강화하고 기성 결과물들의 효율적인 활용방안을 모색

- ▶ 세종 전자사전 모형에 대한 이론적 기반 탐구 등 기초연구를 대폭 강화하여, 세종 전자사전을 근간으로 하는 다양한 응용 전자사전의 모형을 개발해보는 등 세종 전자사전의 효용성을 제고를 위한 제반 연구를 수행
- ▶ 기초 전자사전의 규모와 내용을 확충하고, 검색도구 및 대내외적 실질적 활용방안을 탐색
- ▶ 통합 전자사전의 규모를 확충하고, 검색·서비스 체계를 보완·개선
- ▶ 통합 전자사전의 평가 및 검증 대상의 의미 정보로 확장하여 세종 전자사전의 실제 활용 가능성을 탐색
- ▶ 중규모 사전의 실용성과 효용도를 제고하는데 필수적인 특수어 사전과 복합명사구 사전의 모형을 연구 개발하고, 핵심적인 항목에 대한 표본기술을 수행

• 3단계 연구(2004~2006)의 내용 및 성과

- 사전 구축 작업

- ▶ 체언 및 용언사전을 핵심사전으로 통합 집중 기술
- ▶ 핵심사전의 미시구조를 개선·보완
- ▶ 핵심사전의 정보항목 표상방법을 XML 형식으로 변환
- ▶ 기초전자사전의 미시구조의 표준화 및 XML 형식으로 변환
- ▶ 일부 분과사전 구축 및 확장 구축 완료
 - 구축완료 : 연어사전(2004), 특수어사전(2005), 복합명사구사전(2005)
 - 확장구축 완료 : 관용표현사전(2005)
- ▶ 핵심사전을 제외한 모든 분과사전의 보완·정제 작업 완료(2006)

- 기반 연구

- ▶ 분과 간 미시구조 및 표상체계 표준화 작업을 위한 기초 연구 수행
- ▶ 세종 전자사전 범주체계의 정비·보완 작업 수행
- ▶ 세종계획 내 타분과(특히, 기초자료 구축분과 및 특수자료 구축분과)와의 범주체계 및 표상방법 표준화 작업을 위한 기초 연구 수행
- ▶ 기구축 상세사전의 XML 형식 변환을 위한 기초 연구 수행

- ▶ XML형식의 미시구조를 반영한 통합입력기 개발
- ▶ XML형식의 미시구조 변환에 따른 체언 및 용언사전의 상세검색기, 통합전자사전의 단순검색기 수정, 보완
- ▶ 세종 전자사전의 검증 및 전산적 활용 방안 연구 및 도구 개발
 - 사전별 정보항목의 전산적 활용방안 연구
 - 세종전자사전에 기반한 형태소 분석기 개발을 위한 연구

4. 2007년 사업 목표 및 내용

2007년은 최종단계연도로서 모든 사전의 보완·정제 작업을 완료하여 세종 전자사전 개발을 종료하는 것을 목표로 한다. 즉, 체언 및 용언사전의 보완·정제 작업을 종결하고, 기구축 사전들과의 연계성, 통일성을 확보하여 세종 전자사전 개발을 성공적으로 마무리하는 것이 2007년도 사업의 주안점이다. 따라서, 2007년도 사업기간에는 세종전자사전 전반에 대한 집중적인 보완·정제 작업과 이를 통한 세종전자사전의 품질과 기능의 제고에 모든 역량이 투입될 수 있는 환경을 조성한다.

4.1. 2007년도 연구의 목표

- 2007년 연구는 최종단계 사업으로서 1-3단계 사업의 성과를 기반으로 **세종 전자사전의 완성을 목표로** 한다. 이를 위해, 핵심사전(체언사전, 용언사전)을 중심으로 대대적이고 치밀한 보완·정제 작업을 수행하여, 하위사전의 형식 및 내용을 철저히 검토 검증하고, 정보항목 간, 하위사전 간의 유기적 연계성을 제고하여 세종 전자사전의 효율을 극대화한다.
- 전자사전의 보완·정제 작업과 병행하여 통합입력기, 단순검색기, 상세검색기, 웹 관리기 등 세종 전자사전의 입력, 검색, 검증 및 평가 도구 등을 개선·보완한다. 이와 함께, 세종 전자사전을 기반으로 하는 형태소 분석 및 구문 분석 프로그램을 Web 시연이 가능한 수준으로 보완한다. 이같은 작업들을 통해 3단계 사업이 완료되는 2007년 말에는 **한국어 정보처리와 관련된 다양한 기술 영역에서 가장 보편적으로 그리고 용이하게 응용·활용될 수 있는, 명실상부한 대규모의 범용적 한국어 기반 전자사전이 구축될 수 있도록** 한다.
- 동시에, 향후 출범할 제 2기 세종계획을 전제하여, 제 1기 세종계획의 결과물로서의 세종전자사전의 이론적·기술적 의의를 지속적으로 높이고 급변하는 새로운 인공지능기술/언어처리기술 환경에 유연하게 적응할 수 있는 활용성을 확보하기 위한 연구계획도 구상한다.

4.2. 2007년도 사업 개요

4.2.1. 2007년 사업의 주요 내용 및 특징

○ 핵심 전자사전 (체언사전 및 용언사전)의 보완·정제 작업

- 작업 대상 :
 - 체언 상세전자사전 : 35,100항목
 - 용언 상세전자사전 : 27,700항목
- 작업 내용 :
 - 정보항목별 형식 검증 및 내용 보완
 - 정보항목 간 기술내용의 유기적 연계성 검토 및 보완
 - 하위사전 간 상관 정보항목의 기술 내용 검토 및 통일성 제고

○ 통합 전자사전의 내적 완결성과 외적 활용성 평가·검증 및 개선

- 핵심 전자사전의 각 분과별 보완·정제 결과물 DB 등록
- 각 하위사전 내 기술정보들의 정확성 및 형식적 표상체계 검증·평가 및 보완
- 하위사전 간의 유기적인 연결체계 평가·검증 및 보완·개선
- 독립 기술된 하위사전 정보들 간의 상호 연결성 점검 및 제고
- 통합전자사전 내의 기술정보들의 안정성 및 효율성 검증·평가 및 제고

○ 통합전자사전 관리 시스템 및 입력기, 검색기 보완·개선

- 웹기반의 통합전자사전 관리·검색 시스템 보완·개선 - 세종계획 서버용
- 세종 사전 정보의 통합 관리 시스템 보완·개선
- 단순검색기와 상세검색기 보완·개선
- 통합에디터 보완·개선

○ 통합전자사전 검증·평가 도구 보완·개선

- 세종 전자사전 기반 형태소 및 구분 분석 프로그램 보완·개선

4.2.2. 2007년도 하위분과별 연구 목표 및 내용

분과명	주요 사업 내용
<p>핵심 전자사전 구축 및 통합지원·홍보</p>	<ol style="list-style-type: none"> 1. 체언 전자사전 <ul style="list-style-type: none"> ▶ 상세전자사전 보완·정제 : 35,100 항목 ▶ 체언 전자사전 기술목록 검토 및 정비 ▶ 체언 전자사전 구축지침서 보완·정제 ▶ 용언 전자사전 관련 정보항목의 기술 형식 및 내용에 대한 통일성 제고 2. 용언 전자사전 <ul style="list-style-type: none"> ▶ 상세전자사전 보완·정제 : 27,700 항목 ▶ 용언 전자사전 기술목록 검토 및 정비 ▶ 용언 전자사전 구축지침서 보완·정제 ▶ 체언 전자사전 관련 항목의 기술 형식 및 내용에 대한 통일화 방안 제고 3. 기획·운영 소분과 <ul style="list-style-type: none"> ▶ 세종 전자사전 분과의 연구 기획·관리 ▶ 사업계획서 및 각종 보고서 작성 ▶ 하위분과별 연구 기획·관리 ▶ 워크샵 및 각종 회의 기획 및 진행 ▶ 세종계획 분과간/하위사전분과 및 소과제분과 간 협의회 운영 4. 통합지원·홍보 소분과 <ul style="list-style-type: none"> ▶ 핵심 전자사전 보완·정제 작업 지원 <ul style="list-style-type: none"> - 사전원고의 전산형식 검증, 평가 및 관리 - 세종 전자사전의 각종 언어/어휘자료 및 목록 정비·관리 - 분과 사전간 표준화 및 조정 작업을 통한 통합사전 구축 지원 ▶ 세종전자사전 표준미시구조/표준tag set 보완·정제 ▶ 세종전자사전 표준 기술지침안 보완·정제 ▶ 세종 전자사전 슬어집 보완·정제 ▶ 세종전자사전 활용방안 연구 및 홍보자료 작성

분과명	주요 사업 내용
통합 전자사전 구축	<ul style="list-style-type: none"> ▶ 통합전자사전 구축 <ul style="list-style-type: none"> - 체인 및 용인의 상세전자사전 보완·정제 결과물 DB 등록 ▶ 통합전자사전 관리 <ul style="list-style-type: none"> - 웹 기반의 통합 관리 시스템 보완·정제 ▶ 통합전자사전 검색 <ul style="list-style-type: none"> - 세종 전자사전 단순검색기 및 상세검색기 수정·보완 ▶ 통합전자사전 검증 및 평가 <ul style="list-style-type: none"> - 세종전자사전에 기반한 형태소 및 구문 분석 프로그램 수정·보완 (Web상에서의 공개·시연) ▶ 사전구축분과 지원 <ul style="list-style-type: none"> - 통합전자사전 입력기(통합에디터)의 수정 및 보완

Ⅲ. 연구 추진 계획 및 결과

분 과 명	추진 계획	추진 실적	비고
체인 사전 분과	<input type="checkbox"/> 체언사전 분과 ▶ 상세전자사전 보완·정제 : 35,100 항목 ▶ 체언 전자사전 기술목록 검토/정비 ▶ 체언 전자사전 구축지침서 보완·정제 ▶ 용언 전자사전 관련 항목의 기술 형식 및 내용에 대한 통일화 방안 연구	100 % 100 % 100 % 100 %	35,854 항목 완료 (102%)
용언 사전 분과	<input type="checkbox"/> 용언사전 분과 ▶ 상세전자사전 보완·정제 : 27,700 항목 ▶ 용언 전자사전 기술목록 검토 및 정비 ▶ 용언전자사전 기술목록 검토 및 정비 ▶ 용언 전자사전 구축지침서 보완·정제 ▶ 용언의 범주체계 확립 및 복합용언의 내부구조 유형화로 한국어 용언의 총체적 형태구조에 대한 연구기반 마련 ▶ 문형/의미역/선택제약 정보 표상 방식 개선 연구 및 지침 보완·완성 ▶ 체언 전자사전 관련 항목의 기술 형식 및 내용에 대한 통일화 방안 연구	100% 100% 100% 100% 100% 100%	29,512 항목 완료 (107%)
기획 · 운영 소분과	<input type="checkbox"/> 기획운영 소분과 ▶ 세종 전자사전 분과의 연구 기획·관리 ▶ 사업계획서 및 각종 보고서 작성 ▶ 하위분과별 연구 기획·관리 ▶ 워크샵 및 각종 회의 기획 및 진행 ▶ 세종계획 분과간/하위사전분과 및 소과제 분과간 협의회 운영	100 % 100 % 100 % 100 % 100 %	

분 과 명	추진 계획	추진 실적	비고
통합 · 지원 · 홍보 소분과	<ul style="list-style-type: none"> □ 통합 지원 홍보 분과 <ul style="list-style-type: none"> ▶ 핵심 전자사전 보완·정제 작업 지원 ▶ 세종전자사전 표준미시구조/표준 tag set 보완·정제 ▶ 세종전자사전 표준 기술지침안 보완·정제 ▶ 세종 전자사전 슬어집 보완·정제 ▶ 세종전자사전 활용방안 연구 및 홍보자료 작성 	100 % 100 % 100 % 100 % 100 %	
통합 전자 사전 분과	<ul style="list-style-type: none"> □ 통합 전자사전 분과 <ul style="list-style-type: none"> ▶ 통합전자사전 구축 <ul style="list-style-type: none"> - 체언 및 용언의 상세전자사전 보완·정제 결과물 DB 등록 ▶ 통합전자사전 관리 <ul style="list-style-type: none"> - 웹 기반의 통합 관리시스템 보완·정제 ▶ 통합전자사전 검색 <ul style="list-style-type: none"> - 세종 전자사전 단순검색기 및 상세검색기 수정·보완 ▶ 통합전자사전 검증 및 평가 <ul style="list-style-type: none"> - 세종전자사전에 기반한 형태소 분석 프로그램 수정·보완 (Web상에서의 공개·시연) ▶ 사전구축분과 지원 <ul style="list-style-type: none"> - 통합전자사전 입력기(통합에디터)의 수정·보완 	100 % 100 % 100 % 100 % 100 %	

IV. 세부 추진 내역

1. 핵심사전

1.1. 체언

1.1.1 2007년도 연구 계획 및 추진 결과

1.1.1.1 개요

체언사전 분과의 2007년도 사업은 세종 전자사전 개발 최종단계 사업의 목표에 따라 수행되었다. 즉, 2007년 1-3단계의 성과를 기반으로 세종 전자사전 개발을 성공적으로 마무리하는 것을 목표로 하였다. 우선, 2006에 그 기술을 모두 완료한 체언 상세기술 사전을 대상으로 대대적인 보완·정제 작업을 수행하였다. 이를 통해, 이론적 측면과 실용적 측면 모두의 요구 충족시킬 수 있는 체언 전자사전의 구축을 완료하였다. 그 결과, 총 35,854 항목이 보완·정제 작업을 통해 기술 완료되었다(cf. 계획 항목 수 : 35,100 항목). 또한 이와 더불어 체언사전 구축지침서도 보완·정제하여 최종본을 완성하였다.

1.1.1.2. 연구 내용 및 추진 결과

분 과 명	연 구 내 용	추진 일정						추진 현황		
		6월	7월	8월	9월	10월	11월	12월	진척상황	진척율
체 언 사 전 분 과	▶ 상세기술 사전 보완·정제 : - 35,100 항목								35,854 항목 완료	102%
	▶ 체언 전자사전 기술목록 검토 및 정비									100%
	▶ 체언 전자사전 구축지침서 보완· 정제									100%
	▶ 용언 전자사전 관련 항목의 기술 형식 및 내용에 대한 통일성 제고									100%

1.1.2. 체언 전자사전 연구 수행 내역

1.1.2.1. 체언 전자사전 보완·정제 작업

2007년도 작업에서는 2006년에 상세기술이 완료된 체언 35,100 항목에 대해 보완·정제 작업을 수행한 결과, 총 35,854 항목의 기술을 최종 완료하였다.

체언 상세전자사전	SuperEntry	Entry	Sense
보완·정제	25,458	29,708	35,854

- 진행 방법 : 연구원 및 연구보조원들에게 상세기술 완료된 어휘 항목을 분배하고, 연구원들이 다음 두 단계의 보완·정제 작업을 수행하였다.

- 1단계 작업 : 센스별 의미부류 기술 내용 검토 및 보완·정제 작업
- 2단계 작업 : 보완·정제된 의미부류 항목의 기술 내용에 따른 통사정보구획의 기술 내용에 대한 보완·정제 작업

- 작업 결과 : 1단계 작업과 2단계 작업 모두가 100% 완료되었고, 그 결과 상기 도표에 제시된 바와 같이 35,854 항목의 체언 상세전자사전의 구축이 완료되었다.

1.1.2.2. 체언 상세전자사전 미시구조 및 구축 지침 보완

1.1.2.2.1. 미시구조 및 구축 지침 보완 개요

올해 수행되는 보완·정제 작업은 2006년도 보완된 사전 구축 지침에 따르는 것을 원칙으로 하되, 단계별 작업이 진행되기 전이나 작업 후에 기존 지침의 미비점을 보완하는 방식으로 작업이 진행되었다. 이에 따라 체언 상세전자사전의 미시구조와 구축지침에도 부분적인 보완작업이 진행되었다.

우선, 여러 하위사전들 간의 정보항목 표상의 통일성 제고를 위한 표준화 논의를 통해 미시구조 내 몇몇 정보항목들의 영문 tag와 한글 명칭이 다음과 같이 수정되었다.

- <sem> → <sem_class> (의미 부류)
- <n_n> → <comb_n> (명사 결합 정보)
- <n_aj> → <comb_aj> (형용사 결합 정보)
- <n_v> → <comb_v> (동사 결합 정보 구획)

- <max_n> → <max_n_str> (명사구 최대 구조)
- <sel_res> → <sel_rst> (선택 제약)

이와 함께 단계별 작업과정을 통해 구축지침이 보완·정제되었는데 그 내용은 다음과 같다.

(1) <sem_class> 항목 기술을 위한 상세 지침 보완

- ㉠ 표제어의 의미는 표제어가 술어명사인가 비술어명사인지에 따라 우선 구분한다.
- ㉡ 비술어명사의 의미는 <구체물>, <집단>, <장소>, <추상적대상>에 속하는 하위부류 중에서 선택하여 기술한다.
- ㉢ 이때, <구체물>과 <장소>의 구별과 <집단>과 <구체물>의 구별 및 의미분할에 특히 조심한다.
- ㉣ <구체물>관련 의미기술은 4개 하위부류 중에서 <속성구체물>과 <관계구체물>(N-의 N)을 적극 활용하여 기술한다.
- ㉤ 술어명사의 의미는 <사태>의 하위부류 중 하나를 골라 기술한다.
- ㉥ 표제어 의미가 <행위>나 <사건>에 속할 경우 이 두 의미 간의 규칙적 다의성을 고려하여 센스를 둘로 구별하여 기술한다.
- ㉦ 특히 동사결합 시 구문이 달라지고 논항 관계가 달라질 경우 센스를 구별하는 것을 원칙으로 한다.
 - 동사결합유형에서 '~이/가 V'인 경우 <사건> 혹은 <현상>로
 - 동사결합유형에서 '~을/를 V'인 경우 <행위>로
 예) “싸움”
 - ‘~을 벌이다/하다’와 결합 시 (ex: A와 B 사이에 싸움을 했다/벌었다) :
 ⇨ <행위> 부류 하의 <대칭적행위>로 기술
 - ‘~이 터지다/일어나다/나다’와 결합 시 (ex: A와 B 사이에 싸움이 터졌다/일어났다/났다) : ⇨ <사건> 부류 하의 <물리적충돌>
- ㉧ 단, 다음 예들에서처럼 구문의 차이에도 불구하고 동일한 사행에 속하는 경우에는 센스를 분할하지 않고 한 센스 내에서 기술한다.
 - ~을/를 하다 vs. ~을/를 시키다 : 결합, 배치, 배열 등
 - ~을/를 하다 vs. ~이/가 되다 : 이해, 대응 등
 - ~이/가 나다 vs. ~을/를 내다 : 용기
 예) “용기”
 - ‘~가 있다’와 결합 시 ⇨ <긍정적인성숙성값>
 - ‘~을 내다’와 결합 시 ⇨ <내재적심리행위>
 - ‘~가 나다’와 결합 시 ⇨ 별도의 센스 설정없이 <내재적심리행위> 부여
- ㉨ <추상적대상>에 속하는 의미를 기술할 때는 다음 세 유형을 잘 구별한다.
 - 명명적, 분류적 유형
 - 흔히 추상명사로 보는 단위

- 관계명사 유형 : N-의 ~
- ㉘ 비술어명사 부류인 <속성>과 술어명사 부류인 <속성값> 부류의 구별에 주의한다.
<속성>에 속하는 명사는 주로 다음 구문에 나타난다.
 - X-이 < >-이 <속성값>-이다
 - X-이 < >-이 <속성값>에 속하는 형용사
- ㉙ <속성>의 하위부류 중에서 많은 명사들이 “X-이 < >-이 있다/없다” 구문에 나타날 경우 <속성값> 하위부류의 센스를 따로 구별하여 기술하고, ‘있다/없다’는 기능동사로 기술한다.
예) 성격, 성깔
 - ‘~이 좋다/나쁘다’와 결합 시 ⇨ <속성>
 - ‘~이 있다/없다’와 결합 시 ⇨ <속성값>
- ㉚ <추상적대상>-<속성>의 하위에 있는 <능력>과 별도로, <정적사태>-<속성값>의 하위로 <능력속성값>을 추가하고, <능력>의 보기로는 ‘시력, 청력, 지능지수, 아이큐’ 정도만을 남긴다. 예를 들어 “능력, 힘”의 의미부류는 <능력속성값>이며, <속성값> 부류의 명사는 모두 술어명사므로, 이와 결합하는 ‘~이 있다/없다’는 기능동사로 기술한다.

(2) <syn_grp> 구획 기술을 위한 상세 지침

㉗ <comb_aj> 항목

- type 유형("appr", "magn", "min")에 맞게 기술되었는지 꼭 확인한다.
- “appr”에 “있다/없다”가 기술된 경우 다음과 같이 처리
 - <sem_class>이 <사태>의 하위부류일 때 :
⇨ <comb_v type="npred_vsup">로 이동, 관련사항 기술
 - <sem_class>이 <추상적대상>일 경우 :
⇨ "appr"에서 기술하되 <s_n>정보 기술 여부 꼭 확인
- “magn”에는 전형적인 언어구성 요소 외에도 ‘정도’나 ‘수준’의 극대 표현도 기재
 - ex) (시름)이 깊다, (벼슬)이 높다 vs (욕심)~이 ‘많다’|‘크다’

㉘ <comb_n> 구획

- type 유형, 즉 "lexfunc"와 "compound"의 적절성 검토
- "lexfunc" 유형은 <comb_aj>의 "magn", "min"에 해당되는 어휘들을 구별없이 기재하고, 그 외에도 표제어와 연어를 구성할 수 있는 어휘들을 기재
- 단, "lexfunc" 유형에 속하는 n_n 구성에는 n-의 n 구성도 포함됨에 유의

㉙ <comb_v> 구획

- type 유형이 의미부류에 맞게 설정되었는지를 우선적으로 확인한다.
 - <sem_class>이 <사태>의 하위부류일 때 :
⇨ <comb_v type="npred_X"> 유형 네 가지만 기술
 - <sem_class>이 <사태>이외의 부류일 때 :

- ⇒ <comb_v type="narg_X"> 유형 두 가지만 기술
- 그런 다음 X유형에 맞는 form이 기술되도록 확인 후 수정·보완
 - 표제어가 npred이고 X=vsup일 때 :
 - ⇒ 지침에서 정한 9개 기능동사 중 가능한 것을 기술한다.
 - cf. 주의사항 :
 - (a) ‘시키다’나 ‘주다’가 표제어와의 결합할 때 다른 기능동사 구문과 논항구조에서 차이를 보이면 이 항목에 기재하지 않는다. ‘시키다’의 경우 ‘하다’와 교체가 가능한 경우만 여기에 기술한다.
 - (b) vsup에 기술되어야 할 “있다|없다”가 누락되거나 <n_aj>에 잘못 기술되었는지를 꼭 확인한다.
 - 표제어가 npred이고 X=vsupvar일 때 :
 - ⇒ 사건기능동사, 기능동사 변이형을 모두 기재한다.
 - cf. 주의사항 :
 - (a) 사건기능동사나 변이형기능동사(어휘, 시상)를 vpred로 처리한 경우가 없도록 유의한다.
 - (b) vpred와 vsupvar 구별문제에 유의한다.
 - 표제어가 npred이고 X=vpred일 때 :
 - ⇒ 표제어 술어명사와 특징적으로 빈번히 결합하기는 하나, 술어명사가 표상하는 사태와 일차적인 연관성은 없고, 독자적 의미(술어기능)를 지닌 동사들을 기술
 - cf. 주의사항 : vsupvar와 구별문제
 - ex) 약속을 ‘깨다|파기하다|지키다|연기하다|앞당기다’
 - ⇒ vpred (‘하다’와 대체 불가, 대체 시 의미변화)
 - 표제어가 narg이고 X=vpred일 때
 - ⇒ 표제어를 전형적인 논항으로 취하는 술어동사로서 표제어와 의미적 상관성을 가지는 동사들을 기술한다.
 - cf. 주의사항 :
 - (a) 다면적 접근을 통해 결합 가능한 동사들을 기술한다.
 - ex1) 음식 :
 - 먹는 대상 ⇒ 먹는 방법과 관련된 동사
 - 인공물 ⇒ 준비과정 동사
 - ex2) 약기 :
 - 인공물 ⇒ 연주방법
 - (b) 특징적인 결합이 누락되지 않도록 검토, 보완한다.

㉞ <frame> 구획

- <npred_vsup>과 <npred_vsupvar> 두 항목에 대해서는 필수 기술
- 최대 논항구조로 기술 :
 - (a) 보문절 논항 필수 기술
 - (b) 술어의 내용 및 상황을 나타내는 필수적인 관형절 논항도 기술

- 계획 : X-이 S을 ~이다 : 철수가 유럽여행을 할 계획이다.
S겠다는 ~이다.
- 꿈 : X-이 S을 ~이 있다 : 영희가 유학을 갈 꿈이 있다.
- 소리 : X-이 S는 ~가 나다 : 차가 덜컹거리는 소리가 난다.
- 추억 : X-이 S은 ~이 있다 : 근석은 그녀와 즐겁게 뛰놀던 추억이 있다.
- ☞ 표기방법은 보문절 표기법을 따른다. (시상관련 형태소는 최대한 반영)
- 가능한 모든 구조를 <frame>을 반복하여 기술 (조사가 다를 때도 반복 기술)

㊤ <max_n_str> 구획

- <frame>에 기술된 것들을 바탕으로
- 가능한 모든 명사구 형태를 반복해서 기술
- <frame>에 나타난 보문절 논항의 관형형 형태를 필히 반영
cf) 주의사항 : <s_n>기술과 비교하여 통일

㊦ <sel_rst>

- 의미부류명으로 기재
- 광범위한 상위부류 하나를 기재하기 보다는 세밀한 부류를 열거하는 방식이 필요

㊨ <s_n>

- <sem_class>이 '사태' 이외의 부류로 기술된 경우에 수식절 정보를 기술한다.

1.1.2.2.2. 2007년도 체언 상세전자사전의 미시 구조

2007년도 체언 상세전자사전의 미시구조는 2006도까지 보완된 미시구조와 동일하다. 단, 사전 내에 기술된 정보들이 한국어 전산처리 시에 손쉽게 추출될 수 있도록 항목별 정보기술 방법을 보완하였다. 체언 상세전자사전의 미시구조는 크게 형태정보구획, 의미정보구획 및 통사정보구획으로 구성되는데, 여기에 표제어 항목을 체계적으로 관리하기 위한 관리정보구획이 추가되어 있다. 체언 상세전자사전의 미시구조는 다음과 같다.

ssuperEntry	최상위 표제항 구획
orth	표기 형태
entry+	표제항 구획
mnt_grp	관리 정보 구획
see?	참조 표제항
morph_grp	형태 정보 구획
var*	변이형
str?	내부 구조
org*	원어
hom?	동형어
der*	파생어 형성
comp*	합성어 형성

idm_grp?	속어 구획
idm*	속어
sense+	센스 구획
sem_grp	의미 정보 구획
eg+	응례
trans+	영어 대역어
domain*	전문 영역
sem_class	의미 부류
lr?	어휘 의미 관계 구획
syn*	동의어
ant*	반의어
syn_grp	통사 정보 구획
comb_aj*	형용사 결합 정보
comb_n*	명사 결합 정보
comb_v*	동사 결합 정보 구획
form	동사 결합 형태
frame* 문형	
max_n_str*	명사구 최대 구조
sel_rst*	선택 제약
cl*	단위 표현
prt*	조사 제약
av*	부사적 용법
s_n*	관형절 제약

1.1.2.2.3. 2007년도 체언 상세전자사전 구축 지침

○ 2007년도 체언 상세전자사전의 구축지침은 2004년도에 정비된 상세 기술 지침의 최종 보완본이라고 할 수 있다. 즉, 체언사전의 미시구조를 XML형식으로 변환함에 따라 정비된 지침을 그간의 신규기술 및 정보확장 작업을 통해 보완한 것이다. 2007년도에도 상기 기술한 바와 같이 기기기술 항목에 대한 전면적인 보완·정제 작업을 통해서 지속적으로 지침을 수정·보완해 왔다. 아래에서는 체언 상세전자사전의 최종 구축지침을 항목별 세부 지침을 뺀 핵심 지침의 형식으로 제시하기로 한다.

(1) superEntry (최상위 표제항 구획)

- 동형어 분할 이전의 표제항을 제시하는 구획이다.
- 동형어들을 포괄한 단위를 가리킨다.

(1-1) orth (표기 형태)

- 최상위 표제항의 표기 형태를 한글로 기입한다.

(예) <orth>새나라</orth>

(1-2) entry (표제항 구획)

- 동형어 구분이 된 표제항의 정보를 제시하는 구획이다.
- 동형어 수만큼 반복이 가능하다.
- 단일어, 복합어의 구분을 'pos' attribute의 value로 제시한다.

pos="nng_s" 단일어 명사 (강)
 pos="nng_c" 복합어 명사 (강물, 죽음)

(예) <orth>새나라<orth>
 <entry pos="nng_c">

- 일음절 한자어 어근으로 구성된 이음절 한자어의 경우는 단일어 명사로 분석한다.

(예) <orth>학교<orth>
 <entry pos="nng_s">

(1-2-1) mnt_grp (관리 정보 구획)

- 사전 원고 관리와 관련된 정보 항목들을 제시하는 구획이다.

(1-2-2) see (참조 표제항)

- 관련 entry가 있어서 거기에 제공된 정보가 이 entry의 정보와 동일할 경우를 기입한다.
- entry의 미시 구조는 비워 두고 이 항목에 관련 entry의 표기 형태를 기재한다.

(예) <orth>졸병</orth>
 <see>졸병</see>

(1-2-3) morph_grp (형태 정보 구획)

- 표제항에 대한 형태 정보를 제시하는 구획이다.

(1-2-3-1) var (변이형)

- 표제항의 변이형 형태를 기입한다.
- 변이형의 유형은 'type' attribute의 value로 제시한다.

type="xs"	표준어형인 변이형 ◦축약 또는 단축 관계 (노을 → 놀) ◦센말-여린말, 큰말-작은말 (빨강 → 발강 빨경)
type="xd"	방언형인 변이형 (부추 → 정구지)
type="xx"	어문 규범에 어긋하는 변이형 (골문 → 꼴문)

type="abb"	준말 변이형 (노동조합 → 노조)
type="lng"	본딧말 변이형 (대선 → 대통령선거)
type="spr"	띄어쓰기 변이형 (노동조합 → 노동 조합)

- 변이형의 수만큼 var 요소를 반복한다.

```
(예) <orth>골문</orth>
      <var type="xx">골문</var>

      <orth>사범대학</orth>
      <var type="abb">사범대</var>
      <var type="abb">사대</var>
```

(1-2-3-2) str (내부 구조)

- 표제항 명사가 복합어일 때 그 구성 형태를 기입한다.
- 합성어와 파생어의 유형은 'type' attribute의 value로 제시하고 표제항의 구성 형태를 '+'로 표시한다.
- 'type' attribute의 value는 다음과 같다.

<합성어>

type="N_N"	명사+명사 (가사+노동)
type="NsN"	명사+사이시옷+명사 (나무+ㅅ+잎)
type="R_N"	어근+명사 (국제+관계)
type="N_R"	명사+어근 (화물+선)
type="R_R"	어근+어근 (법조+계)
type="N_Vn"	명사+동사파생명사 (줄+넘기)
type="Vn_N"	동사파생명사+명사 (디딤+돌)
type="R_Vn"	어근+동사파생명사 (토+박이)
type="Vn+R"	동사파생명사+어근 (버팀+목)
type="Ad_N"	부사+명사 (따로+국밥)
type="Am_N"	형용사관형형+명사 (긋은+일)
type="Vm_N"	동사관형형+명사 (굳은+살)
type="fN_N"	어두공통 명사+명사 (수출+수입 : 수출입)
type="N_Nf"	어말공통 명사+명사 (광업+공업 : 광공업)
type="FLW"	사자성어 (이목구비)

<파생어>

type="N_S"	명사+접미사 (계집+질)
type="R_S"	어근+접미사 (항구+적)
type="V_S"	동사+접미사 (죽+ㅁ)

type="A_S" 형용사+접미사 (높+이)
 type="P_N" 접두사+명사 (맨+손)
 type="P_R" 접두사+어근 (초+강경)

<그 외>

type="etc" 상기 value에 속하지 않는 합성어 및 파생어

(예) <orth>가사노동</orth>
 <str type="N_N">가사+노동</str>

 <orth>나뭇잎</orth>
 <str type="NsN">나무+ㅅ+잎</str>

- type 값이 'FLW'일 경우에는 내부 구성 형태로 별도로 기재하지 않고 비워둔다.

(예) <orth>이목구비</orth>
 <str type="FLW"> </str>

- 접미사 'ㅁ'과 '음'의 경우는 단형 'ㅁ'으로 통일한다.
- 내부 구조 분석은 직접성분(IC) 분석에 따라 이분지로만 분석한다.

(예) <orth>긴급연락망</orth>
 <str type="N_N">긴급+연락망</str>

- 이음절 한자어가 두 개의 한자어 어근으로 구성되었을 경우에는 단일어 명사로 간주하므로 내부구조 분석을 하지 않는다. (ex: 학교, 합선, 전함 등)
- 내부 구조 기술을 위한 어근 및 접사 판별은 세종 어근/접사 전자사전의 목록에 준거한다.

(1-2-3-3) org (원어)

- 표제항의 전체 또는 일부가 외래어(한자어 포함)일 경우, 그 외래 요소의 기원이 되는 언어 및 그 언어에서의 표기 형태를 기입한다.
- 외래 요소의 기원이 되는 언어는 'lg' attribute의 value로 기재하되 한자어(si), 영어(en), 기타(etc)의 셋으로만 구분한다.

lg="si" 한자어 (왕권)
 lg="en" 영어 (요트)
 lg="etc" 기타 (앙코르)

(예) <orth>문</orth>
 <org lg="si">門</org>

- 외래 요소의 해당 언어에서의 표기 형태는, 한자어의 경우는 한자로, 영어의 경우는

- 로마자로 표기하고, 기타의 경우는 한글로 표기한다.
- 표제항의 일부만 외래어인 경우 고유어에 해당하는 부분은 밑줄(_)로 표시한다.

(예) <orth>콘크리트길</orth>
<org lg="en">concrete_</org>

- 둘 이상의 외래 요소를 포함하고 있고 이들의 기원 언어가 다른 경우 org 요소를 반복해서 표시해 준다.

(예) <orth>골문</orth>
<org lg="en">goal_</org>
<org lg="si">_門</org>

(1-2-3-4) hom (동형어)

- 해당 표제어의 동형 표제어가 존재할 경우 'exist' attribute의 value를 'yes'로 기입한다.

(예) <orth>김</orth>
<hom exist="yes"/>

- 동형어 존재 여부는 동일 범주 어휘는 물론이고 타범주 어휘의 존재 여부까지 고려하여 기술한다.

(1-2-3-5) der (파생어 형성)

- 표제항 명사를 기어로 하는 파생어의 품사 정보와 결합 요소의 유형 정보를 기입한다.
- 품사 정보는 'pos' attribute의 value로, 결합 요소의 유형 정보는 'type' attribute의 value로 기재한다.

(예) <orth>가위</orth>
<der pos="n" type="suffix">~질</der>

(1-2-3-6) comp (합성어 형성)

- 표제항 명사를 구성 요소로 하는 합성어의 품사 정보와 결합 요소의 위치 정보를 기입한다.
- 품사 정보는 'pos' attribute의 value로, 결합 요소의 위치 정보는 'type' attribute의 value로 기재한다.

(예) <orth>가마</orth>
<der pos="n" type="prefix">꽃~</der>

(1-2-4) idm_grp (숙어 구획)

- 속어 관련 정보를 제시하는 구획이다.

(1-2-4-1) idm (속어)

- 표제항 명사를 포함하는 속어 표현이 있을 경우 속어 유형과 해당 형태를 기입한다.
- 속어의 유형은 'type' attribute의 value로 제시한다.

type="np"	명사+조사
type="na"	명사+형용사
type="nv"	명사+동사
type="da"	명사+이다
type="prv"	속담
type="etc"	기타

- 속어의 수만큼 idm 요소를 반복해서 제시한다.

(예) <orth>도둑</orth>
 <idm type="prv">~의 씨가 따로 없다</idm>
 <idm type="da">~이야</idm>

(1-2-5) sense (센스 구획)

- 표제항이 갖는 각각의 센스를 구별하여 의미 정보 및 통사 정보를 제시하는 구획이다.
- 의미 구분 표지를 n="1"처럼 지수화하여 기입한다.

(1-2-5-1) sem_grp (의미 정보 구획)

- 센스 분할된 표제항의 각종 의미 정보를 제시하는 구획이다.

(1-2-5-1-1) eg (용례)

- 센스 분할된 표제항 명사가 등장하는 예문을 기입한다.
- 표제항 명사가 격조사와 결합된 형태의 문장을 제시한다.
- 표제항 명사의 의미가 가장 잘 드러나는 전형적인 문장을 제시한다.

(예) <orth>도둑</orth>
 <eg>어제 우리집에 도둑이 들었다.</eg>

(1-2-5-1-2) trans (영어 대역어)

- 센스 분할된 표제항 명사의 영어 대역어를 기입한다.
- 분할된 해당 센스에 대응되는 대역어들만 기입한다.

- 이 항목은 센스 분할의 근거로 사용한다.

(1-2-5-1-3) domain (전문 영역)

- 표제항 명사의 해당 의미가 특정 영역에서 사용될 때 그 영역을 기입한다.
- 전문 영역 표기는 세종 전자사전 전문 영역 체계의 영역명을 사용한다.

(예) <orth>집</orth>
<domain>바둑</domain>

(1-2-5-1-4) sem_class (의미 부류)

- 센스 분할된 표제항 명사의 해당 의미를 기입한다.
- 의미는 세종 의미 부류 체계의 의미 부류명으로 표기한다.

(예) <orth>비행기</orth>
<sem_class>공중교통수단</sem_class>

- 센스당 하나의 의미 부류만을 기재하고, 두 개 이상의 의미 부류가 필요할 경우는 센스를 구분한다.

(1-2-5-1-5) lr (어휘 의미 관계 구획)

- 표제항과 밀접한 의미 관계에 있는 어휘들을 제시하는 구획이다.

(1-2-5-1-5-1) syn (동의어)

- 센스 분할된 표제항의 동의어를 기입한다.
- 동의어 유형을 분류하여 다음과 같이 'type' attribute의 value로 제시한다.

type="perfect"	완전 동의어
type="sino"	고유어 표제항에 대한 한자어 동의어
type="kor"	한자어 표제항에 대한 고유어 동의어
type="honor"	비존칭 표제항에 대한 존칭 동의어
type="de_honor"	존칭 표제항에 대한 비존칭 동의어
type="under"	비겸양 표제항에 대한 겸양 동의어
type="de_under"	겸양 표제항에 대한 비겸양 동의어
type="spoken"	非구어 표제항에 대한 구어 동의어
type="de_spoken"	구어 표제항에 대한 非구어 동의어
type="vulgar"	非속어 표제항에 대한 속어 동의어
type="de_vulgar"	속어 표제항에 대한 非속어 동의어
type="etc"	기타

- 동의어가 둘 이상 있을 경우에는 syn 요소를 반복 기재한다.

(예) <orth>구입</orth>
 <syn type="perfect">매입</syn>
 <syn type="perfect">구매</syn>
 <syn type="perfect">매수</syn>

(1-2-5-1-5-2) ant (반의어)

- 센스 분할된 표제항의 반의어 기입한다.
- 반의어 유형을 분류하여 다음과 같이 'type' attribute의 value로 제시한다.

type="compl"	상보적 반의어 (삶-죽음)
type="grad"	정도성 반의어 (더위-추위)
type="dir"	방향성 반의어 (상승-하강)
type="conv"	전환적/관점적 반의어 (판매-구매)
type="etc"	기타 (해-달)

- 반의어가 둘 이상 있을 경우 ant 요소를 반복 기재한다.

(예) <orth>판매</orth>
 <ant type="conv">구매</ant>
 <ant type="conv">매입</ant>

(1-2-5-2) syn_grp (통사 정보 구획)

- 센스 분할된 표제항 명사의 각종 통사 정보를 제시하는 구획이다.

(1-2-5-2-1) comb_aj (형용사 결합 정보)

- 센스 분할된 표제항 명사와 특징적으로 결합하는 형용사를 기입한다.
- 결합 가능한 형용사의 유형을 다음과 같이 'type' attribute의 value로 제시한다.

type="appr"	명사+적정 형용사 (밥-되다, 질다)
type="magn"	명사+극대 형용사 (세금-무겁다, 거짓말-새빨간)
type="min"	명사+극소 형용사 (두통-가볍다, 가능성-희박하다)

- 해당 어휘가 둘 이상 있을 경우 요소를 반복 기재한다.

(예) <orth>가능성</orth>
 <comb_aj type="magn">~이 높다</comb_aj>
 <comb_aj type="magn">~이 크다</comb_aj>
 <comb_aj type="min">~이 낮다</comb_aj>
 <comb_aj type="min">~이 희박하다</comb_aj>

(1-2-5-2-2) comb_n (명사 결합 정보)

- 센스 분할된 표제항 명사와 특징적으로 결합하는 명사를 기입한다.
- 결합 가능한 명사의 유형을 다음과 같이 'type' attribute의 value로 제시한다.

type="lexfunc"	표제항을 핵으로 하는 언어를 구성하는 명사 (흥분의 도가니)
type="compound"	표제항과 결합, 복합명사와 비슷한 성격을 띠는 명사구를 구성하는 명사 (가사노동 vs 가사노동시간)

- 해당 어휘가 둘 이상 있을 경우 요소를 반복 기재한다.

(예) <orth>절망</orth>
<comb_n type="magn">~의 높</comb_aj>
<comb_n type="magn">~의 구렁텅이</comb_aj>

(1-2-5-2-3) comb_v (동사 결합 정보 구획)

- 센스 분할된 표제항 명사와 특징적으로 결합하는 동사와 관련된 각종 정보를 제시하는 구획이다.
- 결합 가능한 동사의 유형을 다음과 같이 'type' attribute의 value로 제시한다.

type="npred_vsup"	술어명사 + 기능동사 (공부를 하다)
type="npred_vsupvar"	술어명사 + 기능동사 변이형 (아양을 떨다)
type="npred_vpred"	술어명사 + 일반동사 (약속을 지키다)
type="npred_vcognate"	술어명사 + 동족동사 (잠을 자다)
type="narg_vcognate"	논항명사 + 동족동사 (신을 신다)
type="narg_vpred"	논항명사 + 일반동사 (밥을 짓다)

- 기능동사는 '하다, 되다, 시키다, 받다, 당하다, 주다, 있다, 없다, 이다'로 한정한다. 즉 이들과 결합하는 동사 문형은 "npred_vsup"로 기입한다.
- 기능 동사 변이형은 다음과 같은 유형이 있다.

어휘변이형 : (아양)을 떨다, (심술)을 부리다, (명령)을 내리다, (인사)를 드리다
시상변이형 : 계속하다, 그치다, 멈추다 등
존칭변이형 : 드리다, 올리다 등

- 결합 가능한 동사 수만큼 요소를 반복해서 기술한다.

(1-2-5-2-3-1) form (동사 결합 형태)

- 표제항 명사와 결합 가능한 동사의 형태를 기입한다.
- 해당 동사가 조사 없이 표제항 명사와 결합하여 합성동사를 구성할 수 있는지의 여부를 'vcompound' attribute의 value로 기재한다.

vcompound="yes" 명사+동사 합성어 가능
 vcompound="no" 명사+동사 합성어 불가능

```
(예) <orth>보초</orth>
      <comb_v type="npred_vsupvar">
        <form vcompound="no">~을 서다</form>

      <orth>공부</orth>
      <comb_v type="npred_vsup">
        <form vcompound="yes">~를 하다</form>
```

(1-2-5-2-3-2) frame (문형)

- 명사+ 기능동사(변이형) 결합이 취하는 논항들 및 그 형태를 기입한다.
- 명사 논항은 X, Y, Z 등으로 표시하고 결합하는 격조사를 다음과 같은 대표 형태를 사용한다. 그리고 복합조사는 붙여쓰는 것을 원칙으로 한다.

격	대표형	대표형으로 포괄되는 이형태
주격	이	가
대격	을	를, ㄹ
여격	에게	한테
구격	로	으로
공동격	와	과
탈격	에서	(으)로부터
	에게서	한테서
	에게	한테
명명격	라고	이라고
	에대해	에대하여, 에대해서
	을위해	을위하여, 을위해서

- 절 논항은 'S고, S것, S읍, S기, S데' 등으로 표기한다.
- 관형절 논항은 보문절 논항의 관형형을 제외하고 기술한다.
- 표제항 명사가 해당 동사와 구성할 수 있는 여러 형태의 문형을, frame 요소를 반복하여 기입한다.

(예) <orth>참견</orth>
 <comb_v type="npred_vsup">
 <form vcompound="yes">~을 하다</form>
 <frame>X-이 Y-에 ~을 하다</frame>
 <frame>X-이 S데-에 ~을 하다</frame>
 <frame>X-이 S것-에 ~을 하다</frame>

(1-2-5-2-4) max_n_str (명사구 최대 구조)

- 술어명사가 자신의 논항들을 명사구 내에서 실현시킬 때의 최대 구조를 기입한다.
- 둘 이상의 명사구 최대 구조가 가능할 경우 요소를 반복한다.

(예) <orth>사랑</orth>
 <max_n_str>X-의 Y ~</max_n_str>
 <max_n_str>X-의 Y-에대한 ~</max_n_str>

 <orth>명령</orth>
 <max_n_str>X-의 Y-에대한 S라는 ~</max_n_str>

(1-2-5-2-5) sel_rst (선택 제약)

- 표제어 술어명사가 논항 X, Y, Z, W 등에 가하는 선택 제약을 기입한다.
- 각 논항이 어떤 의미 영역에 속하는지를 세종 의미 부류 체계의 의미 부류명으로 기입하되, 논항 자리에 나타날 수 있는 명사가 극히 제한적일 경우는 의미 부류명 옆에 괄호를 하고 해당 명사 어휘를 병기한다.
- 여러 의미 부류를 명기할 필요가 있을 때는 ‘|’를 써서 병기한다.
- 이 항목은 명사 논항에 한하며, 논항의 유형은 ‘arg’ attribute의 value로 제시한다.

(예) <orth>공부</orth>
 <sel_rst arg="X">인간|인간집단(학교|교회)</sel_rst>
 <sel_rst arg="Y">학문과목</sel_rst>

(1-2-5-2-6) cl (단위 표현)

- 해당 명사와 결합하는 단위 표현을 기입한다.
- 단위 표현의 유형은 ‘type’ attribute의 value로 제시한다.

type="uni"	낱개 (명, 분, 사람, 마리, 그루 ...)
type="grp"	묶음 (컬레, 벌, 손, 접, 축, 툃, 다발 ...)
type="div"	부분 (조각, 쪽, 토막 ...)
type="qnt"	양 (컵, 접시, 통 ..)

- 단위 표현의 유형과 수만큼 요소를 반복하여 기재한다.

(예) <orth>계란</orth>

```
<cl type="uni">개</cl>
<cl type="uni">알</cl>
<cl type="grp">판</cl>
```

(1-2-5-2-7) prt (조사 제약)

- 표제항 명사와 결합할 수 있는 조사가 제약되어 있는 경우 결합 가능한 조사를 기입한다.

(예) <orth>최근</orth>
<prt>~에</prt>
<prt>~의</prt>

- 제약이 없을 경우 빈칸으로 둔다.

(1-2-5-2-8) av (부사적 용법)

- 표제항 그대로, 또는 조사가 붙어 부사적으로 쓰일 수 있는 경우 해당 형태를 기입한다.

(예) <orth>사실</orth>
<av>~</av>

- 해당 형태의 수만큼 요소를 반복해서 제시한다.

(예) <orth>진정</orth>
<av>~</av>
<av>~으로</av>

(1-2-5-2-9) s_n (관형절 제약)

- 비술어명사인 표제어가 관계절 외의 특정 형태 관형절과 결합할 때 그 형식을 기입한다.

(예) <orth>사실</orth>
<s_n>S다는 ~<s_n>

- 비술어명사인 표제어가 어떤 관형절 수식도 받을 수 없을 경우에는 'yn' attribute를 사용하고 그 value를 'no'로 기재한다.

(예) <orth>간발</orth>
<s_n yn="no"/>

1.1.2.3. 목록 관리 및 형식 검증

이 작업은 단계별 보완·정제 작업과정에서 추가 혹은 제외되는 표제어 항목들을 체언상세전자사전 내에서 빠짐없이 반영함으로써, 기술항목들을 체계적으로 관리하고자 하는 목적으로 수행되었다.

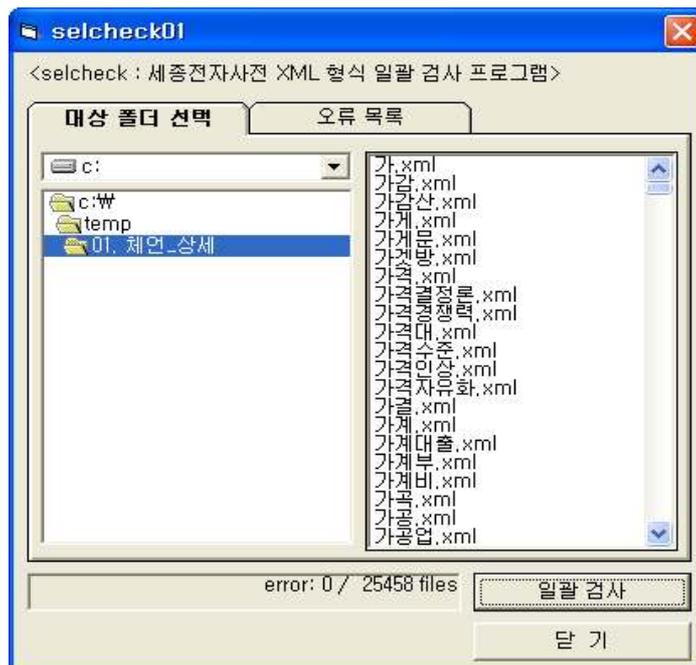
⇒ 기술 목록은 결과물 CD 참조

한편 목록 관리는 형식 검증과 함께 그 결과물 수치를 확인하는 과정을 거치게 되는데 그 과정은 다음과 같다.

1.1.2.3.1. 체언 사전의 형식 검증

체언 사전 분과에서는 체언상세사전의 기술이 사전에 정의한 미시 구조에 맞는지를 검증하기 위해, dtd를 작성하고 이를 기계적으로 검증하는 방식을 사용하였다. 구체적인 검증에는 세종 전자사전 개발 분과에서 자체 개발한 selcheck이라는 프로그램을 이용하였다. 검증 과정은 다음과 같다.

- ㄱ. selcheck 프로그램을 구동시킨다.
- ㄴ. 체언상세 원고가 들어 있는 폴더를 클릭한 후, 일괄검사를 실시한다.



위 화면은 형식 오류가 없는 경우이다. 만약 형식 오류가 있으면 아래와 같이 오류가 있는 파일을 알려준다.



ㄷ. 형식 오류가 있는 경우 이를 수정하고 (ㄴ)과 같이 검증 과정을 다시 거친다.

1.1.2.3.2. 체언 사전 superEntry, entry, sense 항목 수 검증

체언분과에서는 Unix 명령어를 이용하여 superEntry, entry, sense의 개수를 검증하였다. Unix 명령어 중 grep과 wc가 이용되었다. grep은 특정 라인을 찾아주는 기능이 있고, wc는 해당 라인을 카운팅해주는 기능이 있다.

ㄱ. superEntry

```
C:\Wtemp\W01. 체언_상세>grep "</superEntry>" *.xml | wc  
25458 25466 621738
```

ㄴ. entry

```
C:\Wtemp\W01. 체언_상세>grep "</entry>" *.xml | wc  
29708 59417 629734
```

ㄷ. sense

```
C:\Wtemp\W01. 체언_상세>grep "</sense>" *.xml | wc  
35854 71710 790831
```

결과적으로 25458개의 superEntry, 29708개의 entry, 35,854개의 sense가 기술되어 있음을 확인할 수 있었다.

1.2.4. 사전 구축 관련 기반 연구

2007년도 체언사전분과의 기반 연구는 용언 전자사전 관련 항목의 기술 형식 및 내용에 대한 통일성 제고하는데 중점을 두어 수행되었다. 특히, 체언사전과 용언사전의 의미부류 정보(sem_class)와 논항선택제약 정보(sel_rst)의 유기성 확보를 위해 의미부류 체계를 수정하고 보완하는 작업이 수행되었는데, 그 구체적 내용은 다음과 같다.

- (1) <관계속성값> 하에 <상이관계속성값>, <유사관계속성값>, <우호관계속성값>, <적대관계속성값> 등의 하위부류를 설정한다.
 - <상이관계속성값> : 상이하다, 다르다, 이질적이다
 - <유사관계속성값> : 같다, 유사하다, 비슷하다, 가깝다

- <우호관계속성값> : 친하다, 절친하다, 가깝다, 친분(이 있다)
- <소원적대관계속성값> : 멀다, 소원하다, 견원지간이다

(2) 이와 함께 <중성적관계속성값> 설정도 고려할 수 있음.

⇒ 설정근거 : <관계>에 속하는 명사들이 '있다'와 결합하는 경우의 기술 시 필요함.

(ex) X-가 Y-와 (인연, 관계)-가 있다

(3) [속성] > [크기] > [길이], [넓이], [무게]를 [길이속성], [넓이속성], [무게속성]으로 변경을 검토.

(4) [단위] > [길이], [넓이], [무게]의 부류명 변경 검토 :

⇒ [길이단위], [넓이단위], [무게단위]

(5) [단위] > [부피], [액체부피]의 부류명 변경 검토 :

⇒ [부피단위], [액체부피단위]

(6) [정도속성값]과 병행해서 [수량속성값]을 설정하고 그 하위부류에 [대], [중], [소]를 설정하는 방안 검토.

(7) <대칭적행위>를 <물리적대칭행위>로 변경하고, 그 하위부류로 <만남>외에 <이별>, <물리적충돌행위>를 설정.

(8) 이와 병행하여 <추상적행위> 하에 <추상적대칭행위>를 설정하고 그 하위부류로 <교류>, <협정>, <화해>를 설정을 검토. 단, <교류>, <협정>, <화해>를 추상적행위의 범주에 넣을 수 있을 지를 아울러 고려할 필요가 있음.

1.1.2.5. 회의

☞ 결과물 CD의 『회의록』 참조

1.2. 용언

1.2.1. 2007년도 연구 계획 및 추진 결과

1.2.1.1. 개요

용언사전분과의 2007년도 사업은, 세종전자사전 개발의 마지막 해인 만큼, 사전의 양을 늘리기보다는 포함된 정보의 정확성과 질을 높이고 전체적인 통일성을 높이는 데 중점을 두었다. 또한 자연언어처리에서 사전의 실제적인 효용성을 높이기 위한 구체적인 작업을 수행하였다.

첫번째로 역점을 둔 것은 용언의미부류 정보의 추가이다. 사업의 마지막 단계에 와서 정보항목을 추가한다는 것이 사업 진행에 상당한 부담이 될 수도 있으나, 용언의미부류 정보가 부사, 보조용언, 어미 등과의 결합 제약을 정확하게 기술하는 데에 필수불가결한 정보이기 때문에, 늦게나마 이 정보를 추가하기로 했다. 그래서, 우선 명사의미부류 체계 중 [사태] 부류의 하위분류 체계를 출발점으로 삼아 기초적인 작업을 수행하였는데, 그 결과 명사에서는 제기되지 않았던 많은 문제들이 제기되어 훨씬 더 상세하고 세분화된 용언의미부류 체계를 구성하게 되었다. 이렇게 해서 구성된 용언의미부류 체계 및 지침을 바탕으로 하여, 지금까지 작성된 모든 용언 항목에 대해 의미부류 정보를 부착하였다.

두번째로 역점을 둔 것은 선택제약 정보의 보완이다. 용언사전의 정보항목 중 자연언어처리에서 당장 요긴하게 써먹을 수 있는 핵심 정보항목은 문형 정보와 선택제약 정보라 할 수 있다. 문형 정보는 지침이 매우 정교화되어 있어서 사전 기술자 간의 변이의 폭이 크지 않았고, 여러 차례의 검토 및 형식검증 작업을 통해 상당한 정도로 정밀화, 세련화되어 있다고 자부할 수 있다. 반면에 선택제약 정보는, 명사의미부류 체계 자체가 계속 수정·보완되어 왔고, 그 체계의 방대함으로 말미암아 사전 기술자가 완전히 숙지하여 사전 기술에 적용하기에 벅찬 면이 있어서, 기술자 간의 변이의 폭도 크고 보완되어야 할 측면이 많이 있었다. 여러 차례 검토 작업을 거치기는 했으나, 구문분석 등의 실제적인 활용을 염두에 두기보다, 순수하게 언어학적인 측면에서 기술의 타당성을 검토하는 데에 머물렀었다. 그래서 올해에는, 구문분석이라는 매우 구체적인 활용 영역에 초점을 맞추어, 기존 용언사전의 선택제약 정보의 정확성 및 활용도를 점검하였다. 세종전자사전 분과에서 세종전자사전의 전산적 활용성 검증을 위해 자체 개발한 구문분석기를 이용하여 세종용언사전에 수록된 용례들을 구문분석하였는데, 이 과정에서 input 문장에 등장하는 논항 명사의 체언사전 내 의미부류 정보와 그 명사를 논항으로 취하는 용언의 용언사전 내 선택제약 정보가 일치하는지를 점검하였다. 이 두 정보가 일치하지 않는 경우를 살펴보고, 체언사전에 수록되어야 할 명사들을 찾아내고, 용언사전의 선택제약 정보가 잘못되었거나 미진한 경우 보완하는 식으로 작업을 진행하였다. 이러한 작업은 구문분석기의 성능을 향상시키는 데에도 도움이 될 것이다.

세번째로 역점을 둔 것은 의미역 정보의 보완이다. 기존 용언사전 원고에서 의미역과

관련하여 가장 큰 문제점은 일관성의 부족이었다. 특정 용어의 특정 논항에 대해 어떤 의미역을 부여할지 애매한 경우가 많이 있는데, 비슷한 문제를 야기하는 많은 용어들에 대해 일관된 지침에 따라 일관되게 의미역을 부여하는 것이 중요하다. 그런데 기존 작업에서는 여러 연구원이 분담하여 작업을 하다 보니, 그런 일관성을 기하기가 어려웠던 것이다. 그래서 올해에는 용어분과 연구원 중 한 명이 의미역을 전담하여, 용어사전 원고 전체에 걸쳐서 의미역 정보의 일관성과 정확성을 높이는 작업을 수행하였다.

의미역 정보 수정 작업과 기타 작업(용어의미부류 및 선택제약 수정 작업)이 별도로 동시에 진행되었기 때문에, 나중에 이 두 작업의 결과를 하나로 통합해야 했다. 이 일을 기계적으로 수행하기 위해서는 두 결과물의 구조가 일치해야 한다. 그런데, 작업 수행 과정에서 문형정보에 논항을 추가하거나 삭제한 경우도 있었고, 새로운 subsense, frame, sense, entry 등을 추가하여 사전의 구조가 달라지는 일도 종종 있었다. 이런 경우 각 작업자가 명시적으로 기록을 해 두었고, 이 기록을 바탕으로 두 사전의 구조를 일치시키는 작업을 수작업으로 수행하였으며, 그 뒤에 두 작업 결과를 기계적으로 통합하였다.

1.2.1.2. 연구 내용별 추진 결과

분 과 명	연구 내용	추진 일정						추진 현황		
		6 월	7 월	8 월	9 월	10 월	11 월	12 월	진척 상황	진척율
용 언 사 전 분 과	▶ 기성 결과물 보완 정제: 27,700 항목								29,512항목 완료	107%
	◦ 의미부류 및 선택계약 27,700항목									107%
	◦ 의미역 27,700항목									107%
	▶ 용언전자사전 기술목록 검토 및 정비			▶						100%
	▶ 용언전자사전 구축지침서 보완 및 정제				▶					100%
	▶ 용언의 범주체계 확립 및 복합용언의 내 부구조 유형화로 한국어 용언의 총체적 형태구조에 대한 연구기반 마련									100%
	▶ 문형/의미역/선택계약 정보 표상 방식 개 선 연구 및 지침 보완·완성									100%
	▶ 체언 전자사전 관련 항목의 기술 형식 및 내용에 대한 통일화 방안 연구									100%

1.2.2. 용언 전자사전 연구 수행 내역

1.2.2.1. 용언 전자사전 보완·정제 작업

2007년에는 27,700항목에 대해 기성 결과물을 보완·정제하는 것을 목표로 하였으며, 이것을 의미역 정보 수정 작업과 기타 정보(용언의미부류 및 선택계약에 중점을 둔) 수정 작업으로 나누어 동시 진행하였다.

의미역 정보 수정 작업 및 기타 정보 수정 작업 둘 다 29,512개(107%)의 sense에 대한 작업이 완료되었다. 동사는 23,268개, 형용사는 6,244개이다.

용언 상세전자사전		SuperEntry	Entry	Sense
의미역 및 기타 정보 수정 작업	동사	15,180	16,530	23,268
	형용사	4,398	4,530	6,244
	합계	19,578	21,060	29,512

보완·정제 결과물은 결과물 CD 참조

1.2.2.1.1. 의미역 정보 보완·정제 작업

의미역 정보 수정 작업은 다음 사항에 중점을 두고 진행되었다.

- 기존 사전 원고에서 의미역 관련 지침 숙지의 미흡으로 인한 기술의 오류를 수정하고 보완한다.
- 의미역과 관련하여 논란이 될 수 있는 문제 항목들을 찾아내어 유형화한다.
- 유형화된 문제 항목들 각각의 유형에 대해 처리 지침을 마련하고, 이것을 상세하게 기록한다.
- 지침에 따라 문제 항목들에 대해 일관되게 의미역 정보를 부여한다.

1.2.2.1.2. 용언의미부류 부착 작업

용언의미부류는 용언과 부사, 본용언과 보조용언, 용언과 어미의 결합 제약을 기술하는 데에 매우 유용한 정보이다. 예컨대, ‘새삼’은 [인식]의 의미를 갖는 용언만 수식할 수 있고, “어떤 느낌이 느껴지다”를 의미하는 ‘-어 오-’는 감각 용언과만 결합할 수 있다. 이런 정보가 체계적으로 사전에 기술되면 자연언어처리에서 요긴하게 쓰일 수 있다. 그래서 용언의미부류 정보를 추가하기로 결정하였다.

용언 의미부류를 부착 작업은 다음 사항에 중점을 두고 진행되었다.

- 명사의미부류 체계 중 [사태]의 하위 부류 체계를 작업의 출발점으로 삼는다.
- 기초 작업을 통해, 명사의미부류 체계 중 [사태]의 하위 부류 체계의 미비점을 발견하여 새로운 부류를 추가함으로써 이를 보완한다.
- 새로 추가되는 부류의 정의, 판정 기준, 예 등을 풍부하게 기술한다.
- 문제가 되는 부류의 처리 지침을 마련한다. 예: 피동적행위, 사동적행위

이 작업을 위해 용언의미부류 체계를 구축하는 기반연구를 수행하였다.

☞ 용어의미부류 체계 구축에 대해서는 『1.2.2.4.1. 용언의미부류 관련 연구』를 참조

1.2.2.1.3. 선택제약 보완·정제 작업

선택제약 수정 작업은 다음과 같은 순서로 진행되었다.

- 세종용언사전에 수록된 약 5만개의 용례들을 추출하여, 맞춤법·띄어쓰기 등의 오류를 수정한다. 과도하게 붙여 쓴 복합명사는 띄어쓰기의 오류는 아니지만 구문분석시에 문제를 야기할 수 있으므로 구성 단위별로 띄어 쓴다.

- 구문분석기 개발 소분과에서 개발한 구문분석기로 위의 용례들을 구분분석한다.
- 구문분석기의 output을 검토하되, input 문장에 등장하는 논항 명사의 체언사전 내 의미부류 정보와 그 명사를 논항으로 취하는 용언의 용언사전 내 선택제약 정보가 일치하는지에 초점을 맞추어 점검한다.
- 이 두 정보가 일치하지 않는 경우, 이런 현상이 발생하게 된 원인을 찾는다.
- 체언사전에 수록될 만한 명사인데 수록되지 않은 것들은 기록해 두었다가 체언사전 분과에 전달하여 체언사전에 수록될 수 있도록 한다.
- 체언사전의 명사의미부류 정보가 잘못된 경우 기록해 두었다가 체언사전 분과에 전달한다.
- 용언사전의 선택제약 정보가 잘못되었거나 미비한 경우, 용언사전 원고를 수정한다.
- 용례에 출현한 명사가 해당 용언의 논항으로서 매우 빈번히 쓰인다고 판단되면, 그 명사를 괄호 안에 제시하고, 그 명사가 속하는 의미부류를 괄호 앞에 써 준다.

◦ 예: ‘가결되다’의 주어 논항 ‘안건’, ‘법안’ 등

- 해당 명사가 해당 용언의 논항으로서 그리 빈번히 쓰이는 것은 아니지만, 그 명사가 속한 의미부류가 해당 용언의 선택제약 정보로서 유용하다고 판단되면, 그 의미부류를 써 준다.
- 용례에 출현한 명사가 고유명사라서 논항 매칭이 안 된 경우, 그 고유명사가 속하는 의미부류를 검토하여, 해당 용언의 선택제약 정보로서 유용하다고 판단되면, 그 의미부류를 써 준다.
- 구문분석 결과에 오류가 있을 경우, 특히 어떤 명사가 어떤 용언의 논항인데 다른 용언의 논항으로 판정된 경우, 용례를 간략화하여 구문분석이 올바른 결과를 낼 수 있게 만든 뒤에 선택제약을 검토한다.

◦ 예1: ‘가능되다’의 용례

⇒ ‘우리는 일에 대한 사회적 중요도의 여부가 돈으로 가능되는 사회에 살고 있다.’에서 ‘여부가’가 ‘가능되다’의 주어로 판정되지 않고 있음.

◦ 예2: ‘가다’의 용례 ‘할멈은 작년에 먼저 갔고 이제 내 차례지.’

- 용례에 맞춤법이나 띄어쓰기의 오류가 있는 경우, 이것을 수정한 뒤 구문분석 결과를 검토한다.

◦ 예1: ‘가느다래지다’의 용례 ‘나무가지가 갈수록 가느다래진다’

◦ 예2: ‘가다’의 용례 ‘철수는 무작정 부산으로 가는 버스를 탔다.’

◦ 예3: ‘가누다’의 용례 ‘지하철역에서 교실까지 한숨에 달려온 슬기는 숨을 가누려고 책상에 엎드려 있었다.’

- 띄어쓰기의 오류라고는 할 수 없지만 복합명사를 붙여썼기 때문에 체언사전에서 해당

명사를 찾지 못한 경우가 많이 있다. 이런 경우에는 해당 복합명사를 띄어 썼을 경우의 구문분석 결과를 검토한다.

◦ 예: 밤샘간호, 진행상황

- 본용언 보조용언 구성도 마찬가지로이다.

◦ 예: '가능하다'의 용례

⇒ '이번 기회는두 사람의 실력을 가늠해볼 수 있는 좋은 기회이다.'

- 접미사 등의 형태 정보를 활용하는 방안: '가결되다'의 주어의 경우 '파병안', '결의안', '정부안', '여당안'처럼 '-안'으로 끝나는 명사들이 많은데, 이들 명사 모두를 체언사전에 신기도 어렵고 용언사전의 선택제약 정보로 제시하기도 어렵다. 이런 경우 용언사전 선택제약 구획에 '-안' 같은 식으로 표시하여, 구문분석시에 '-안'으로 끝나는 명사를 만나면 이 선택제약 정보와 매칭되는 것으로 판정할 수 있도록 한다.

◦ 예: -량을 늘리다, -수를 늘리다, -교를 mind, -신을 mind, -로를 타다

- [구체물], [추상적대상]처럼 너무 큰 부류를 준 경우, 좀 더 작은 부류로 제한할 수 있는지를 검토한다.
- 구문분석기의 오류는 기록해 두었다가 구문분석기 개발 소분과에 전달하여 구문분석기 성능 향상 작업에 참고하도록 한다.

1.2.2.2 사전 구축 지침 및 보완

1.2.2.2.1. 지침 보완 개요

2007년에 미시구조가 달라진 것은 없으나, 용언의미부류 정보를 부착하는 작업이 본격적으로 진행됨에 따라, 이와 관련된 지침이 대폭 추가되었다. 이것을 용언사전 구축 일반지침에 포함시키기에는 양이 너무 많으므로 별도 파일로 관리하고 있다. 용언의미부류 관련 지침은 위의 1.2.2.1.2.를 참고하기 바라며, 아래에는 일반지침을 제시한다.

한편, 여러 하위사전들 간의 통일성 제고 및 표준화 논의를 통해, 용언의미부류의 태그를 <class>에서 <sem_class>로, 의미 정의(의미 표시)의 태그를 <sem_def>에서 <sem_rep>으로, 제약 용법 용례 태그를 <eg>에서 <defect_eg>로 수정하였다.

1.2.2.2.2. 2007년도 용언사전 미시 구조

superEntry	최상위 표제항 구획
orth	표기 형태
entry+	표제항 구획
mnt_grp	관리 정보 구획
see?	참조 표제항
morph_grp	형태 정보 구획
var*	변이형
cntr?	축약 정보
str?	내부 구조
org*	원어
infl?	굴절 정보
comp*	합성어 형성
der*	파생어 형성
sense*	센스 구획
sem_grp	의미 정보 구획
sem_class	의미 부류
trans+	영어 대역어
sem_rep?	의미 표시
domain?	전문 영역
lr	어휘 의미 관계 구획
syn*	동의어
ant*	반의어
hyper*	상위어
hypo*	하위어
holo*	전체어
mero*	부분어
rel*	관련어
frame_grp+	문형 구성 구획
frame	문형
subsense+	하위 센스 구획
sel_rst*	선택 제약
n_appr*	적정 명사
arg_rst?	논항 제약
ord_rst?	어순 제약
eg+	용례
com	결합 정보 구획
morph_rst?	형태론적 제약 구획
end_rst1*	어말어미 제약
end_rst2*	선어말어미 제약
neg_rst*	부정 제약
col_grp?	연어 구획
col*	연어 구성
morph_rel?	구성 관계 구획
ad_rel*	부사구 구성
np_rel*	명사구 구성
aux_rst*	보조용언 제약
defect*	제약된 활용형 용법 구획
infl_rst+	활용 제약
trans+	영어 대역어
sel_rst*	선택 제약
defect_eg+	제약 용법 용례

idm_grp?
idm+

숙어 구획
숙어

1.2.2.2.3. 2007년도 용언사전 구축 지침

(1) superEntry (최상위 표제항 구획)

- 동형어 분할 이전의 동사 표제항, 즉 동형어들을 포괄하는 구획이다.

(1-1) orth (표기 형태)

- 최상위 표제항의 표기 형태를 한글로 기입한다.

(예) <orth>공부하다</orth>

(1-2) entry (표제항 구획)

- 동형어 구분이 된 표제항의 정보를 제시하는 구획이다.
- 동형어 수만큼 반복이 가능하며, n attribute의 숫자 값으로 제시한다.
- 'pos' attribute의 value로 동사는 "vv", 형용사는 "va"을 기입한다.

(1-2-1) mnt_grp (관리 정보 구획)

- 사전 원고 관리와 관련된 정보 항목들을 제시하는 구획이다.

(1-2-2) see (참조 표제항)

- 관련 entry가 있어서 거기에 제공된 정보가 이 entry의 정보와 동일할 경우, 이 entry의 미시 구조는 비워 두고 여기에 관련 entry의 표기 형태를 기입한다.

(예) <orth>번쩍거리다</orth>
<see>번쩍이다</see>

- 다른 항목을 참조하고 미시 구조의 정보를 비워 두는 경우라 하더라도, 형태 정보는 제공한다.
- 'X거리다, X대다, X이다'의 경우 일반적으로 'X거리다'를 기본형으로 간주하여 'X거리다'만 자세히 기술하고 나머지는 'X거리다'를 참조하도록 한다. 그러나, 특수한 경우에 'X대다'나 'X이다'가 더 일반적으로 사용된다고 판단되는 경우에는 그것을 자세히 기술하고 나머지가 그것을 참조하도록 한다. 위의 '번쩍이다'가 그러한 경우이다.
- 'X뜨리다', 'X트리다'의 경우, 'X뜨리다'를 기본형으로 간주하여 자세히 기술하고 'X트리다'에서는 'X뜨리다'를 참조하도록 한다.

- ‘들르다’를 ‘들리다’로 잘못 쓰는 경우가 많은데, superEntry ‘들리다’에서 “(귀에 소리가 들리다)”, “(물건이 손에 들리다)” 등의 용법들을 각각 entry로 기술하고 마지막에 ‘들르다’의 비표준어 변이형으로서의 용법을 첨가하여 ‘들르다’를 참조하도록 한다. 기타 비표준어 변이형도 비슷한 방식으로 처리한다.

(1-2-3) morph_grp (형태 정보 구획)

- 표제항의 형태론적 정보를 제시하는 구획이다.

(1-2-3-1) var (변이형)

- 표제항의 변이형을 유형별로 구별하여 기입한다.
- 변이형의 하위 유형은 다음과 같이 구별한다.

type="abb"	준말
type="lng"	본딧말
type="spr"	격조사에 의한 분리형

(주격, 대격 이외의 조사에 의한 분리형도 제시. 예: 주옥과 같다)

type="etc" 기타. 모음변이->자음변이의 순서로 제시

- 변이형이 둘 이상일 경우 var 요소를 반복하여 제시한다.

(예) <orth>반짝거리다</orth>
 <var type="etc">번쩍거리다</var>
 <var type="etc">뽀쩍거리다</var>
 <var type="etc">뽀쩍거리다</var>

(1-2-3-2) cntr (축약 정보)

- 어간말 모음과 어미 첫머리의 모음 사이에 일어날 수 있는 축약 현상을 유형화하여 기입한다.
- attribute ‘opt’의 value에는 ‘opt’와 ‘obl’이 있다. 축약이 수의적이면 ‘opt’로, 필수적이면 ‘obl’로 표시한다.
- attribute ‘type’의 value는 다음과 같이 유형화하여 제시한다.

type="a"	ㅏ + 아 → ㅑ
type="ai"	ㅙ + 어 → ㅞ
type="e"	ㅘ + 어 → ㅚ
type="eo"	ㅓ + 아 → ㅕ
type="u"	ㅜ + 어 → ㅠ
type="i"	ㅣ + 어 → ㅓ

type="oi"	ㅛ + ㅕ → ㅞ
type="VDelCCon"	-하 + 게/도록 → -케/토록
type="CDeIVCon"	놀 + 아 → ㉠

(1-2-3-3) str (내부 구조)

- 표제항의 형태론적 내부 구조를 기입한다.
- 구성 요소의 범주를 다음과 같이 부호화한다.

단일 동사:	V
단일 형용사:	A
명사:	N
어근:	R
의성의태어근:	I
접두사:	P
(파생)접미사:	S
부사:	D
관형사:	M

- 단, 다음 고빈도 기능 용언 및 접미사는 범주 부호 대신 실제 형태를 밝혀서 써준다.

하다:	.하
당하다:	.당
시키다:	.시
되다:	.되
지다:	.지
뜨리(트리):	.뜨
대다:	.대
거리다:	.거
이다(반복):	.이3
이다(지정사):	.이4

- 사동 접미사 ‘이, 히, 리, 기, 우, 구, 추’도 범주 부호 대신 실제 형태를 밝혀 적는다. ‘이, 히, 리, 기’는 피동 접미사와의 구별을 위해 ‘이1, 히1, 리1, 기1’로 표기한다.
- 피동 접미사 ‘이, 히, 리, 기’도 범주 부호 대신 실제 형태를 밝혀 적는다. 사동 접미사와의 구별을 위해 ‘이2, 히2, 리2, 기2’로 표기한다.
- 구성 요소와 구성 요소 사이에는 하이픈(-) 또는 점(.)을 사용하여 구별한다. 용언 어간과 어미 사이에는 하이픈을 사용하고 그 외의 경우에는 점을 사용한다.

(예) 가공적.이	N.이4
가공.하	R.하
가깝	A
가깝-디.가깝	A-디A
가능성.없	NA
가느.다랗	AS
가느스름.하	Ru.하
가당.찮	RS

- 어미 ‘-아/어’는 ‘어’로 통일하여 표기한다.

- 표제항 용언의 직접 구성 요소(immediate constituent, IC)만을 반영한다.
 - ‘-적’, ‘-성’, ‘-口’ 등 명사화 접미사를 분석하지 않는다.
(이하의 예에서는 범주 부호로 코드화하기 이전의 실제 형태로 예시함.)
(예) 가공적.이 / 거리낌.없
 - 어근에 상응하는 복합 요소의 내부 구조를 분석하지 않는다.
(예) 보잘것.없 / 하잘것.없 / 물샐틈.없 / 비할데.없 / 쉴새.없 / 쓸데.없
 쏟살.갈 / 먹음직.하 / 온데간데.없 / 할일.없 / 난데.없 / 그럴듯.하 / 그럴싸.하
 웬만.하 / 그만.하 / 게딱지만.하 / 자그만.하
 - 분석하지 않은 어근 상당 단위가 순수한 A나 공인된 R이 아닐 때 그 뒤에 u를 붙여
 주어(Au, Ru) A나 R과 구별한다. (동사의 경우 V도 마찬가지로(Vu))
 - 체언에 상당하는 복합 요소의 내부 구조를 분석하지 않는다.
(예) 간단명료.하 / 고래등.갈 / 감개무량.하
 - 의미나 기능이 확인되지 않는 유일형태소(unique morpheme)에 가까운 요소는 따로
 분석하지 않는다.
(예) 곁맞 / 강마르 / 고부랑.하 / 고부스름.하 / 꼬무레.하 / 끈질기 / 기똥차 / 곱상.하
 날짜지근.하 / 늙수그레.하 / 꼬부라지 / 흐벅지 / 하릴.없 / 가마노르께.하
 하찮 / 애꿎 / 애달프 / 애달 / 날카롭
 - 의미가 완전히 바뀌었다고 생각되는 경우는 분석하지 않는다.
(예) 깎듯.하 / 되바라지
 cf. 되-지.못.하 / 되-어.먹 / 마-지.못.하
 - 의성의태 어근이 중첩된 경우는 분석해서 제시한다.
(예) 고불탕.고불탕.하 / 까슬.까슬.하
 - 의성의태 어근이 독립적으로 부사로 쓰일 수 있더라도 ‘-거리다/대다/이다’ 앞에 쓰
 인 것은 의성의태 어근 I로 표기한다. 의성의태 어근이 반복된 경우 I를 반복해서 써
 준다.
(예) 꼬불탕꼬불탕하다 II.하
 - ‘-(으)락/-(으)르락’으로 중첩된 구성은 분석하지 않는다.
(예) 들릴락말락.하
 - 용언에 상당하는 복합 요소의 내부 구조도 분석하지 않는다.
(예) 기.막히 / 대수롭.잖 / 얼토당토.아니하 / 멋.들어지 / 속.시원하

외.떨어지 / 가느다랗-어.지

- 축약형은 본래 형태로 환원시키지 않고 그대로 놔둔다.

(예) 가당.참 / 남.부럽.잖 / 머-지.않 / 마-지.못.하

- 어간과 어미가 줄어든 경우에는 각각을 밝혀 준다.

(예) 크-어.다랗 / 되-어.먹 / 막.되-어.먹

(1-2-3-4) org (원어)

- 표제항의 전체 또는 일부가 외래어(한자어 포함)일 경우, 그 외래 요소의 기원이 되는 언어 및 그 언어에서의 표기 형태를 기입한다.
- 외래 요소의 기원이 되는 언어는 한자어(si), 영어(en), 기타(etc)의 셋으로만 구분한다.

lg="si"	한자어
lg="en"	영어
lg="etc"	기타

- 외래 요소의 해당 언어에서의 표기 형태는, 한자어의 경우는 한자로, 영어의 경우는 로마자로 표기하고, 기타의 경우는 한글로 표기한다.
- 표제항의 일부만 외래어인 경우 고유어에 해당하는 부분은 밑줄(_)로 표시한다. 어미 '-다'에 해당하는 부분은 밑줄을 긋지 않는다.

(예) <orth>헛공부하다<orth>
<org lg="si">_工夫_</org>

- 둘 이상의 외래 요소를 포함하고 있고 이들의 기원 언어가 다른 경우 org 요소를 반복해서 표시해 준다.

(예) <orth>강슛하다</orth>
<org lg="si">强_</org>
<org lg="en">_shoot_</org>

(1-2-3-5) infl (굴절 정보)

- 용언의 활용 유형, 즉 규칙 활용인지 불규칙 활용인지, 불규칙 활용의 경우 어떤 불규칙 활용인지를 기입한다.
- 활용 방식을 다음과 같이 유형화하여 attribute 'type'의 value로 제시한다.

규칙/불규칙 유형	값	규칙/불규칙 유형	값
규칙형	reg	‘리’ 불규칙	leo
‘ㄷ’ 불규칙	d	‘여’ 불규칙	yeo
‘ㄹ’ 불규칙	l	‘우’ 불규칙	u
‘ㅅ’ 불규칙	s	‘ㅎ’ 불규칙	h
‘ㅂ’ 불규칙	b	‘거라’ 불규칙	geora
‘으’ 불규칙	eu	‘너라’ 불규칙	neora
‘르’ 불규칙	leu		

(1-2-3-6) comp (합성어 형성), (1-2-3-7) der (파생어 형성)

- 표제항을 포함한 합성어나 파생어를 제공해 주는 구획이다.

(1-2-4) sense (센스 구획)

- 표제항이 갖는 각각의 센스를 구별하여 의미 정보 및 통사 정보를 제시하는 구획이다.

(1-2-4-1) sem_grp (의미 정보 구획)

- 의미 관련 정보들을 제시하는 구획이다.
- 형식이 동일한 관용표현이 여러 의미를 가지는 경우 n attribute의 value에 번호를 붙여 구별한다.

(1-2-4-1-1) sem_class (의미 부류)

- 표제항 용언의 해당 센스의 의미 부류를 기입한다.
- 용언 의미 부류 체계는 별도 지침 파일 “온vv_2007_05_02.hjt”를 참조한다.

(1-2-4-1-2) trans (영어 대역어)

- 해당 센스에 대응하는 영어 대역어를 로마자로 기입한다.
- 둘 이상일 경우 trans 요소를 반복해서 제시한다.

(예) <trans>consider</trans>
<trans>take into consideration</trans>

(1-2-4-1-3) sem_rep (의미 표시)

- 의미를 자연 언어로 자유롭게 기입한다.

- 모든 표제항에 대해 꼭 기입해야 하는 정보는 아니다. 필요한 경우에만 제시한다.
- 주로, 둘 이상 제시되는 sense가 서로 어떻게 차이나는지를 알아보기 쉽게 제시하기 위한 목적으로 사용한다.

(예) <orth>올리다</orth>
 <sense n="1">
 <sem_rep>아래에서 위로 움직이다; 위쪽으로 들다</sem_rep>
 <sense n="2">
 <sem_rep>수치가 높아지거나 많아지게 하다</sem_rep>

(1-2-4-1-4) domain (전문 영역)

- 표제항이 전문용어일 경우 해당 전문 영역을 기입한다.
- 전문용어가 아닌 일반 표제항은 아무 정보도 제시하지 않는다.
- 전문 영역은 전문 영역 소분과에서 작성한 체계에 따른다.

(1-2-4-1-5) lr (어휘 의미 관계 구획)

- 표제항과 밀접한 의미 관계에 있는 단어들을 제시하는 구획이다.

(1-2-4-1-5-1) syn (동의어)

- 표제항의 동의어를 다음과 같이 구별하여 기입한다.

type="perfect"	완전 동의어(고유어-고유어 or 한자어-한자어)
어)	
type="sino"	고유어 표제항에 대한 한자어 동의어
type="kor"	한자어 표제항에 대한 고유어 동의어
type="honor"	비존칭 표제항에 대한 존칭 동의어
type="de_honor"	존칭 표제항에 대한 비존칭 동의어
type="under"	비겸양 표제항에 대한 겸양 동의어
type="de_under"	겸양 표제항에 대한 비겸양 동의어
type="spoken"	非구어 표제항에 대한 구어 동의어
type="de_spoken"	구어 표제항에 대한 非구어 동의어
type="vulgar"	非속어 표제항에 대한 속어 동의어
type="de_vulgar"	속어 표제항에 대한 非속어 동의어
type="etc"	기타

- 동의어가 둘 이상일 경우 syn 요소를 반복해서 제시한다.

(예) <syn type="etc">연구하다</syn>
 <syn type="etc">탐구하다</syn>

(1-2-4-1-5-2) ant (반의어)

- 표제항의 반의어를 다음과 같이 구별하여 기입한다.

type="compl"	상보적 반의어 (죽다-살다)
type="grad"	정도성 반의어 (덥다-춥다)
type="dir"	방향성 반의어 (올라가다-내려가다)
type="conv"	전환적/관점적 반의어 (사다-팔다)
type="etc"	기타

- 위 4가지 이외의 유형인 경우 또는 어느 유형인지 판정하기 어려운 경우 etc로 처리한다.
- '죽다 - 죽이다', '생산하다 - 생산되다'와 같이 능동사-피동사의 쌍은 전환적/관점적 반의어로 볼 수 있기는 하지만, 이렇게 형태론적으로 관련된 동사 쌍은 어휘의미관계 구획에서 제시하지 않는다. '때리다-맞다'처럼 형태론적으로 관련이 없는 것만 제시한다.
- 'X-어지다', 'X-어뜨리다' 쌍도 형태론적으로 관련된 것이므로, 어휘의미관계 구획에서 제시하지 않는다.

(1-2-4-1-5-3) hyper (상위어)

- 표제항의 상위어를 기입한다.
- 상위어가 둘 이상인 경우 hyper 요소를 반복해서 제시한다.

(1-2-4-1-5-4) hypo (하위어)

- 표제항의 하위어를 기입한다.
- 하위어가 둘 이상인 경우 hypo 요소를 반복해서 제시한다.

(예) <orth>죽다</orth>
<hypo>동사하다</hypo>
<hypo>아사하다</hypo>
<hypo>압사하다</hypo>
<hypo>고사하다</hypo>
<hypo>병사하다</hypo>

(1-2-4-1-5-5) holo (전체어)

- 표제항이 나타내는 사태를 포함하는 더 큰 사태를 기입한다.

(예) <orth>다녀오다</orth>
<holo>오다</holo>

(1-2-4-1-5-6) mero (부분어)

- 표제항이 나타내는 사태에 포함되는 작은 사태를 기입한다.

(예) <orth>불합하다</orth>
<holo>수술하다</holo>

(1-2-4-1-5-7) rel (관련어)

- 동의어, 반의어, 상위어, 하위어, 전체어, 부분어 이외의 기타 관련어를 기입한다.

(1-2-4-2) frame_grp (문형 구성 구획)

- 문형 및 논항에 관한 정보를 제시하는 구획이다.
- 하나의 센스 아래에서 표제항 용언이 규정하는 문형의 차이에 따라 문형 구획들을 설정한다.
- 이 문형의 통사적 특성을 다음과 같이 유형화하여 'type' attribute의 value로 기입한다.

type="FIN"	일반자동사
type="FTR"	일반타동사
type="FIT"	일반자타동사
type="SIN"	기능자동사
type="STR"	기능타동사
type="SIT"	기능자타동사
type="CIN"	숙어자동사
type="CTR"	숙어타동사
type="CIT"	숙어자타동사
type="Vaux"	보조동사
type="FA"	일반형용사
type="SA"	기능형용사
type="CA"	숙어형용사
type="Aaux"	보조형용사

(예) <frame_grp type="FIN">

- 두 문형 구획이 서로 일정한 통사·의미상의 대응 관계에 있을 경우, 그 대응 관계를 'cor' attribute에 대한 value로 제시해 준다. 대응 관계의 유형은 다음과 같다.

cor="sym"	대칭 (symmetric)
cor="reci"	상호 (reciprocal)

cor="nut" 중립 (neutral)
cor="cross" 교차 (cross)

- 서로 대응 관계에 있는 문형 구획들은 cor 값과 n 값을 일치하게끔 부여한다.

(예) <orth>비슷하다</orth>
 <frame_grp type="FA" cor="sym" n="1">
 <frame>X=N0-이 Y=N1-와 (서로) V</frame>
 <eg>이것은 저것과 비슷하다.</eg>
 </frame_grp>
 <frame_grp type="FA" cor="sym" n="1">
 <frame>X+Y=N0-이 (서로) V</frame>
 <eg>이것과 저것이 비슷하다.</eg>
 </frame_grp>

- 대칭 구문은 용언의 어휘 의미 속에 대칭성의 의미 성분이 들어 있어서 부사 ‘서로’가 수의적으로 사용되는 구문이다. 대칭 구문의 문형 및 구문 대응 관계는 위의 ‘비슷하다’처럼 표시하면 된다.
- 상호 구문은 용언의 어휘 의미 속에 대칭성/상호성의 의미 성분이 들어 있지는 않지만 부사 ‘서로’와 공기함으로써 구문이 대칭성/상호성의 의미를 획득하게 되는 경우이다. 상호 구문의 문형 및 구문 대응 관계는 다음과 같이 표시한다.

(예) <orth>사랑하다</orth>
 <frame_grp type="FIN" cor="reci" n="1">
 <frame>X=N0-이 Y=N1-와 서로 V</frame>
 <eg>철수는 영희와 서로 사랑한다.</eg>
 </frame_grp>
 <frame_grp type="FIN" cor="reci" n="1">
 <frame>X+Y=N0-이 서로 V</frame>
 <eg>철수와 영희는 서로 사랑한다.</eg>
 </frame_grp>

- 교차 구문은 ‘병에 물을 채우다’ - ‘병을 물로 채우다’처럼 주어 이외에 용기(container)와 내용물(content)의 두 보어 논항을 취하는데, 용기가 처격, 내용물이 대격으로 실현될 수도 있고, 용기가 대격, 내용물이 구격으로 실현될 수도 있는 구문이다. 상호 구문의 문형 및 구문 대응 관계 및 각 논항의 의미역은 다음과 같이 표시한다.

(예) <orth>채우다</orth>
 <frame_grp type="FTR" cor="cross" n="1">
 <frame>X=N0-이 Z=N2-에 Y=N1-을 V</frame>
 <subsense>
 <sel_rst arg="X" tht="AGT">인간</sel_rst>
 <sel_rst arg="Y" tht="THM">구체물</sel_rst>
 <sel_rst arg="Z" tht="LOC">장소|용기</sel_rst>
 <eg>그는 병에 물을 채웠다.</eg>
 </subsense>
 </frame_grp>

```

<frame_grp type="FTR" cor="cross" n="1">
  <frame>X=N0-이 Y=N1-을 Z=N2-로 V</frame>
  <subsense>
    <sel_rst arg="X" tht="AGT">인간</sel_rst>
    <sel_rst arg="Y" tht="THM">장소|용기</sel_rst>
    <sel_rst arg="X" tht="INS">구체물</sel_rst>
    <eg>그는 병을 물로 채웠다.</eg>
  </subsense>
</frame_grp>

```

(1-2-4-2-1) frame (문형)

- 각 의미 논항들이 어떤 표면 형태로 실현되는지를 기입한다.
- 의미 논항은 X, Y, Z, W, U로 나타낸다. 이 의미 논항의 식별 기호 X, Y, Z, W, U는 선택 제약에서도 사용된다.
- 표면의 통사 논항은 명사구 논항과 절 논항을 구분해서 전자는 N으로, 후자는 S로 나타낸다. 간접 의문절은 특히 Q로 나타낸다.
- 표제항 용언이 기능동사로서 술어명사와 결합할 때, 이 술어명사는 Npr로 나타낸다.
- 표제항 용언이 숙어 구문을 형성하는 경우 고정된 명사 논항은 C로 나타낸다.
- 통사 논항에는 통사적 위계에 따라 숫자를 부여한다. 주어는 0을 부여받고, 나머지 논항들은 동사와의 긴밀도에 따라 순서대로 1, 2, …… 등을 부여받는다. 즉 직접목적어는 1을 부여받고, 나머지 사격 논항은 2 이하의 숫자를 부여받는다.

(예) <orth>주다</orth>
 <frame>X=N0-이 Z=N2-에게 Y=N1-을 V</frame>

- 의미 논항 식별 기호 X, Y, Z, W, U는 통사적 위계에 따라 순서대로 부여한다. 통사적 위계가 동일할 경우, 즉 사격 논항들끼리는 다음과 같이 의미역 위계에 따라 순서대로 부여한다.

· THM > CNT > GOL/DIR/FNS > SRC > PUR
 · THM > LOC/CRT > PUR

- 주어를 제외하면 직접 목적어가 위계상 가장 높으므로, 직접 목적어가 대개 1을 부여받는다.

(예) <frame>X=N0-이 Z=N2-에|에게 Y=N1-을 V</frame>

- 직접 목적어가 없을 때에는 다른 논항이 1을 부여받을 수 있다.

(예) <orth>인사하다</orth>
 <frame>X=N0-이 Y=N1-에게 V</frame>

- 절 논항 S, Q 및 술어명사 Npr이나 숙어 구문의 고정 명사 C도 이 위계에 따라 숫자를 부여받는다.
- GOL/DIR/FNS를 실현하는 논항은 SRC를 실현하는 논항에 비해 위계상 높은 위치를

차지한다.

(예) <frame>X=N0-이 Z=N2-에서 Y=N1-로 V</frame>

- 표면 논항들의 순서는 가장 자연스러운 어순을 반영하도록 한다.
- 명사구 논항에 붙는 격표지를 하이픈 뒤에 써 준다.

(예) <frame>X=N0-이 Y=N1-을 V</frame>

- 명사구 논항에 붙는 격표지는 대표형으로 표시해 준다.

격	대표형	대표형으로 포괄되는 이형태
주격	이	가
대격	을	를, 르
여격	에게	한테
구격	로	으로
공동격	와	과
탈격	에서	(으)로부터
	에게서	한테서
	에게	한테
명명격	라고	이라고
	에 대해	에 대하여, 에 대해서
	을 위해	을 위하여, 을 위해서
	에 의해	에 의하여, 에 의해서
	을 놓고	을 놓고서

- ‘을위해’ 논항은 ‘노력하다, 투쟁하다’와 같이 동사의 어휘 의미(lexical meaning) 속에 [목적/목표]가 포함되는 경우에 한하여 문형에 기입한다.
- ‘에 의해’ 논항은 피동문의 AGENT를 표시하는데, 피동사 중 이 AGENT가 자주 특징적으로 나타나는 것에 한정하여 사용한다.
- ‘을놓고’ 논항은 ‘다투다’, ‘경쟁하다’, ‘경합하다’와 같이 동사의 어휘 의미 속에 [경쟁]이 포함되며 경쟁의 목표물이 ‘-을 놓고’와 결합하여 나타나는 경우에 한하여 문형에 기입한다.
- 단체 주격 조사 ‘-에서’(예: 학교에서 운동회를 개최했다)는 문형에 반영하지 않는다. 그냥 주격조사 ‘-이’로 반영한다.
- GOL을 실현하는 논항은 ‘-에|에게|로|에게로’의 넷 중에서 가능한 것들을 골라서 이 순서대로 기입한다.

(예) <frame>X=N0-이 Y=N1-에|에게|로|에게로 V</frame>
 <frame>X=N0-이 Y=N1-에|에게 V</frame>
 <frame>X=N0-이 Y=N1-에|로 V</frame>

- SRC를 실현하는 논항은 ‘-에서|에게서|에게’의 셋 중에서 가능한 것들을 골라서 이 순서대로 기입한다.
- ‘그를 아버지라고 부르다’, ‘그를 형이라고 부르다’와 같이 명사구에 ‘-라고/이라고’가 붙은 경우 절 논항 S로 처리하지 않고 명사구 논항 N으로 처리한다. 문형 및 의미역은 다음과 같이 표시한다.

(예) <orth>부르다</orth>
 <frame_grp type="FTR">
 <frame>X=N0-이 Y=N1-을 Z=N2-라고 V</frame>
 <subsense>
 <sel_rst arg="X" tht="AGT">인간</sel_rst>
 <sel_rst arg="Y" tht="THM">온</sel_rst>
 <sel_rst arg="Z" tht="FNS">온</sel_rst>
 </subsense>
 </frame_grp>

- 간접 의문절 논항인 Q에도 명사구 논항과 마찬가지로 격표지를 붙인다.

(예) <orth>묻다</orth>
 <frame>X=N0-이 Z=N2-에게 Y=Q1-을 V</frame>

- 간접 의문절 논항 Q의 보문소가 ‘-(으)리지’, ‘-(으)리지’, ‘-(으)니까’, ‘-(으)니까’, ‘-는지’, ‘-나’, ‘-는가’ 중 일부만 가능할 경우 이 사실을 주석으로 제시해 준다.
- 일반 절 논항 S 중에서 명사화소 ‘-것|기|음’이 붙는 경우에는 명사구 논항과 마찬가지로 방법으로 격표지를 붙인다.

(예) <orth>알다</orth>
 <frame>X=N0-이 Y=S것1|S기1|S음1-을 V</frame>

- 의존명사 ‘데, 줄, 수’를 핵으로 갖는 절 논항은 각각 ‘S데’, ‘S줄’, ‘S수’로 표기한다.
- 의존명사 앞에 오는 관형사형 어미가 ‘-(으)ㄴ’, ‘-(으)ㄹ’, ‘-는’ 중 일부만 가능할 경우, 이 사실을 주석으로 제시해 준다.
- 그 외의 절 논항에는 S 바로 뒤에 숫자를 붙이고 하이픈을 하고 그 뒤에 보문자를 붙인다.

(예) <orth>말하다</orth>
 <frame>X=N0-이 Z=N2-에게 Y=S1-고 V</frame>

- 보문자 ‘-고’가 붙는 절 논항의 경우 의향법(평서, 의문, 명령, 청유, 약속)의 제약이 없으면 그냥 ‘S1-고’ 식으로 써 주고, 의향법의 제약이 있을 때에는 의향법을 나타내는 종결어미의 대표형을 ‘-고’ 앞에 기입한다.

(예) <orth>명령하다</orth>
 <frame>X=N0-이 Z=N2-에게 Y=S1-라고 V</frame>

- 의향법 어미가 둘 이상 가능한 경우 ‘|’로 병기한다.

(예) <frame>X=N0-이 Z=N2-에게 Y=S1-라고|자고 V</frame>

- 문형에서는 명사구 논항 N으로 표기하고 subsense를 둘 이상 설정하여 그 중 하나에 서는 이 명사구 논항에 대한 선택 제약을 보통의 경우처럼 제시하고, 다른 하나의 subsense에서는 이 명사구 논항 자리에 절 논항 S나 Q가 올 수 있다는 사실을 선택 제약으로 제시해 줄 수 있다. 명사구 논항과 절 논항이 완전히 계열 관계를 이루어 대등한 자격을 가진다고 판단될 때 이 방법을 택한다.

(예) <orth>알다</orth>
 <frame>X=N0-이 Y=N1-을 V</frame>
 <subsense>
 <sel_rst arg="X" tht="EXP">인간</sel_rst>
 <sel_rst arg="Y" tht="THM">온</sel_rst>
 </subsense>
 <subsense>
 <sel_rst arg="X" tht="EXP">인간</sel_rst>
 <sel_rst arg="Y" tht="THM">S것|S음</sel_rst>
 </subsense>

(1-2-4-2-2) subsense (하위 센스 구획)

- 하위 센스별로 각 논항의 선택 제약을 제시하는 구획이다.
- 하나의 문형 아래에서 각 의미 논항에 대한 선택 제약이 차이 나는 경우들을 하위 센스로 설정한다.

(1-2-4-2-2-1) sel_rst (선택 제약)

- 표제항 용언의 의미 논항의 의미역 및 선택 제약을 기입한다.
- 의미역은 'tht' attribute에 대한 value로 제시한다. 의미역의 목록은 다음과 같다.

tht="AGT"	행위주
tht="EXP"	경험주
tht="MAG"	심리행위주
tht="COM"	동반주
tht="THM"	대상
tht="LOC"	장소
tht="DIR"	방향
tht="GOL"	도착점
tht="FNS"	결과상태
tht="SRC"	출발점
tht="INS"	도구
tht="EFF"	영향주
tht="CRT"	기준치
tht="PUR"	목적

tht="CNT" 내용

- 의미 부류 MAG, DIR, CNT, PUR, COM, CRT는 다음과 같은 경우 기입한다.
- 심리행위주(mental agent, MAG): 인지/지각/감정 행위의 주체. EXP의 성격과 AGT의 성격을 동시에 지님. 자신의 의지에 의한 통제가 가능한 인지/지각/감정 행위를 수행하는 주체.

(예) <orth>존경하다</orth>
<sel_rst arg="X" tht="MAG">인간</sel_rst>

- 방향(direction, DIR): 공간적/물리적 행위가 이루어질 때 지향하는 방향. 지금까지 GOL 의미역을 부여하던 논항 중 실제로 도달하는 것이 아니라 지향하기만 하는 논항이 DIR에 해당됨. 대개 격표지 ‘-로/에게로’를 달고 나타남. 추상적인 상태변화의 중착점으로서의 FNS는 변동 사항 없음.

(예) <orth>향하다</orth>
<sel_rst arg="Y" tht="DIR">장소|방향</sel_rst>

- 내용(content, CNT): 발화/사유/인지 행위 등의 내용. 대개 ‘S고’ 보문으로 실현됨.

(예) <orth>말하다</orth>
<sel_rst arg="Y" tht="CNT"></sel_rst>

- 목적(purpose, PUR): 행위주가 행위를 할 때 마음 속에 가지고 있는 목적. 대개 ‘S리’ 보문으로 실현됨.
- ‘같다’, ‘다르다’, ‘비슷하다’ 등의 ‘-와’ 보어는 COM을 부여한다. COM은 그 구체적 인 성격이 다른 논항에 의존적인 특수 의미역으로 파악한다. CRT 의미역은 ‘-보다 낫다’, ‘-만 못하다’ 등의 논항에 부여한다.

- 각 논항의 선택 제약은 명사 의미 부류 체계에 따라 제공한다.
- 절 논항(S)에 대해서는 의미역만 제시한다. 선택 제약은 명사구 논항(N)에 대해서만 제시한다.
- 의미 부류명은 괄호 밖에, 실제 명사는 괄호 안에 제시한다.
- 특정 의미 부류에 속하는 실제 명사는 해당 의미 부류명 바로 뒤에서 괄호 안에 제시한다.
- 의미 부류명과 의미 부류명 사이, 실제 명사와 명사 사이에는 수직바(|)를 사용한다.

(1-2-4-2-2-2) n_appr (적정 명사)

- 선택 제약 정보 항목에 제시된 논항(주로 X 논항)과 관련된 적정 명사가 있는 경우

기입한다.

- 'N-이 길이가 길다', 'N-이 무게가 무겁다', 'N-이 색깔이 파랗다'와 같은 구문에서 '길이', '무게', '색깔' 등을 적정 명사라고 한다.
- 적정 명사는 표면에 실현되어 2중주어문을 만들 수도 있고, 생략될 수도 있다.
- 이런 적정 명사는 문형에는 제시하지 않고 n_appr 구획에 제시해 준다.
- 적정 명사가 표제항 용언의 어떤 의미 논항과 관련된 것인지를 arg="X"와 같은 식으로 표시해 준다.
- 적정 명사가 둘 이상일 경우에는 n_appr 요소를 반복해서 제시한다.

(예) <orth>파랗다</orth>
<sel_rst arg="X" tht="THM">구체물</sel_rst>
<n_appr arg="X">색</n_appr>
<n_appr arg="X">색깔</n_appr>
<n_appr arg="X">빛깔</n_appr>

(1-2-4-2-2-3) arg_rst (논항 제약)

- 논항의 한정성, 수식어의 필수성, 생략 불가능성 등에 대한 정보를 기입한다.

(예) <orth>올리다</orth>
<frame>X=N0-이 Z=N2-에서 Npr1-을 V</frame>
<sel_rst arg="X" tht="AGT">인간</sel_rst>
<sel_rst arg="Npr1">(성과)</sel_rst>
<sel_rst arg="Z" tht="LOC">(대회)</sel_rst>
<arg_rst>Npr1=mdef</arg_rst>

(1-2-4-2-2-4) ord_rst (어순 제약)

- 논항들 사이의 순서에 대한 제약을 기입한다.
- 한국어에서는 어순이 비교적 자유로우나, 간혹 그렇지 않은 경우가 있다. 이렇게 논항의 순서에 제약이 있어서 바꿀 수 없는 경우에는 해당 정보를 기입한다.

(예) <orth>몰다</orth>
<frame>X=N0-이 Y=N1-을 Z=N2-로 V</frame>
<sel_rst arg="X" tht="AGT">인간</sel_rst>
<sel_rst arg="Y" tht="THM">인간</sel_rst>
<sel_rst arg="Z" tht="FNS">부정적속성인간</sel_rst>
<ord_rst>Y=N1-을 Z=N2-로 V</ord_rst>

(1-2-4-2-2-5) eg (용례)

- 해당 센스의 해당 문형의 해당 선택 제약을 반영하는 전형적인 용례를 기입한다.
- 특수한 경우가 아니면 온전한 문장 형태로 제시한다. 문장 끝에는 마침표를 찍는다.
- 둘 이상일 경우 eg 요소를 반복해서 제시한다.

(1-2-4-3) com (결합 정보 구획)

- 각종 결합 제약을 제시해 주는 구획이다.

(1-2-4-3-1) morph_rst (형태론적 제약 구획)

- 표제항 동사의 활용이나 부정문 형성 등에 제약이 있을 경우 해당 제약을 제시하는 구획이다.

(1-2-4-3-1-1) end_rst1 (어말어미 제약)

- 표제항 용언의 활용에 제약이 있어서 특정 활용형만 가능한 경우(이른바 불구 활용, defective conjugation), 가능한 활용형을 기입한다.
- 가능한 활용형이 둘 이상일 경우 end_rst1 요소를 반복해서 제시해 준다.

(예) <orth>데리다</orth>
 <end_rst1>데리고</end_rst1>
 <end_rst1>데려</end_rst1>
 <end_rst1>데리러</end_rst1>

- 불구 용언이 아니라서 대부분의 활용형이 가능하지만 소수 활용형만 불가능한 경우에는 불가능한 활용형을 제시해 주고 impossible이라는 attribute에 “yes”라는 value를 준다.
- 불가능한 활용형이 둘 이상일 경우 end_rst1 요소를 반복해서 제시해 준다.

(예) <orth>걱정되다</orth>
 <end_rst1 impossible="yes">걱정되어라</end_rst1>
 <end_rst1 impossible="yes">걱정되자</end_rst1>

- 명령형이 불가능할 경우 ‘-어라’형 하나만 대표로 제시하고 나머지 명령형(‘-십시오’, ‘-게’, ‘-소서’ 등)은 제시하지 않는다. 청유형 등 다른 경우도 마찬가지이다.

(1-2-4-3-1-2) end_rst2 (선어말어미 제약)

- 표제항 용언이 선어말어미와의 결합에 있어서 특수한 제약이 있을 경우, 즉 특정 선어말어미와의 결합이 불가능할 경우 그 선어말어미를 기입한다.
- 결합이 불가능한 선어말어미가 둘 이상일 경우 end_rst2 요소를 반복해서 제시해 준다.
- end_rst1의 경우와는 달리 불가능한 경우만 있으므로 impossible 같은 attribute는 써 주지 않는다.
- end_rst1의 경우와는 달리 활용형 전체를 써 주지 않고 선어말어미만 써 준다. 활용형 전체를 써 주려고 한다면 어떤 어말어미를 써 주어야 할지 분명치 않기 때문이다.

(1-2-4-3-1-3) neg_rst (부정 제약)

- 표제항 용언이 부정형으로 쓰일 수 없을 경우, 불가능한 부정형을 기입한다.
- ‘-지 않-’, ‘-지 못하-’, ‘-지 말-’ 등의 부정형들 중 어느 하나만 불가능하면 그것만

제시하고, 불가능한 것이 둘 이상이면 모두 제시한다. 후자의 경우 neg_rst 요소를 반복해서 제시한다.

- 불가능한 경우만 있으므로 impossible 같은 attribute는 써 주지 않는다.

(1-2-4-3-2) col_grp (연어 구획)

- 연어 정보를 제시하는 구획이다.

(1-2-4-3-2-1) col (연어)

- 표제항 용언과 관습적으로 공기하는 요소, 즉 연어 관계(collocational relation)에 있는 요소를 유형별로 나누어 기입한다.
- 속어의 성격을 띠는 것은 idm_grp에 제시한다.
- 연어의 유형은 다음과 같다.

type="magn"	극대·극소값을 나타내는 수식어
type="time"	시간을 나타내는 수식어
type="manner1"	의성어 부사
type="manner2"	의태어 부사
type="manner3"	기타 방식 부사어(manner adverbial)
type="etc"	기타

(예) <orth>취하다</orth>
<col type="magn">곤드레만드레</col>

(1-2-4-3-3) morph_rel (구성 관계 구획)

- 표제항 용언을 포함하는 전형적인 부사구 구성이나 명사구 구성을 제시하는 구획이다.

(1-2-4-3-3-1) ad_rel (부사구 구성)

- 표제항 용언이 부사적으로 쓰이는 전형적인 구성을 기입한다.

(예) <orth>간드러지다</orth>
<ad_rel>간드러지게 웃다</ad_rel>

(1-2-4-3-3-2) np_rel (명사구 구성)

- 표제항 용언을 포함하는 전형적인 명사구 구성을 기입한다.

(예) <orth>겉먹다</orth>
<np_rel>겉먹은 얼굴</np_rel>

(1-2-4-3-4) aux_rst (보조용언 제약)

- 표제항 용언이 보조용언과의 결합에 있어 특수한 제약을 보일 경우, 즉 특정 보조용언과의 결합이 불가능할 경우 그 보조용언을 기입한다.
- 보조용언 앞의 연결어미까지 함께 기입한다.

(예) <aux_rst>고 있</aux_rst>

- 결합이 불가능한 보조용언이 둘 이상일 경우 aux_rst 요소를 반복해서 제시해 준다.
- 불가능한 경우만 있으므로 impossible 같은 attribute는 써 주지 않는다.

(1-2-5) defect (제약된 활용형 용법 구획)

- 표제항 용언이 제약된 활용형으로만 쓰이면서 다른 센스들과 구별되는 용법을 보일 때 이 용법에 관한 정보를 제시하는 구획이다.
- 이 구획에서는 infl_rst+, trans+, sel_rst*, defect_eg+ 만을 하위 요소로 제공한다.
- 선택 제약은 필요하고 유용한 경우에만 제시한다.

(예) <infl_rst>이른</infl_rst>
<trans>early</trans>
<sel_rst arg="X" tht="THM">순환시간</sel_rst>

(1-2-5-1) infl_rst (활용 제약)

- 해당 defect 용법에서 가능한 활용형을 기입한다.

(예) <orth>이르다</orth>
<defect>
<infl_rst>이른</infl_rst>
<trans>early</trans>

<orth>곱다</orth>
<defect>
<infl_rst>곱게</infl_rst>
<trans>easily</trans>

- 둘 이상일 때는 infl_rst 요소를 반복해서 제시한다.

(예) <orth>따르다</orth>
<defect>
<infl_rst>따른</infl_rst>
<infl_rst>따라</infl_rst>

(1-2-6) idm_grp (숙어 구획)

- 표제항 용언을 포함하는 숙어들을 제시해 주는 구획이다.

(1-2-6-1) idm (숙어)

- 표제항 용언을 포함하는 숙어의 형태를 기입한다.
- 둘 이상의 숙어 표현이 있을 경우, idm 항목을 반복해서 제시해 준다.
- 의미가 같고 조사 등 극히 일부분의 변이가 있을 경우에도, idm 항목을 반복해서 제시해 준다.

(예) <orth>펼리다</orth>
<idm_grp>
<idm>배알이 펼리다</idm>
<idm>뺨이 펼리다</idm>

1.2.2.3. 목록 관리 및 형식 검증

2007년은 기성결과물에 대한 보완·정제를 목표로 하는 만큼 기기술된 27,700 항목을 대상으로 하였다. 단, 보완·정제 과정에서 일부 항목들을 삭제되거나 추가되는 경우가 발생하여 최종 29,512 항목이 기술 완료되었다. 기술 항목이 늘어난 것은 세종전자사전 분과 내 사자성어 연구팀에서 작성한 사자성어 목록을 검토하여, 여기에 ‘하다’ 등이 붙어서 용언으로 자주 사용된다고 판단되는 것들을 추가로 기술하였기 때문이다.

⇒ 기술 목록은 결과물 CD 참조

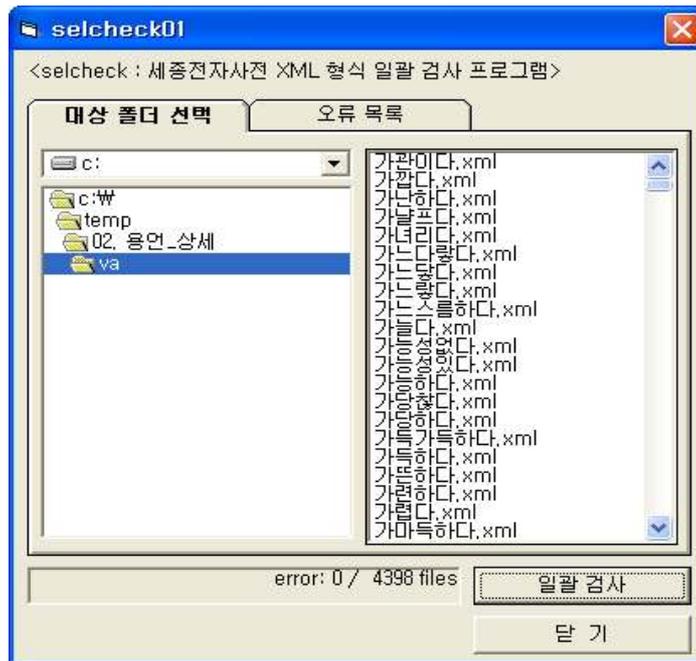
한편 목록 관리의 형식 검증과 함께 그 결과물 수치를 확인하는 과정을 거치게 되는데 그 과정은 다음과 같다.

1.2.2.3.1. 용언 사전의 형식 검증

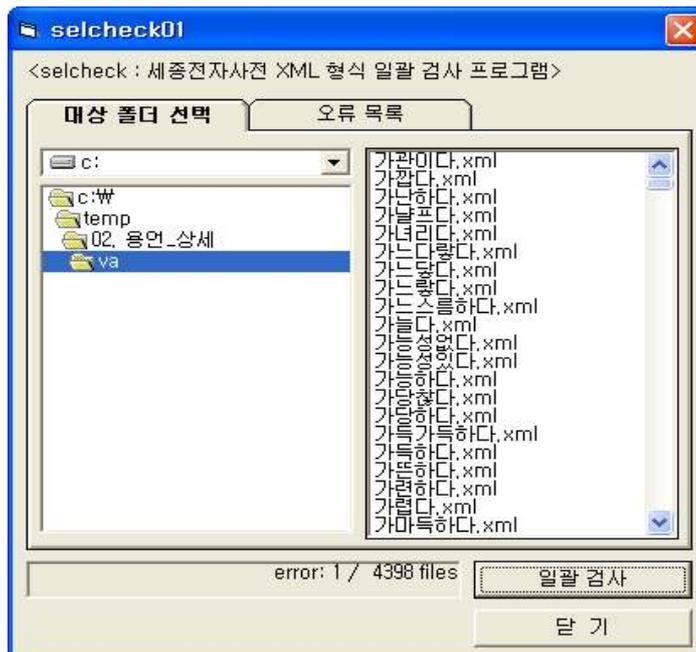
용언 분과에서는 용언상세사전의 기술이 사전에 정의한 미시 구조에 맞는지를 검증하기 위해, dtd를 작성하고 이를 기계적으로 검증하는 방식을 사용하였다. 구체적인 검증에는 세종전자사전에서 만든 selcheck라는 프로그램을 이용하였다. 검증 과정은 다음과 같다. 형용사 사전을 대상으로 검증 과정을 보이면 다음과 같다.

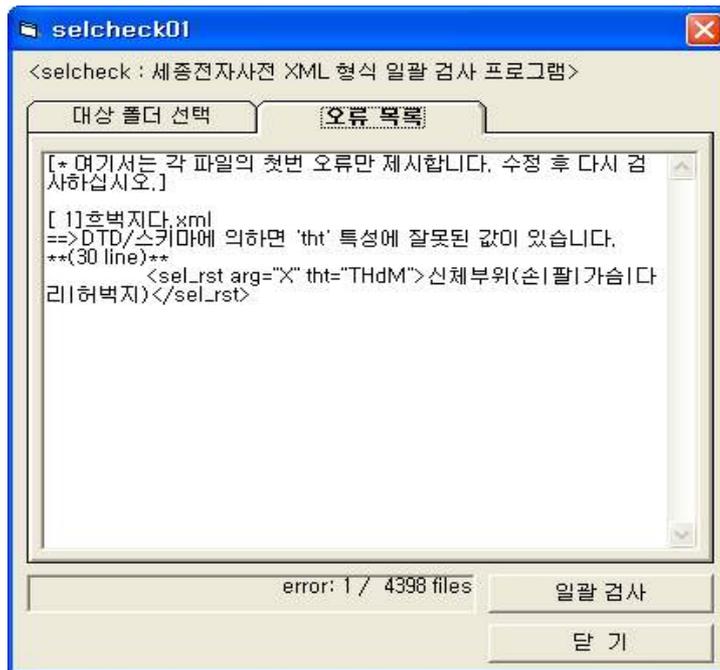
ㄱ. selcheck 프로그램을 구동시킨다.

ㄴ. 상세 원고가 들어 있는 폴더를 클릭한 후, 일괄검사를 실시한다.



위 화면은 형식 오류가 없는 경우이다. 만약 형식 오류가 있으면 아래와 같이 오류가 있는 파일을 알려준다.





ㄷ. 형식 오류가 있는 경우 이를 수정하고 (ㄴ)과 같이 검증 과정을 다시 거친다.

1.2.2.3.2. 체언 사전 superEntry, entry, sense 항목 수 검증

용언분과에서는 Unix 명령어를 이용하여 superEntry, entry, sense의 개수를 검증하였다. Unix 명령어 중 grep과 wc가 이용되었다. grep은 특정 라인을 찾아주는 기능이 있고, wc는 해당 라인을 카운팅해주는 기능이 있다. 동사와 형용사 사전의 작업 과정을 각각 보이면 다음과 같다.

ㄱ. 형용사 사전의 superEntry, entry, sense

```

C:\#temp#02. 용언_상세#va>grep "</superEntry>" *.xml|wc
4398      4398      126574

C:\#temp#02. 용언_상세#va>grep "</entry>" *.xml|wc
4545      9090      112487

C:\#temp#02. 용언_상세#va>grep "</sense>" *.xml|wc
6224      12448     158388

```

ㄴ. 동사 사전의 superEntry, entry, sense

```
C:\temp\#02. 용언_상세\#00>grep "</superEntry>" *.xml|wc
15181 15181 432612
C:\temp\#02. 용언_상세\#00>grep "</entry>" *.xml|wc
16517 33034 402738
C:\temp\#02. 용언_상세\#00>grep "</sense>" *.xml|wc
23268 46536 582619
```

결과적으로 형용사 사전의 경우 4398개의 superEntry, 4545개의 entry, 6224개의 sense가 기술되어 있음을 확인할 수 있었다. 동사 사전의 경우 15181개의 superEntry, 16517개의 entry, 23268개의 sense가 기술되어 있음을 확인할 수 있었다.

1.2.2.4. 사전 구축 관련 기반 연구

2007년에 이루어진 용언사전 관련 기반 연구는 크게 용언의미부류 관련 연구와 구문 분석 관련 연구가 있다.

1.2.2.4.1. 용언의미부류 관련 연구

용언사전에 용언의미부류 정보를 새로 부착하기로 함에 따라, 이 작업의 기초를 이루는 용언의미부류 체계를 구축하는 작업을 수행하였다. 이 작업은 이미 구축된 명사의미부류 체계 중 [사태]의 하위부류체계를 출발점으로 삼았다. 여기에서 이미 [사태]의 여러 하위부류들이 꽤 상세하게 설정되어 있기는 하나, 그것은 명사만을 대상으로 하여 만들어진 것이기 때문에, 용언에 적용하기에는 부족한 점이 많이 있었다. 그래서 용언의 검토 결과 많은 부류가 추가되었고, 새로 추가된 부류에 대해서는 정의, 예, 판정기준 등에 대한 지침이 마련되었다. 용언의미부류 기술 지침에 대한 상세한 사항은 1.2.2.1.2.에서 제시되었으므로 여기서는 용언의미부류 구축과 관련된 사항들만을 간략히 제시하도록 한다.

(가) 용언의미부류의 구축

용언의미부류는 이미 구축되어 있는 명사의미부류체계를 바탕으로 용언의 의미기술을 위해 필요한 부류들을 추가 또는 조정하는 방법으로 구축되었다. 명사의미부류 중 특히 사태에 속하는 부류들로 용언의미부류가 구성이 되는데 이를 위해 추가되거나 조정된 부류들을 다음과 같다.

<사태>

<정적사태>

<속성값>: <속도속성값>, <평가속성값>, <촉감속성값>, <소리속성값>

<상태>: <관계상태>, <시간상태>, <기상상태>

<행위>

<물리적 행위>

<단독 행위>: <감각 행위>, <방식 행위>, <소리내기>, <섭취 행위>, <배설 행위>, <반복적 행위>

<방향성 행위>: <차별>, <영향>, <제약>, <간섭>, <수여>

<대칭적 행위>: <만남>

<비의도적 행위>: <생리 행위>, <죽음>, <출생>, <인간의 변화>, <움직임>, <피동적 행위>

<추상적 행위>

<인지적 행위>: <비교>, <평가>

<조사>

<소통 행위>: <허락>, <반대>

<현상>: <반복적 현상>

<상태 변화>: <종료>, <변형>, <악화>, <개선>

상기 작업을 통해 마련된 용언의미부류 체계는 명사의미부류 체계의 [사태] 부류와 비교할 때, 특히 [행위] 부류의 하위 체계가 상당히 확장·보완되었다.

(나) 용언의미부류의 체계

☞ 결과물 CD의 『기반 연구 결과물』 참조

1.2.2.4.2. 구문분석 관련 연구

세종전자사전 분과에서는 세종전자사전의 전산적 활용성을 검증 및 그 응용 방안 연구의 일환으로 세종전자사전을 기반으로 하는 구문분석기 개발을 위한 연구를 수행하였다. 구문분석 시에는 용언 전자사전의 정보가 핵심적이고도 필수적인 만큼 용언 전자사전 분과에서 이 연구에 적극 참여하였다. 이를 통해 용언 전자사전의 문형정보, 선택제약 정보 등 각종 정보의 활용 방안을 제시하여 세종 전자사전의 관련 정보들을 효과적으로 활용하는 구문분석기의 개발을 가능하게 하였다. 그 결과는 제 18회 한글 및 한국어 정보처리 학술대회에서 논문으로 발표되었다.

☞ 결과물 CD의 『기반 연구 결과물』 참조

1.2.2.5. 회의

☞ 결과물 CD의 『회의록』 참조

1.3. 기획·운영 소분과

1.3.1. 2007년도 연구 계획 및 추진 결과

1.3.1.1. 개요

기획·운영 소분과의 2007년도 사업은 크게 다음의 세 가지 업무를 중심으로 추진하였다. 첫째, 세종 전자사전 개발 분과 전체 및 하위사전 분과들의 연구를 총괄 기획·관리하고, 둘째, 워크숍 및 세미나, 그리고 분과 전체 또는 분과간 각종 회의를 기획·진행하며, 셋째, 사업계획서 및 각종 보고서를 작성·관리하는 것이다. 그 구체적인 업무 내용별 추진 결과는 다음과 같다.

1.3.1.2. 연구 내용별 추진 결과

분과명	연구내용	추진일정						추진현황		
		6월	7월	8월	9월	10월	11월	12월	진척상황	진척율
기획·운영 소분과	▶ 세종전자사전 분과의 연구 기획·관리								완료	100%
	▶ 사업계획서 및 각종 보고서 작성								완료	100%
	▶ 하위분과별 연구 기획·관리								완료	100%
	▶ 워크숍 및 각종 회의 기획 및 진행								완료	100%
	▶ 세종계획 분과간/하위사전분과 및 소과제분과 간 협의회 운영								완료	100%

1.3.2. 기획·운영 소분과 연구 수행 내역

1.3.2.1. 세종전자사전 분과 전체 및 하위분과별 연구 기획·관리

2007년은 최종단계 사업연도로서 세종전자사전 구축 작업이 완료되는 해이다. 따라서, 세종전자사전 전반에 대한 집중적인 보완·정제 작업과 이를 통한 세종전자사전의 성능 및 제고에 모든 역량을 투입하였다.

기획·운영 소분과는 이러한 시각에서 세종 전자사전 개발 분과의 당해연도 연구계획을

수립하는 한편 그 추진을 총괄 관리하였다. 또한 이와 병행하여 각 하위분과별 연구도 기획·관리하였다.

⇒ 상세한 내용은 본보고서 「 II. 사업개요」 및 「 III. 연구 추진 계획 및 결과」을 참조할 것.

1.3.2.2. 계획서 및 각종 보고서 작성

기획·운영 소분과는 세종 전자사전 개발 분과 사업과 관련된 계획서 및 각종 보고서 작성을 총괄, 담당한다. 전체 사업의 방향과 세부 내용, 추진 일정 등을 수립하고, 이를 관리한 내용과 그 결과들을 담고 있다.

(자세한 내용은 2007년 사업계획서, 중간보고서 및 본보고서 참조)

1.3.2.3. 워크숍, 세미나 및 각종 회의 기획과 진행

기획·운영 소분과에서는 세종 전자사전 개발 분과의 각종 연구를 기획·관리할 뿐 아니라 연구 추진을 지원하기 위해 각종 관련 회의들도 기획하고 진행하는 업무도 담당하였다.

1.3.2.2.1. 전체회의 및 전문가 초청 세미나 개최

기획·운영 소분과는 2007년도에도 참여연구진 전원이 참여하여 최종 보완·정제작업 과정과 현황에 대해 발표하고 논의하는 전체회의를 정례적으로 개최하였다. 또한 수차례에 걸쳐 국내외 전문가들을 초청한 세미나를 개최함으로써 연구진 전원이 전자사전 개발을 위한 이론적 기반을 다질 수 있는 기회를 마련하였다. 이 과정을 통하여, 각 분과의 작업 결과들을 검토, 정리하고 이에 따라 향후 작업 방향을 수정, 보완하였다. 또한, 전자사전 구축 관련 분야 최신 이론을 습득하고 및 응용 현황을 파악하여 세종 전자사전 구축에 적용하는 가능성을 모색하는 기회를 제공하였다. (전체회의 및 전문가 초청 세미나 개최 현황 및 구체적 내용은 결과물 CD를 참조)

1.3.2.2.2. 분과사전 책임자, 연구원 회의 및 세미나

세종 전자사전 개발 사업의 유기적이고 효율적이고 추진을 위하여 하위분과 및 소분과 책임자들과 실무 연구원들이 참여하는 각종 회의와 세미나를 개최하였다. 이 과정을 통해 각 분과의 사업 추진 시 부딪치는 제반 문제들에 대한 심도 있는 검토와 논의가 이루어졌으며, 최선의 해결방안을 마련 공유할 수 있도록 하였다. 또한, 각 분과의 책임자들과 실무 연구진들이 해당 분과 외에 타 분과의 연구과제에 대한 이해를 심화함으로써 분과 간에 보다 유기적인 연계체계를 유지할 수 있도록 하였다. (자세한 사항은 결과

물 CD의 회의록을 참조)

1.3.2.2.3. 핵심사전분과 회의 및 세미나

기획·운영 분과는 체언 및 용언 사전, 통합지원 분과 등을 포괄하는 핵심사전분과의 각종 회의와 세미나를 기획·운영하였다. 이를 통해 핵심사전 구축에서의 각종 현안들을 논의하고 해결방안을 마련함으로써 세종 전자사전 내의 정보내용과 항목표상의 통일성을 제고하기 위한 시안들을 마련하였다. (자세한 사항은 결과물 CD의 회의록을 참조)

1.3.2.2.4. 분과 간 협의회 개최

기획·운영 분과는 또한 현안별 관련 분과들의 실무자를 중심으로 한 분과 간 협의회를 운영하였다. 2007년도는 특히 세종 전자사전 각 분과의 미시구조 및 범주체계의 표준화 방안을 확정짓고 이를 반영하는 것과 기구축 사전들에 대한 정보확장 및 보완·정제 작업을 실질적으로 마무리 짓은 것이 주된 과제임에 비추어, 이와 관련한 분과 간 협의회를 상설 운영하였다. (이와 관련된 결과는 결과물 CD의 회의록을 참조)

1.3.2.4. 회의

☞ 결과물 CD의 『회의록』 참조

1.4. 통합지원 홍보

1.4.1. 2007년도 연구 계획 및 추진 결과

1.4.1.1. 개요

당해연도 통합지원 및 홍보분과의 연구는 1)핵심전자사전 보완·정제 작업 지원, 2)세종전자사전 표준미시구조/표준tagset 보완·정제, 3)세종전자사전 표준 기술지침안 보완·정제, 4)세종전자사전 슬어집 보완·정제, 5)세종전자사전 활용방안 연구 및 홍보자료 작성 등으로 요약된다.

1.4.1.2. 연구 내용별 추진 결과

분 과 명	연구 내용	추진 일정						추진 현황		
		6 월	7 월	8 월	9 월	10 월	11 월	12 월	진척 상황	진척율
통 합 지 원 · 홍 보 소 분 과	▷ 핵심 전자사전 보완·정제 작업 지원									100%
	▷ 세종전자사전 표준미시구조/표준 tag set 보완·정제									100%
	▷ 세종전자사전 표준 기술지침안 보완·정제									100%
	▷ 세종 전자사전 슬어집 보완·정제									100%
	▷ 세종전자사전 활용방안 개발 및 홍보자료 작성									100%

1.4.2. 연구 수행 내역

1.4.2.1. 핵심사전분과 보완·정제작업 지원

1.4.2.1.1. 사전 데이터의 형식 검증, 평가 및 관리

체인사전, 용언사전 등 수합된 결과물의 XML 형식 검증 작업은 일차적으로 검증도구를 이용하여 하위사전분과 내부에서 이루어졌다. 그 결과물에 대한 2차 검증을 통합지원분과에서 담당하였고, 2차 검증을 거친 결과물을 통합전자사전분과로 전달함으로써 핵심사전분과의 보완·정제 작업을 지원하였다. 최종보고 시점을 기준으로 볼 때, 특기할 만한 형식오류는 2차 검증 시에 발견되지 않았다.

1.4.2.1.2. 각종 어휘자료 수집 및 표제어 목록 관리

금년도에는 핵심사전인 체언사전과 용언사전의 상세전자사전을 대상으로 한 보완·정제 작업만이 수행되었으므로, 사전 구축을 위한 기초데이터의 수요는 없었다고 할 수 있다. 대신, 핵심사전의 보완·정제 과정에서 추가 및 삭제되는 표제어들의 목록을 전달받아 이를 반영하는 것이 주요 작업이라 할 수 있다.

1.4.2.2. 세종전자사전의 표준미시구조 및 tag set 보완·정제

세종전자사전의 미시구조 및 tag set 표준화 작업은 주로 미시구조 설계와 관련하여 하위사전간 정보 표상의 통일성과 일관성을 담보하기 위한 것이다. 세종전자사전은 하위사전들에 미시구조 설계의 독자성을 어느 정도 인정함으로써 각 범주 특유의 정보항목을 체계적, 효율적으로 기술할 수 있도록 설계되었다. 표준화작업은 이러한 조건 하에서 전체 사전의 미시구조를 조정함으로써 사전 간 공통 정보와 유사 정보들의 상호연계성을 강화하는 것을 목표로 한다. 따라서 표준화 작업은 사전데이터의 저장과 관리에 있어 매우 효율적인 XML 기술 체제의 적용에 초점이 맞추어져 있다.

이와 관련된 주요 작업들은 2006년도까지의 사업을 통해 대부분이 수행되었는데 그 결과물 중 핵심사전의 보완·정제 결과를 반영하는 것이 금년도 사업의 주목표이다. 그 내용은 다음과 같다. 1) 핵심사전에 대해 가독성과 확장성 측면에서 보완작업을 수행. 2) 그 결과를 기반으로 기구축 사전에 대해 그 유형에 따라 체언 혹은 용언 두 핵심 사전의 틀로 보완하여 표준화 작업을 완료하였다. 3) 하위사전간 유사정보 항목은 '가능한 범위 내로 제한하여' 동일한 Tag를 적용하였다. 4) 동일정보 항목의 위계 및 배열은 미시구조가 대대적으로 변동되지 않는 범위 내로 최소화하였다. 5) 유사정보 항목은 하위사전별 특성을 고려하여 조정 여부를 별도로 논의하였다.

실질적인 표준화 작업은 기본 방안과 틀을 마련한 다음 핵심사전의 보완·정제 작업의

최종결과물이 나온 후반부에 수행하였다.

☞ 결과물 CD의 『기반 연구 결과물』 참조

1.4.2.3. 세종전자사전의 표준 기술 지침안 및 술어집 보완·정제

2007년 세종전자사전의 표준 기술지침안 보완·정제 작업은 세종 전자사전 하위사전 분과의 미시구조와 정보항목 영문 tag 및 한글 명칭 표준화 작업과 구축 지침서 표준화 작업에 중점을 두었다. 이는 올해가 세종전자사전 개발 사업을 마무리 짓고, 그 결과를 공개하는 만큼 사용자들의 이해를 돕기 위한 표준적인 미시구조와 구축지침서를 제공하기 위함이었다(1.4.2.1. 참조). 이와 함께, 세종전자사전 활용방안 연구의 일환으로 수행된 ‘세종전자사전의 전산적 활용을 위한 가이드 북(가칭)’ 작성 방안 연구와 함께 수행되었다(1.4.2.4. 참조). 한편, 술어집 보완·정제는 각 하위분과 사전에 기존 술어집의 관련 항목들을 전달하여 추가 및 삭제, 또는 보완할 사항들을 정리하여 제출하게 하여 이를 종합하고 정리하는 방식으로 수행하였다.

☞ 결과물 CD의 『기반 연구 결과물』 참조

1.4.2.4. 세종전자사전 활용방안 연구

1.4.2.4.1. 개요

세종전자사전 구축사업의 종료를 앞두고 결과물 데이터의 전산적 활용도를 제고하기 위한 목적으로 ‘세종전자사전의 전산적 활용을 위한 가이드 북(가칭)’을 작성한다.

(가) 작성의 이유

	사전구축지침서	가이드 북
관점	개발자 관점	사용자 관점
목적	기술의 체계성/일관성 유지	정보의 즉각적 활용 및 평가
특징	하위사전별 해설	정보항목별 통합 해설
	방대함, 난해함(?)	간결하고 쉽게 정리. 구체적 적용사례 예시
	열거식	인덱스 형식

(나) 작성 목적 및 효용성

- 세종전자사전의 활용도 제고.

자연언어처리, 검색, 문서요약 등의 알고리즘 개발에서 세종전자사전의 ‘유용성’ 검증.

‘개별정보’의 전산적 활용도 평가.

- 코퍼스 상의 확률에 기초한 ‘통계 및 기계학습 기반’의 자연어처리 기법을 대신할 ‘사전기반’의 자연어처리 기법 확산을 위한 사전사용자용 가이드북 기능.
- 사전 데이터 활용에 앞서 사용자들에게 요구되는 ‘학습’ 부담 경감 효과.

1.4.2.4.2. 작성 지침

가) 체제

- 가이드북은 크게 2부로 나누되,
- 1부는 첫째, 세종전자사전의 구축 목적, 특성과 구축 원리 등에 대한 압축된 해설과, 둘째, 하위사전(언어, 어미, 관용표현...)별 구축의 당위성 및 활용성 등에 대한 설명으로 이루어진다.
- 2부는 ‘한글정보태그’ 또는 ‘정보처리 유형’별 인덱스 형식으로 작성한다. (4-1. <예시1>과 4-2. <예시2> 참조)
- 하위사전의 해당 정보항목들은 통합 해설한다. (정보항목 구성에 대해서는 논의 필요)

나) 정보 항목의 구성

- 개별 정보 항목의 전산적 활용 방안

● XXX	한글정보태그 (또는 정보처리 유형)
1. 부사	해당 품사범주
ㄱ) 정보 항목의 내용	간략 설명
ㄴ) 정보 표상 방식	간략 설명
ㄷ) 활용가능 분야	구문분석/ 형태소분석/ key word 추출...
ㄹ) 전산적 활용 방안	예시
2. 연어	
...	

	‘한글정보태그별’ 항목 구성 방식	‘정보처리 유형별’ 항목 구성 방식
장점	기존 자료를 직접 활용할 수 있으므로 작성작업의 부담이 작음	사용자의 needs에 직접 대응 가능
단점	사용상의 불편	이른바 ‘기술수요조사’ 같은 조사과정을 거쳐 정보항목 리스트를 만들어야 하므로, 제한된 기간에 작업하는 데

	<p>어려움이 예상된다.</p> <p>*** 그러나, 'exhaustive list' 작성을 위한 출발 작업이라 생각하고 1단계 list를 작성하는 것을 목표로 삼으면 가능한 작업.</p>
--	---

- 복수 정보 항목의 복합적 활용 방안: 자유 서술

다) 기타

- 2차 사전 개발자는 물론 응용프로그램 개발자, 일반 사용자 등이 쉽게 이해할 수 있는 용어와 개념을 써서 작성한다.
- 항목의 해설은, 가령 전산 알고리즘 개발자가 '미해결 문제'를 해결하기 위해서는 어떤 '언어학적 개념'을 고려하고 어떤 'clue'를 써야 하는지 등의 정보를 구체적 사례를 통해 제시함으로써 세종전자사전의 직, 간접적 활용도를 높일 수 있게끔 작성한다.
- 기술항목의 선정과 작성은 각 하위사전분과에서 맡고, 형식 통일과 편집은 별도 인력이 맡는다.

1.4.2.4.3. 작성 예시

*** 부사사전 구축지침서의 '통사정보구획a'의 시안

<pre> <syn_grp_a> <comb pos="v" exist="yes"></comb> <comb pos="adj" exist="no"/> <comb pos="det" exist="temp"></comb> <comb pos="av" exist="temp"></comb> <comb pos="n" exist="temp"></comb> <prt></prt> <argdist type=""></argdist> <col></col> </pre>	<p>-통사 정보 구획a</p> <p>-동사 결합 정보</p> <p>-형용사 결합 정보</p> <p>-관형사 결합 정보</p> <p>-부사 결합 정보</p> <p>-명사 결합 정보</p> <p>-조사 결합 정보</p> <p>-논항 분포</p> <p>-연어</p>
---	--

가) 개별 정보 항목의 활용

<예시1>

●	<p>통사정보구획A :<syn_grp_a></p> <ul style="list-style-type: none"> - 정보 항목의 내용: 통사 정보 구획은 다양한 결합 제약 정보를 기술하는 정보 구획이다. 어떤 어휘 범주를 수식하는가, 표제어 부사가 어떤 조사와 결합하는가 하는 점 등이 기술된다. - 정보 항목의 표상 방법: 결합하는 범주의 exist의 속성값에 "yes"라고 표시하고 해당 어휘 형태를 제시한다. - 전산적 활용 방안 : 각 결합 정보는 형태소 분석과 구문 분석 과정에서 유용하게 활용될 수
---	---

있다. 정확한 품사 부착 결과인지를 확인할 때에도 부사가 포함되어 있는 경우에는 그 부사의 결합 제약을 품사 부착 결과와 비교해 보는 방식으로 활용될 수 있다. 가령 “빨리 감기는 줄”의 경우 “빨리”라는 부사가 다른 형용사 등과는 결합이 불가능하고 동사와 결합한다는 결합 정보를 이용하면 “감기는”이 동사임을 검증하는 데에 도움이 될 것이다. 구문 분석의 과정에서도 “빨리”가 동사를 수식하는 부사이므로 “감기다”라는 동사를 수식하는 구조임을 확인할 수 있다.

나) 복수 정보 항목의 복합적 활용 (자유 서술)

<예시2>

● **문장부사와 성분부사의 중의성과 판별**

- 부사 사전의 정보들을 이용하여, 아래와 같은 문장에서 “확실히”가 문장부사인지 성분부사인지를 판단해 보도록 하자.

- (1) 가. 확실히 영희는 아름답다.
나. 영희는 확실히 아름답다.
- (2) 가. 확실히 철수는 밥을 먹었다.
나. 철수는 확실히 밥을 먹었다.
다. ?철수는 밥을 확실히 먹었다.
- (3) 가. 확실히 철수는 그 내용을 이해하고 있다.
나. 철수는 확실히 그 내용을 이해하고 있다.
다. 철수는 그 내용을 확실히 이해하고 있다.

(1)과 (2)는 문장부사 “확실히”의 예이다. (3)의 경우는 좀 복잡한 양상이다. (3가)은 일차적으로 문장부사로 해석되고 (3나)은 성분부사인 것으로 여겨진다. (3나)은 문장부사와 성분부사의 중의성을 드러내는 예라 할 수 있다.

현재 부사 사전에서 “확실히”의 기술 내용은 다음과 같다.

```
<superEntry>확실히
<sense n="01"> @eg=[~ 영희는 아름답다] @sem=[문장부사/양상부사/추측]
<synt_a> @comb=[v={~달리다}; adj={~아름답다}; det={}; av={}; n={~천재}]
@prt=[@argdist=[np={};vp={};ap={}] @col=[] </synt_a> <synt_b> @conc_conj=[] @conc_mood=[]
@conc_mod=[] @conc_ta=[] @conc_neg=[] @conc_agr=[] @distr=[~S~OV] </synt_b> </sense>
<sense n="02"> @eg=[그 내용을 ~ 이해해야 한다] @sem=[성분부사/양태부사/방식]
<synt_a> @comb=[v={~이해하다}; adj={*}; det={}; av={}; n={}] @prt=[@argdist=[np={};vp={};ap={}]
@col=[] </synt_a> <synt_b> @conc_conj=[] @conc_mood=[] @conc_mod=[] @conc_ta=[] @conc_neg=[]
@conc_agr=[] @distr=[S~O~V] </synt_b> </sense> </entry> </superEntry>
```

위의 기술 내용을 바탕으로 예문 (1)-(3)의 예문을 정확하게 분석해 내려면, 우선 (1)의 경우는 결합하는 용언이 형용사 “아름답다”가 출현하였다는 점을 확인하면 <sense1>의 문장 부사임을 확인할 수 있다. <sense2>의 경우는 형용사와 결합이 불가능하기 때문이다(adj={*}). (2)의 예는 현재 사전 기술 내용으로는 확인할 방법이 마땅치 않다. 가능한 한 가지 방법은 <sense1>의 결합가능동사 “달리다”와 <sense2>의 결합가능 동사 “이해하다”를 형태로 보지 않고, 부류 정보로 보아 “먹다”는 “달리다”류의 동사 부류라는 점을 포착하거나 “이해하다”류와는 다른 부류의 동사임을 파악할 수 있으면 가능하다. 앞서 언급한 바와 같이 용언의 유형 분류와 연계성이 필요함을 다시 한번 느낄 수 있다. (3)의 예는 어느 정도 가능하다. 일단 “이해하다” 동사와 결합을 하고 있으니 일단은 <sense2>일 가능성이 높다. 그러나 (3가)의 경우는 부사의 위치가 <sense2>의 경우와 일

치하지 않고 있다. 따라서 <sense2>가 아니라 <sense1>로 판단하게 할 수 있다. <3c>의 경우는 <sense2>의 기술 내용과 일치하므로 성분부사라는 판단이 가능하다. <3n>의 경우도 경우도 <sense2>의 기술내용과 일치하므로 성분부사로 판정될 것이다. 그러나 실제 <3n>의 경우는 문장 부사와 성분부사 모두로 해석될 수 있는 중의성을 지닌 문장이라는 점에서 정확한 판정은 아닌 셈이다. 어쨌든 이처럼 여러 가지 정보를 조합시키는 방식으로 사전의 활용 가능성을 높이는 방안도 가능할 것이다.

☞ 결과물 CD의 『기반 연구 결과물』 참조

1.4.2.5. 회의

☞ 결과물 CD의 『회의록』 참조

2. 통합전자사전 구축 분과

2.1. 2007년도 연구 계획 및 추진 결과

2.1.1. 개요

세종 전자사전은 대규모 한국어 어휘를 수집하고 체계적으로 분석·기술한 범용적 전자사전으로서 각종 언어 정보 처리 분야에 필수적인 자원으로서 활용될 수 있도록 구현된 언어자원이다. 전자사전 개발 분과에서는 매년 각 사전 분과별로 신규 기술 목록과 확장 기술 목록을 정하고 꾸준히 사전을 구축해 왔고, 올해에는 세종 계획에 따라 최종 마무리 및 기존 목록의 보완을 주 활동사항으로 하였다. 본 분과에서는 이렇게 각 사전 분과별로 구축한 사전들을 하나의 통합된 전자사전으로 만들고, 이를 사용자들이 체계적으로 검색 및 관리할 수 있도록 지원하였다. 이밖에도 세종 전자사전을 다양한 자연언어처리 분야에 적용해 봄으로써 세종 전자사전의 활용성 및 유용성을 검증하는 역할을 함께 수행하였다.

본 분과에서는 각 사전 분과에서 기술한 방대한 규모의 한국어 전자사전이 형식적 일관성과 내용적 적합성을 갖추고 있는지 검사하고, 매년 새로 기술되는 어휘의 추가 및 기존 어휘의 수정, 검색, 삭제 등을 쉽고 빠르게 할 수 있는 체계적인 통합 관리 시스템을 구축하기위해 노력하였다. 다음 그림은 본 분과에서 구축한 통합 환경을 나타내며, 각 구성요소는 다음과 같은 기능을 수행한다.

- 1) 통합전자사전 입력기 : 각 사전별 목록과 작성규칙(DTD파일)을 참조하여 구조화된 문서(XML 파일)를 작성
 - 통합 에디터 : 구조화된 문서를 기술하기위한 도구
 - 외형 검증 도구 : 기술된 구조화 문서의 외형을 검증 및 통합
- 2) 통합전자사전 관리 시스템 : 구조화된 문서들을 DB화하여 관리
 - XML 파일 전처리기 : 구조화 문서의 외형 재검증 및 전처리
 - XML 파일 관리자 : 구조화 문서들의 통합 관리
 - 색인 및 검색기 : 데이터의 색인 및 관리 보조
- 3) 통합전자사전 검색 시스템 : 단순검색기, 상세검색기, Web검색기 및 관리자
- 4) 통합전자사전 검증 : 구축된 전자사전에 대한 평가·검증 및 이를 수행하기위한 방법론 연구

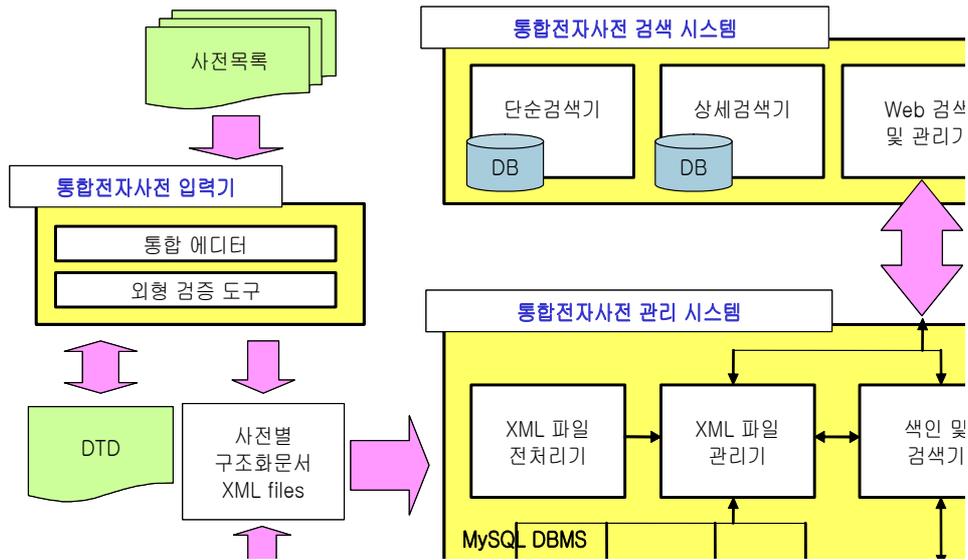


그림 1 통합 환경의 전체 구성도

2.1.2. 연구 내용별 추진 결과

분과명	연구 내용	추진 일정						추진 현황	
		7월	8월	9월	10월	11월	12월	진척상황	진척율
통합 전자사전 분야	▷ 통합전자사전 구축 - 체언 및 용언의 상세전자사전 보완·정제 결과물 DB 등록						→		100% 100%
	▷ 통합전자사전 관리 - 웹 기반의 통합 관리시스템 보완·정제						→		100% 100%
	▷ 통합전자사전 검색 - 세종 전자사전 단순검색기 및 상세검색기 수정·보완						→		100% 100%
	▷ 통합전자사전 검증 및 평가 - 세종전자사전에 기반한 형태소 분석 프로그램 수정·보완 (Web상에서의 공개·시연)						→		100% 100%
	▷ 사전구축분과 지원 - 통합입력기(통합에디터)의 수정·보완						→		100% 100%

2.2. 통합전자사전 연구 수행 내역

2.2.1. 통합전자사전 구축

통합전자사전의 구축은 세종계획이 3단계에 접어들어 이후에도 계속된 수정과 보완 작업을 통해 새로운 결과물을 통합하고 관리하는 방식으로 이루어졌다. 최종단계에서도 기존에 구축된 통합 환경을 바탕으로 각각의 처리단계가 일괄적으로 수행되며 보다 자동화된 방법을 적용함으로써 효과적인 통합전자사전 구축을 수행하였다. 즉, 새롭게 추가되거나 보완된 상세기술사전 및 확장기초전자사전의 내용을 수합한 후 순차적으로 수행되어야 할 처리단계들을 일괄적으로 전산처리하여 하나의 통합된 전자사전을 구축하였다.

통합전자사전은 Linux 서버용 통합전자사전 관리시스템을 위해 Mysql DBMS의 데이터베이스로 만들어졌으며, 이를 Microsoft사의 Access용 데이터베이스 파일인 mdb파일로 읽어온 후 필요한 형태로 변환하여 단순 검색기, 상세검색기 등에서 참조하는 데이터베이스로 사용하였다.

다음 그림은 본 분과에서 구축한 통합 환경을 나타내며, 각 구성 요소는 다음과 같은 기능을 수행한다.

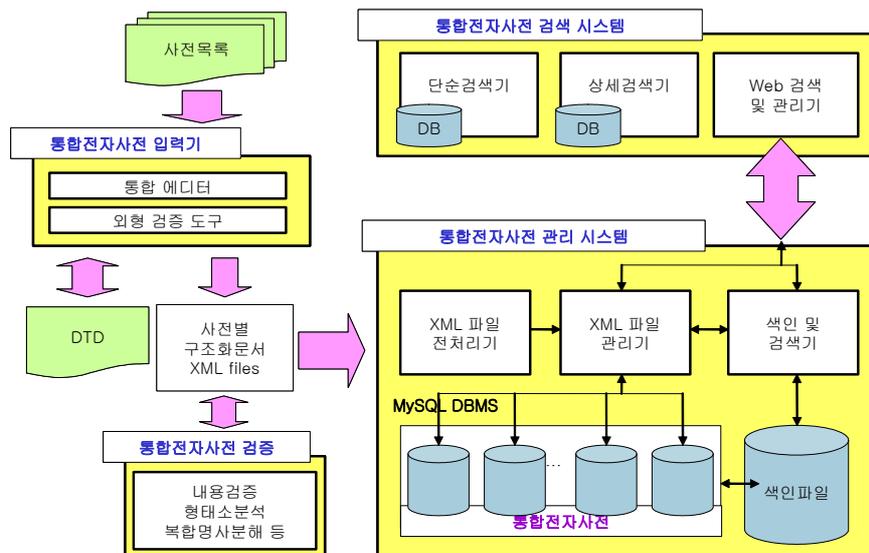


그림 1 통합 환경의 전체 구성도

- 통합전자사전 입력기 : 각 사전별 목록과 작성 규칙(DTD파일)을 참조하여 구조화된 문서(XML 파일)를 작성
 - 통합 에디터 : 구조화된 문서를 기술하기 위한 도구
 - 외형 검증 도구 : 기술된 구조화 문서의 외형을 검증 및 통합

- 통합전자사전 관리 시스템 : 구조화된 문서들을 DB화하여 관리
 - XML 파일 전처리기 : 구조화 문서의 외형 재검증 및 전처리
 - XML 파일 관리자 : 구조화 문서들의 통합 관리
 - 색인 및 검색기 : 데이터의 색인 및 관리 보조
- 통합전자사전 검색 시스템 : 단순검색기, 상세검색기(체언, 용언, 부사, 의대수, 관용 표현, 특수어), Web검색기 및 관리자(웹 기반 통합 관리 시스템)
- 통합전자사전 검증 : 구축된 전자사전에 대한 평가·검증 및 이를 수행하기 위한 방법론 연구

2.2.1.1. 각 분과의 상세기술사전 결과물 통합 및 관리

상세기술사전은 일괄적인 전산처리를 통해 하나의 통합사전으로 만들어져 관리한다. 각 분과에서 기술한 전자사전(XML 파일들)을 다른 응용 프로그램에서 쉽게 활용될 수 있도록 데이터베이스(DB)화하였고, 사전 작업자 및 사용자, 관리자들은 이러한 데이터베이스를 사용하여 필요한 정보를 쉽게 검색, 수정, 삭제, 추가할 수 있다. 특히 이번 최종 단계에서는 체언 및 용언의 상세기술사전의 보완·정제 결과물을 데이터베이스에 등록하고 다른 분과의 수정된 결과물들을 취합하여 기존의 데이터들을 갱신하는 마무리 단계에 와있다. 다음 그림은 전자사전 결과물을 통합 및 관리하기 위한 전산처리과정을 도식화한 것이고 각 구성요소들의 기능을 나열하면 다음과 같다.

- (1) DB 스키마 생성기 : 데이터베이스를 구성하는 테이블의 스키마 자동 생성
- (2) 전자사전 전처리기 : 기술된 사전의 전처리 수행 (유니코드화)
- (3) DB 입력기 : 생성된 스키마와 전처리된 사전을 DB에 입력
- (4) 색인 생성 및 XML 파일 저장 : 검색기를 위한 색인 생성 및 XML 파일 저장

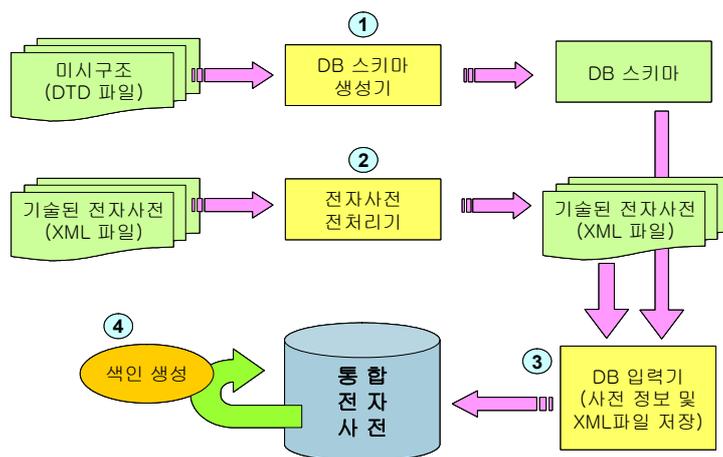


그림 2 통합전자사전 전산 처리과정 (1)

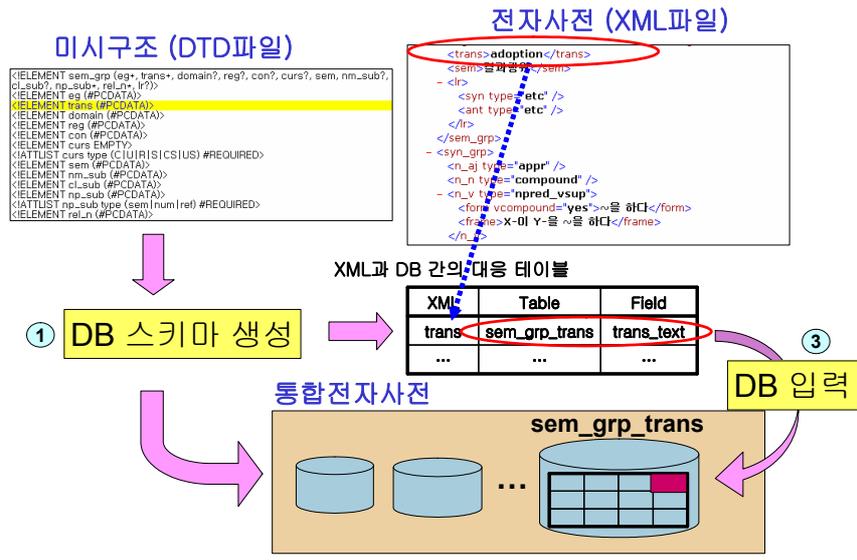


그림 3 통합전자사전 전산 처리과정 (2)

세종전자사전 XML 파일들을 DB에 넣는 과정은 크게 XML 파일의 전처리, DTD파일을 통한 DB 구조 결정 및 생성, XML 파일의 분석 및 DB로의 삽입 등으로 나누어지며 이러한 과정들을 차례대로 수행할 때 통합전자사전의 전산처리가 완료된다. 보다 구체적으로 살펴보면, 첫 번째 단계인 전처리 과정은 XML 파일을 DB에 넣기 전의 준비 단계로서 각 사전분과에서 기술된 XML 파일의 기본 인코딩인 EUC-KR 코드를 코드변환 유틸리티를 사용하여 유니코드 인코딩 UTF-8 코드로 변환하는 작업을 한다. 두 번째 단계인 DB 스키마 생성 과정은 각 사전의 미시구조가 저장된 DTD 파일을 이용해 DB에 들어갈 테이블의 구조를 기술한 스키마 파일과 각 사전의 어떤 정보항목(element)들이 실제 DB의 어떤 테이블, 어떤 필드로 들어가야 하는지를 기술한 사상관계(mapping) 파일을 만든다. 또한 본 과정에서 생성한 스키마 파일을 참고하여 필요한 테이블들을 DB에 생성한다. 이 과정이 모두 끝나면 최종적으로 XML 파일을 DB에 넣는 과정을 진행한다. 각 XML 파일을 XML 파서로 읽어 들이고, 각 정보항목을 사상관계(mapping) 파일을 참조하여 적절한 테이블의 적절한 위치에 기록한다. 이로써 통합전자사전은 모두 완성된다.

추가적으로 위의 과정들이 완료된 후에 윈도우용 단순검색기에서 사용할 테이블을 만드는 작업을 수행하게 되는데, 표제어 색인 테이블과 실제 각 사전 분과에서 기술한 XML 원본 파일을 저장한 XML테이블이 이 과정을 통해 만들어진다.

2.2.1.2. 각 분과의 확장기초전자사전 결과물 통합 및 관리

확장기초전자사전도 앞의 상세기술사전에서의 일괄적인 전산처리를 따른다. 즉 각 분과에서 기술한 수정된 전자사전(XML 파일들)을 기존 데이터베이스의 데이터와 교체하여 다른 응용 프로그램에서 쉽게 활용될 수 있도록 하고, 이를 사용하여 필요한 정보를 간편하게 검색, 수정, 삭제, 추가할 수 있도록 하였다.

2.2.2. 통합전자사전 관리

2.2.2.1. 웹 기반 통합 관리 시스템 수정 보완

웹 기반 통합 관리 시스템은 관리자, 사전 작업자, 일반 사용자 등 모든 사용자들을 위한 인터페이스를 제공하여 좀 더 원활하고 활용도가 높은 작업 환경을 제공한다.



그림 4 웹 기반 통합 관리 시스템

위의 그림을 통해서 알 수 있듯이 웹 기반 통합 관리 시스템은 체언, 용언, 부사, 관형사, 복합명사구, 연어, 고유명사, 의존명사, 조사, 어미, 관용표현, 감탄사, 특수어, 어근, 접사 상세기술사전에 대해서 검색, 추가, 수정, 삭제 기능이 구현되어있다. 특히 추가, 수정, 삭제 기능은 관리자 및 사전 작업자들에게만 사용할 수 있는 권한을 부여하였고 검색 기능은 본 시스템에 로그인한 모든 사용자들에게 그 권한을 부여하였다. 웹 기반 통합 관리 시스템의 상세기술사전 정보는 통합전자사전이 구축될 때마다 지속적으로 업데이트를 수행하고 있으며 시스템을 보완 및 수정하는 과정에서 관리자 및 사전 작업자의 편의를 위해 XML 파일을 직접 선택하여 해당 데이터를 상세기술사전에 추가할 수 있는 기능을 추가하였다. 추가할 XML 파일이 해당 사전에서 요구하는 형식을 만족하지 검사하여 해당 형식을 만족하지 못할 경우에는 오류메시지를 출력한다.

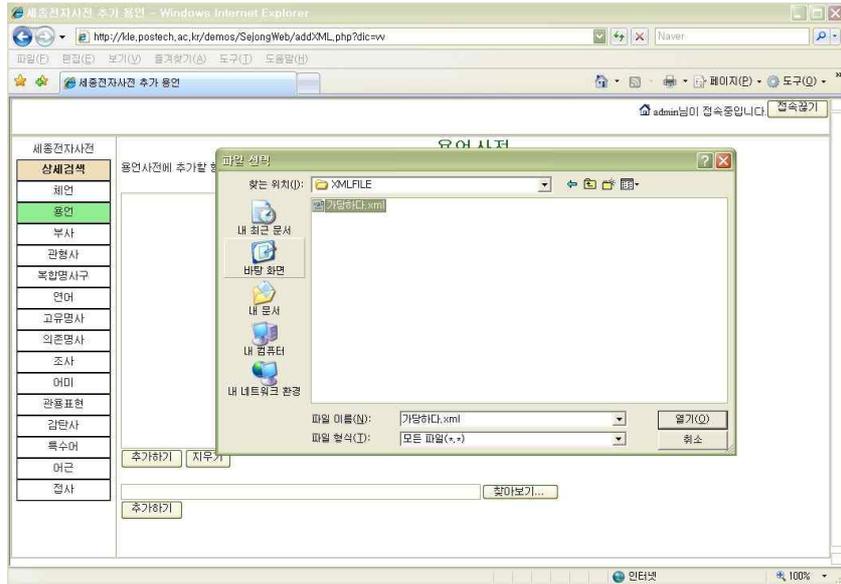


그림 5 웹 파일 추가 실행 화면

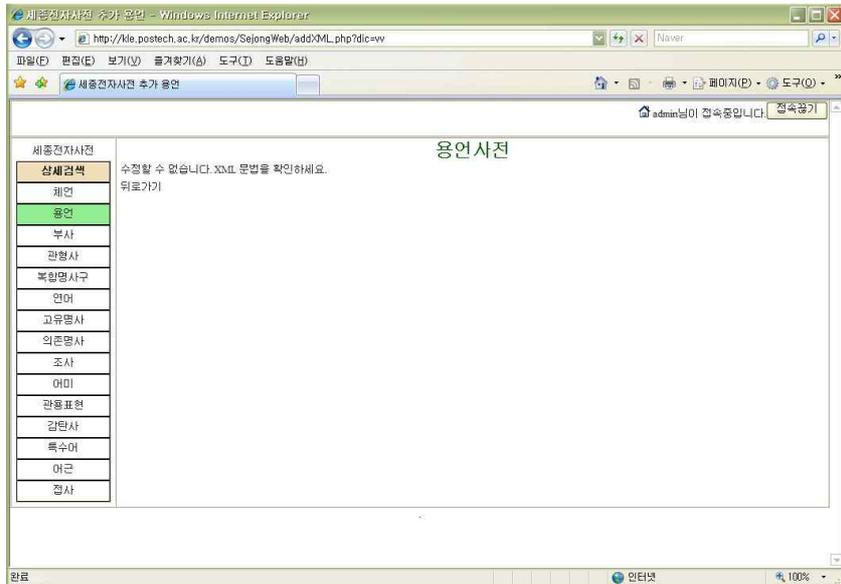


그림 6 웹 파일 추가 오류 화면

2.2.3. 통합전자사전 검색

2.2.3.1. 단순검색기 및 상세검색기 수정 보완 및 사전 업데이트

단순검색기와 체언 및 용언 상세검색기는 2003년에 처음으로 개발된 이후에 지속적인 수정과 보완, 그리고 해당 사전데이터들에 대한 업데이트를 수행하고 있다. 2006년에 개발된 부사, 관용표현, 특수어 상세검색기는 해당 분과에서 요청하는 검색 조건 및 기능들을 최대한 반영하여 수정 및 보완을 하고 있으며 인터페이스에 대한 변경도 부분적으로 적용하고 있다. 또한 기존 사전들의 변경된 내용들을 계속해서 반영하기 위한 사전 업데이트를 수행하고 있다.

이러한 검색기들은 해당 분과에서 각 사전들을 수정하거나 미시구조를 변경할 경우 이를 반영한 데이터베이스 구축 및 해당 데이터를 올바르게 처리할 수 있는 기능을 구현해야 하기 때문에 최종 단계에서도 이에 대한 수정 보완 및 업데이트가 수행되었다.

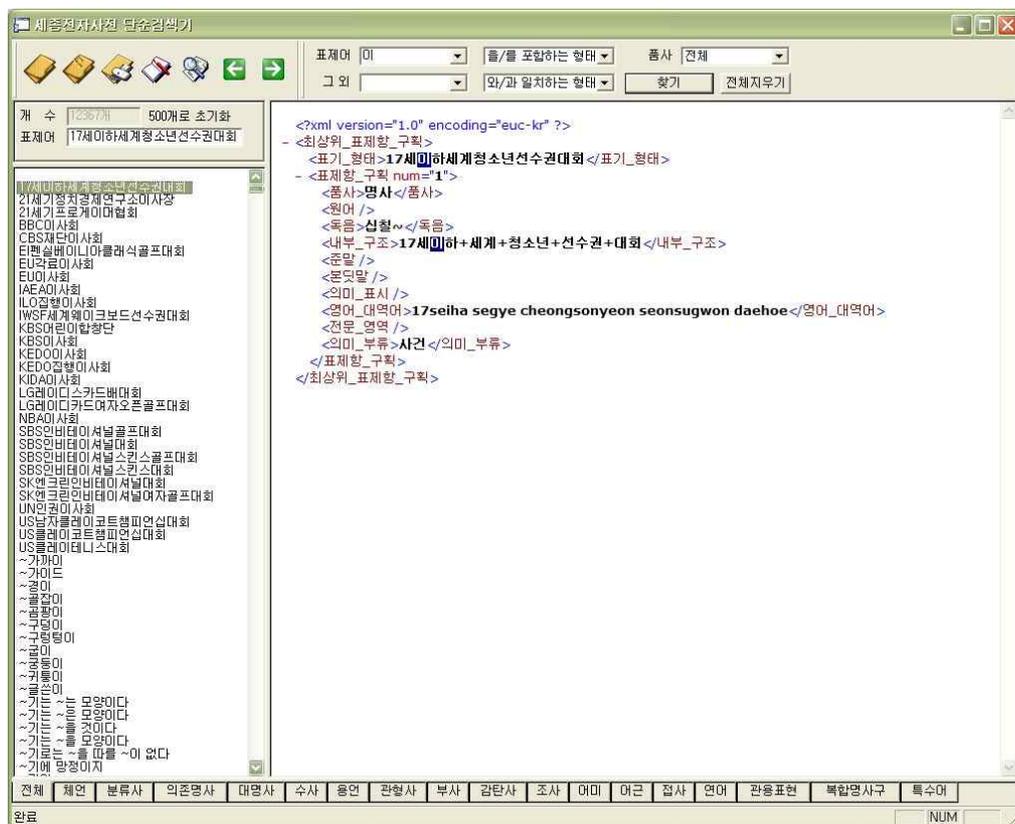


그림 7 단순검색기 기본 창

2.2.3.2. 의대수 상세검색기 개발

의대수 상세검색기는 2002년에 작업이 완료된 의존명사, 분류사, 대명사, 수사 사전의 내용을 간단한 검색 조건 및 세부적인 상세 설정을 통하여 검색할 수 있는 윈도우즈 응용프로그램으로 개발되었다. 의대수 상세검색기는 상세검색기에 공통적으로 적용되는 요구사항인 검색(완전일치 검색, 부분일치 검색), 저장(표제어 목록 저장, 표제어 XML 파일 저장), 출력(표제어 XML 파일 출력) 기능을 만족시킨다.



그림 8 의대수 상세검색기

위 그림의 각 항목의 기능은 다음과 같다.

- 표제어 형태 : 검색할 표제어를 입력한다. (대상 사전들의 공통된 검색조건)
- 용례 : 검색할 용례를 입력한다. (대상 사전들의 공통된 검색조건)
- 통합 : 모든 대상 사전들에 대한 통합검색을 수행하도록 설정한다.
- 의존명사 : 의존명사 사전에 대한 검색을 수행하도록 설정한다.
- 분류사 : 분류사 사전에 대한 검색을 수행하도록 설정한다.
- 대명사 : 대명사 사전에 대한 검색을 수행하도록 설정한다.
- 수사 : 수사 사전에 대한 검색을 수행하도록 설정한다.
- 상세설정 : 상세설정 창을 활성화시킨다. (통합검색시 비활성화)
- 질의어 : 현재 적용하고 있는 검색 조건을 화면에 출력한다.
- 검색조건초기화 : 검색 조건을 초기화한다.
- 검색시작 : 현재 적용하고 있는 검색 조건으로 검색을 수행한다.
- 표제어 목록 저장 : 현재 화면 출력된 표제어 목록을 저장한다.
- XML 파일 저장 : 현재 화면 출력된 표제어 XML 파일을 저장한다.
- XML 파일 출력 : 현재 화면 출력된 표제어 XML 파일을 출력한다.

(가) 기본 검색

기본 검색 조건은 표제어와 용례를 사용하며 보조 검색 조건으로서 완전일치, '시작'부분일치, '가운데'부분일치, '끝'부분일치를 설정할 수 있다. 표제어와 용례를 기본 검색 조건으로 선언한 이유는 의존명사, 분류사, 대명사, 수사 사전에서 동일하게 사용하는 검색 조건이기 때문이다. 또한 검색을 수행하기 전에 검색할 대상 사전을 선택하게 되는데, 다음 그림은 기본 검색 기능을 사용하여 '가'를 포함하고 있는 표제어를 가진 분류사들을 검색한 결과이다. 기본적으로, 본 상세검색기는 검색된 모든 표제어의 수를 표제어

목록 상단에 표시해주고 검색된 표제어들 중 하나를 선택하여 더블클릭하면 해당 표제어의 XML 파일 내용이 화면에 출력된다.

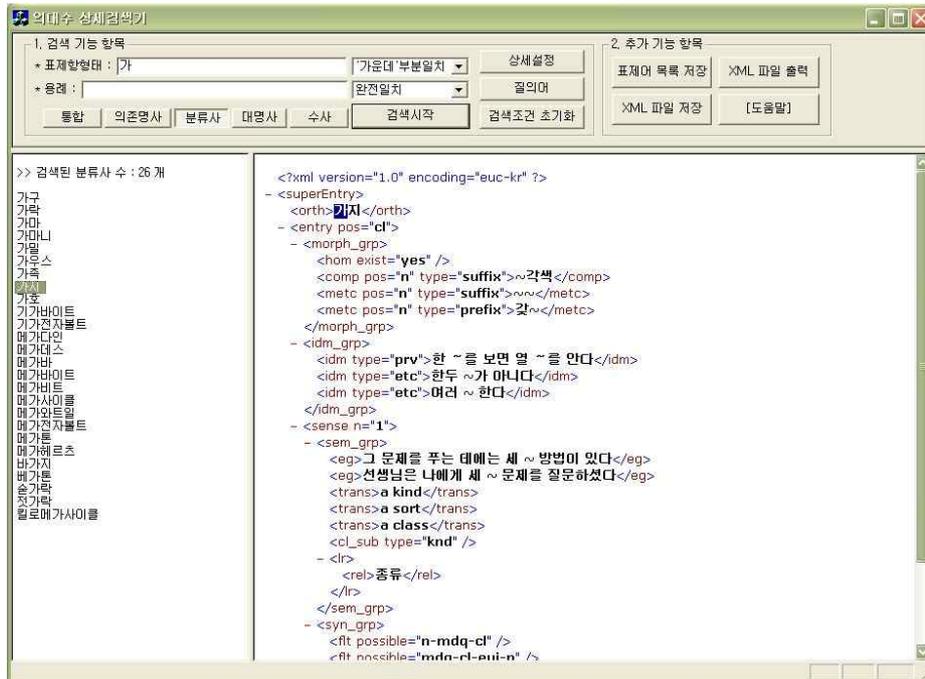


그림 9 의대수 상세검색기 기본 검색 결과

(나) 상세설정 검색

상세설정 검색 조건은 기본 검색 조건 이외의 세부적인 검색 조건을 설정할 때 사용한다. 검색할 대상 사전에 따라 상세설정 창의 내용이 변경되며 통합검색을 수행할 경우에는 상세설정 검색을 수행할 수 없다. 설정된 각 조건들은 AND연산을 적용하여 검색 조건을 갱신한다. 대상 사전별 각 항목의 검색 조건 기능은 다음과 같다.

1) 의존명사 상세설정

- 레지스터 : 문어, 구어, 아어, 고어, 비속어 중에서 하나를 선택하여 검색할 수 있다.
- 조사제약 : 조사제약 내용을 입력하여 검색할 수 있다.
- 동사결합정보 : 동사결합정보(동사 종류, 내용)를 입력하여 검색할 수 있다.
- 형용사결합정보 : 형용사결합정보(형용사 종류, 내용)를 입력하여 검색할 수 있다.
- 이다자유구성 : '이다' 결합에 대한 정보를 입력하여 검색할 수 있다.
- 한정사제약 : 한정사제약 내용을 입력하여 검색할 수 있다.

- 수식어명사 : 수식어명사 내용을 입력하여 검색할 수 있다.
- 관형절정보 : 관형절 내용을 입력하여 검색할 수 있다.
- 반복구성 : 해당 어형의 반복구성 내용을 입력하여 검색할 수 있다.
- 기타구성 : 반복구성 이외의 결합양상 내용을 입력하여 검색할 수 있다.

그림 9 의존명사 상세설정 창

2) 분류사 상세설정

그림 10 분류사 상세 설정 창

- 변이형 : 변이형의 완전일치, '시작'부분일치, '가운데'부분일치, '끝'부분일치를 설정하여 검색할 수 있다.
- 원어 : 원어의 완전일치, '시작'부분일치, '가운데'부분일치, '끝'부분일치를 설정하여 검색할 수 있다.
- 기호 : 해당 분류사의 기호를 입력하여 검색할 수 있다.
- 하위분류정보 : 하위분류정보의 종류(uni, grp, div, lnt, wdt, wgt, blk, sum, tim,

fqn)와 내용을 입력하여 검색할 수 있다.

- 분류사관련명사 : 분류사관련명사를 입력하여 검색할 수 있다.
- 결합가능정보 : 'n-mdq-cl', 'mdq-cl-eui-n', 'mdq-cl' 중에서 하나를 선택하여 검색할 수 있다.
- 기수/서수정보 : 기수정보(ko, si, etc)와 서수정보(ko, si, etc)를 입력하여 검색할 수 있다.

3) 대명사 상세설정



그림 11 대명사 상세 설정 창

- 변이형 : 변이형의 완전일치, '시작'부분일치, '가운데'부분일치, '끝'부분일치를 설정하여 검색할 수 있다.
- 내부구조 : 내부구조 내용의 완전일치, '시작'부분일치, '가운데'부분일치, '끝'부분일치를 설정하여 검색할 수 있다.
- 원어 : 원어의 완전일치, '시작'부분일치, '가운데'부분일치, '끝'부분일치를 설정하여 검색할 수 있다.
- 의미항구획 : 의미항구획을 이루고 있는 sem(1, 2, 3, interrogative, negative, reflexive, reciprocal, definite, indefinite), num(singular, plural, etc), ref(human, thing, time, place)항목을 입력하여 검색할 수 있다.
- 레지스터 : 문어, 구어, 아어, 고어, 비속어 중에서 하나를 선택하여 검색할 수 있다.
- 내포의미 : 내포의미 내용을 입력하여 검색할 수 있다.
- 조사제약 : 조사제약 내용을 입력하여 검색할 수 있다.
- 한정사제약 : 한정사제약 내용을 입력하여 검색할 수 있다.
- 수식어명사 : 수식어명사 내용을 입력하여 검색할 수 있다.
- 기타구성 : 반복구성 이외의 결합양상 내용을 입력하여 검색할 수 있다.

4) 수사 상세설정



그림 12 수사 상세 설정 창

- 변이형 : 변이형의 완전일치, '시작'부분일치, '가운데'부분일치, '끝'부분일치를 설정하여 검색할 수 있다.
- 원어 : 원어의 완전일치, '시작'부분일치, '가운데'부분일치, '끝'부분일치를 설정하여 검색할 수 있다.
- 기호 : 해당 수사의 기호를 입력하여 검색할 수 있다.
- 의미항구획 : 의미항구획을 이루고 있는 type(car, ord), def(def, indef)항목을 입력하여 검색할 수 있다.

2.2.4. 통합전자사전 검증 및 평가

2.2.4.1. 세종전자사전 기반 형태소 분석 프로그램 개발·보완

한국어 형태소 분석기는 그동안 수많은 대학과 기업들이 연구해 왔으며, 각 기관의 특성에 맞게 자신들 나름대로의 형태소 분석기를 개발하여 이용해 오고 있다. 통합전자사전 구축분과에서 개발한 형태소 분석기는 본 분과를 맡고 있는 포항공대 KLE 연구실에서 이미 보유하고 있는 형태소 분석기 KoMA를 활용하였으며 기존의 형태소 분석용 사전에 세종 전자사전을 결합하여 사전 정보를 확충하는 방법을 택했다.

본 형태소 분석기는 대부분의 형태소 분석기처럼 어절을 그 처리 단위로 삼고 있으며, 음절 단위의 CYK 알고리즘을 사용하여 형태소 분석을 수행한다. 이때 대상 문장의 각 어절 내 형태소들 간의 접속 정보와 어절 간의 접속 정보가 필요하다. 즉 어절을 이루고 있는 각 형태소에 품사 정보를 할당하고 이들이 서로 접속 가능한지 검사하여 만족하지 못할 경우에는 해당 후보를 제거함으로써 분석 후보의 수를 줄이고, 계산 복잡도도 함께 줄일 수 있다. 어절 간의 접속 정보는 임의의 어절의 마지막에 위치한 형태소와 다음 어절의 처음 형태소의 태그 정보들의 쌍을 가리키는 것으로, 이러한 정보를 사용하면 각 어절의 형태소 분석 결과들 중에서 어절 간의 접속 정보를 만족하지 못하는 후보를 다시금 제거할 수 있다.

이러한 형태소 분석 과정을 통해 형태소 분석 후보들이 모두 모였다면, 이 중에서 가장 가능성이 높은 형태소 분석 결과를 선택하는 태깅 과정을 수행한다. 본 태깅 과정은 말뭉치에서 추출한 빈출 정보를 바탕으로 통계 기반 태깅 방식을 적용함과 동시에 언어 규칙들을 이용한다. 태깅 과정에서 모호성 해소를 위해 사용한 통계 정보는 어휘 확률과 품사 바이그램(Bigram)이다. 품사 바이그램은 어절 내 품사 전이 정보와 어절 간 품사 전이 정보로 나뉘어져 각각 태깅을 위한 문맥 확률로 활용된다. 즉 어절 내 품사 전이 정보를 사용하여 각각의 어절을 분석하고, 분석된 후보 결과들을 어절 간 품사 전이 정보를 통해 재분석하여 가장 올바르게 분석된 태깅 결과를 출력한다. 본 형태소 분석기는 위의 두 가지 전이 정보를 구별하여 적용시키기 위해 띄어쓰기를 고려한 품사 태깅 모델을 사용한다.

2.2.5. 사전 구축 분과 지원

2.2.5.1. 통합 에디터 수정 및 보완

통합 에디터는 모든 상세전자사전에 관한 에디터 기능을 제공한다. 하지만 기존의 통합 에디터는 각 DTD 파일을 통해 자동으로 XSD 파일을 생성할 수 없었기 때문에 이를 위한 별도의 처리과정을 거쳐야했다. 하지만 이번에 새롭게 수정 및 보완된 통합 에디터는 DTD 파일에서 XSD 파일, Template XML 파일, DML 파일을 자동으로 생성할 수 있도록 구현하였다.

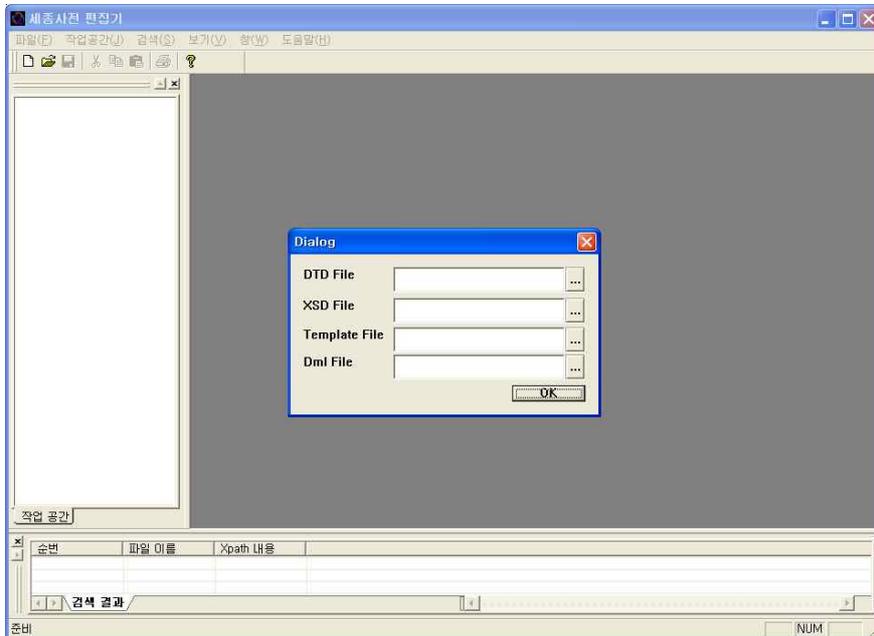


그림 13 통합 에디터의 파일 설정 화면

XSD 파일은 XML Schema Description 파일로서 DTD 파일과 같은 역할을 수행하며, Template XML 파일은 새로 XML 파일을 생성할 때 Template를 제공한다. 그리고 DML 파일은 XML 파일을 작성할 때 각 필드의 내용을 쉽게 입력할 수 있도록 도움을 주기위한 인터페이스를 정의하는 역할을 한다.

다음 그림은 통합 에디터의 실행 화면으로서, 그림에서 볼 수 있는 것처럼 새 파일을 작성할 때 여러 사전의 XML 파일을 선택하여 작업할 수 있다.

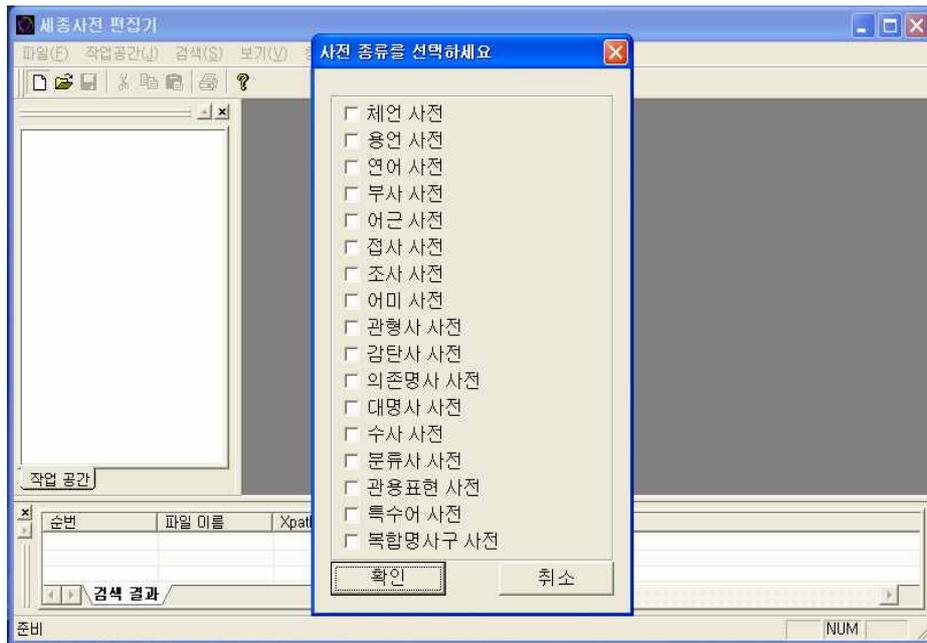


그림 14 통합 에디터의 새 파일 작성 화면

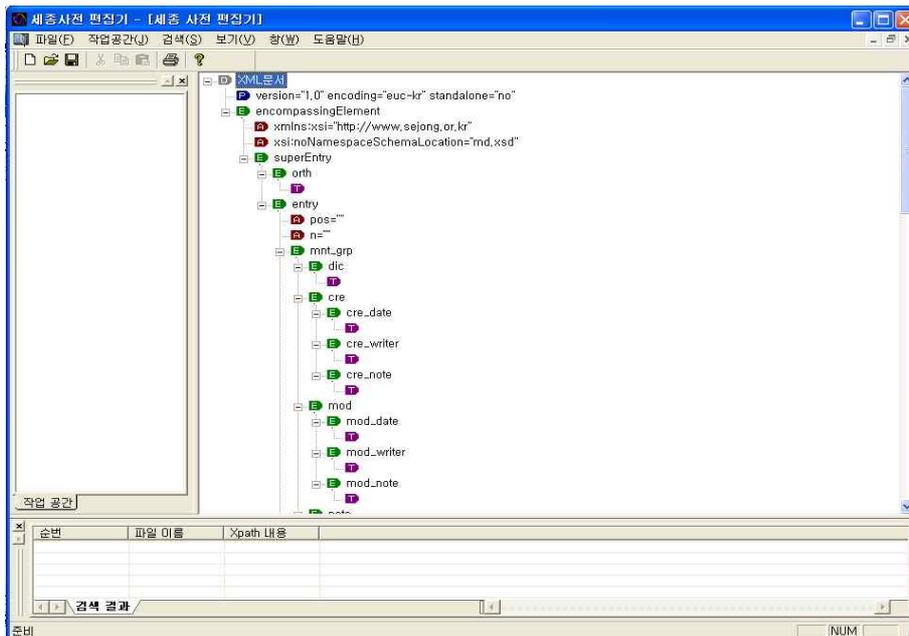


그림 15 통합 에디터의 관형사 사전 편집 화면

2.2.5.2. 외형 검증 도구 개발

외형 검증 도구는 작성된 XML 파일이 해당 DTD 파일에 정의된 규칙을 만족하는지 검사하는 도구로서 사전 구축을 수월하게 처리하기 위한 보조 도구로서 개발된 것이다. 검증할 XML 파일들이 위치한 폴더의 경로를 입력한 후 통합버튼을 클릭하면 해당 폴더에 있는 모든 XML 파일들을 동일한 폴더에 위치한 DTD 파일을 참조하여 외형 검증을 수행한다. 외형 검증을 수행하는 도중에 잘못된 XML 파일이 발견된 경우에는 해당 정보(파일명, 오류 위치)를 화면에 출력하며, 모든 처리가 무사히 완료되었을 경우에는 처리된 XML 파일의 표제어 목록과 superEntry, entry, sense의 수, 그리고 해당 XML 파일들을 통합한 최종 결과물을 생성한다.



그림 16 외형 검증 도구

2.2.6. 2007년도 중간 및 최종 결과물 CD 제작

2.2.6.1 2007년도 확장본 통합전자사전

각 세부사전 구축분과에서 2007년도에 기술한 전자사전 기술목록들은 최종적으로 본 분과에서 데이터베이스로 변환하게 된다. 결과물 CD에 들어가는 데이터베이스는 Microsoft사의 Access파일 포맷인 mdb형태로 만들어지며, 이렇게 만들어진 데이터베이스는 각종 언어처리 프로그램에서 쉽게 사용될 수 있게 된다. 실제로 이렇게 만들어진 데이터베이스는 본 분과에서 개발한 각종 윈도우즈용 검색기 및 웹기반의 통합 전자사전 관리기에서 그대로 사용되고 있다.

아래 표는 올해 각 하위사전 분과에서 기술한 기술목록을 XML파서를 통해 분석하여, 데이터베이스화한 최종 결과물의 superEntry, entry, sense의 수이다.

○ 상세전자사전 구축

구분		상세전자사전								
		계획 항목 수			실제 구축 항목 수					
		SuperEntry	Entry	Sense	SuperEntry	Entry		Sense		
보완 · 정제	체언			35,100	25,458		29,708		35,854	
	용언	형용사		27,700	4,398	19,578	4,530	21,060	6,244	29,512
		동사			15,180		16,530		23,268	
처리항목					45,036		50,768		65,366	
합계				62,800	65,366					

[참고 1] 체언 및 용언 상세전자사전 총 구축량 산출 단위 : Sense 단위

[참고 2] 실제 구축 항목 수는 보완·정제 과정에서 표제항 삭제 및 추가를 통해 계획 항목 수를 초과하게 되었다.

2.2.6.2 기타 통합전자사전용 도구

2006년도에 개발되었거나, 그 이전에 개발된 도구 중 수정 보완이 이루어진 통합전자사전용 도구에는 윈도우즈 기반 단순검색기, 상세검색기(체언, 용언, 부사, 관용표현, 특수어), 통합 전자사전 입력기(통합에디터)가 있다. 단순검색기와 체언 및 용언 상세검색기는 세종계획 2단계가 끝난 2003년에 중규모 전자사전용으로 개발되기 시작해서 지속적인 수정보완이 이루어졌으며, 통합 전자사전 입력기 또한 꾸준한 업데이트가 이루어졌다. 부사, 관용표현, 특수어 상세검색기는 2006년에 개발된 후 검색 조건 및 인터페이스

에 대한 부분적인 수정보완이 이루어졌으며 기존 사전들의 변경된 내용들을 계속해서 반영하였다. 이들 프로그램들은 설치파일 형태로 2007년도 결과물 CD에 포함되어 있다.

기타 세종전자사전을 이용한 검증 및 평가 도구는 그동안 꾸준히 여러 형태로 개발되어 웹페이지에서 실제로 사용자가 이용해 볼 수 있도록 제공되어 왔으며, 이는 본 분과의 홈페이지를 통해 접근할 수 있도록 하였다. 올해는 의대수 상세검색기 및 세종전자사전을 이용한 형태소 분석 프로그램을 개발·보완하였으며, 사전분과의 편의를 위해 XML 파일의 외형 검증 도구를 개발 지원하였다.

2.2.7. 회의

☞ 결과물 CD의 『회의록』 참조

V. 결론 및 향후 전망

1. 문제점 및 개선방안

이상에서 정리한 바와 같이 세종 전자사전 개발 본과의 2007년도 사업은 당초 계획대로 차질 없이 추진되었다. 여전히 완전하다고 말하기는 어려운 사항들이 있기는 하지만, 금년도 사업이 10년간의 사업을 최종 마무리하는 것인 만큼 사업기간 동안 집중적인 관심과 노력을 투자하여 최대한의 해결하였다고 할 수 있다.

2007년 사업의 종료 시에도 여전히 문제점으로 남은 사항이란 결국 세종 전자사전의 완성도와 활용도를 극대화하는 것이라고 할 것이다. 이는 추후 장기적 전망 속에서 지속적으로 노력에 의해서만 해결될 수 있을 것이다.

문제점	향후대책
세종 전자사전의 완성도와 활용도 제고	→ 추후 별도의 중·장기 사업으로 지속적인 노력 투여

2. 전망 및 향후 계획

한국어에 대한 다양한 전산 처리에 활용될 목적으로 개발이 추진된 세종 전자사전은 대규모 범용 현대 한국어 전자사전을 지향하고, 이에 부합하는 규모와 구성, 기능을 갖는 전산 어휘부로 구축되었다. 이는 무엇보다도 언어학 및 언어공학, 컴퓨터 공학 등 관련 분야의 다양한 기존 연구 성과들을 총체적이면서도 치밀하게 검토, 분석하는 한편 한국어의 작용 원리를 면밀하게 관찰, 분석한 결과를 바탕으로 전자사전을 구축함으로써 이르게 된 결과라고 할 수 있다. 특히, 현재의 기술 환경을 반영하는 데 그치지 않고, 미래의 기술 환경을 예견하여 사전 구조의 유연성과 확장성, 연동성을 추구한 것은 세종 전자사전 개발이 한국어 전산 처리 분야의 요구에 답하는 수준에 머무르지 않고, 나아가 이 분야의 기술 혁신과 발전을 촉진할 기반을 조성하는 데에까지 이른 것이라고 할 수 있을 것이다.

물론, 세종 전자사전의 완성도를 높이기 위해서는 아직도 보완해야 할 사항들이 많이 있다. 기초전자사전에만 기술된 어휘(군)들에 대한 상세 기술 작업, 기술된 어휘(군)들에 대한 형식 및 내용의 보완·정제 작업, 항목별 하위 사전들 간의 통일성과 연계성을 제고하기 위한 표준화 작업, 한국어의 특성이 반영된 지침과 규약이 만들어지도록 국제 표준화 기구(ISO)의 활동에 지속적이고 적극적으로 참여하는 노력 등은 세종 전자사전 개

발 사업의 공식적 종료와는 무관하게 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 하나 둘 축적될 때, 세종 전자사전은 그 유연성과 확장성에 힘입어 더욱 개선·개량될 것이고 그만큼 세종 전자사전 개발의 의의도 커지게 될 것이다.

VI. 참고문헌

- 강명윤, (2003), 『언어와 세계』, 한신문화사.
- 강범모, (1983), “보문명사 구문의 의미 특성”, 『어학연구』 19-1.
- _____, (1991), “접속과 논항구조”, 『언어학』 13.
- _____, (1999), “빈도와 언어 기술”, 『언어정보의 탐구』 1, 연세대학교 언어정보 개발 연구원.
- _____, (2000), “서술명사의 기준과 의미 구조”, 『한국언어정보학회 여름학술대회 발표논문집』 pp.71-83.
- 강수언, (1989), “한-영어의 문장 부사 비교”, 『한국어와 영어의 비교 분석』, 서울: 한신문화사.
- _____, (1998), “국어관용표현 연구”, 『세종 우리말 총서4』, 세종출판사.
- 강승식, (1993), 『음절 정보와 복수어 단위정보를 이용한 한국어 형태소 분석 기법』, 서울대 박사학위논문.
- _____, (1994), “다층 형태론과 한국어 형태소 분석 모델,” 제 6회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- _____, (1996), “한국어 형태소 분석을 위한 단어 유형 분류와 자료구조”, 제 8회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 강승식, 이하규외, (2001), “조사 유형 및 복합명사 인식에 의한 용어 가중치 부여 기법”, 『정보과학회 2001년 추계학술대회』 VOL.28 NO.02 pp.0196-0198.
- 강위규, (1988), “관용어의 특성에 대하여”, 『부산 한글』 7(한글학회 부산지회).
- 강은국, (1993), 『조선어 문형연구』, 도서출판 박이정.
- 강인수, (1999), 『대역어 가중치 재조정에 기반한 질의 변환 방식의 교차 언어 문서 검색』, 포항공과 대학교 컴퓨터공학과 석사학위논문.
- 강현규, 박세영외, (1995), “자동 키워드망과 2단계 문서순위결정에 의한 자연어 정보 검색 모델”, 제 7회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 강현화, (1998a), “‘체인+용어’ 꼴의 언어 구성에 대한 연구”, 『사전편찬학 연구』 8, 연세대 언어 정보개발 연구원.
- _____, (1998b), “국어의 동사연결 구성에 대한 연구”, 『말뭉치 기반 국어 연구 총서2』, 연세대학교 언어 정보개발 연구원, 한국문화사.
- 강호관, (1998), 『한국어 의존관계 파싱에 적합한 구문단위의 정의』, 포항공과대학교 컴퓨터공학과 석사학위논문.
- 고광주, (1994), “한국어의 논항구조와 그 유형”, 『한국어문교육』 7.
- 고동혁, (1994), 『조선어 문형 개론』, 한국문화사.
- 고석주, (1996), “ ‘있다’ 구문에 대한 연구”, 『국어문법의 탐구 III(남기심편)』, 태학사.

- 고영근, (1972), “현대 국어의 접미사에 대한 구조적 연구(I) - 확립기준을 중심으로”, 『서울대논문집』 18.
- _____, (1986), “능력성과 국어의 통사구조”, 『한글』 192.
- _____, (1987), 『표준 중세 국어 문법론』. 서울: 탐출판사.
- 고영근, 성광수의(공편), (1992). 『국어학연구백년사 I』, 서울: 일조각.
- 고영진, (1997), 『한국어의 문법화 과정』, 국학자료원.
- 고창수, (1987), “후치사 재고”, 『어문논집』(고려대) 27.
- _____, (1992a), “국어의 통사적 어형성”, 『국어학』 22.
- _____, (1992b), “국어의 격이론”, 『남사 이관수선생 화갑기념논총』.
- _____, (1992c), “동사구의 지정어”, 『언어』 17-2.
- 구연미, (1994), 『우리말 임의 성분 연구』, 부산대 박사학위논문.
- 국어연구회(편), (1990), 『국어연구 어디까지 왔나』, 서울: 동아출판사.
- 김경옥, (1986), 『독일어 Valenz 문법』, 청록출판사.
- 김경학, (1997), “수동태, 중간동사 및 능력동사 구문의 도식에 대해”, 『담화와 인지』 4-1.
- 김경훈, (1977), “국어의 부사수식 연구”, 『국어 연구』(서울대) 37.
- 김계곤, (1980), “국어의 위치 제약에 대하여”, 『현평효 박사 회갑기념논총』, 서울: 형설출판사.
- _____, (1981), “국어의 양상 구조에 대하여”, 『개신 어문 연구』(충북대) 1.
- _____, (1990), “부사 및 부사화”, 『국어연구회(편)』.
- _____, (1996a), 『현대국어의 조어법 연구』, 도서출판 박이정.
- _____, (1996b), 『현대 국어 부사어 연구』, 서울대 박사학위논문.
- 김광해, (1981), “‘의’의 분포에 대한 조사 연구”, 『김형규박사 고회기념논총』.
- _____, (1982), “복합명사의 신생과 어휘화 과정에 대하여”, 『국어국문학』 88.
- _____, (1984), “‘-의’의 의미”, 『문법연구』 5, 문법연구회, 탐출판사, pp. 161-228.
- _____, (1993), 『국어 어휘론 개설』, 집문당.
- _____, (1994), “한자 합성어”, 『국어학』 24, 국어학회, pp. 467-484.
- 김광해, 김동식, (1993), 『국어 사전에서의 합성어 처리에 관한 연구』, 국립국어연구원.
- 김귀화, (1994), 『국어의 격 연구』, 한국문화사.
- 김기혁, (1988), “국어문법에서의 통사구조와 의미구조”, 『언어연구』 8.
- _____, (1990), “관형구성의 통어현상과 의미관계”, 『한글』 209.
- _____, (1991), “형태·통어적 구성과 중간범주”, 『동방학지』 71·72 합집, 연세대 국학연구원[김기혁(1995)에 재수록된 것을 참고함].
- _____, (1995), 『국어 문법 연구 - 형태·통어론-』, 도서출판 박이정.
- 김동식, (1993a), “복합 명사를 찾아서”, 『제20회 국어학회 공동연구회 발표논문』.
- _____, (1993b), 『현대국어 동사의 통사적 특성에 관한 연구』, 서울대 박사학위논문.

- 김두봉, (1922), 『조선말본』, 김민수, 하동호, 고영근(공편)(1986) 재록.
- 김명희, (1987), “{의}의 의미 기능”, 『언어』 12-2, 한국언어학회, pp.248-260.
- 김문창, (1990), 『관용어』, 국어연구 어디까지 왔나, 동아출판사.
- 김미영, (1998), 『국어 용어의 접어화』, 한국문화사.
- 김미영, 강신제외, (2000), “단위분석과 의존문법에 기반한 한국어 구문분석”, 제 27회 『정보과학회 봄 학술대회 논문집』.
- 김미형, (1995), 『한국어 대명사』, 서울: 한신문화사.
- 김민수, (1960), 『국어 문법론 연구』, 통문관.
- _____, (1971), 『국어문법론』, 서울: 일조각.
- 김민수, 하동호외(공편), (1986), 『역대문법대계』, 서울: 탑출판사.
- 김민수, 김태연외, (1995), “국어사전을 이용한 한국어 명사에 대한 상위어 자동 추출 및 WordNet의 프로토타입 개발”, 『한국정보처리학회 논문지』 A, VOL.02 NO.06 pp.0847-0856.
- 김민정, 권혁철, (1991), “한국어 형태소 분석에서의 수사처리”, 제 3회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 김병수, 이용훈, 나승훈, 김준기, 이종혁, (2006), “부트스트래핑 알고리즘을 이용한 한국어 격조사의 의미역 결정”, 『한국컴퓨터종합학술대회』.
- 김병수, 이용훈, 나승훈, 김준기, 이종혁, (2006), “비지도 학습을 기반으로 한 한국어 부사격의 의미역 결정”, 『제 18회 한글 및 한국어 정보처리 학술대회』.
- 김병일, (2000), 『국어 명사구의 내적 구조 연구』, 부산대 박사학위논문.
- 김석득, (1992), 『우리말 형태론; 말본론』, 서울: 탑출판사.
- 김선호, 윤준태, 송만석, (2002), “한국어 문서처리를 위한 동적 생성 로컬 사전 기반 미등록어 분석”, 『한국정보과학회 논문지』 B, VOL.29 NO.06 pp.0407~0416
- 김선희, (1985), “체인 수식 부사의 의미 분석”, 『한글』 187.
- _____, (1987), 『현대 국어의 시간어 연구』, 연세대 박사학위논문.
- _____, (1988), “시간어와 시간적 의미”, 『목원어문학』(목원대) 7.
- 김승곤 외, (1996), “우리말 통어 연구”, 『우리말 연구3, 한말연구 모임, 도서출판 박이정.
- 김양순, (1993), “Adverbs and Head Movement in English”, 『생성 문법 연구』, 3권 1호. 서울: 한국문화사.
- 김영석, 이상억, (1992), 『현대형태론, 학연사.
- 김영희, (1974), “한국어 조사류어의 연구”, 『문법연구』 1.
- _____, (1976), “형용사의 부사화 구문”, 『어학연구』 12-2.
- _____, (1985), “셈술말로서의 정도부사”, 『한글』 187.
- _____, (1988), 『한국어 통사론의 모색』, 탑출판사.
- 김용하, (1990), 『국어 명사구의 구조 연구』, 계명대 석사학위논문.
- 김윤경, (1948), 『나라말본』. 김민수/하동호/고영근(공편)(1986) 재록.

- 김응모, (1992), 『국어 평행이동 자동사 낱말밭』, 서광학술자료사.
- 김인균, (2002), 『국어의 명사 연결 구성 연구』, 서강대 박사학위논문.
- _____, (2003), “관형 명사구의 구조와 의미 관계”, 『국어학』 41, 국어학회, pp.197-223.
- 김일웅, (1994), “폴이말의 결합가와 격”, 『한글』 186.
- 김재영, (1996), 『성능중심 어휘론』, 국학자료원.
- 김재윤, (1994), “국어 조사 ‘-으로’의 통사·의미상의 특징,” 『박갑수선생 회갑 기념논집』, 태학사.
- 김재훈, 김준홍외, (2000), “여과 및 분리 기법을 이용한 한국어 기준 명사 추출”, 『제 12회 한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 김지은, (1996), 『우리말 양태용언 구문에 대한 연구』, 연세대 박사학위논문.
- _____, (1998), “우리말 양태용언 구문 연구”, 『말뭉치 기반 연구 총서』, 연세대학교 언어정보개발 연구원, 한국문화사.
- 김지홍, (1992), 『국어 부사형어미 구문과 논항구조에 대한 연구』, 서강대 박사학위논문.
- _____, (1994), “수량사를 가진 명사구의 논항구조”, 『배달말』 19.
- _____, (1995), “명사구의 확장과 그 논항구조에 대하여”, 『배달말』 20.
- _____, (1997), “명사구를 필요로 하는 기능범주에 대하여”, 『배달말』 22, 배달말학회, pp.1-45.
- 김진규, (1985), “시간 부사 ‘벌써,’ ‘이미’와 ‘아직’의 상과 통사 제약”, 『한글』 189.
- _____, (1999), 『한국어 품사 태그』, 보성.
- 김진해, (1995), 『관용어의 통사·의미론적 제약 연구』, 경희대 석사학위논문.
- _____, (1999), “언어(collocation)의 의미 관계에 대하여”, 『한국어 의미학』 4, 한국어 의미학회.
- _____, (2000a), 『언어연구』, 한국문화사.
- _____, (2000b), “‘기능 동사’는 어휘적 의미가 없는가?”, 『한국문화연구』 3, 221-241, 경희대학교 민속학 연구소.
- 김창섭, (1994), 『국어의 단어형성과 단어구조』, 서울대 박사학위논문.
- _____, (2005), “합성어 연구의 최근 쟁점”, 『한국어 합성어 연구의 최근 쟁점 워크숍 발표 논문집』.
- 김철완, 장재우, (1994), “형태소 네트워크를 이용한 한글 문헌의 자동 키워드 추출”, 제 6회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 김충효, (1992), 『국어의 의문사와 부정사에 관한 연구』, 한양대 박사학위논문.
- 김태현, 이현숙외, (2000), “데이터집합을 이용한 고유명사 추출”, 제 12회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 김택구, (1984), “우리말 부사어의 통어 기능”, 『두메 박지홍 선생 회갑기념논문집』.
- 김현준, 김정화, 강승식, 우종우, 윤보현, (2001), “패턴 학습을 이용한 고유명사 추출”,

- 『정보과학회 2001년 추계학술대회』 VOL.28 NO.02 pp.0184~0186.
- 김형근,서광준,권오욱,정성영,최기선, (1994), “구문태깅의 자동화와 복합명사 인식”, 제 6회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 김홍규,성광수,홍종선, (1998), “대규모 한국어 데이터베이스의 다원적 통계 분석 연구”, 『한국어전산학』 2, 1-169, 한국어 전산학회.
- 김홍수, (1993), 『한국어 심리동사 구문 연구』, 탑출판사.
- 나찬연, (1994), “어찌말의 ‘뚜렷이하기’ 기능에 대하여”, 『우리말 연구』(우리말 연구회), 4.
- 남기심,고영근, (1985a), 『표준국어문법론』, 탑출판사.
- _____, (1985b), 『표준 국어 문법론』, 서울: 탑출판사.
- _____, (1986), “ ‘이다’ 구문의 통사론적 분석”, 『한글연구』 7.
- _____, (1990), “토씨 ‘-와/과’ 의 쓰임에 대하여”, 『동방학지』 66.
- _____, (1993a), 『국어조사의 용법 - ‘-에’ 와 ‘-로’ 를 중심으로-』, 서광학술자료사.
- _____, (1993b), 『표준 국어문법론(개정판)』, 탑출판사.
- _____, (1995), “어휘의미와 문법”, 『동방학지』 88.
- _____, (1996a), 『국어문법의 탐구 I, II』, 태학사.
- _____, (1996b), 『국어문법의 탐구 III』, 태학사.
- 남세진,이지연외, (1996), “복합명사의 통계적 처리에 대한 평가,” 제 8회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 남지순, (2002), “고유명사 자동 처리를 위한 전자 데이터베이스의 구축”, 『어학연구』 제38권 제1호, 서울대학교 어학연구소.
- 남지순,최기선, (1998), “한국어 ‘백과 명사’ 전자 사전의 구축 (I) - <인명 관련 백과 명사의 연구 >,” 제10회 『한글 및 한국어 정보처리 학술대회』.
- 노대규, (1983), “부사의 의미와 수식 범위: ‘그만’ 과 ‘이만’ 을 중심으로”, 『한글』 180.
- _____, (1988), 『국어 의미론 연구』, 서울: 국학자료원.
- _____, (1997) 『한국어의 감탄문』, 국학자료원.
- 노동부 편집부, (1998), 『99한국직업사전』, 노동부 중앙고용 정보보관소.
- 노명희, (1998), 『현대국어 한자어의 단어 구조 연구』, 서울대 박사학위논문.
- 노용균, (1997), “한국어 동사와 명사 사이의 하위 범주화에 있어서의 평행성”, 『언어와 정보』 1.
- 노태길,이상조, (2000), “규칙 기반의 기계학습을 통한 고유명사의 추출과 분류”, 『정보과학회 2000년 추계학술대회』 VOL.27 NO.02 pp.0170~0172.
- 민현식, (1990), 『중세 국어의 시간 부사 연구』, 서울대 박사학위논문.
- _____, (1991), 『국어의 시간과 시간 부사』. 서울: 개문사.
- _____, (1995), “양태부사의 의미에 대하여”, 『태릉 어문 연구』(우산 이인섭 교수 회갑 기념 특집, 서울여대) 5, 6합집.
- 류민홍,나동열외, (2000), “통계정보와 Wordnet 을 이용한 복합명사 분석”, 제 12회

- 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 류법모, 이종혁외, (1996), “연결주의/기호주의 기법에 기반한 한국어 의존 파서”, 『인지과학』 제 7권 제 4호, pp.141-162,
- 목정수, (2005), “전자사전 구축과 합성어 선정 문제”, 『한국어 합성어 연구의 최근 쟁점 워크숍 발표 논문집』.
- 목정수·임유중, (2005), “전자사전의 구축 및 활용: 세종전자사전을 중심으로”, 2005 제6회 국어정보화 아카데미 강의자료.
- 박경현, (1984), “국어 명사 복합형의 기능과 의미 구조”, 『명지어문학』 16, pp. 15-56.
- 박근우, (1971), “국어와 영어의 수식 구조 비교”, 『언어교육』 (서울대) 3-2.
- 박덕유, (1996), 『국어의 동사상 연구』, 한국문화사.
- 박만규, (2005), “합성어의 정체성 확립을 위하여: 벨축 형태론 관점에서”, 『한국어 합성어 연구의 최근 쟁점 워크숍 발표 논문집』.
- 박만규·이선웅외, (2001), “21세기 세종계획 관용표현 전자사전 구축에 대하여”, 제 13회 『한글 및 한국어정보처리 학술대회 논문집』, pp. 334-340, 한국정보과학회·한국인지과학회.
- 박만규, 호정은, 신서인, (2006), “세종전자사전에서 속담의 범주적 정체성 문제에 관하여”, 『2006년 대한언어학회-한국언어학회 가을 공동 학술대회 발표논문집』, 대한언어학회·한국언어학회
- 박만수, (1988), 『국어 자리말 연구』, 태화출판사.
- 박봉래, 임해창, (1998), “한국어 정보 처리 시스템의 전처리를 위한 미등록어 추정 및 절차 오류의 자동 교정”, 『한국정보처리학회 논문지』 A, VOL.05 NO.10 pp.2591~2599.
- 박봉래, 황영숙외, (1996), “확장 정의된 유사어절의 분석에 근거한 실시간 미등록어 인식”, 제 8회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 박병수, (1976), “양태부사에 대하여”, 『언어』 1-1.
- 박수준, 이현아외, (1996), “효율적인 색인을 위한 복합명사의 분해”, 제 8회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 박석문, (1992), “부사 형성 제약의 변천”, 『기곡 강신항 박사 정년 퇴직 기념 국어 국문학 논총』, 서울: 태학사.
- 박선자, (1983), 『한국어 어찌말 연구』, 부산대 박사학위논문.
- _____, (1994), “월어찌말과 말재어찌말”, 『새결 박태권선생 회갑 기념 논총』.
- _____, (1996), 『한국어 어찌말의 통어 의미론』, 서울: 집문당.
- 박성숙, (1997), “한불 사전에서의 연어 처리”, 『불어불문학연구』 34, 한국불어불문학회.
- 박영순외, (1997), 『한국어 의미학 1』, 한국어 의미학회.
- _____, (1998a), 『한국어 의미학 2』, 한국어 의미학회.
- _____, (1998b), 『한국어 의미학 3』, 한국어 의미학회.

- 박영순, (1991), 『현대 한국어 통사론』, 서울: 집문당.
- 박영찬, 최기선, (1996), “통계적 명사패턴 분류를 이용한 복합명사 검색모델”, 제 8회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 박인철, 배우정의, (1995), “개념분류기법을 적용한 한국어 명사분류”, 제 7회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 박정오, 황도삼, (2000), “전문용어 추출시스템”, 『정보과학회 2000년 춘계학술대회』 VOL.27 NO. 01 pp.0381~0383.
- 박창해, (1990), 『한국어 구조론』, 서울: 탐출판사.
- 박지원, (1995), “한국어 명사의 통사·의미론적 특성에 관한 연구 - 논항 구조를 중심으로”, 『언어학연구』(서울대 언어학과 석사학위논문) 제28호.
- 박형익, (1989), “동사 ‘주다’ 의 3가지 용법”, 『한글』 203.
- 백봉자, (1965), 『우리말의 어찌씨 연구』, 연세대 석사학위논문.
- _____, (1974), “한국어의 부사격 토 연구”, 『언어문화』 1.
- 백춘범, (1992), 『단어 결합과 단어 어울림 연구』, 북한 사회과학출판사.
- 서상규, (1984), “부사의 통사적 기능과 부정의 범위”, 『한글』 186.
- _____, (1991), “정도부사에 대한 국어학사적인 조명과 그 분류에 대하여”, 『연세어문학』 23.
- _____, (1993), “현대 한국어의 시늉말의 문법적 기능에 대한 연구”, 『조선학보』 149.
- _____, (1997a), “‘단어 결합’ 과 ‘단어 어울림’ 에 대한 연구”, 『동방학지』 98, 419-468, 연세대학교 동방학연구소.
- _____, (1997b), “언어 연구의 도구로서의 컴퓨터: 국어정보학과 사전편찬학의 응용”, 『언어 정보의 탐구』 1, 271-309, 월인.
- _____, (1998), “말뭉치 분석에 기반을 둔 낱말 빈도의 조사와 그 응용: ‘연세 말뭉치’ 를 중심으로”, 『한글』 242, 한글학회, 225-270.
- 서정목, (1978), 『국어 구문론 연구』, 서울: 탐출판사.
- _____, (1998), 『문법의 모형과 핵 계층 이론』, 태학사.
- 서정수, (1984a), “의문사와 의문 보문자의 호응”, 『국어학』 13.
- _____, (1984b), 『준대법의 연구』, 서울: 한신문화사.
- _____, (1985), “접속문의 의문사와 의문 보문자”, 『국어학』 14.
- _____, (1987), 『국어 의문문 연구』, 서울대 박사학위논문.
- _____, (1990), 『현대 국어 문법의 연구 I. II』, 서울: 한국문화사.
- _____, (1991), 『국어 문법 연구의 개관』, 서울: 한국문화사.
- _____, (1992), 『지배와 결속 이론에 의한 국어 연구』, 고영근 외 3인 공편(1992).
- _____, (1993), “한국어의 구절 구조와 엑스바 이론”, 『언어』 18-2.
- _____, (1994), 『국어문법』, 뿌리깊은나무사.
- _____, (1996a), 『수정증보 국어문법』, 서울: 한양대 출판부.
- _____, (1996b), 『현대 국어 문법론』, 서울: 한양대 출판부.
- 서태룡, (1985), “통사”, 『국어국문학 연구사』, 서울: 도서출판 우석.

- 성광수, (1972), “국어 관형격 구성”, 『국어국문학』 58-60.
- _____, (1974), “국어 주어 및 목적어의 중출 현상: 격문법론적 고찰을 중심으로”, 『문법연구』 1.
- _____, (1979), 『국어 조사의 연구』, 대구: 형설출판사.
- _____, (1985), “국어 격형과 의미 자질 ‘-에’ 와 ‘-로’ 에 대한 점검”, 『어문논집』 (고려대), 24-25.
- _____, (1986), “국어의 격과 조사”, 『국어생활』 4.
- _____, (1990), “단어의 조성법과 어휘 규범화”, 『이중언어학회지』 7.
- 성기철, (1974), “경험의 형태 {있}에 대하여”, 『문법연구』 1.
- _____, (1979), “경험과 추정: ‘졌’ 과 ‘을’ 을 중심으로”, 『문법연구』 4.
- _____, (1984), 『현대 국어의 대우법 연구』, 서울: 개문사.
- _____, (1987), “문서술어 복합문,” 『국어학』 16.
- _____, (1992), “국어 어순 연구”, 『한글』 218.
- 손광주, (1995), 「신경망과 제약만족 알고리즘을 이용한 한국어 구문분석」, 포항공과대학교 컴퓨터공학과 석사학위논문.
- 손남익, (1995), 『국어 부사 연구』, 서울: 박이정.
- _____, (1996), “국어 부사와 수식 대상”, 『한국어학』 (한국어학회) 4.
- 손세모돌, (1988), “‘좀’ 의 상황적 의미”, 『한국학논집』 (한양대) 14.
- _____, (1994), “보조 용언의 의미에 관한 연구”, 『한글』 22.
- _____, (1996), 『국어 보조용언 연구』, 서울: 한국문화사.
- 손숙자, (1990), 『국어 부가어 연구』, 서울대 박사학위논문.
- 송병학, (1994), “담화 불변화사 ‘아니’ ”, 『우리말연구의 썬터, 대전: 문경출판사.
- 송복승, (1995), 『국어의 논항구조 연구』, 보고서.
- 송인동, (1994), "Collocation and Extend Meaning", 『어학연구』 30-4. 서울대학교 어학연구소, 761-797.
- 송철의, (1989), 『국어의 파생어 형성 연구』, 서울대 박사학위논문.
- 시정곤, (1991), “국어 관형구성의 형태-통사적 양면성”, 『어문론집』 30.
- _____, (1992a), “국어의 기능 범주에 대하여”, 국어학연구 백년사, 서울: 일조각.
- _____, (1992b), “국어의 논항구조의 성격에 대하여”, 『한국어문교육』 6.
- _____, (1993), 『국어의 단어형성 원리』, 고려대 박사학위논문.
- _____, (1994), 『국어의 단어 형성 원리』, 서울: 국학자료원.
- 시정곤·고광주의, (2000), 『논항구조란 무엇인가』, 월인.
- 신명철, 이용훈, 김미영, 정유진, 이종혁, (2005), “기능동사 구문과 개념 유사도를 이용한 한국어 부사격의 의미역 결정”, 『한국컴퓨터종합학술대회』.
- 신명철, 이용훈, 김미영, 정유진, 이종혁, (2005), “세종전자사전을 이용한 한국어 부사격의 의미역 결정”, 『제 17회 한글 및 한국어 정보처리 학술대회』.
- 신선경, (1999), “명사구 형성과 속격 표지 ‘-의’ ”, 『울산어문논집』 13·14 합집,

- 울산대 국어국문학과, pp.55-80.
- _____, (2001), “ ‘-의’ 의 실현 양상에 관한 일고찰-서술 관계 명사 연결 구성을 중심으로-”, 『국어 연구의 이론과 실제』(이광호교수 회갑기념논총), 태학사, pp.1091-1113.
- 신수송, (1991), 『통합 문법 이론의 이해: 어휘 기능 문법』, 서울: 한신문화사.
- _____, (1999), “최소자원 최대효과의 구문분석”, 『한글 및 한국어 정보처리 학회』, pp.242-248.
- 심재기, (1982), 『국어 어휘론』, 서울: 집문당.
- _____, (1986a), “속어의 화용론적 고찰”, 『꼭 읽어야 할 국어학 논문집』, 집문당.
- _____, (1986b), “한국어 관용 표현의 화용론적 연구”, 『관악어문연구』 11, 서울대 국어국문학과.
- 안경화, (1986), 『한국어 속어 유형에 대한 분석적 연구』, 서울대 석사학위논문.
- 안성현, 장재우, (1995), “문법형태소 네트워크를 이용한 자동색인 시스템의 설계”, 제 7회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 안주호, (1997), 『한국어 명사의 문법화 현상 연구』, 한국문화사.
- 양동휘, (1989), 『지배 결속 이론의 기초』, 서울: 신아사.
- 양승원, (1991), 『고유 명사와 지시의 문제』, 서울대학교 철학과 석사학위논문.
- 양영희, (1995), “관용 표현의 의미 구현 양상”, 『국어학』 26, 국어학회.
- 양인석, (1975), “한국어 부사의 의미(1)”, 『어학교육』(전남대) 7.
- _____, (1977), “한국어 부사의 의미 분석”, 『이승녕 선생 고회 기념 논총』.
- 양정석, (1991), “재구조화를 특징으로 하는 문장들”, 『동방학지』 71-72.
- _____, (1992), 『한국어 동사의 어휘구조 연구』, 연세대 박사학위논문.
- _____, (1995), 『국어동사의 의미분석과 연결이론』, 도서출판 박이정.
- _____, (1996a), “ ‘-와/과’ 문장의 통사구조”, 국어문법의 탐구 III(남기심편), 태학사.
- _____, (1996b), “ ‘이다’ 구문의 해석”, 『동방학지』 91.
- 엄정호, (1989), “소위 지정사 구문의 통사 구조”, 『국어학』 18.
- _____, (1993), “ ‘이다’ 의 범주 규정”, 『국어국문학』 110.
- _____, (1995), “국어 양태 부사의 연구”, 『태릉어문연구』 5-6.
- 연재훈, (1989), “국어 중립동사 구문에 대한 연구”, 『한글』 203.
- 오승신, (1995), 『국어의 간투사 연구』, 이화여대 박사학위논문.
- 오중훈, 이경순외, (2002), “분야간 유사도와 통계기법을 이용한 전문용어의 자동 추출”, 『한국정보과학회 논문지』 B, VOL.29 NO.04 pp.0258~0269.
- 왕문용, (1989), “명사 관형 구성에 대한 고찰”, 『주시경학보』 4, 탑출판사, 139-157.
- 우순조, (1996), “자유 어순 언어의 형상성;한국어의 경우”, 『언어』 21-3.
- 우인혜, (1990), “시능 부사의 구문론적 제약;용언과의 어울림 관계를 바탕으로”, 『한국학논집』(한양대) 17.

- _____. (1997), 『우리말 피동 연구』, 서울: 한국문화사.
- 우형식, (1994), “동사의 결합가 기술에 대한 방법론적 접근”, 『한글』 225.
- _____. (1995), “연결이론에서의 격표지 교체 분석”, 『애산학보』 17.
- _____. (1996a), 『국어 타동구문 연구』, 도서출판 박이정.
- _____. (1996b), “국어 보충어의 범위”, 『배달말』 16.
- _____. (1998), 『국어 동사 구문의 분석』, 태학사.
- 원대성, (1985), “명사의 상적 특성에 대한 연구”, 『국어연구』 65, 국어연구회.
- 원명옥 역, (1998), 『어휘론의 이론과 응용』, 한국문화사.
- 유경중, (1995), 『근대 국어 피동과 사동 표현의 연구』, 한양대 박사학위논문.
- 유구상, (1968), “국어의 후치사”, 『어문논집』 (고려대) 11.
- _____. (1981), “부사와 ‘는’의 결합관계”, 『한글』 173-174.
- 유동석, (1995), 『국어 매개변인 문법』, 서울: 신구문화사.
- 유목상 외(공편), (1986), 『국어학 신연구』, 서울: 탑출판사.
- 유목상, (1993), 『한국어 문법의 이해』, 서울: 일조각.
- 유필재, (1991), “대칭 서술어의 통사, 의론적 특성-형용사를 중심으로”, 『관악어문 연구』 6.
- 유현경, (1991), “논항과 부가어: 서술 논항의 정립을 위하여”, 『우리말글연구』 (우리말학회) 1.
- _____. (1994), “논항과 부가어”, 『우리말글연구』 1.
- _____. (1997), “심리형용사 구문에 대한 연구”, 『말』 22.
- _____. (1998), 『국어 형용사 연구』, 한국문화사.
- 윤준태, (1999), “구문분석을 위한 말뭉치로부터의 어휘 정보 획득 및 응용”, 『언어 정보의 탐구』 1, 143-141, 연세대학교 언어정보개발 연구원.
- 이관규, (1992), “서술어와 서술관계”, 『주시경학보』 10.
- _____. (1994), “합성동사의 구성에 대한 고찰”, 『한국어학』 1.
- 이광호, (1988), 『국어의 격조사 ‘-을/를’의 연구』, 탑출판사.
- 이기동, (1992), 『영어 동사의 문법』, 신아사.
- 이기오, 이근용외, (1996), “효율적인 한국어 분석을 위한 확장된 최장일치법”, 제 8회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 이남순, (1998), “격표지의 비실현과 생략”, 『국어학』 31, 국어학회, pp.339-360.
- 이도길, 류원호외, (2000), “분석 배제 정보와 후절어를 이용한 한국어 명사 추출”, 제 12회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 이도길, 이상주외, (2003), “명사 출현 특성을 이용한 효율적인 한국어 명사 추출 방법”, 『한국정보과학회 논문지』 B, VOL.30 NO.02 pp.0173-0183.
- 이동혁, (1998), 『국어의 언어적 의미 연구』, 고려대 석사학위논문.
- 이병규, (1994), 『한국어 동사구문의 잠재 논항 실현에 대하여』, 연세대 석사학위논문.
- _____. (1996), “문장 구성 성분의 향가 의존성 검토”, 『국어문법의 탐구 III(남기심편)』, 태학사.

- 이병근, 박진호, (2000), “설명 결합사전의 어휘 기술 방법론”, 『인문논총』 43, 155-2000, 서울대 인문학 연구소.
- 이병근, 박진호외, (2001), “어휘함수를 이용한 한국어 어휘 관계의 기술”, 『어학연구』 37:3, 서울대 인문학 연구소, 507-531.
- 이병찬, (1984), “보족어와 첨가어 설정의 문제점”, 『독일문학』 34
- 이상규, (1993), 『국어 의성의태어 연구』, 한양대 석사학위논문.
- 이상복, (1977), “한국어 부사류에 대하여”, 『연세어문학』 9,10집.
- 이상섭, (1988), “뭉치 언어학적으로 본 사전편찬의 실제 문제”, 『사전편찬연구』 2.
- _____, (1995a), “말뭉치 : 그 개념과 구현”, 『사전 편찬학 연구』 5·6.
- 이상섭 외, (1997), 『사전 편찬학 연구』, 연세대학교 한국어 사전 편찬회, 한국문화사.
- 이상억, (1993), “관용표현과 합성어의 분석 및 어휘부 내에서의 처리”, 『어학연구』 29권 3호. 서울대학교 어학연구소.
- _____, (1995), “국어 관용 표현의 분석과 어휘부 내에서의 처리”, 『인문논총』 33, 서울대학교 인문대학 인문학연구소, 1-44.
- 이석주, (1989), 『국어형태론』, 도서출판 한샘.
- 이선아,전중남외, (2002), “정보추출 기법을 이용한 서열정보분석 데이터베이스 구축 시스템 설계”, 『정보과학회 2002년 추계학술대회』, VOL.29 NO.02.
- 이선용, (2000), “국어의 한자어 ‘관형명사’ 에 대하여”, 『한국문화』 26, 서울대 한국문화연구소, pp.35-58.
- _____, (2005a), “세종전자사전의 구성 체계”, 『성심어문논집』 27, 성심어문학회.
- _____, (2005b), 『국어 명사의 논항구조 연구』, 월인.
- 이성현, (2001), “전자사전 구축을 위한 언어기술의 한 방법 : 대상부류”, 『언어학』 30, 한국언어학회.
- _____, (2004), “전자사전에서의 기능동사 구문 처리 문제 : 세종 체언 전자사전의 경우”, 『한국사전학』 4호, 한국사전학회.
- _____, (2005), “전자사전 구축과 의미부류 - 세종 명사 의미부류 체계의 예”, 『한국사전학』 5호, 한국사전학회.
- _____, (2006), “세종전자사전의 전산적 활용방안”, 『2006 제7회 국어정보화 아카데미 강의자료』.
- _____, (2006), “세종 전자사전의 의미부류 체계 - 언어기반 온톨로지로서의 전망”, 『제1회 언어중립적 온톨로지 워크숍 자료집』.
- 이승녕, (1956), 『고등 국어 문법』, 김민수, 하동호, 고영근(공편)(1986) 재록.
- 이승선,송주원외, (1995), “Compact TRIE Index(CompTI): A Database Index Structure for the Korean Electronic Dictionary”, 『한국정보과학회』, Vol.22, No.1, pp.3-12.
- 이승우,이근배, (2000), “국소 문맥과 공기 정보를 이용한 비교사 학습 방식의 명사 의미 중의성 해소”, 『한국정보과학회 논문지』 B, VOL.27 NO.07 pp.0769-0782.

- 이용주, (1969), “부사, 접속사의 기능에 대하여”, 『논문집』(서울대) 6.
- 이원근, (1996), 『우리말 도움말토씨 연구』, 연세대 박사학위논문.
- 이익섭, (1973), “국어 수량사의 통사 기능에 대하여”, 『어학연구』 9-1.
- 이익섭, 이상억외, (1997), 『한국의 언어』, 서울: 신구문화사.
- 이익섭, 채완, (1999), 『국어 문법론 강의』, 학연사.
- 이재성, 최병진외, (1996), “텍스트 및 전자사전 관리시스템을 위한 표준사전 표기언어의 설계”, 『인지과학』 제 7권, 제 4호, pp.75-92.
- 이정미, (1993), 『현대 국어의 강조 표현에 관한 연구』, 한양대 석사학위논문.
- 이주행, (1996), 『한국어 문법 연구』, 서울: 중앙대 출판부.
- 이종인, 한광록외, (1999), “한국어 명사의 시소러스 구축을 위한 시스템 설계 및 구현”, 『한국정보처리학회 논문지』 A, VOL.06 NO.02 pp.0347-0356.
- 이준영, 강상배외, (1996), “다중색인에 의한 정보검색 시스템 구현”, 『제 8회 한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 이지선, 전현경외, (2000), “WordNet을 이용한 한국어 명사 의미지표 자동 구축”, 『정보과학회 2000년 춘계학술대회』, VOL.27 NO.01 pp.0333-0335.
- 이충우, (1986), “국어 정도부사의 동사 수식에 대하여”, 『국어국문학논문집』(서울대) 24.
- 이태승, (1996), 『술어 중심 제약 전파를 통한 한국어 파싱 알고리즘』, 포항공과대학교 컴퓨터공학과 석사학위논문.
- 이필영, (1981), “국어의 관계 관형절에 대한 연구”, 『국어 연구』(서울대) 7.
- _____, (1986), “호격 및 감탄 조사에 대한 연구”, 유목상 외(공편)(1986).
- _____, (1989), “상 형태와 동사의 상적 특성을 통한 상의 고찰”, 『주시경학보』 3.
- _____, (1993), 『국어의 인용 구문 연구』, 서울: 탐출판사.
- 이현민, 박혁로, (2000), “복합명사의 역방향 분해 알고리즘”, 제 12회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- _____, (2001), “복합명사의 역방향 분해 알고리즘”, 『한국정보처리학회 논문지』 B, VOL.8-B NO.04 pp.0357-0364.
- 이현아, 이종혁외, (1995), “구문분석과 공기정보를 이용한 개념기반 명사구 색인방법”, 제 7회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 이현우, (1995), 『현대국어의 명사구의 구조 연구』, 서울대 박사학위논문.
- 이흥식, (1990), 현대국어 관형절 연구, 『국어연구』 98.
- 이환목, (1974), “영어 부사에 관한 연구: 문장 부사와 술부 부사의 차이를 중심으로”, 『영어영문학』, 51-52.
- _____, (1975), “양상부사의 통어적 특성에 대한 의미론적 접근”, 『어학교육』(전남대) 7.
- _____, (1976), “문장부사 ‘다행히’ 에 대하여”, 『언어』 1-2.
- _____, (1979), “부사론(1)”, 『어학교육』(전남대) 10.
- _____, (1981), “‘또’ 의 의미분석”, 『한글』 173-174.
- _____, (1982), 『국어 함수 표현에 관한 연구: 아니, 또, -도』, 서울대 박사학위

논문.

- 이휘봉, (1998), 『구문의존구조에서 중간언어 방식 기계번역을 위한 개념그래프의 생성』, 포항공과대학교 컴퓨터공학과 박사학위논문.
- 이희승, (1949), 『초급 국어 문법』, 김민수, 하동호, 고영근(공편)(1986) 재록.
- _____, (1956), 『중등 문법』. 김민수, 하동호, 고영근(공편)(1986) 재록.
- 이희자, (1995), “현대 국어 관용구의 결합 관계 고찰”, 제 6회 『한글 및 한국어 정보처리 학술대회』, 333-352.
- 이희자 외, (1998), “사전식 텍스트분석적 국어 조사의 연구”, 『말뭉치 기반 국어 연구 총서 1』, 연세대학교 언어정보 개발연구원, 한국문화사.
- _____, (1999), “사전식 텍스트 분석적 국어 어미의 연구”, 『말뭉치 기반 국어 연구 총서 5』, 연세대학교 언어정보개발연구원, 한국문화사.
- 인터미디어,(엮음), (1994), 『일-한 부사사전』, 서울: 홍신문화사.
- 임경순, (1980), “관용어와 언어 구조:Chafe의 ‘의미와 언어 구조’ 를 중심으로”, 『연암 현평효박사 회갑기념 논총』, 형설출판사, 549-574.
- 임동훈, (1991), “현대국어 형식명사 연구”, 『국어연구』 103.
- _____, (2002), “한국어 조사 연구의 현황과 전망”, 『한국어학』 16.
- 임소영, (1997), 『한국어 식물이름의 연구』, 한국문화사.
- 임유중, (1995), “좀/조금’ 에 대하여”, 『한양어문연구』 (한양대) 13.
- _____, (1996), “부사의 하위 구분 체계에 대한 연구”, 『한양어문연구』 (한양대) 14.
- _____, (1997a), “문장 수식어에 관하여”, 『한국학논집』 (한양대) 27.
- _____, (1997b), 『국어 부사의 범주 정립과 호응 및 어순에 관한 연구』, 한양대 박사학위 논문.
- _____, (1998a), “국어 부사의 하위 분류”, 『어학 연구』 (서울대) 34-1.
- _____, (1998b), “개화기 국어의 부사에 관한 연구”, 『한민족 문화 연구』 3.
- _____, (1998c), “부사의 수식 범위와 성분통어”, 『한양어문연구』 16.
- 임지룡, (1987), “정도 그림씨의 의미 대립 특성”, 『언어』 12-1.
- 임홍빈, (1981), “존재 전제와 속격 표지 {의}”, 『언어와 언어학』 7, 한국외대 언어연구소 [임홍빈(1998)에 재수록된 것을 참고함].
- _____, (1987a), 『국어 재귀사 연구』, 서울: 신구문화사.
- _____, (1987b), “국어의 명사구 확장 규칙에 대하여”, 『국어학』 16.
- _____, (1989), “통사적 파생에 대하여”, 『어학연구』 25-1.
- _____, (1999), “국어 명사구와 조사구의 통사구조에 대하여”, 『관악어문연구』 24, 서울대 국어국문학과, pp.1-62.
- _____, (2000), “가변 중간 투사론,” 『21세기 국어학의 과제』, 월인, 1279-1320.
- 임홍빈, 송철의, (1998), “한국어 정보 처리를 위한 어절 분석 표지의 표준화 연구”, 『국어 기초자료 구축』, 193-337, 문화관광부.
- 임홍빈, 이홍식, (1999), “구문 분석 표지의 권장 표준안 연구”, 『국어 기초자료 구축』, 305-475, 문화관광부.

- 임흥빈·이홍식외, (2002), 『한국어 구문 분석 방법론』, 한국문화사.
- 왕문용, 민현식, (1993), 『국어 문법론의 이해』, 서울: 개문사.
- 장경준, (1998), “현대국어 ‘-어 하다’ 문의 유형 분류”, 『연세학술논집』 27.
- 장경희, (1980), “지시어 ‘이, 그, 저’ 의 의미론적 분석”, 『어학 연구』(서울대) 16-2.
- _____, (1985), 『국어 양태 범주의 연구』, 서울: 탐출판사.
- _____, (1988), “국어의 간접 표현”, 『주시경학보』 1, 주시경학회.
- _____, (1989), “지시사 ‘이, 그, 저’ 의 범주 지시”, 『인문논총』(한양대) 17.
- _____, (1990), 『조용 표현』, 국어연구회(편)(1990).
- 장동현, 맹성현, (1996), “효율적인 색인어 추출을 위한 복합명사 분석방법”, 제 8회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』
- 장세경, 장경희, (1994), “국어 관용어에 대한 연구”, 『한국학논집』 25, 한양대 한국학연구소.
- 정교환, (1988), 『국어 문장 부사의 연구』, 동아대 박사학위논문.
- 정길남, (1987), 『개화기 국역 성서의 표기법과 문법 형태』, 서울: 개문사.
- _____, (1992), 『19세기 성서의 우리말 연구』, 서울: 서광학술자료사.
- 정래정, 김준태, (1996), “고유명사 출현 패턴을 이용한 색인의 성능 향상에 관한 연구”, 제 8회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 정영미, (1993), 『정보 검색론』, 구미부역(주) 출판부,
- 장영준, (1998), “On the Structure of NP”, 한국언어학회 가을연구회 발표문(유인물).
- 정정덕, (1986), 『국어 접속 어미의 통사 의미론적 연구』, 한양대 박사학위논문.
- 정주리, (1995), “국어 보문 동사의 통사·의미론적 연구”, 『한글』 228.
- 정태구, (2001), 『논항구조와 영어 통사론』, 한국문화사.
- 정호성, (2000), “표준국어대사전 수록 정보의 통계적 분석”, 『새국어생활』 10-1.
- 정희정, (1996), “자동사/타동사 분류에 대한 비판적 고찰 -사전처리와 관련하여-”, 『국어 문법의 탐구 III』(남기심편), 태학사.
- _____, (2000), 『한국어 명사 연구』, 한국문화사.
- 조규설, (1958), “첨용 부사의 고찰”, 『어문학』(대구) 3.
- 조용준, (1996), “서술성 명사의 피동화에 대한 고찰”, 『한말연구』 2, 한말연구모임, 박이정, pp.231-262.
- 조의성, (1997), “현대한국어의 단어결합에 대하여”, 『朝鮮學報』 163, 1-36.
- 차정원, 이근배외, (1996), “미등록어를 고려한 TAKTAG의 개선”, 제 8회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 차현실, (1986), “양상부사의 통사적 특성에 따른 의미분석: ‘아마’, ‘글세’, ‘혹시’ 를 중심으로”, 『말』 11.
- 채숙희, (2002), “연결어미 상당의 명사구 보문 구성 연구”, 『국어연구』 170.
- 채완, (1977), “현대 국어 특수 조사의 연구”, 『국어 연구』(서울대) 39.
- _____, (1985), “병렬의 어순과 사고 방식”, 『국어학』 14.

- 최경봉, (1996), 『국어 명사의 의미구조 연구』, 고려대 박사학위 논문.
- 최경숙, (1997), 『언어 구성의 기능 동사에 관한 연구』, 울산대 교육대학원 석사학위 논문.
- 최기용, (1996), “의미역 배경과 관련된 명사의 성격에 대하여”, 『생성문법연구』 6-1, pp.85-119.
- 최승권, (1995), “한국어와 여러 언어 통합기반 기계 번역: 기능동사 구문을 중심으로”, 『어학연구』 31-1, 서울대학교 어학연구소.
- 최재혁, (1996), “음절수에 따른 한국어 복합 명사 분리 방안”, 제 8회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 최재희, (1994), “양화사 ‘다’의 양화 유형에 대하여”, 우리말 연구의 쟁점, 대전: 문경출판사.
- 한 길, (1983), “정도 어찌씨에 관한 의미론적 연구”, 『새국어교육』 (한국 국어 교육 학회), 37-38.
- 한동완, (1996), 『국어의 시제 연구』, 서울: 태학사.
- 한송화, (1997), 『국어 자동사 연구』, 연세대 박사학위 논문.
- 한영균, (1997), “‘명사+동사’ 합성구의 형태론적 특징”, 『울산어문논집』 12, 95-123, 울산대 인문대 국어국문학과.
- 한정길, (1986), “숙어표현에 대하여”, 『어학연구』 22권 1호. 서울대학교 어학연구소.
- 허윤영, 권혁철, (1994), "'의미적 한 단어' 유형 분석 및 형태소 분석 기법", 제 6회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 홍사만, (1977), “국어 정도 부사와 상태 부사의 비교연구”, 『동양 문화 연구』 (경북대) 4.
- 홍재성, (1987), “한국어 사전 편찬과 문법정보”, 『어학연구』 23-1, 서울대학교 어학연구소.
- _____, (1988), “한국어사전에서의 동사항목의 기술과 통사정보”, 『인문과학』.
- _____, (1989), “내려가다/내려오다와 그 사전적 처리”, 『애산학보』 제 7집.
- _____, (1990a), “견디다 구문의 기술을 위하여”, 『한글』 208호.
- _____, (1990b), “한국어 자동사/타동사 구문의 구별과 사전”, 『사전편찬학연구』 제 3집.
- _____, (1992), “동사 먹다의 사전적 처리를 위한 몇 가지 논의”, 『새국어생활』 2-4. 국어연구소.
- _____, (1993a), “약속의 문법: 서술명사의 어휘 통사적 기술과 사전편찬”, 『동방학지』 81, 연세대학교 한국학연구소.
- _____, (1993b), “먹다 숙어동사구문의 통사적 기술”, 『어학연구』 29권 3호. 서울대학교 어학연구소.
- _____, (1995a), “어휘함수 개념에 의한 한국어 어휘기술과 사전편”, 『해방 50주년-세계 속의 한국학』, 인하대학교 한국학연구소.
- _____, (1995b), “의미·텍스트 대응모형에 따른 불어어휘의 연구: 어휘함수 개념의

- 활용을 중심으로”, 『인문논총』 33, 서울대학교 인문학연구소.
- _____, (1995c), “한불사전의 명사항목구성”, 『불어불문학연구』 제 30집.
- _____, (1997a), 『현대 한국어 동사구문사전(공저)』, 두산동아.
- _____, (1997b), “이동동사와 기능동사”, 『말』 22, 연세대학교 한국어학당, pp.121-140.
- _____, (1997c), “술어명사 사전과 ‘-이다’ 술어명사 구문의 기술”, 제 6회 『국제 한국어 학술대회 발표논문집』, 한글학회, pp.131-149.
- _____, (1998), “동사, 형용사의 사전기술”, 『새국어생활』, 국립국어연구원.
- _____, (1999), “기능동사 연구의 한 시각 : 어휘적 접근”, 『인문논총』 41, 서울대학교 인문학연구소.
- _____, (2001), “한국어의 명사 (I), (II)”, 『새국어생활』 2(3-4), 국립국어연구원.
- _____, (2002), “한국어의 동사(I)”, 『새국어생활』 12(3), 국립국어연구원.
- _____, (2003), “불한 관계형용사 어휘대조연구 서설”, 『인문논총』 제48집, 서울대학교 인문학연구소.
- _____, (2005), “합성어 연구의 기본 문제”, 『한국어 합성어 연구의 최근 쟁점 워크숍 발표 논문집』.
- 홍중선, (1990), 『국어 체언화 구문의 연구』, 고려대학교 민족 문화연구소.
- _____, (1998), “명사의 사전적 처리”, 『새국어생활』 8-1, 국립국어연구원, pp.111-130.
- 홍중선, 강범모외, (2000), “한국어 연어 정보의 분석·응용에 관한 연구”, 『한국어학』 11, 73-158, 한국어학회.
- 홍중선, 이동혁, (2000), “국어 말모듬의 연어 관계 표지화 방안”, 『21세기 국어학의 과제』, 1078-1110, 월인.
- 홍중화, (1994), “부사어의 기능”, 『말』 (연세대) 19.
- 황병순, (1984), “국어 부사에 대하여”, 『배달말』 9.
- 황이규, 이근용외, (2000), “구문형태소를 이용한 색인어 추출”, 제 12회 『한글 및 한국어 정보 처리 논문대회 발표 논문집』.
- 황화상, (1996), 『국어 체언서술어의 연구』, 고려대 석사학위논문.
- Aarts, J., Meijs, W., (1986), *Corpus Linguistics II. New Studies in the Analysis and Exploitation of Computer Corpora*, Editions Rodopi B.V.
- Abeillé, A., (1987), *Grammaire des noms de métier*, Mémoire de DEA, Université de Paris 7.
- _____, (1993), *Les nouvelles syntaxes*, A. Colin.
- Abney, S., (1987), *The English Noun Phrase in its Sentential Aspect*, Ph.D. Dissertation, MIT.
- Achard, M., (1998), *Representation of Cognitive Structures*, Mouton de Gruyter.
- Ackerman, F., Webelhuth, G., (1998), *A Theory of Predicates*, CSLI publications.
- Ahlsén, E., (2006), *Introduction to Neurolinguistics*, John Benjamins Pub Co.
- Aijamer, K., Alterberg, B., (1991), *English Corpus Linguistics*, Longman Group

- UK Limited.
- Akkerman, E. *et al.*, *Designing a computerized lexicon for Linguistic purposes*, ASCOT Report No 1, Rodopi.
- Albertari, L. eds., (2000), *Meaning and Cognition*, John Benjamins.
- Allerton, D.J., (1982), *Valency and the English Verb*, Academic Press.
- Allwood, J., Gårdenfore, P. eds.,(1998), *Cognitive Semantics*, John Benjamins.
- Alsina, A. eds., (1997), *Complex Predicates*, CSLI Lecture notes number 64, CSLI Publications.
- Anderson, M., (1979), *Noun Phrase Structure*, Ph.D. Dissertation, University of Connecticut.
- Arnaud, P., Bjoint, H., (1992), *Vocabulary and Applied Linguistics*, Macmilan.
- Aroui, J-L., (2001), *Le sens et la mesure*, Honoré champion.
- Bach, E., (1968), Noun and Noun Phrase, in Bach, E. and R. Harms(eds), *Universal in Liguistic Theory*, Holt, Rinehart and Winston, pp.91-122.
- Baker, M. C., (1988), *Incorporation*, The University of Chicago Press.
- Bartsch, R., (1998), *Dynamic Conceptual Semantics, A logico-philosophical investigation into concept formation and understanding*, CSLI Publications & Folli.
- Bateman, J. A., (1991). The Theoretical Status of Ontologies in Natural Language Processing, *Proceedings of the workshop on Text Representation and Domain Modeling - Ideas from Linguistics and AI*, Technical University Berlin, Oct. 9-11, KIT Report 97, cmp-lg/9704010.
- Baurer, L., (1983), *English word-formation*, Cambridge, Cambridge University Press
- Beale, S., (1997), *Hunter-Gatherer: Applying Constraint Satisfaction, Branch-and-Bound and Solution Synthesis to Computational Semantics*, Ph.D. dissertation, Program in Language and Information Technologies, School of Computer Science, Carnegie Mellon University.
- Benson, N. *et al.*, (1986), *The BBI Combinatory Dictionary of English*, John Benjamins, Amsterdam.
- Blanco, J., (1997), Un dictionnaire électronique des déterminants nominaux en espagnol, BULAG, numéro spécial Linguistique et Informatique, *Actes du colloque intermational FRACTAL 97*, Univrsité de Franche-Compte, Besançon.
- Blumenthal, P., (2006), *Wortprofil im Französischen*, Max Niemeyer Verlag.
- Bolinger, D., (1976), Meaning and Memory, *Forum Linguistium* no.1: 1-14.
- Boone, A., (1989), Les noms de maladies, *Recherches linguistiques* 13, Klincksieck, Paris.
- Boons, J.-P., (1985), Préliminaires à la classification des verbes locatifs : Les compléments de lieu, leurs critères, leurs valeurs aspectuelles, *Lingvisticae Investigationes*, Vol. IX-2.

- _____, (1992), *La structure des phrases simples en français. Constructions transitives locatives*, Droz, Genève.
- Bouillon, P., (1998), *Traitement automatique des langues naturelles*, Editions.
- _____, (2001), *The Language of Word meaning*, Cambridge University Press.
- Boutet, J., (1994, 1997), *Construire le sens*, 2e édition, Peter Lang.
- Boyan, A. Onyshkevych., (1997), *An Ontological-Semantic Framework for Text Analysis*, Ph.D. dissertation, Program in Language and Information Technologies, School of Computer Science, Carnegie Mellon University, CMU-LTI-97-148.
- Braasch, A., Povlsen, C. eds., (2002), *Proceedings of the Tenth EURALEX International Congress*, EURALEX Copenhagen: Center for Sprogteknologi, Københavns Universitet, Copenhagen, Denmark, August 13-17, 2002.
- Browning, M. A., (1993), Adverbial Reflexives, *NELS* no.23.
- Buvet, P.-A., (1993), *Les déterminants nominaux quantificateurs*, Thèse de doctorat, Paris 13.
- _____, (1994), Déterminants nominaux, *Lingvisticae Investigationes* no.18:1, John Benjamins B. V., Amsterdam.
- _____, (1995), Particularités syntaxiques des noms de contenants, *Suvremena Lingvistika* no.40, Zagreb.
- _____, (1998), Détermination et classes d'objets, *Langages* no.131, Paris, Larousse.
- Buvet, P.-A., Lim, J. H., (1996), Les déterminants nominaux aspectuels, *Lingvisticae Investigationes* no.20:2, John Benjamins B. V., Amsterdam.
- Cacciari, C., Tabosi, P. eds., (1993), *Idioms: Processing, Structure and Interpretation*, Lawrence Erlbaum Associates, NJ. Hillsdale.
- Cahiers de lexicologie, (2003), No1~4*, Didier Erudition.
- Candel, D. *et al.*, (1990) Aspects de la documentation scientifique et technique dans un grand dictionnaire de langue in *Autour d'un dictionnaire :le « Trésor de la langue française ». témoignage d'atelier et voies nouvelles*, ed. B. Quemada, Didier édition, Paris.
- Carnie, A., (2002), *Syntax*. Blackwell Publishing.
- Cattel, R., (1984), Composite Predicates in English, *Syntax and Semantics* Vol.17, Academic Press.
- Chafe, W.L., (1968), Idiomaticity as an Anomaly in the Chomskyan Paradigm, *FL* 4: 109-127.
- Chang, S.-J., (1996), *Korean*, John Benjamins Publishing Company.
- Choi, H.-S., (1985), Remarks on Configurationality Parameters, *Harvard Studies in Korean Linguistics*, Seoul: Hanshin Publishing.
- _____, (1988), *Restructing Parameters and Complex Predicates Transformational Approach*, Doctoral dissertation, MIT.
- Choi-Jonin *et al.*, (2005), *Questions de classification en linguistique*, Groupe

- Editorial Peter Lang.
- Chomsky, N., (1965), *Aspects of the Theory of Syntax*, MIT Press.
- Cinque, G., (1997), *Adverbs and Functional heads, A Cross-linguistic Perspective*, Oxford University Press.
- Cormack, A., (1998), *Definitions, Implicational for Syntax, semantics, and the Language of Thought*, Garland Publishing.
- Cormier, M. C *et al.*, (2003), *Les dictionnaires Le Robert*, PU Montréal.
- _____, (2006), *Les dictionnaires Larousse*, PU Montréal.
- Corréar, M.-H. ed., (2002), *Lexicography and Natural Language Processing : A Festschrift in Honor of B.T.S. Atkins*. EURALEX. Info: <http://www.ims.uni-stuttgart.de/euralex/>.
- Corum, C., (1974), Adverbs ... Long and Tangled Roots, *CLS* no.10.
- Cottez, H.,(1994) « La place des termes techniques et scientifiques dans un dictionnaire de langue » in *Français scientifique et technique et dictionnaire de langue*, ed. Candell, D. Paris. Didier Erudition.
- Coursil, J., (2000), *La fonction muette du langage*, Ibis rouge.
- Covington, M. A., (1990), A Dependency Parser for Variable-word-order Languages, Research Report AI-1990-01, *Artificial Intelligence Programs*, Univ. of Georgia.
- Cruse, A. *et al.* eds., (2002), *Lexicology I*, Berlin, New York: Walter de Gruyter.
- Culicover, P. W., (1993), The Adverb Effect : evidence against ECP accounts of the that-t effect, *NELS*, 23.
- _____ et Jackendoff, R., (2005), *Simpler Syntax*, Oxford University Press USA.
- Danblon, E., (2005), *La fonction persuasive*, Armand Colin.
- Danlos, L., (1981), *La morphosyntaxe des expressions figées*, Thèse de doctorat, Université Paris VII.
- Danlos, L. *et al.*, (1988), Les expressions figées, *Langages* no.63, Larousse, Paris.
- Delorme, E., R. Dougherty. (1972), Appositive NP Constructions, *Foundations of Language* 8-1.
- Denhiere, G. *et al.*, (1991), Text and Text Processing, *Advances in Psychology* no.79, Elsevier science publishers.
- Dirven, R. eds., (1995), *Current approaches to the lexicon*, Duisburg Papers on Research in Language and Culture, Peter Lang.
- Downing, P., (1977), On the creation and use of English compound nouns, *Language* no.53, 810-42.
- Dunbar, G., (1991), *The cognitive lexicon*, Tübingen: Guster Nan Verlag.
- Dziemianko, A., (2006), *User Friendliness of verbe syntax in pedagogical dictionaries of English*, Max Niemeyer Verlag.
- Ellis, D., (1996), *Progress and Problems in Information Retrieval*, Library Association Publishing.
- Encelberg, S., (1996), Event structure and lexical semantics, *MIT Working papers*

in linguistics.

- Erik F. *et al.*, (1999), Representing Text Chunks, *Proceedings of EACL '99, Association for Computational Linguistics*.
- Everaert, M. *et al.*, eds., (1995), *Idioms: Structural and Psychological Perspectives*, Lawrence Erlbaum Associates, NJ. Hillsdale.
- Fauconnier, G., (1997), *Mapping in Thought and Language*, Cambridge University Press.
- Fauconnier, G. *et al.*, (1996), *Spaces, Worlds & Grammar*, The University of Chicago Press.
- Fellbaum, C. ed., (1998), *Wordnet. An Electronic Lexical Database*, The MIT Press.
- Fiala, P. *et al.* eds., (1997), *La locution entre lexique, syntaxe et pragmatique*, Klincksieck.
- Fillmore, C., (1968), *The case for case, Universals in Linguistic Theory*, Georgetown University Press.
- Fiumara, G. C., (1995), *The metaphoric process, Connections between language and life*, Routledge.
- Fontenelle, T., (1997), *Turning a Bilingual Dictionary into a Lexical-Semantic Database*, Max Niemeyer Verlag.
- _____, (1998), *ACTES EURALEX'98 PROCEEDINGS*, Communications soumises à ERALEX'98 (Huitième Congrès International de Lexicographie) à Liège, Belgique / Papers submitted to the Eighth EURALEX, International Congress on Lexicography in Liège, Belgium, English and Dutch Departments, University of Liège.
- Foucou, P.-Y., (1997), Une classe d'objets <hypertexte> pour l'exploration automatique du Web, BULAG, *numéro spécial Linguistique et Informatique, Actes du colloque international FRACTAL 97*, Université de Franche-Compte, Besançon.
- Francis, E. J. et Michaelis. L. A., (2002), *Mismatch*, Center for the Study of Language and Inf.
- Fraser, J. B., (1970), Idioms within a Transformational Grammar, *FL* 6: 22-42.
- Froon, J., De Jong, F. M. G., (2002), The Benefits of a Product-Independent Lexical Database with Formal Word Features, *Kernerman Dictionary News* 10:4-8. also available at <http://kdictionaries.com/newsletter.html>.
- Fujii, A., (1998), *Corpus-Based Word Sense Disambiguation*, Tokyo Institute of Technology Department of Computer Science.
- Gamper, J., Knapp, J., (2001), New Media in the Design of a Learners' Dictionary in *Proceedings of the World Conference on Educational Multimedia, Hypermedia & Telecommunications* (ED Media 2001). Tampere, 25-30 June 2001, 541-6. Also available at [http://www.eurac.edu/websites/eurac/collaborators/homepage/KnJ/HTML Files:publications.html](http://www.eurac.edu/websites/eurac/collaborators/homepage/KnJ/HTML%20Files:publications.html).
- Gariilidou, Z., (1995), *Etude comparative des suites binominales N N*

- métaphoriques en français et en grec*, DEA en sciences du langage, Université Paris 13.
- Gerot, L., Wignel, P., (1994), *Making Sense of Functional Grammar*, Gerd Stabler Antipodean Educational Enterprises.
- Giorgi, A., Longobardi. G., (1991), *The Syntax of Noun Phrase: Configuration, Parameters and Empty Categories*, Cambridge University Press.
- Giry-Schneider, J., (1978), *Les nominalisation en français. L'opérateur faire dans le lexique*, Droz.
- _____, (1987), *Les prédicats nominaux en français. Les phrases simples à verbe support*, Droz.
- _____, (1988), L'interprétation événementielle des phrases en il y a, *Linguisticae Investigaciones* no.12:1, John Benjamins, Amsterdam /Philadelphia.
- _____, (1991), Relation entre le sens des noms et leur structure prédicative, *Revue Québécoise de Linguistique* Vol.20-1, Presses de l'Université du Québec à Montréal.
- Givon, T., (1984), *Syntax, A Functional-Typological Introduction (I)*, John Benjamins Publishing Company.
- Goldberg, A. E., (1996), *Conceptual Structure, Discourse and Language*, CSLI Publications.
- Gorfein, D. S., (1989), *Resolving Semantic Ambiguity*, Springer-Verlag.
- Gosselin, L., (1996), *Sémantique de la temporalité en français, Un modèle calculatoire et cognitif du temps et de l'esprit*, Duculot.
- Graeme, D. R. et al., (1992), *Computational Morphology*, A Bradford Book, The MIT Press.
- Granger, S. et Altenberg, B., (2002), *Lexis in contrast*, John Benjamins Pub Co.
- Greenbaum, S., (1969), *Studies in English Adverbial Usage*, London: Longmans.
- Grefenstette, G., (1998), *Cross-Language Information Retrieval*, Kluwer Academic Publishers.
- Grimshaw, J., (1990), *Argument structure*, MIT Press, Cambridge, MA.
- Grimshaw, J., Mester, A., (1988), Light Verb and θ -Marking, *LI* 19-2. MIT.
- Gross. G., (1988), Le degré de figement des noms composés, *Langages* no.90, Larousse, Paris.
- _____, (1992a), *Formes d'un dictionnaire électronique, L'environnement traductionnel*, Silley, Presses de l'Université du Québec.
- _____, (1992b), Classes d'objets et enseignement, *Actes du colloque sur le traitement automatique*, Séoul.
- _____, (1993a), Les phrases figées en français, *L'information grammaticale* no.59.
- _____, (1993b), Trois applications de la notion de verbe support, *L'information grammaticale* no.9, Paris.
- _____, (1994a), *Classes d'objets et synonymie*, Annales littéraires de l'Université de Besançon.
- _____, (1994b), Un outil pour le FLE : les classes d'objets, Actes du colloque du FLE, Lille, Presses Universitaires de Lille.

- _____. (1994c), Classes d'objets et description des verbes, *Langages* 115, Paris, Larousse.
- _____. (1995), Les Classes d'objets, Les concepts opératoires de l'INaLF, *rapport de l'INaLF*, pp. 83-88.
- _____. (1996), *Les expressions figées en français, noms composés et autres locutions*, Ophrys.
- _____. (to appear), Pour une véritable Fonction "synonymie" dans un traitement de textes.
- Gross, G., Vivès, R., (1986), Les constructions nominales et l'élaboration d'un lexique-grammaire, *Langue Française* no.69, Larousse, Paris.
- Gross, M., (1975), *Méthodes en syntaxe*, Hermann.
- _____. (1976), *Sur quelques groupes nominaux complexes, Méthode en grammaire française*, Klincksieck, Paris.
- _____. (1981), Les bases empiriques de la notion de prédicat sémantique, *Langages* no.63, Larousse.
- _____. (1982), Une classification des phrases <figées> du français, *Revue québécoise de linguistique* vol.11 no. 2.
- _____. (1984), Une famille d'adverbes figés: les constructions comparatives en COMME, *Revue québécoise de linguistique* vol. 11 no. 2, Uqam, Montréal.
- _____. (1985a), Sur les déterminants dans les expressions figées, *Langages* no. 79, Larousse, Paris.
- _____. (1985b), Sur les phrases figées complexes du français, *Langages* no.79, Larousse, Paris.
- _____. (1986a), Les Adjectifs composés du français - in rapport de recherches du PRC, *Informatique et linguistique*, CNRS, Paris.
- _____. (1986b), *Grammaire transformationnelle du français, 3 -Syntaxe de l'adverbe*, Asstril.
- _____. (1986c), *Typologie des noms composés : le lexique électronique des noms composés du français*, Université Paris 13.
- _____. (1988a). Les phrases élémentaires et ses composants. Une discussion de quelques exemples, *Travaux de Linguistique* no.17, Duculot.
- _____. (1988b). Les limites de la phrase figée, *Langages* no.90, Larousse.
- _____. (1990), *Grammaire transformationnelle du français 3: Syntaxe de l'adverbe*, Asstril.
- _____. (1991a), *Les expressions figées, Une descriptions des expressions françaises et ses conséquences théoriques*, Université Paris 7.
- _____. (1991b), Les principes de la représentation des formes figées dans le lexique- grammaire du français (manuscrit).
- _____. (1995), La notion de lieu argument du verbe, Tendances récentes en linguistique française et générale, *LIS* 20. J. Benjamins B.V.
- _____. (1996), Les verbes supports d'adjectifs et le passif, *Langages* no.121, Larousse.
- _____. (1998), La fonction sémantique des verbes supports, *Travaux de linguistique* no.37, Duculot.

- _____. (1999), Sur la définition d'auxiliaire du verbe, *Langages* no.135, Larousse.
- Gruber, J. S., (1976), *Lexical Structure in Syntax and Semantics*, North Holland Publishing Company.
- Guillet, A., Leclère, C., (1981), Restructuration du groupe nominal, Formes syntaxiques et prédicats sémantiques, *Langages* no.63, Larousse, Paris.
- _____. (1992), *La structure des phrases simples en français. Constructions transitives locatives*, Droz.
- Guiraud, P., (1961), *Les locutions françaises*, Paris, P.U.F., Que sais-je?
- Haegeman, L., (1991), *Introduction to Government and Binding Theory*. Basil Blackwell, Inc.
- _____. (1997), *The New Comparative Syntax*, Longman Linguistics Library.
- Han, H.-S., (1987), *The Configurational Structure of the Korean Language*, Doctoral dissertation, University of Texas at Austin.
- Hardin, C. L. eds., (1997), *Color categories in thought and language*, Cambridge University Press.
- Heid, U., (1997), *Zur Strukturierung von einsprachigen und kontrastiven elektronischen Wörterbüchern*, Max Niemeyer Verlag, Tübingen.
- _____. (2000), *Proceedings of the Ninth, EURALEX International Congress*. EURALEX 2000. Stuttgart. Germany.
- Heinz, M., (1993), *Les locutions figurées dans le < Petit Robert >*, Max Niemeyer Verlag.
- Herpiö, M., (2002), Benedict: an EU Project for an Intelligent Dictionary, Kernerman Dictionary News 10:9. Also available at <http://kdictionaries.com.newsletter.html>.
- Heylen, D. eds., (1993), Collocations, *Technical report ET_10175*, Taaltechnologie, Utrecht.
- Heylen, D. et al., (1993), Collocations, Dictionaries and MT, AAA, *93 Spring Symposium Series: Building Lexicons for Machine Translation*, Stanford.
- Hintikka, J. ed., (1994), *Aspect of Metaphor*, Kluwer Academic Publishers.
- 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.
- Hopper, P., Thomson, S. ed., (1982), *Syntax and Semantics* Vol. 15, Studies in Transitivity, Academic Press.
- Ide, K. (1993), (in Japanese) ['Denshi Jiten Nyuumon Q&A.']=A question and answer introduction to electronic dictionaries] [Gengo=Language] 22.5:18-21.
- Im, Y.-J., (1984), *Case-Tropism: The Nature of Phrase and Clausal Case*, Doctoral dissertation, University of Washington.

- Indurkha, B., (1992), *Metaphor and Cognition, An Interactionist Approach*, Kluwer Academic Publishers.
- Iwan'ska, T. M., S. Shapiro. eds., (2000), *Natural Language Processing and Knowledge Representation*, AAA/Pun/ The MIT Press.
- Jackendoff, R., (1972), *Semantic Interpretation in Generative Grammar*, MIT Press.
- _____, (1977), *X' Syntax: A Study of Phrase Structure*. The MIT Press.
- _____, (1987), The status of Thematic Relations in Linguistics Theory, *Linguistic Inquiry* no.21.
- _____, (1990). *Semantic Structures*, MIT Press.
- _____, (2002). *Foundations of language*, Oxford University Press.
- Jacquet-Pfaut, C. et al., (2005), *Mais que font les linguistes*, L'Harmattan.
- Jenkins, L., (2006), *Variation and Universals in Bilingualistics*, Elsevier Health sciences.
- Jespersen, O., (1971), *La Philosophie de la grammaire*, Les Editions de Minuit, Paris.
- Kang , M.-Y., (1988), *Topics in Korean, Phrase Structure, Variable Binding and Movement*, Doctoral dissertation, MIT.
- Kang , Y.-S., (1986), *Korean Syntax and Universal Grammar*, Doctoral dissertation, Harvard University.
- Katz, J.J. (1973), Compositionality, Idiomaticity, and Lexical Substitution, in Anderson & Kiparsky (eds.).
- Katz, J.J., Postal, P.M., (1963), Semantic Interpretation of Idioms and Sentences Containing Them, *QPR* 70: 275-82.
- Kavi, M., (1995), *Syntax-Semantics Interaction in Sentence Understanding*, Ph.D. diss., Computer Science, Georgia Institute of Technology, GIT-CC-95/10.
- _____, (1996), Ontology Development for Machine Translation: Ideology and Methodology, *Technical Report MCCS-96-292, Computing Research Laboratory*, New Mexico State University, Las Cruces, NM.
- Kay. P., (1997), *Words and the grammar of context*, CSLI Publications.
- Kayne, R. S., (1994), *The Antisymmetry of Syntax*, MIT Press.
- Kennedy, Ch., (1999), *Projecting the adjective*, New York/London: Garland Publishing.
- Kepser, S., (2005), *Linguistic Evidence*, Mouton de Gruyter.
- Keyser, S. J., (1968), Review of Sven Jacobson, Adverbial Position in English, *Language* no.44.
- Kiefer, F., Kiss, G., Pajzs, J., (1996), *Papers in Computational Lexicography Complex'96*, Research Institute for Linguistics Hungarian Academy of Sciences.
- Kittay, E., (1987), *Metaphor, Its cognitive force and linguistic structure*, Clarendon Press Oxford.
- Kleiber, G., (1999), *Problèmes de sémantique. La polysémie en questions*,

P.U.S.Nancy.

- Knowles, F. E., (1990), The Computer in Lexicography in F. J. Hausmann *et al.* eds. 1989-91:1645-72.
- Koenig, J-P.,(1999), *Lexical Relations*, CSLI Publications.
- Korfhage, R., (1997), *Information Storage and Retrieval*, Wiley Computer Publishing.
- Kronning, H., (1996), *Modalité, cognition et polysémie*, Acta Universitatis Upsaliensis.
- _____, (2001), *Langage et référence*, Uppsala: Uppsala Universit.
- Kulikov, L. eds., (1998), *Typology of Verbal Categories*, Max Niemeyer Verlag.
- Lagae, V., *et al.* eds., (2002), *Temps et aspect: de la grammaire au lexique*, Amsterdam: Rodopi.
- Lakoff, G., (1968), Instrumental Adverbs and the Concept of Deep Structure, *Foundation of Language*.
- _____, (1970), *Pronominalization, Negation, and the Analysis of Adverbs*, English Transformational Grammar. Mass.: Ginn and Co.
- Lambrecht, K., (1994), *Information Structure and Sentence Form : Topic, Focus and the Mental Representation of Discours Referent*, Cambridge University Press.
- Lappin, S. ed.,(1988), The syntax and semantics of NPs, Special issue of *Linguistics* no.26.
- Lawers, P., (2004), *La Description Du Francais entre La Tradition grammaticale et La Modernité* *Linguistique*, Peeters.
- Le Pesant, D., (1994), Les compléments nominaux du verbe <lire>, une illustration de la notion de <classes d'objets>, *Langages* 115, Paris, Larousse.
- Lee, S.-H., (2001), *Les classes d'objets d'événements. Pour une typologie sémantique des noms prédicatifs d'événements en français*, thèse de doctorat, LLI-Univ. Paris XIII.
- _____, (2003), Some Problems in Constructing Semantic Classes for Korean Predicate Nouns, *Language Research Sepecial Issue*, Language Education Institute, Seoul National University.
- Leech, G.N., (1974), *Semantics*, Harmondworth, Penguin.
- Leffel, K., Bouchard, D. eds., (1991), *Views on Phrase Structure*. Kluwer Academic Publishers.
- Ligozat, G., (1994), *Représentation des connaissances et linguistique*, A. Colin.
- Lin, D., (1995), A dependency-based method for evaluating broad-coverage parsers, *Proceedings of IJCAI-95*.
- Loses, R. L., (1998), *Text Retrieval and Filtering: Analytic Models of Performance*, Kluwer Academic Publishers.
- Lyons, J.,(1977), *Semantics*, Cambridge University Press.
- Maienborn C. et Wöllstein. A., (2005), *Event Arguments*, Max Niemeyer Verlag.
- Makini, S., M. Tsutsui., (1986), *A Dictionary of Basic Japanese Grammar*, The Japan Times.

- _____. (1995), *A Dictionary of Intermediate Japanese Grammar*,
The Japan Times.
- Makkai, A., (1972), *Idiom Structure in English*, The Hague: Mouton.
- Malim, T., (1994), *Cognitive Processes , Attention, Perception, Memory, Thinking
& Language*, The Macmillan Press LTD.
- Marantz, P., (1984), *On the Nature of Grammatical Relations*, MIT Press.
- Marcono, D., (1997), *Lexical Competence*, The MIT Press.
- Mathieu, Y.Y., (2000), *Les verbes de sentiment*, Paris: Editions CNRS.
- Martin, J. R. eds., (1998), *Reading Science, critical and functional perspectives
on discourses of science*, Routledge.
- Martin, S. E., (1992). *A Reference Grammar of Korean*, Charles E. Tuttle
Company.
- Martins-Baltar, M. et al., (1990), *Langue naturelle et logique, La sémantique
intensionnelle de Richard Montague*, Peter Lang.
- Mathieu-Colas, M. et al., (1998), *Les Marques de domaines dans le dictionnaires
électroniques*. Rapport technique no.1 de LLI, Université Paris
13.
- Mauldin, M. L., (1991), *Conceptual Information Retrieval: A Case Study in
Adaptive Partial Parsing*, Kluwer Academic Publishers.
- McLeod, C. et al., (1998), *COMPLEX Syntax*, *Computers and Humanities* no.31.
- Mckeown, M. G., (1987), *The Nature of Vocabulary Acquisition*, Lawrence
Erlbaum Associates Publishers.
- Mejri, S., (1996), *Le figement lexical*, Université de Tunis.
- Mel'čuk, I.-A., (1988), *Dependency Syntax: Theory and Practice*, State Univ. of
New York Press, Albany.
- Mel'čuk, I.-A. et al., (1984), *Dictionnaire explicatif et combinatoire du français
contemporain I*, Les presses de l'université de Montréal.
- _____, (1988), *Dictionnaire explicatif et combinatoire du français
contemporain II*, Les presses de l'université de Montréal.
- _____, (1992), *Dictionnaire explicatif et combinatoire du français
contemporain III*, Les presses de l'université de Montréal.
- _____, (1995), *Introduction à la lexicologie explicative et combinatoire*,
Editions Duculot.
- Michael, L., (2000), *Teaching collocation*, Klette.
- Michael, Z. et Carroll, J., (2004), *Les dictionnaires électroniques (Traitement
automatique des langues)*,
Hermes Science
Publications.
- Mitkov, R. eds., (1997), *Recent Advances in Natural Language Processing*, John
Benjamins Publishing Company.
- _____, (2003), *The Oxford Handbook of Computational Linguistics*,
Oxford University Press.
- Mollier, J-Y., Ory, P., (2003), *Pierre Larousse et son temps*, Larousse.
- Monceau, A., (1996), *La formation des noms composés de structure NOM
ADJECTIF : élaboration d'un lexique électronique*, Thèse de

- doctorat, Université de Paris 7.
- Moon, R., (1998), *Fixed Expressions and Idioms in English, A corpus-based approach*, Clarendon Press, Oxford.
- Murasugi, K., (1991), *Noun Phrases in Japanese and English*, Storrs: University of Connecticut.
- Nam, J.-S., (1994), *Classification Syntaxique des Constructions Adjectivales en Coréen*, these de doctorat, LADL-Univ. Paris VII.
- Napoli, D. J., (1993), *Syntax*. Oxford Univ. Press.
- Nehaniv, C. L. eds., (1999), *Computation for Metaphors, Analogy, and Agents*, Lecture Notes in Artificial Intelligence 1562, *Subseries of Lecture Notes in Computer Science*, Springer.
- Newmeyer, F. J., (1974), The Regularity of Idiom Behavior, *Lingua* no.34: 327-42.
- , (1988), *Linguistics: The Cambridge Survey*, Cambridge University Press.
- Nolke. H., (2001), *Le regard du locuteur 2*, Kimé.
- O'Grady, W. D.,(1992), *Contemporary Linguistic Analysis*. Copp Clark Pitman Ltd.
- P. Bouillon et A. Clas. eds., (1993), *La traductique*, Montréal, PUM , Editions Duculot, Belgique.
- Pak, M.-G., (1997), Traduction automatique et classes d'objets : le problème de <porter un vêtement> en français et en coréen, *Meta* no.42-1, Presses de l'Université de Montréal.
- Pearson, J., (1998), *Terms in Context*, John Benjamins.
- Piotrowsky, D. ed., (1996), *Lexicographie et informatique : autour de l'informatisation du Trésor de la Langue Française*, Didier Erudition.
- Pruvost, J., (2002), *Pierre Larousse. du grand dictionnaire au petit Larousse*, Honoré Champion.
- Puhvel, J. ed., (1969), *Substance and Structure of Language*, Berkeley: University of California Press.
- Pustejovsky, J., (1995), *The Generative Lexicon*, MIT Press.
- Pustejovsky, J., Bergler, S. eds., (1992), *Lexical Semantics and Knowledge Representation*, Springer Verlag.
- Radford, A., (1981), *Transformational Syntax*. Cambridge Univ. Press.
- , (1988), *Transformational Grammar*. Cambridge Univ. Press.
- Radman, Z., (1997), *Metaphors: Figures of the mind*, Kluwer Academic Publishers.
- Ramat, P., (1987), *Linguistic Typology*, Mouton de Gruyter.
- Rat, M., (1957), *Dictionnaire des locutions françaises*, Paris, Larousse.
- Retif, A., (1975), *Pierre Larousse et son oeuvre 1817-1875*, Larousse.
- Ricardo Baeza-Yates, Berthier Ribeiro-Neto, (1999), *Modern Information Retrieval*, Addison-Wesley.
- Riemsdijk, H., van E. Williams, (1981), NP structures, *The Linguistic Review* no.1-2, pp.171-218.

- Rijkhoff, J., (2002), *The Noun Phrase*, Oxford-New York: Oxford University Press.
- Rizzi, L., (1992), Argument/Adjunct (A)symmetries, *NELS* no.22.
- Rogers, M. A., Ahmad, K., (1998), The Translator and The Dictionary : Beyond Words? in *B.T.S. Atkins ed. Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*, Tübingen: Max Niemeyer Verlag.
- Sag, I., Wasow, Th., (1999), *Syntactic Theory, A Formal Introduction*, CSLI Publications.
- Sager, J. C., (1990), *A Practical Course in Terminology Processing*, John Benjamins publishing company.
- Saint-Dizier, P., (1999), *Predicative Forms in Natural Language & in Lexical Knowledge Bases*, Kluwer Academic Publishers.
- Salton, G., (1988), *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley Publishing Company.
- Salton, G., McGill, M. J., (1983), *Introduction to Modern Information Retrieval*, McGraw-Hill Book Company.
- Sampson, G., Haigh, R., Atwell, E., (1989), Natural language analysis by stochastic optimization: a progress report on Project APRIL, *Journal of experimental and theoretical Artificial Intelligence* no.1.
- Schank, R. C., (1975), *Conceptual Information Processing*, Amsterdam, North Holland.
- Schörebeld, D., (2001), *Where Lexicon and Syntax meet*, Berlin/New York: Mouton de Gruyter.
- Schreiber, P. A., (1971), Some Constraints on the Formation of English Sentence Adverbs, *Linguistic Inquiry* no.2.
- Schryver, G. M., (2003), Lexicographers dreams in the electronic-dictionary age, *International Journal of lexicography* no.16-2.
- Schtze, H., (1997), *Ambiguity resolution in Language learning, Computational and Cognitive models*, CSLI Publications.
- Schwanenflugel, P. eds., (1991), *The Psychology of word meanings*, Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Schwarz, M., (1994), *Kognitive Semantik/ Cognitive Semantics, Ergebnisse, Probleme, Perspektiven*, Gunter Narr Verlag.
- Searle, J. R., (1979), *Speech Acts: An Essay in the Philosophy of Language*, Cambridge Univ. Press, London.
- Selva, T., Chnier, T., (1998), Apport de l'informatique pour l'accès lexical dans les dictionnaires pour apprenants, projet Alexia in T. Fontenelle et al. eds., 631-42. Also available at <http://lifc.univ-fcomte.fr/RECHERCHE/P7/>.
- Sharman, R., (1990), *Evaluating a grammar as a language model for speech*.
- Shibatani, M. eds., (1995), *Approaches to Language Typology*, Oxford University Press.

- _____, (1996), *Grammatical Constructions , Their Form and Meaning*, Clarendon Press, Oxford.
- Shin, H.-P., (1999), The VP-Barrier Algorithm for Robust Syntactic Parsing in Head-Final Languages, *Proceedings of Natural Language Processing Pacific Rim Symposium*.
- Shyldkrot, H. B-Z., et Le Querler, N., (2005), *Les Périphrases Verbales*, John Benjamins Pub Co.
- Sinclair, J., (1991), *Corpus, Concordance, Collocation*, Oxford University Press.
- Siouffi, G. et Raendonck, D. V., (2000), *100 Fiches pour comprendre la linguistique*, Bréal.
- Smadja, F., (1991), *Extracting collocations from text. An application : Language generation*, UMI.
- Somers, H. L., (1987), *Valency and Case in Computational Linguistics*, Edinburgh University Press.
- Sowa, J. F., (1984), *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley, Massachusetts.
- _____, (1991), *Principles of Semantic Networks, Explorations in the Representation of Knowledge*, Morgan Kaufmann Publishers, INC.
- Sowa, J. F., E. C. Way, (1986), Implementing a semantic interpreter using conceptual graphs, *IBM Journal of Research & Development* Vol. 30, No.1, pp.57-69.
- Sproat, R., (1992), *Morphology and Computation, A Bradford Book*, The MIT Press
- Srinivas B., Sarkar A., Doran C. & Hockey B., (1998), Grammar & Parser Evaluation in the XTAG Project, *Proceedings of the International Conference on Language Resources and Evaluation*.
- Stede, M., (1999), *Lexical Semantics & Knowledge Representation in Multilingual Text Generation*, Kluwer Academic Publishers.
- Steele, J. eds., (1990), *Meaning-text theory, Linguistics, lexicography, and implications*, University of Ottawa Press.
- Steven A., (1991), *Parsing by Chunks In Principle Based Parsing*, Kluwer Academic Publishers.
- Svartvik, J., (1992), Directions in Corpus Linguistics, *Proceedings of Nobel Symposium 82 Stockholm*, 4-8 August 1991, Mouton de Gruyter.
- Szczepaniak, R., (2006), *The Role of Dictionary Use in the Comprehension of Idiom Variants*, Max Niemeyer Verlag.
- Tartter, V. C., (1998), *Language and It's Normal Processing*, Sage Publications.
- Teresa Cabré, M., (1999), *Terminology, theory, methods and applications*, John Benjamins Publishing Company.
- Thomason, R. H., R. C. Stalmaker, (1973), A Semantic Theory of Adverbs, *Linguistic Inquiry* no.4.

- Valentin, P. *et al.*, (1998), *Lexique et Cognition*, Presses de l'Université de Paris-Sorbonne
- Van Eynde, F., Gibbon, D. eds., (2000), *Lexicon Development for Speech and Language Processing*, Kluwer Academic Publisher.
- Van Halteren, H. eds., (2000), *Syntactic Wordclass Tagging*, Kluwer Academic Publishers.
- Van Valin, R., LaPolla, R., (1997), *Syntax*. Cambridge University Press.
- Veenstra, J., Sabine. B., (1998), *Fast NP Chunking Using Memory-Based Learning Techniques*, ILK(Induction of Linguistic Knowledge)-9807.
- Verspoor, M. eds., (1995), *Lexical & Syntactical Constructions and the Construction of Meaning*, John Benjamins P.C.
- Vivès, R., (1983), *Avoir, Prendre, Perdre : Constructions à verbe support et Extensions aspectuelles*, Thèse de 3ème cycle, Université de Paris 8.
- _____, (1988), Lexique-grammaire, nominalisation et paraphrase, *Lexique* no.6, P.U.L.
- _____, (1990), Les composés nominaux par juxtaposition, *Langue Française* no.87, Larousse.
- _____, (1993), La prédication nominale et l'analyse par verbes supports, *L'information grammaticale* no.59.
- Vosen P. ed., (1998), *EuroWordNet*, Dordrecht : Keuwer Academic Publishers.
- Walter A., Cook, S. J., (1989), *Case Grammar Theory*, Georgetown University Press.
- Weigand, E., (1998), *Contrastive Lexical Semantics*, University of Mnster.
- Weinreich, U., (1969), Problems in the analysis of idioms, in *Puhvel* (ed.) *Substance and Structure of Language*, Berkeley and Los Angeles: University of California Press.
- Whitman, (1989), Topic, Modality, and IP Structure, *Harvard Studies in Korean Linguistics* III.
- Williams, E., (1980), Predication, *LI* no.11-1, MIT.
- _____, (1981), Argument Structure and Morphology, *The Linguistic Review* no.1.
- Williams, E., (1980), The NP Cycle, *Linguistic Inquiry* no.11, pp.277-295.
- Willems, D. et Defrancq, B., (2004), *Contrastive analysis in language*, Palgrave Macmillan.
- Yun, J.-R., (1990), *Korean Syntax and Generalized X-bar Theory*, Doctoral dissertation, University of Texas at Austion.
- Yun, J.-T., (1999), Three Types of Chunking in Korean and Dependency Analysis Based on Lexical Association, *Proceeding of the 18th International Conference on Computer Processing Languages* (ICCPOL 99).
- Zubizarreta, M.L., (1982), *On the Relationship of the Lexicon to Syntax*, MIT Press.

◇ 별첨 결과물 CD 내역

- 2007년 최종 결과물

1. 상세전자사전 : 체언/용언
2. 전자사전 목록 : 체언/용언
3. 사전구축 지침서 : 체언/용언
4. 통합전자사전 :
- MDB데이터베이스/전자사전 관련 프로그램
5. 기반 연구 결과물
6. 회의록
7. 결과보고서

이 보고서는 국립국어원의 21세기 세종계획 ‘전자사전 개발’ 연구 용역 결과 보고서이다.

이 연구 보고서의 내용을 대외적으로 발표할 때에는 반드시 국립국어원의 21세기 세종계획 연구 결과물임을 밝혀야 한다.

21세기 세종계획 전자사전 개발

2007. 12. 03. 인쇄

2007. 12. 10. 발행

발행자 : 국립국어원

서울시 강서구 방화3동 827

전화 2669-9775

인쇄처 : 한울기획

전화 877-0815

이 보고서는 국립국어원에서 시행한
21세기 세종계획 연구 용역 결과 보고서입니다.