

발 간 등 록 번 호

11-1370000-000087-14

연구보고서 [2004년]

21세기 세종계획 국어 기초자료 구축

연구책임자 : 김흥규 (고려대학교)

문 화 관 광 부
국 립 국 어 원

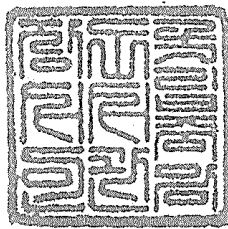
연구보고서 【2004년】

21세기 세종계획 국어 기초자료 구축

총괄 연구책임자

김 흥 규

문 화 관 광 부
국 립 국 어 원



3373040

제 출 문

문화관광부장관 귀하

21세기 세종계획 “국어 기초자료 구축”에 관하여 귀 부와 체결한 연구
용역 계약에 의하여 연구 보고서를 작성하여 제출합니다.

2004년 12월 12일

총괄 연구 책임자 : 김홍규(고려대학교)

■ 세부 과제명 : 현대국어 기초 말뭉치 개발

연구 책임자 : 김홍규(고려대학교)

연구 원 : 강범모 외 12명(고려대학교)

연구 보조원 : 강인범 외 28명

보 조 원 : 김종성 외 9명

요 약 문

1. 사업명

국어기초자료 구축 - 기초자료 분과

2. 연구의 목적

이 연구는 국어정보화 중장기 발전계획의 일환으로 시작된 <21세기 세종계획>의 10년간(1998-2007) 사업 중 제3단계의 1차년도 사업이다.

제1단계(1998-2000) 사업이 국어 정보화를 위한 기반을 다지는 데 주력했다고 한다면, 제2단계(2001-2003) 사업에서는 이를 바탕으로 보다 성숙된 결과를 확보하는 것을 주목적으로 삼았다. 즉, 제1단계 사업에서는 기초 말뭉치의 개발과 표준화, 그리고 이 말뭉치를 활용하기 위한 도구의 개발이 주된 목표였던 것에 비해, 제2단계 사업에서는 기초 말뭉치의 구축뿐 아니라 기초 말뭉치에 언어 정보를 주석한 2차 가공 말뭉치의 개발과 이를 활용하기 위한 도구의 개발이 주된 목표였다. 그리고, 올해부터 시작된 제3단계(2004-2007) 사업에서는 각종 말뭉치 구축 사업을 지속적으로 진행하는 동시에, 그간 구축된 말뭉치의 통합, 조정, 분석 및 서비스를 위한 종합적 연구를 목표로 삼는다.

이러한 목표 아래 2004년도에는 지난해에 이어 저작권 승인을 받은 현대국어 문어 말뭉치 500만 어절과, 형태분석 말뭉치 200만 어절, 형태의미분석 말뭉치 250만 어절을 개발한다. 아울러, 2003년도에 개발한 구문분석 말뭉치 종합 관리도구를 기반으로 20만 어절의 구문분석 말뭉치를 개발한다.

3. 연구의 필요성

우리의 국어 생활을 과학적 토대 위에서 향상시키고 국어의 특성에 적합한 정보 문화와 새로운 정보 처리 이론 및 기술을 발전시키기 위해서는 그 기반으로서 대규모의 국어 텍스트 말뭉치가 필요하다.

어문 생활과 정보 처리의 선진화 및 효율화를 추구하는 국가들은 모두 자국어론 중심으로 한 기초 언어 자료 데이터베이스를 구축하고 있으며, 이를 통해 언어 정책의 수립과 교육의 효율화는 물론 정보처리의 생산성을 높이는 데 커다란 효과를 얻고 있다. 우리나라에서도 정보화가 국가적 과제로 요구되는 상황에서, 국민들 모두가 우리말로 쉽고 편리하게 정보를 찾고 활용할 수 있게 하기 위해서는 우수한 품질과 대용량의 규모를 갖춘 국가적 언어자료 기반의 지속적 축적이 필수적이다.

아울러, 현대 국어는 국어의 역사적 발전·변화의 산물이며, 공간적으로는 국내 각 지역의 방언을 포함하고, 북한 지역 및 국외에 있는 동포들의 한국어까지를 포용하는 것이다. 이에 대한 심층적 연구와 한민족의 언어적·문화적 통합을 위한 기반을 마련하기 위한 언어 자원의 확보 또한 필요하다. 특히 각종 언어 정보가 부착된 분석 말뭉치의 개발은 보다 전문적이고 효과적인 언어 정보처리에 기여할 수 있으며, 한국어 연구에도 기여하는 바가 크다. 이와 같은 언어자료의 효과적 이용을 가능하게 하는 도구의 개발도 병행될 필요가 있다.

이와 같은 총괄적 필요성에 입각한 현대국어 말뭉치 개발의 필요성을 기술하면 다음과 같다.

국어 정보화를 위해서는 그 자료 기반으로서 대규모의 텍스트 말뭉치가 필수적이다. 각종 언어 처리를 위한 정보를 대규모의 말뭉치에서 통계적 방법으로 추출하는 것은 언어 연구와 정보처리의 선진화에 필수적인 과제이며, 따라서 우수한 품질과 대용량의 규모를 갖춘 말뭉치의 지속적인 구축이 필요하다. 이와 함께 정밀한 언어 분석 및 이론, 기술 개발의 기반으로 쓰일 수 있도록 원시 말뭉치에 여러 분석 표지를 부착한 분석 말뭉치의 개발도 이루어져야 한다. 아울러 이와 같은 사업의 성과를 연구자와 일반인이 손쉽게 이용할 수 있도록 지원할 필요도 있다.

4. 연구 내용 및 결과물

2004년도 세종계획에서는 현대국어 기초 말뭉치, 형태분석 말뭉치, 형태이미분석 말뭉치, 구문분석 말뭉치의 구축이 이루어진다.

1) 기초 말뭉치 구축

- ① 저작권 승인을 받은 현대국어 기초 말뭉치 500만 어절
- ② 표준화 지침에 따라 각종 문서 정보 기입
- ③ 오류 0.1% 이내
- ④ 결과물 : 전산 파일(CD-ROM) 10부, 보고서 200부, 저작권 승인 문서 일체

2) 형태분석 말뭉치 구축

- ① 품사 표지 부착된 말뭉치 200만 어절
- ② 세종 말뭉치에 대한 품사 표지 부착 표준 지침서
- ③ 오류 0.5% 이내
- ④ 결과물 : 전산 파일(CD-ROM) 10부, 보고서 200부

3) 형태이미분석 말뭉치 구축

- ① 품사 표지 및 의미 변별 표지가 부착된 말뭉치 250만 어절
- ② 세종 말뭉치에 대한 의미 표지 부착 표준 지침서
- ③ 오류 0.5% 이내
- ④ 결과물 : 전산 파일(CD-ROM) 10부, 보고서 200부

4) 구문분석 말뭉치 구축

- ① 구문 분석 표지가 부착된 말뭉치 20만 어절
- ② 세종 말뭉치에 대한 구문 분석 표지 부착 표준 지침서
- ③ 오류 1% 이내
- ④ 결과물 : 전산 파일(CD-ROM) 10부, 보고서 200부

5. 연구 결과의 활용

본 연구의 성과는 이후의 기초 언어자료 구축 연구에 계속 반영될 것이며, 적절한 과정을 거쳐 전문 연구기관, 학자, 산업체 및 국어 연구 교육 분야에 보급됨으로써 국어 정보화의 전 사회적 역량을 증진하는 데 기여하도록 한다. 보급은 ‘국어정보관리센터’를 통해 자료의 특성과 저작권 등의 사항을 고려하여 이루어지도록 한다.

1) 세종계획 홈페이지

- 현재 운영되고 있는 세종계획 홈페이지를 통해 국어 정보화 사업인 <21세기 세종계획>의 홍보와 함께 자료들을 공개한다.
- 홈페이지를 통해 기초자료뿐 아니라 <21세기 세종계획>의 다른 분과의 결과물도 제공된다.

2) 전문기관 연구용 배포

- 매년 세종계획을 통해 구축되는 자료들은 주관기관(국립국어연구원)이 판단하여 전문기관에 한정적으로 배포하고 있고, 배포된 자료들은 사업의 평가나 <21세기 세종계획> 내 분과간 협력 등의 목적으로 이용되고 있다.

3) 일반 사용자 배포

- 저작권자로부터 사용 승인을 얻은 말뭉치 자료와 분석 정보, 각종 분석도구를 일반 연구자와 사용자들에게 배포하여 국가적 자원의 활용도를 높이고 우리 사회 전체의 정보처리 역량을 증진하는 데 기여한다.

6. 참여 연구진

◦ 연구책임자	김홍규	고려대학교 국어국문학과
◦ 연구원	강범모	고려대학교 언어과학과
	김원경	고려대학교 국어국문학과
	김의수	고려대학교 민족문화연구원
	도원영	고려대학교 민족문화연구원
	배성우	고려대학교 국어국문학과
	서태길	고려대학교 민족문화연구원
	이동석	고려대학교 국어국문학과
	이봉원	고려대학교 민족문화연구원
	이상원	고려대학교 국어국문학과
	이상혁	고려대학교 민족문화연구원
	이정식	고려대학교 민족문화연구원
	정명숙	고려대학교 국어국문학과
	황화상	고려대학교 민족문화연구원
◦ 연구보조원	강인범	고려대학교 민족문화연구원
	고은숙	고려대학교 국어국문학과
	고은지	고려대학교 국어국문학과
	김대희	고려대학교 국어국문학과
	김윤희	고려대학교 국어국문학과
	김일환	고려대학교 민족문화연구원
	김정현	고려대학교 국어국문학과
	김지혜	고려대학교 국어국문학과
	남경완	고려대학교 민족문화연구원
	노석은	고려대학교 민족문화연구원
	박미경	고려대학교 언어과학과
	박봉곤	고려대학교 국어국문학과
	박지혜	고려대학교 민족문화연구원
	박현아	고려대학교 언어과학과
	성지연	고려대학교 국어국문학과
	손다정	고려대학교 국어국문학과

송경화	고려대학교 언어과학과
송원호	고려대학교 국어국문학과
안정아	고려대학교 민족문화연구원
이경숙	고려대학교 국어국문학과
이경호	고려대학교 민족문화연구원
이동혁	고려대학교 민족문화연구원
이영배	고려대학교 국어국문학과
이영제	고려대학교 국어국문학과
이지용	고려대학교 국어국문학과
조경임	고려대학교 국어국문학과
조경철	고려대학교 국어국문학과
차준경	고려대학교 민족문화연구원
하태석	고려대학교 국어국문학과

◦ 보조원

권대용	고려대학교 국어국문학과
김종성	고려대학교 국어국문학과
김태완	고려대학교 국어국문학과
박준희	고려대학교 국어국문학과
서일수	고려대학교 국어국문학과
신길선	고려대학교 국어국문학과
이두현	고려대학교 국어국문학과
정양석	고려대학교 언어과학과
조민영	고려대학교 국어국문학과
최동익	고려대학교 국어국문학과

연구 추진 실적

1. 1단계 연구추진 실적(1998~2000)

구분 \ 연도			1998	1999	2000	합계/비고
말 뽕 치	세종말뽕치 통합처리		(7,500만 어절)			(누계 합산 제외)
	현대 국어	기초말뽕치	1,600만 어절	1,600만 어절	800만 어절	4,000만
		형태소분석		150만 어절	200만 어절	350만
개발 규모 합계			1,600만 어절	1,750만 어절	1,000만 어절	4,350만
방법론, 표준화 연구			-어절분석표지 표준 화 연구 -병렬말뽕치 구축 기 초연구 -말뽕치 구성방법론 연구	-구문분석 방법, 표지 표준 연구	-형태 분석 말뽕치를 이용한 국어의 문 법 현상에 대한 계 량적 연구	
소프트웨어 개발			-각종 기초 도구 -용례추출기	-지능형 형태소 분석기	-말뽕치 통합 응용 시스템	
사업 진도 (%)			100	100	100	100
연구비 (천원)			459,000	510,000	376,000	1,345,000 ¹⁾

1) 1998, 1999년도 연구비는 구어자료, 역사, 구비, 병렬말뽕치 구축 사업까지 포함한 것임

2. 2단계 연구추진 실적(2001~2003)

구분연도			2001	2002	2003	합계/비고
말 뭉 치	현대 국어	기초말뭉치	700만 어절	500만 어절	500만 어절	1,700만 어절
		형태 분석	200만 어절	250만 어절	200만 어절	650만 어절
		형태 의미 분석	150만 어절	200만 어절	200만 어절	550만 어절
		구문 분석	기초 연구	3만 어절	12만 어절	15만 어절
개발 규모 합계			1,050만 어절	953만 어절	912만 어절	2,915만 어절
방법론 표준화 연구			체언 어절을 구성 하는 형태소의 빈도 (분포별 빈도, 개별 빈도 추출)	국어 연구에서 의 말뭉치 활용방 법 연구		
소프트웨어 개발			구문 태그 부착 말뭉치 구축도구	반자동 구문분석 말뭉치 구축도구 개발	구문분석 말뭉치 종합 관리 도구 개발	
사업 진도 (%)			100	100	100	100
연구비 (천원)			410,000	420,000	395,000	1,225,000

3. 3단계 연구추진 실적 및 계획(2004~2007)

구분연도			2004	2005	2006	2007	합계/비고
말뭉치	현대 국어	기초말뭉치	500만 어절	500만 어절	500만 어절	통합 조정 (1차년도 후처리분 중 2,800만 어절 흡수)	4,300만 어절
		형태 분석	200만 어절	400만 어절	1,000만 어절	1,400만 어절	3,000만 어절
		형태 의미 분석	250만 어절	200만 어절	-	-	450만 어절
		심층형 의미분석 말뭉치	-	50만 어절	150만 어절	100만 어절	300만 어절
		구문 분석	20만어절	20만 어절	45만 어절	50만 어절	135만 어절
개발 규모 합계			970만 어절	1,170만 어절	1,695만 어절	4,350만 어절	8,185만 어절
방법론, 표준화 연구			-	말뭉치를 이용한 언어 연구	말뭉치를 이용한 언어 연구	말뭉치를 이용한 언어 연구	
소프트웨어 개발			-	자동형 구문 분석 도구 개발	완전 자동 형태분석 태거 개발	말뭉치 분석 시스템 개발	
사업 진도 (%)			100				
연구비 (천원)			335,900	400,000	440,000	400,000	1,575,900

목 차

1. 기초 말뭉치 구축	1
가. 목표	1
나. 구축 설계 및 원칙	1
1) 기초 말뭉치의 설계	1
2) 말뭉치 구성의 원칙	2
3) 장르별 균형을 고려한 말뭉치 설계	2
다. 구축 방법 및 결과	3
1) 구축 방법	3
2) 구축 결과	5
라. 저작권 승인	5
참고문헌	8
<부록 1> 세종계획 기초 말뭉치 구축 지침 (2004 ver. 1.0)	9
<부록 2> 2004년도 세종계획 기초 말뭉치 목록	68
<부록 3> 2004년도 세종계획 기초 말뭉치 원전 소장 정보	74
2. 형태분석 말뭉치 구축	83
가. 연구 배경 및 목표	83
1) 연구 배경	83
2) 연구 목표	84
나. 구축과정	84
1) 형태분석 말뭉치 구축 준비	84
가) 대상 말뭉치 선정	84
나) 장르별 구성	86
다) 분석 지침 수정 및 보완	86
2) 형태분석 말뭉치 구축	86
가) 말뭉치 전처리	86
나) 자동 tagger를 이용한 표지 부착	87
다) 후처리 과정	87
다. 통계 결과	90
참고문헌	124
<부록> 형태분석 말뭉치 구축 지침 (2004 ver. 1.0)	125

3. 형태이미분석 말뭉치 구축	169
가. 연구의 목적과 필요성	169
1) 연구 목적	169
2) 연구의 필요성	169
나. 연구 내용	170
1) 개요	170
2) 분석 지침 수정	171
3) 작업 대상 선정	171
가) 대상 자료	171
나) 대상 범주	173
4) 분석 도구 개발	175
5) 말뭉치 구축 절차	175
가) 형태분석 말뭉치의 통계 추출	176
나) 동음이의어 사전과 비교	176
다) 작업대상 어휘의 용례 추출	176
라) 분석 도구를 이용한 분석	176
마) 원본 말뭉치에 결과 적용	176
바) 오류 수정	177
다. 구축 결과의 통계	177
참고문헌	203
<부록> 형태이미분석 말뭉치 구축 지침 (2004 ver. 1.0)	204
4. 구문분석 말뭉치 구축	207
가. 연구의 목적과 필요성	207
1) 연구 목적	207
2) 연구의 필요성	207
3) 관련 연구 검토	208
나. 연구 내용	209
1) 분석 방법론 연구	209
2) 작업 대상 선정	210
가) 대상 자료	210
나) 분석 단위	211
다) 말뭉치 구축 절차	214
다. 연구 결과	215
참고문헌	221
<부록> 구문분석 말뭉치 분석 지침 (2004 ver. 1.0)	223

현대국어 기초 말뭉치 개발

연구책임자 : 김홍규

연구 원 : 강범모 외 12명

연구보조원 : 강인범 외 28명

1. 기초 말뭉치 구축

가. 목표

국어 정보화를 위해서는 우수한 품질의 현대 국어 말뭉치가 대규모로 구축되어야 한다. 또한 각종 자연 언어 처리 및 응용 소프트웨어의 장기적인 발전을 도모하기 위해서는, 그 기반이 되는 현대 국어 말뭉치가 지속적으로 구축될 필요가 있다. 구축된 기초 말뭉치는 그 자체로 언어 연구를 위한 소중한 자료가 될 수도 있으며, 여기에 형태 정보나 의미 정보를 주석하여 보다 정밀한 2차 말뭉치로도 가공될 수 있다. 아울러 이러한 사업의 성과를 국어학이나 전산학 연구자들뿐만 아니라 일반 사용자들이 쉽게 이용할 수 있도록 지원할 필요가 있다.

이러한 필요성에 의해 진행된 세종계획 기초자료 구축분과의 지난 6년간(1998~2003) 작업에서는 약 1억 3,200만 어절 규모의 기초 말뭉치¹⁾가 구축되었다. 세종계획 제3단계의 1차년도인 2004년도에는 그간 구축된 말뭉치의 장르별·시대별 구성 비율을 세밀하게 검토하여, 부족한 부분의 텍스트를 우선적으로 구축하였다.

기초 말뭉치의 보급 및 활용을 원활히 하기 위해서는 원본 텍스트의 저작권 승인 문제를 해결해야 한다. 따라서 2001년도 이후 구축되는 기초 말뭉치는 구축 작업과 병행하여 원본 텍스트의 저작권 승인 문제도 해결하였다.

나. 구축 설계 및 원칙

1) 기초 말뭉치의 설계

-
- 1) 이 수량은 기존에 이미 구축된 국립국어연구원과 과학기술원의 말뭉치를 통합, 후처리한 것도 포함한 것이다. 통합, 후처리한 어절수는 약 7,500만 어절이므로 세종계획에 의해 새롭게 구축된 것은 약 5,700만 어절이 된다.

20세기 초 이후부터 현재까지의 한국어 문어 자료를 대상으로 한다.

2) 말뭉치 구성의 원칙

- 지난 6년간 세종계획을 통해 구축된 말뭉치의 시대별·장르별 분포를 참조하여, 말뭉치의 균형을 확보한다.
- 연령, 사회 계층 및 각 분야의 언어 현실을 충분히 드러낼 수 있도록 텍스트의 다양성과 포괄성을 확보한다.

3) 장르별 균형을 고려한 말뭉치 설계

세종계획의 기초 말뭉치는 3단계까지 총 1억 어절 규모의 균형 말뭉치 구축을 목표로 하고 있다. 1단계와 2단계에서는 5,700만 어절을 구축했고, 3단계 3년차까지 구축되는 1,500만 어절과 1차년도 기존 말뭉치 후처리 분 가운데 2,800만 어절을²⁾ 통합하여 총 1억 어절 규모의 균형 말뭉치를 구축할 예정이다. 따라서, 3단계에서는 1억 어절의 균형 말뭉치 구축을 위한 구축 계획에 따라 진행될 것이다.

말뭉치의 장르별 균형에 대해서는 이미 1998년도에 방법론이 제시된 바 있다. 1-2단계에서 구축된 말뭉치와 3단계에서 일부 통합될 7500만 말뭉치의 장르별 분포를 도표로 제시하면 다음과 같다.

2) 기존 말뭉치 후처리분은 전체 7,500만 어절이지만, 순구어나 준구어에 해당되지 않는 순수 문어 말뭉치는 6,100만 어절이다. 문어 말뭉치 6,100만 어절 가운데 2,800만 어절을 통합하게 될 것이다.

제1분류	제2분류	제3분류	7500만 구축 어절	1-2단계 구축 어절	1억 어절 대비 구축 비율	3단계 구축 목표량
구어 (10%)	순구어(5%)			3,430,153	3.4%	
	준구어(5%)					
문어 (90%)	신문(20%)		0	15,725,030	15.7%	649만
	책,상상(20%)		2,634만	10,104,466	10.1%	1,211만
	책,정보 (35%)	교육자료(10%)	218만	2,240,302	2.2%	164만
		사회(15%)	722만	2,391,027	2.4%	344만
		예술/취미/생활(15%)	312만	3,029,496	3.0%	280만
		인문(20%)	870만	4,924,745	4.9%	285만
		자연(10%)	151만	882,059	0.9%	300만
		체험기술(15%)	486만	3,120,191	3.1%	271만
		총류(15%)	0	1,978,725	2.0%	385만
		잡지(10%) 합계		735만	7,603,024	7.6%
	기타(5%) 합계		6만	1,775,110	1.8%	378만
총계		6,134만	57,204,328	57.1%	4,617만	

[표 1] 1-2단계 구축 말뭉치와 7500만 말뭉치의 장르별 분포

3단계 1차년인 2004년에는 기존 말뭉치 구축 현황을 파악하여 3단계 전체 구축 계획을 수립하고, 1차년도 구축 목표를 설정했다. 표에서 3단계 구축 목표량과 7500만 구축 어절 수를 비교해 보면, <신문>과 <책-정보-자연>, <책-정보-예술, 취미, 생활>, <책-정보-총류>, <기타> 장르는 기구축 어절 수가 부족함을 알 수 있다. 따라서, 2004년에는 <책-정보-자연>과 <책-정보-예술, 취미, 생활> 장르의 자료를 확충하는 데에 중점을 두었다.

다. 구축 방법 및 결과

1) 구축 방법

기초 말뭉치 구축과 병행하여 저작권 승인을 받기 위해서 다음과 같은 작업 과정을 거쳤다.

■ 구축과정

말뭉치 지도 검토 ⇒ 수집 장르 및 텍스트 선정 ⇒ 출판사 협조 요청
⇒ 저작권 승인 ⇒ 파일 입수 ⇒ 표준화 및 마크업 ⇒ 헤더 부착
⇒ 마크업 및 TEI 헤더 오류 수정

◦ 말뭉치 지도 검토

지난 6년간 구축된 말뭉치의 세부 장르를 검토하여 구축량이 부족한 장르를 파악하였다.

◦ 수집 장르 및 텍스트 선정

말뭉치 지도를 검토한 결과 <책-정보-자연>, <책-정보-예술, 취미, 생활> 장르의 자료가 부족한 것으로 확인되었다. 따라서 이 분야의 구축 목표량을 상향 조정하였다.

◦ 출판사 협조 요청

원시 말뭉치 구축에 있어 출판사의 협조는 매우 중요하다. 저작권자의 연락처 및 책 파일을 받아야 하고, 저작권자의 사용 승인을 획득하는 데에도 출판사의 이해가 큰 도움이 된다. 구축량이 부족한 장르의 책을 출판하는 출판사에 협조 요청을 하였다.

◦ 저작권 승인

저작권자에게 직접 연락을 하여 사용 승인과 승인서를 요청하였다.

◦ 파일 입수

자료 사용 승인서를 획득하면, 출판사에 관련 파일을 요청하였다.

◦ 표준화 및 마크업

지침에 따라 자료의 표준화와 본문 마크업을 실시하였다.

◦ 헤더 부착

TEI 지침에 따라 마련된 21세기 세종계획의 TEI 헤더 형판을 부착한 후, 텍스트에 따라 관련 내용을 기입하거나 수정하였다.

◦ 마크업 및 TEI 헤더 오류 수정

마크업 및 헤더에 포함될 수 있는 오류를 수정하기 위해 1998년도 세종계획의 기초 도구 중 하나인 “Chk_Tei”를 이용하였다. 이 도구의 구동 결과 나타난 오류를 일일이 수정하였다.

2) 구축 결과

21세기 세종계획 전체 기초 말뭉치의 장르별 분포를 고려하여 설계된 2004년도 기초 말뭉치는 다음과 같이 구성되어 있다.

분 류		목표 어절수	구축어절수	구축비율	비 고
신문		2,000,000	2,159,965	41.3%	
잡지		800,000	804,829	15.4%	
책-상상		2,200,000	298,026	5.7%	
책-정보	교육		7,189	0.2%	
	체험		116,671	2.2%	
	인문		315,503	6.1%	
	사회		319,209	6.1%	
	자연		524,208	10.0%	
	예술/취미/생활		679,853	13.0%	
합 계		5,000,000	5,225,453	100%	

[표 2] 2004년도 세종계획 현대국어 기초 말뭉치의 장르별 비율

라. 저작권 승인

말뭉치의 폭넓은 보급과 활용을 위해서는 원 텍스트의 저작권자로부터 텍스트 사용 승인을 받아야 한다. 특히 최근 들어 저작권에 대한 인식이 사회적으로 더욱 높아지면서 원 텍스트의 저작권 사용 승인 문제는 매우 시급한 해결 과제가 되었다. 이에 따라 기초 말뭉치는 말뭉치 구축과 병행하여 저작권 사용 승인을 확보함으로써 배포 및 활용에 어려움이 없도록 하였다. 원본 텍스트의 저작권자로부터 저작권 사용 승인을 받는데 사용한 문서 양식은 다음과 같다.

<양식 1>

국가 언어 자원(말뭉치) 구축 및 활용 승낙서

저작물 :

저 자 :

저희 고려대학교 민족문화연구원 전자텍스트연구소에서는 문화관광부가 추진하는 “21세기 세종계획(국어정보화 중장기 사업)” 중 대규모 언어 자원 구축과 관련하여 귀하의 저작물을 우리말 어휘·문장 단위의 분석 자료인 말뭉치(corpus)로 구축하여 국어 연구 및 자연 언어 처리 응용 연구 등 공익적 목적을 위해 사용하고자 합니다. 본 사업의 관심 영역은 귀하가 만든 저작물의 내용이나 표현이 아니라 개별 단어 및 문장에 대한 국어학적인 정보 추출과 분석임을 말씀드리며 귀하의 저작물을 이용하여 구축한 말뭉치가 귀하의 저작·출판에 관한 어떠한 권리에도 손상을 입히지 않을 것임을 약속드립니다. 모쪼록 21세기 지식 정보화 시대에 귀하의 저작물을 우리말 정보 처리 발전의 기초가 되는 국가적 언어 자원의 구축에 유용하게 활용할 수 있도록 협조하여 주시면 감사하겠습니다.

본인은 위의 저작물에 대하여 연구 목적에 한해 아래의 내용에 해당하는 저작물의 사용을 승인한다.

1. 위의 저작물 텍스트를 일정한 형식으로 전자적 기록 매체(CD-ROM 등)에 담아 보존하는 일
2. 자모, 음절, 어휘, 구절, 문장 및 텍스트 단위의 국어연구와 언어 정보처리 분야에 응용하기 위해 저작물을 전자적으로 복제·변형(문장의 형태소별 분리, 문법·의미주석 부착 등)하고, 이를 연구용으로 학계·연구기관 등에 배포하는 일

2004년 월 일

저자 _____

(자필 서명 또는 날인)

문화관광부

민족문화연구원 전자텍스트연구소 귀중

<양식 2>

전자적 텍스트 사용에 관한 서약서

자 료 :

사용승락자(혹은 기관):

고려대학교 민족문화연구원 전자텍스트연구소는 위의 자료를 공익적 연구 목적을 위한 <한국어 기초 자료>의 일부분으로 수용하여, 어휘·문장 단위의 국어 연구에 활용되도록 할 것입니다. 본 연구소는 귀하(귀 기관)께 약속한 사용 조건 및 관리상의 규범을 충실히 이행함은 물론, 저작·출판자의 권익에 조금의 손상도 없도록 만전을 기할 것임을 서약합니다.

2004 년 월 일

고려대학교 민족문화연구원 전자텍스트연구소

소 장 강 범 모

○○○ 귀하(중)

참고문헌

강범모 편역 (1997) 전자텍스트 부호화 개설, (버나드, 스퍼버그머퀸 저), 서울: 고려대학교 민족문화연구소.

강범모, 장효현, 윤재민 (1998) “한국학 문헌의 전산화를 위한 TEI 표준안 응용 및 확장 방안 연구.” 『한국어전산학』 2. 한국어 전산학회.

국립국어연구원 (1995) 문헌자료 입력지침.

김흥규, 강범모 (1996) “고려대학교 한국어 말모듬 1” 『한국어학』 3. 한국어학회.

김흥규/강범모 (2000) “한국어 형태소 및 어휘 사용 빈도의 분석 1”, 서울: 고려대 민족문화연구원.

문화관광부 (1998) “21세기 세종계획 국어 기초자료 구축”분과 연구 보고서.

문화관광부 (1999) “21세기 세종계획 국어 기초자료 구축”분과 연구 보고서.

문화관광부 (2000) “21세기 세종계획 국어 기초자료 구축”분과 연구 보고서.

문화관광부 (2001) “21세기 세종계획 국어 기초자료 구축”분과 연구 보고서.

문화관광부 (2002) “21세기 세종계획 국어 기초자료 구축”분과 연구 보고서.

문화관광부 (2003) “21세기 세종계획 국어 기초자료 구축”분과 연구 보고서.

서상규 (1998) “연세대학교 언어정보 개발연구원 말뭉치 입력 지침”, 내부 문서.

Sperberg-McQueen, M. and L. Burnard (1995) TEI P3.

〈부록 1〉 세종계획 기초 말뭉치 구축 지침 (2004 version 1.0)

1. 개 관

1.1. 목적

여기서 제시하는 지침은 <21세기 세종 계획> 말뭉치 구축 분야에서 이용할 말뭉치 자료의 작성 지침이다. 내용상 크게 입력에 관한 부분과 TEI 규정에 따른 각종 표시 부분으로 나뉘어진다. 이 지침은 <21세기 세종계획 국어 기초자료 구축> 분과의 제1단계 사업 결과로 작성된 지침을 바탕으로 수정, 보완한 것이다.

1.2. 적용 대상

이 지침은 주로 현대국어 문어와 구어의 말뭉치 구축을 중심으로 마련된 것이며, 북한 및 해외 한국어 자료, 역사 문헌, 병렬 말뭉치 등의 특수 말뭉치 등도 기본적인 사항은 이 지침에 의거한다. 단, 1999년의 1단계 2차년도 사업과 같이 순구어 전사를 위한 별도 지침은 제외하였으며, 특수 말뭉치에 대한 상세 지침도 해당 분과의 보고서에 수록되었으므로, 여기에서는 다루지 않는다.

1.3. 지침 작성 방법

이 지침은 국립국어연구원, 고려대 민족문화연구원, 연세대 언어정보개발연구원이 독자적으로 마련해 온 각 지침을 중심으로 TEI 관련서적을 참고하여 종합·보완한 1차년도(1998) 지침에, 그간의 실무 작업에서 발견된 문제점 등을 수정, 추가, 보완한 것이다. 여기서 사용한 TEI header나 마크업은 가장 기본적인 사항들을 중심으로 제시하고 추가된 부분은 별도로 정리하였다.

1.4. 기본 원칙

- 원문 유지 : 문헌의 경우 원문을 그대로 유지함을 원칙으로 한다. 단, 역사 문헌의 경우는 별도로 규정한다.
- 문서 정보 표기 방식 : SGML/TEI, 특히 TEI Lite를 수용하여 문서 정보를 표기한다.
- 헤더 : 헤더에 기본적인 정보는 자세히 표기한다.
- 본문 마크업 : 본문에는 최소한의 본문 정보 마크업을 붙인다.
- 표본 : 헤더의 경우 표본 형판 파일을 제공하여 그곳에 채워 넣는 것을 원칙으로 하며, 본문도 텍스트의 종류에 따른 마크업의 예를 제공한다.

2. 파일 이름

2.1. 파일 이름의 기본 형식

도스 파일 명칭을 준수한다. 파일 이름은 8자리를 모두 사용하며, 3자리의 확장자를 붙인다. 파일의 첫 숫자는 세종계획의 사업연도를 표시하는 것이다.

따라서 7차년도인 올해 구축되는 파일의 형식은 다음과 같다. (나머지 7자리의 구성은 부록 참조.)

7xxxxxxx.hwp

2.2. 확장자

작업 중간 단계의 가공 파일은 일정한 확장자를 붙여 정리하는 것을 권장한다. (*.h01, *.h02, *.h03 등)

최종 파일의 확장자는 *.hwp로 했다.

2.3. 책의 파일화

한 권을 한 파일로 만드는 것을 원칙으로 한다.

※ 고서(古書)는 권(卷)과 책(冊)의 구분이 다른 경우가 적지 않다. 즉 동일한 권(卷)이 분책된 경우나 여러 권(卷)이 한 책으로 묶이는 것이다. 이를 고려하여 한 권(卷)이 하나의 파일이 되도록 한다.

2.4. 파일 이름 작성

- 1) 현대어 자료 파일의 첫 자리에는 세종계획 연차 표시를 한다¹⁾. 둘째 자리에 파일 작성 기관의 코드를 명시한다. 단, 현대어 자료 이외의 북한어, 연변어, 역사문헌자료, 구비문학자료는 첫 자리에 자료 코드를 표시한다. 북한어 및 해외한국어의 경우, 두 번째 자리 이후부터는 현대어 세 번째 자리 이후에 해당하는 분류 체계를 사용하며, 역사문헌자료와 구비문학자료는 별도의 분류 체계를 사용한다. 역사문헌자료와 구비문학자료의 분류 체계에 대해서는 6), 7) 참조.

<분류 코드>

- A: 국어연구원
- B: 고려대학교
- C: 연세대학교
- D: 한국과학기술원
- Z: 개인
- N: 북한 자료
- O: 연변 자료
- P: 역사문헌 자료 (국어사, 고전문학)
- Q: 구비문학 자료

- 2) 세 번째(북한 및 해외 한국어는 두 번째) 자리에는 매체 구분과 책의 중요 장르 구분을 혼합하여 반영한다.

<분류 코드>

- A: 신문

1) 1999년 2차년도 작업부터 각 연차별 구별을 위해 새로 도입했다.(예: 2bl94001.hwp)

- B: 잡지
- C: 책-교과서
- D: 책-사전
- E: 책-소설
- F: 책-시
- G: 책-기타 상상적 텍스트 (수필 등)
- H: 책-정보(인문, 사회, 자연, 예술), 기타 비상상적 텍스트
- I: 기타 출판물
- J: 기타 비출판물
- K: 화면 있는 방송 녹화 전사
- L: 화면 없는 방송 녹화 전사
- M: 기타 녹음 전사
- N: 전자출판물

3) 네 번째/다섯 번째(북한 및 해외 한국어는 세 번째/네 번째) 자리에는 각 분야별로 규정된 방식의 표기를 기입한다.

① 신문

연도를 두 자리로 표시한다.

② 잡지

연도를 두 자리로 표시한다.

③ 문학작품(소설, 시, 수필, 기타 상상적 텍스트)

연도를 두 자리로 표시한다. 개정판 등으로 조판이 달라진 경우에는 구축된 자료의 조판의 1쇄 연도를 표시한다.

④ 교과서

과정과 학년을 표시한다.

<분류 코드>

▷ 과정(4번째 자리-북한 및 해외 한국어는 3번째)

A: 초등학교

B: 중학교

C: 고등학교

D: 대학교

Z: 기타

▷ 학년(5번째 자리-북한 및 해외 한국어는 4번째)

1: 초등/중고등 1학년

- 2: 초등/중고등 2학년
- 3: 초등/중고등 3학년
- 4: 대학의 교양과정
- 5: 대학의 전공과정
- 0: 학년 구분이 없거나 뚜렷하지 않은 경우

【예】

- A1 : 초등학교 1학년
- B3 : 중학교 3학년
- C2 : 고등학교 2학년

- 4) 부가 정보 명시 (6번째 자리-북한 및 해외 한국어는 5번째; 선택적)
 추가적 정보가 필요한 일부 매체의 경우 6번째(북한 및 해외 한국어는 5번째) 자리에 표시한다.

① 신문

종류를 표시한다. 신문 종류의 별도 목록을 작성, 관리한다.

<분류 코드>

- A: 조선일보
- B: 동아일보
- C: 한국일보
- D: 중앙일보
- E: 한겨레신문
- F: 경향신문
- G: 문화일보
- H: 국민일보
- I: 서울신문
- J: 세계일보
- K: 경제지
- L: 스포츠지
- M - U: 향후 신문 종류 표시를 위해 보유
- V: 로동신문
- W - Y: 북한, 연변 신문에 배정
- Z: 기타

② 잡지

종류를 표시한다. 잡지 종류는 대분류로 하고, 별도 목록을 작성, 관리한다.

<분류 코드>

- A: 시사월간지
- B: 시사주간지
- C: 대중/연예지
- D: 여성지
- E: 경제지
- F: 문화/예술지
- G: 과학/기술지
- H: 취미/실용지
- I: 초등용 학생잡지
- J: 중고등용 학생잡지
- ...
- Z: 기타

③ 교과서

과목을 표시한다.

<분류 코드>

- A: 국어
- B: 역사
- C: 외국어
- D: 사회
- E: 수학
- F: 과학
- G: 예체능
- ...
- Z: 기타 과목

5) 일련 번호 (6-8번째 자리)

각 자료에 일정한 일련번호를 붙인다. 6번째 자리에 부가 정보를 표시하는 경우 7-8번째 자리를 사용한다.

6) 역사문헌 자료(국어사/고전문학)

① 두 번째 자리에는 시기를 표시한다.

<분류 코드>

- 1 : 11세기 및 그 이전 시기
- 2 : 12세기
- 3 : 13세기
- 4 : 14세기
- 5 : 15세기
- 6 : 16세기
- 7 : 17세기
- 8 : 18세기
- 9 : 19세기
- 0 : 연대 불명 및 연대 혼합 자료
- A : 20세기 초

② 세 번째 자리에는 문헌의 성격에 따른 구분을 표기한다.

<분류 코드>

- A : 한문자료
- B : 원국문본
- C : 언해/번역자료
- D : 음역(音譯)자료
- E : 이두(吏讀)자료
- F : 구결(口訣)자료
- G : 향찰(鄉札)자료
- ...
- Z: 기타

③ 네 번째 자리에는 텍스트 유형에 따른 구분을 표기한다.

<분류 코드>

- A : 시가류
- B : 소설류
- C : 산문기록류 : 일기 등
- D : 종교서류 : 불경, 유교서, 도가서, 성서류 등
- E : 기술서류 : 의학서(醫學書), 병서(兵書), 농서(農書) 등
- F : 역학서류 : 한어, 몽어, 만주어(청어), 왜어 등
- G : 교민서류 : 윤음, 학습서, 교화서 등

H : 역사서류
I : 운서(韻書)류
J : 사서(辭書)류
K : 언간 및 문서류
Z : 기타

④ 5~8자리에는 일련번호를 붙인다(일련번호는 역사자료 전체를 하나로 한다).

7) 구비문학 자료

① 두 번째 자리에는 종류를 표기한다.

<분류 코드>

A: 설화
B: 민요
C: 민속극
D: 무가
Z: 기타

② 세 번째 자리와 네 번째 자리에는 지역을 표기한다. 세 번째 자리는 도 단위 구별 기호를, 네 번째 자리에는 군 단위 구별 기호를 붙인다.

<분류 코드 ; 도 단위 구별 기호>

A: 서울
B: 경기
C: 충남
D: 충북
E: 전남
F: 전북
G: 경남
H: 경북
I: 강원
J: 제주
K: 평안
L: 함경

M: 황해

O: 구별 불가능 또는 불명확

<군 단위 구별 기호>

별도 목록 작성, 관리 필요. 명시가 없을 경우 'O' 표시.

③ 5~8자리에는 일련번호를 표기한다.

2.5. 파일 이름의 예

- 신문 : 연차-기관-매체(신문)-연도(2자리)-종류-일련번호
 ▷ 2CA96E01.HWP (2차년도 연대 구축 한겨레신문 1996년)
- 잡지 : 연차-기관-매체(잡지)-연도(2자리)-종류-일련번호
 ▷ 2BB97B01.HWP (2차년도 고대 구축 시사주간지 1997년)
- 책-교과서 : 연차-기관-매체(교과서)-과정-학년-과목-일련번호
 ▷ 2BCA1A01.HWP (2차년도 고대 구축 국민학교 1학년 국어교과서)
- 책-소설 : 연차-기관-매체(소설)-일련번호
 ▷ 2BE00001.HWP (2차년도 고대 구축 소설)
- 북한(신문) : 북한자료-매체(신문)-연도(2자리)-종류-일련번호
 ▷ NA95V001.HWP (로동신문 1995년)
- 역사문헌 자료 : 역사문헌자료-시기-문헌유형-종류-일련번호
 ▷ P8BA0001.HWP (18세기 원국문본 고전 시가)
- 구비문학 : 구비문학자료-종류-지역(2자리)-일련번호
 ▷ QAD00001.HWP (충남 지방의 설화)

3. 입력 환경

- 한글 3.x 이상에서 입력한다.²⁾
- 탭 등 포맷 조판과 관련된 기능을 사용하지 않는다.
- 전체를 바탕글 스타일로 하며, 바탕글은 다음 속성을 가진다.

2) 올해부터 구축되는 기초 말뭉치는 유니코드로 제출될 계획이었으나 옛한글 문자, 특수 문자의 유니코드 통일안이 마련되지 않았으므로 기초 말뭉치 자료에 한해 기존의 방식대로 한글 3.0 포맷으로 구축하였다.

① 글자모양

- 명조
- 10 포인트
- 보통 모양
- 글자간격 0%
- 장평 100%

② 문단모양

- 왼쪽여백, 오른쪽여백, 문단간격, 낱말간격 모두 0
- 들여쓰기 2
- 줄간격 150%
- 정렬방식 왼쪽

③ 편집용지

- A4
- 용지방향 좁게
- 위/아래/왼쪽/오른쪽/머리말/꼬리말 모두 15.00mm, 제본 0mm

④ 페이지

- 페이지 표시는 하지 않는다.

⑤ 머리말/꼬리말

- 머리말/꼬리말은 달지 않는다.

4. 본문의 자료 표준화

4.1. 입력 내용

- 본문을 입력하는 것을 원칙으로 한다.
- 앞표지 부분, 목차 등과 부록, 색인, 뒷표지 등의 내용은 역사문헌 자료 등 특별한 목적이 있지 않는 한 입력하지 않는 것을 원칙으로 한다. 입력할 경우 앞, 뒷부분에는 적절한 태그(<front>, <back>)를 부착하여 본문(<body>)과 구분한다.

1) 역사 문헌 자료의 입력 내용

- 문헌 자료에 기재되어 있는 것은 가능한 한 모두 입력하는 것을 원칙으로 한다. : 내사기(內賜記), 간기(刊記), 발(跋), 서문(序文), 목차(目次) 등

4.2. 텍스트 외형 정보

- 현대어 텍스트는 인쇄된 텍스트의 외형을 중요시하지 않는다. 즉, 페이지 숫자, 페이지 바꿈이나 줄바꿈 등 인쇄된 형태의 정보를 줄 필요는 없다. 그 외 태그를 사용하여 표시할 가능성은 열어둔다.
- 활자 크기, 밑줄, 드러냄표 등은 무시한다.
- 원문에서 장식 등의 이유로 어절의 내부 음절을 띄운 경우 모두 붙인다.

【예】 장 길 산 → 장길산

4.3. 맞춤법

- 특별한 이유가 없는 한, 원문의 맞춤법을 그대로 유지한다.
- 단, 상당한 오류가 있어 교정이 필요한 경우에는 현행 맞춤법에 따라 교정한다. 수정 정보는 헤더에 반영한다.

4.4. 띄어쓰기

- 가능한 한 원문의 띄어쓰기를 유지한다. 단, 띄어쓰기에 상당한 오류가 있어 교정이 필요한 경우에는 현행 띄어쓰기 규정에 따라 교정한다. 수정 정보는 헤더에 반영한다.
- 교정 시 현행 띄어쓰기 규정에서 허용 규정으로 된 것은 원문의 띄어쓰기를 따른다.

4.5. 문장 부호 및 특수 기호

원문 그대로 유지하는 것을 원칙으로 한다.

1) 자판에 있는 부호 및 기호는 자판의 것을 사용한다.

① 자판에서 입력 가능한 기호를 전각 기호나 반각 기호를 사용하여 입력하지 않도록 한다.

② 겹낫표는 한글 자판 상태에서 입력하는 겹낫표는 표준코드에 없는 것이므로 전각으로 입력하기로 함.

③ 분수 표시는 전각 기호를 사용하지 않는다.

【예】 1/2, 3/4 (⅓ 등을 사용하지 않는다.)

④ 단, 따옴표, 가운데점 등은 모두 전각 기호로 사용한다.

2) 자판에 없는 부호 및 기호는 일차적으로 전각 기호를 선택하여 입력한다. 전각 기호가 없을 경우, 한글의 반각 기호로 입력한다.

① 따옴표는 전각 기호(3410과 3411)를 사용한다. 열고 닫는 따옴표를 구분하여 사용한다.

② 가운데점은 전각 기호(3404)를 사용한다. 한글 2.0 이상에서 가능한, 'Ctrl'과 'F10' 키를 동시에 눌러 문답상자를 부른 후 마침표를 쳐서 입력하는 방식을 취하지 않도록 한다.

【예】 국가·민족

③ 말줄임표는 전각 기호(3406)를 사용한다.

【예】 그러나...

- 말줄임표는 원문에서 점 3개 짜리나 6개 짜리가 섞여서 쓰이고 있다. 이 경우 원문대로 전각문자로 점 3개 짜리 또는 점 6개 짜리로 입력하기로 한다. 다만 준구어의 시나리오와 같은 경우 점 2개 짜리부터 7개 - 10개 짜리 등이 섞여서 쓰이고 있으므로 이 경우에만 전각문자를 사용하여 점 2개, 7개, 10개 등으로 표현된 것을 3개 혹은 6개로 바꾸기로 한다.

- 말줄임표 다음에 마침표가 있기도 하고 없기도 하는데, 문장이 마침표로 끝나는 상황이라면 마침표를 붙이기로 한다.

- 텍스트에서 말줄임의 의미로 쓰인 줄표의 경우 자판에 있는 붙임표(하이픈)로 입력하지 않고 전각 기호에 있는 긴 줄로 입력하는 것도 허용한다.

④ 영문자, 한글 자모, 괄호문자는 자판에 있는 기호로 간주하여 전각 기호를 사용하지 않는다.

【예】 ㄱㄴㄷㄹ, (1) (2) (3)

⑤ 원문자는 전각 기호를 그대로 사용되 원문자 숫자가 16 이상일 경우 전각 기호가 없으므로 두 숫자의 원문자를 나란히 붙여 사용한다.

【예】 ① ② ②①

3) 부호 또는 기호가 연이어 사용되되 어절이 구분되는 경우 그 사이를 띄운다.

【예】 …한다.” “그리고…

4.6. 한자

- 특별한 허용 규정이 없는 한, 원문의 한자를 모두 유지한다.
- 가능한 한 독음대로 해당 한자를 입력한다.
- 표준 한자 이외의 한자는 <한글>의 확장한자로 입력한다. 사용자 정의문자는 본 사업 수행을 위해 향후에 공인되는 것 이외에는 사용하지 않는다.
- 표준한자/확장한자에 없는 한자는 다음과 같이 외부 典據 중의 하나로써 표시한다.
- 외부 전거는 가능한 한 UCS 코드 또는 국내 간행 사전을 우선적으로 사용한다.
- 외부 전거를 사용했을 경우에는 그 사실과, 전거의 종류를 헤더에 기록한다.

외부 전거 표시의 형식 및 본보기는 다음과 같다.

<형식>

- ① 외부 전거 시작 표지로 &를, 마침 표지로 ;를 쓴다
- ② 위의 표지 사이에 전거에 따른 출처 표시를 넣는다.
- ③ 외부 전거 표시 내부에는 빈칸이 들어가지 않는다.
- ④ 외부 전거 표지는 그 전체가 한 글자에 해당하며, 앞뒤의 입력 상태에 별도의 스페이스 등을 추가하지 않는다.

【예】

孟子謂軻&U-9F03;曰: “子之辭靈丘而請士師, 似也, 爲其可以言也; … ”
淳于&U-9AE0;은 <孟子>에 등장하는 인물로서…

<외부 전거 종류와 표현 형식>

■ 국내 사전 및 자전류

- 東亞出版社 刊, 東亞 漢韓大辭典(李家源 외 감수, 1982)

예: &XD0873.05; (D=출처, 0873=해당 한자 수록면, 05=해당 면에서의 수록 순차)

- 民衆書林 刊, 漢韓大字典(民衆書林編輯局 編, 전면개정 · 증보판, 1997).

- 예: &XE0792.09; (E=출처, 0792=해당 한자 수록면, 09=해당 면에서의 수록 순차)
- 明文堂 刊, 明文漢韓大字典(車柱環 외 감수, 1984)
 - 예: &XY2133.09; (Y=출처, 2133=해당 한자 수록면, 09=해당 면에서의 수록 순차)

■ 국외 사전 및 자전류

- 大修館書店 刊, 大漢和辭典(諸橋轍次, 修訂版, 1986)
- 예: &XM08269; (M=출처, 08269=모로하시 사전 수록 한자의 일련번호)
- 四川辭書出版社・湖北辭書出版社 刊, 漢語大字典(漢語大字典編輯委員會 編, 1986).
- 예: &XH4.2619.01; (H=출처, 04=권차, 2619=수록면, 01=해당 면에서의 수록 순차)
- 中華書局 刊, 中華字海(冷玉龍 외 편, 1994).
- 예: &XZ2133.09; (Z=출처, 2133=해당 한자 수록면, 09=해당 면에서의 수록 순차)

■ 코드북

- UCS 코드(Unicode) : CJK 영역 한자 목록
- 예: &U-67E4;(67E4는 해당 한자의 16진수 코드값)
- CNS 코드 : 대만에서 제정한 Chinese National Standard 1~7수준 한자
- 예: &C3-76A3;(3은 CNS 문자 수준, 76A3은 16진수 코드값. Big5 코드는 'C0-'으로 표시)
- SuperCJK 자료집 : IRG에서 최근에 제안된 Version 3.0 CJK 슈퍼세트 한자 자료집. 68,323자 수록.
- 예: &S-13456;(13456은 SuperCJK 코드번호)

4.7. 영문자 및 외국문자

- ① 원문의 외국문자를 그대로 유지한다.
- ② 일차적으로는 자판에서 입력하되, 자판 입력이 불가능한 경우에는 전각 기호를 사용한다. 전각 기호가 없을 경우에는 반각 기호를 이용한다.
- ③ 글자 겹침은 사용하지 않는다.

4.8. 기타

- ① 각 입력기(한글)에서 제공하는 틀(수식, 표, 그림) 등은 사용하지 않는다.

- ② 원문에 도표/그림이 있을 경우 생략하고 그 사실을 명시적으로 밝혀 줌을 원칙으로 한다. 이때 <gap> 태그를 사용하고 reason 속성으로 그 이유를 밝힌다.
- ③ 각주는 원칙적으로 생략하나, 보존할 필요가 있을 경우 <note> 태그를 사용하여 표기한다.
- ④ 북한어의 경우, 대화를 표시하는 각괄호는 전각 기호(《(3416), 》(3417))를 사용한다.
- ⑤ 역사 문헌의 협주는 <note> 태그를 사용하여 구분한다.
- ⑥ 이 외의 비표준 문자의 경우 키보드 입력이 불가능하므로 음절단위로 &와 ; 사이에 모든 자소를 분리하여 입력한다.(예: &ㅈㅇㅏ;)⊕

5. 파일 전체 구조 및 SGML/TEI 태깅 일반 원리

이 연구에서 사용한 문서의 구조와 부호화는 SGML/TEI를 (보완) 수용한 것이다.

5.1. TEI(Text Encoding Initiative; 텍스트 부호화 주도 기구)

학문적 관점에서 전자 문서의 표준을 제정함으로써, 전자 문서의 교환 및 연구를 촉진시키기 위한 국제적 연구 프로젝트이다. 이 작업은 1987년부터 시작되었으며, 1994년에 보고서(TEI P3)가 제출되었다. TEI는 텍스트의 특성을 명시적으로 만드는 부호화 방법에 SGML(Standard Generalized Markup Language)을 사용한다.

SGML은 국제 표준으로(ISO 8879) 정보 처리 산업 전반에 걸쳐서 널리 사용되고 있으며, TEI는 SGML의 하위 문서 규약이다.

TEI 라이트는 1995년에 간행되었으며, TEI의 권고안에 대한 소개서로서 전체 TEI 부호화 방법 중 일부를 기술한 것이다.

5.2. 문서 구조

TEI 방식에 따르는 텍스트는 TEI 헤더(<teiHeader>)와 본문 텍스트(<text>)로 이루어진다.

모든 문서는 다음의 형식을 갖추도록 했다.

```
<!DOCTYPE tei.2 SYSTEM "c:\Wsgml\Wdtd\Wtei2.dtd" [
    <!ENTITY % TEI.corpus "INCLUDE">
    <!ENTITY % TEI.extensions.ent SYSTEM "sejong1.ent">
    <!ENTITY % TEI.extensions.dtd SYSTEM "sejong1.dtd">
]>
<tei.2>

    <teiHeader>
    헤더 내용
    </teiHeader>

    <text>
    본문 내용
    </text>

</tei.2>
```

- 1) 헤더: 헤더에는 전자 문서에 대한 기본적인 정보, 문서가 전자화된 경위, 텍스트의 분류적·상황적 정보, 수정 기록 등이 포함되는데, 이 지침에서는 TEI P3의 코퍼스 관련 태그(teicorp2.dtd)를 충분히, 그러나 단순화하여 이용한다.
- 2) 본문: TEI 지침을 단순화하여 수용, 부호화한다.

5.3 SGML 태깅

- 1) SGML의 기본 개념 : SGML은 텍스트의 마크업(부호화)에 대한 규약이며, 마크업은 '텍스트의 해석을 명확하게 하기 위해 표시하는 모든 방법'을 말한다. SGML은 개체, 요소, 문서유형 선언의 세 가지 중심 개념으로 이루어진다. 개체란 텍스트의 일부를 말하며, 문서 전체, 문자열, 특정 문자 모두 개체가 될 수 있다. 요소란 텍스트의 여러 층위에서의 구성 요소를 말하는데, 제목, 문단, 문장, 낱자 등의 다양한 텍스트 구성 요소이며, 특정한 태그로 그 시작과 끝을 표시하게 된다. 이러한 요소들 간의 관계, 요소와 속성 간의 관계를 규정해 주는 것이 DTD(document type declaration; 문서유형 선언)이다.
- 2) 태그 : SGML의 태그는 시작부분과 끝부분을 명시하는 것을 원칙으로 한다: <tagN>.....</tagN> 단, 공태그(empty)들--<pb>, <lb>, <gap> 등--은 끝 태그가 없다.

6. 텍스트 분류

6.1 원칙

- 1) 이 분류는 헤더에 텍스트 종류를 표시하기 위해 사용한다(<profileDesc>)

```
<profileDesc>
  <textClass>
    <catRef scheme='SJ21' target='M1331'></catRef>
  </textClass>
</profileDesc>
```

- 2) 네 자리 코드를 사용하고, 그 앞에 말뭉치 종류를 밝힐 수 있는 로마자 코드를 넣는다.

<분류 코드>

M: 현대어 (남한)
 N: 북한
 O: 연변 자료
 P: 역사문헌 자료
 Q: 구비문학 자료
 ...

【현대어(남한) 텍스트 예】

▶ 문어

- 신문 사회면 기사: M1113
- 컴퓨터 잡지의 기사: M1273
- 백과사전: M1300
- 국민학교 교과서: M1324
- 단행본 소설: M1331

- 잡지에 실린 시: M1233
- 시청 민원 안내문: M1400
- 중학교 학생 작문: M1527
- 여성 월간지 일년치 전체: M1290
- 시사월간지 일년치 전체: M1200

▶ (준)구어

- 일상 담화를 녹음하여 전사한 것: M2801
- TV에서 예술에 관하여 인터뷰한 내용을 전사: M2684
- 환경문제에 관한 좌담을 잡지에 수록한 것: M2263
- 방송 드라마 대본 (비출판): M2437
- 즉석 연설을 녹음하여 전사한 것: M2807
- 책으로 출판된 희곡: M2336

3) 현대어 이외의 자료들은 현대어의 4자리 코드를 그대로 사용할 수도 있고, 독자적 코드를 개발하여 사용할 수도 있다. 북한어도 남한 현대어와 동일한 코드로 표시하는데, 분류코드는 M 대신 N을 사용한다. 예를 들어

- 북한 신문 사회면 기사: N1113
- 북한 단행본 소설: N1331

4) 역사 문헌 자료(고전 자료)

네 자리 코드를 사용하고, 말뭉치 종류를 밝힐 수 있는 로마자 코드를 넣는다.

- 첫 번째 자리에는 역사자료 코드(P)를 표시했다.
- 두 번째 자리에는 시기를 표시한다.

<분류코드>

- 1 : 11세기 및 그 이전 시기
- 2 : 12세기
- 3 : 13세기
- 4 : 14세기
- 5 : 15세기
- 6 : 16세기
- 7 : 17세기
- 8 : 18세기
- 9 : 19세기
- 0 : 연대 불명 및 연대 혼합 자료

A : 1901년부터 1919년 사이의 자료(특수한 경우만 사용)

- 세 번째 자리에는 문헌의 성격에 따른 구분을 표기한다.

<분류코드>

A : 한문 자료
 B : 원국문본
 C : 언해/번역 자료
 D : 음역(音譯) 자료
 E : 이두(吏讀) 자료
 F : 구결(口訣) 자료
 G : 향찰(鄉札) 자료
 ...
 Z: 기타

- 네 번째 자리에는 텍스트 유형에 따른 구분을 표기한다.

<분류코드>

A : 시가류
 B : 소설류
 C : 산문기록류 : 일기 등
 D : 종교서류 : 불경, 유교서, 도가서, 성서류 등
 E : 기술서류 : 의학서(醫學書), 병서(兵書), 농서(農書) 등
 F : 역학서류 : 한어, 몽어, 만주어(청어), 왜어 등
 G : 교민서류 : 훈음, 학습서, 교화서 등
 H : 역사서류
 I : 운서(韻書)류
 J : 사서(辭書)류
 K : 연간 및 문서류
 Z : 기타

6.2 총론

텍스트 분류는 4개의 숫자로 표시하되, 각 자리는 1) 문/구어 구분, 2) 전달 매체에

따른 구분 (신문, 잡지, 방송 등), 3) 텍스트의 내용에 따른 구분 (인문, 사회, 자연 등), 4) 문어의 경우 더욱 자세한 내용의 구분 / 구어의 경우 담화 상황 (전화, 대담 등)을 표시하도록 한다.

6.3 분류체계

6.3.1. 문어/구어의 분류

- 1: 문어
- 2: 구어

6.3.2. 매체에 따른 분류

- 1: 신문
- 2: 잡지 (주간지, 월간지, 계간지)
- 3: 책
- 4: 기타 출판물 (안내문, 소책자, 정부 문서 등)
- 5: 기타 비출판물 (일기, 편지, 작문 등)
- 6: 화면이 있는 방송 (TV) 녹화 > 전사
- 7: 화면이 없는 방송 (라디오) 녹음 > 전사
- 8: 기타 녹음 > 전사
- 9: 전자출판물 (CD-ROM 등)

예) 어떤 문제에 대한 좌담을 기록한 것이 잡지에 출판된 경우, 구어로 분류하고, 매체는 잡지로 한다.

예) 방송극/시나리오/희곡의 경우 (준)구어로 인정하고, 매체는 책 또는 기타 출판물로 한다.

6.3.3. 내용에 따른 분류

내용에 따른 구분은 2단계이다. 첫번째 단계는 대략적인 구분이고, 두번째 단계는 상

제한 구분이다. 문어의 경우 이 두단계의 구분이 모두 필요하지만, 불확실할 경우 '0'으로 표시한다. 구어의 경우는 첫단계의 분류만 사용한다.

▷ 0: 총류 - 백과사전 등 여러가지 내용이 섞여 있거나, 그 구분이 불확실할 경우. 구어의 경우는 대개 내용적으로 여기에 속한다. 또, 시사지 등 종합지적 성격의 잡지 전체 내용도 총류로 파악할 수 있다.

0: 일반 - 총류의 경우 더욱 자세한 구분을 할 수 없다.

▷ 1: 신문 - 신문에만 독립적으로 적용되는 내용 코드

- 1: 보도·해설·정치 (종합)
- 2: 보도·해설·사회
- 3: 보도·해설·경제
- 4: 보도·해설·문화
- 5: 보도·해설·생활·과학
- 6: 보도·해설·국제·외신
- 7: 보도·해설·취미·연예·스포츠
- 8: 사설
- 9: 칼럼
- 0: 기타 (독자투고, 인물, 화제, 용어 등)

▷ 2: 교육자료

- 0: 일반
- 1: 아동 도서/상상적 산문
- 2: 아동 도서/비허구적 산문
- 3: 아동 도서/운문
- 4: 교과서/초등
- 5: 교과서/중등 (중고등학교)
- 6: 작문/초등
- 7: 작문/중등
- 8: 작문/대학

▷ 3: 상상적(예술적) 텍스트

- 0: 일반
- 1: 소설/일반
- 2: 소설/역사

3: 설화

4: 민요

5: 시

▷ 4: 체험 기술적 텍스트

0: 일반

1: 수필

2: 전기

3: 실기, 르뽀

4: 일기

▷ 5: 인문

0: 일반

1: 한국문화 일반

2: 언어

3: 철학, 사상

4: 신앙, 종교

5: 역사

6: 민속, 인류학

7: 심리

8: 교육

9: 지리

▷ 6: 사회

0: 일반

1: 사회학, 사회론, 사회운동

2: 정치학, 정치론

3: 법, 행정, 공문서

4: 외교, 국제

5: 경제, 금융, 경영

6: 군사, 국방

7: 환경, 공해

8: 매스컴, 대중문화

9: 여성

▷ 7: 자연

- 0: 일반
- 1: 기초 자연과학
- 2: 공학, 공업, 기술
- 3: 컴퓨터, 정보 통신
- 4: 농업, 임업, 축산
- 5: 해양, 수산업
- 6: (현재 비어있는 코드)
- 7: 보건, 의학

▷ 8: 예술론

- 0: 일반
- 1: 문학론, 문학사
- 2: 연극, 영화, 방송극론
- 3: 음악론, 음악사 (한국 음악, 외국 음악)
- 4: 무용
- 5: 미술
- 6: 건축, 디자인
- 7: 대중 연예

▷ 9: 생활

- 0: 일반 (여성 잡지/주간지의 내용도 전체적으로 여기에 해당)
- 1: 가정, 가족 (육아, 청소년, 노인)
- 2: 의식주
- 3: 취미, 오락
- 4: 스포츠
- 5: 만화
- 6: 광고문
- 7: 편지, 실용문
- 8: 설명서

6.3.4. 구어의 담화 상황에 따른 분류

문어의 경우 마지막 두 개 (3, 4 자리) 코드가 내용을 나타내지만, 구어의 경우 셋째

자리만이 대체적인 내용을 나타내고 네번째 자리는 담화 상황을 표시한다.
셋째 자리는 다음과 같이 문어의 경우와 같은 내용 코드를 사용한다.

- 0: 총류 (일상 대화 등 대부분의 구어에 지정된다)
- 1: 신문 (신문에 난 인터뷰 기사 등)
- 2: 교육자료
- 3: 상상적(예술적) 텍스트 (희곡, 드라마 대본 등)
- 4: 체험기술적 텍스트
- 5: 인문 (인문학 주제 강연 등)
- 6: 사회
- 7: 자연
- 8: 예술론
- 9: 생활 (스포츠 중계 등)

넷째 자리는 구어에 고유한 코드이다.

- 1: 대화/일상
- 2: 대화/전화
- 3: 대화/대답, 좌담
- 4: 대화/인터뷰
- 5: 토론, 회의
- 6: 연극(희곡)
- 7: 방송극, 영화
- 8: 연설, 강의, 강연
- 9: 즉흥적 독백 (즉흥적 연설 등)

6.3.5. 예

▷ 문어

- 1) 신문 사회면 기사: 1113
- 2) 컴퓨터 잡지의 기사: 1273
- 3) 백과사전: 1300
- 4) 국민학교 교과서: 1324
- 5) 단행본 소설: 1331
- 6) 잡지에 실린 시: 1233

- 7) 시청 민원 안내문: 1400
- 8) 중학교 학생 작문: 1527
- 9) 여성 월간지 일년치 전체: 1290
- 10) 시사월간지 일년치 전체: 1200

▷ (준)구어

- 1) 일상 담화를 녹음하여 전사한 것: 2801
- 2) TV에서 예술에 관하여 인터뷰한 내용을 전사: 2684
- 3) 환경문제에 관한 좌담을 잡지에 수록한 것: 2263
- 4) 방송 드라마 대본 (비출판): 2437
- 5) 즉석 연설을 녹음하여 전사한 것: 2807
- 6) 책으로 출판된 희곡: 2336

7. 헤더

7.1. 일반 규정

- 1) 헤더는 텍스트에 대한 자세한 정보를 포함하므로, 말뭉치를 구축할 때에 서지 정보 뿐만 아니라 텍스트 내용에 대한 정보도 충실히 기록해 두어야 한다.
- 2) 헤더는 TEI P3의 관련 태그를 이용한다.
- 3) 헤더는 개념적으로 다음과 같은 정보들을 제공한다. 각 정보에 대한 구체적인 예를 이 절 말미에 제시한다.

- ① 컴퓨터 파일에 관한 전자 서지 사항(제목, 배포기관)과 원전에 대한 서지사항(제목, 저자, 출판사, 판 등)

<fileDesc>

<titleStmt>	저작 제목과 저자 정보
<title>	제목
<author>	저자
<sponsor>	스폰서
<respStmt>	지적 내용 책임 정보

<resp>	지적 책임 성격
<name>	고유 명사
<extent>	크기
<publicationStmnt>	발행, 배포 정보
<distributor>	배포 책임 개인/기관
<idno>	파일 이름
<availability>	사용/배포 제약, 저작권
<notesStmnt>	부가 정보
<note>	주석
<sourceDesc>	원 텍스트의 서지 정보
<bibl>	원전 서지 사항
<pubPlace>	출판지
<publisher>	출판사
<date>	출판일

② 전자화와 관련된 방법 명시

<encodingDesc>	
<ProjectDesc>	전자 텍스트 구축 목적
<samplingDecl>	텍스트 선택 기준과 방법
<editorialDecl>	편집상의 원리와 실행 원칙

③ 텍스트의 유형에 대한 정보 (구어 전사의 경우, 담화 상황에 대한 정보 포함)

<profileDesc>	
<creation>	텍스트 생산 정보
<langUsage>	텍스트에 나타나는 언어
<textClass>	표준분류체계에 의한 분류 정보
<catRef>	분류체계에서의 범주

④ 전자 텍스트를 입력하고 수정하는 과정에 대한 정보

<revisionDesc>	
<change>	텍스트 수정 정보
<date>	수정 날짜
<respStmnt>	수정자 책임 정보
<resp>	수정자 책임 성격
<name>	수정자
<item>	각 항목

7.2. 구어 전사 자료 헤더

- 1) 구어 헤더에는 전사자 정보가 포함되어야 한다. <revisionDesc>에 기록한다.
- 2) 구어 헤더에는 대화 참여자와 대화 장면에 대한 정보가 명시적으로 제공되어야 한다. 따라서 헤더 내의 <profileDesc>은 다음의 두 가지 요소를 포함해야 한다.

- ① <particDesc> : 대화 참여자에 대한 정보. <person> 으로 대화 참여자에 대해 기술하거나, <p>로 산문적으로 기술할 수 있다.
 - <person> : 대화참여자. sex(m, f, u 중 택일), age 속성을 사용할 수 있지만 꼭 필요하지는 않다.(5.1 참조)
 - 속성 ID를 지정하여 전사 본문 내에 <u who=xx>의 xx에 해당하는 값으로 사용한다. (5.2 참조)
- ② <settingDesc> : 대화 장면에 대한 정보. <p>를 사용하여 산문적으로 표현 가능하다.

【 예 】

```

<particDesc>
  <person id='A' sex='m' age='30대'>
    <p>30대 남자, 고등교육 받음, 교수.</p>
  </person>
  <person id='B' sex='m' age='20대'>
    <p>20대 남자, 대학생</p>
  </person>
  ...
</particDesc>

<settingDesc>
  <p>1994년 여름, 교수 연구실에서 상담.</p>
</settingDesc>

<particDesc>
  <p>20대 여자들 6명.</p>
</particDesc>

```

7.3. 헤더 부착용 형판(Template)

■ 문어 헤더

```
<!DOCTYPE tei.2 SYSTEM "c:\Wsgml\Wtd\Wtei2.dtd" [  
    <!ENTITY % TEI.corpus "INCLUDE">  
    <!ENTITY % TEI.extensions.ent SYSTEM "sejong1.ent">  
    <!ENTITY % TEI.extensions.dtd SYSTEM "sejong1.dtd">  
>
```

```
<tei.2>
```

```
<teiHeader>
```

```
<fileDesc>
```

```
<titleStmt>
```

```
<title>【자료제목】</title>
```

```
<author>【저자】</author>
```

```
<sponsor>대한민국 문화관광부</sponsor>
```

```
<respStmt><resp>【자료구축내용】</resp>
```

```
<name>【기관명】</name>
```

```
</respStmt>
```

```
</titleStmt>
```

```
<extent>【어절수】 어절</extent>
```

```
<publicationStmt>
```

```
<distributor>국립국어연구원</distributor>
```

```
<idno>【파일명】</idno>
```

```
<availability><p>배포 불가</p></availability>
```

```
</publicationStmt>
```

```
<notesStmt>
```

```
<note><p>【기타주석】</p></note>
```

```
</notesStmt>
```

```
<sourceDesc>
```

```
<bibl><author>【원전저자】</author>
```

```
<title>【원전제목】</title>
```



```

        <pubPlace>【출판지】 </pubPlace>
        <publisher>【출판사】 </publisher>
        <date>【출판일】 </date>
    </bibl>
</sourceDesc>
</fileDesc>
<encodingDesc>
    <projectDesc><p>21세기 세종계획 3차년도 말뭉치 구축</p>
    </projectDesc>
    <samplingDecl><p>【구축방법】 </p>
    </samplingDecl>
    <editorialDecl><p>21세기 세종계획 말뭉치 문헌 입력 지침에 따름</p>
    </editorialDecl>
</encodingDesc>
<profileDesc>
    <creation><date>【텍스트생성일】 </date></creation>
    <langUsage>
        <language id=KO usage=99>한국어, 표준어</language>
    </langUsage>
    <textClass>
        <catRef scheme='SJ21' target='【분류코드】'>【텍스트종류】 </catRef>
    </textClass>
</profileDesc>
<revisionDesc>
    <change><date>【전자텍스트구축및수정일자】 </date>

<respStmt><resp>【지적책임성격】 </resp><name>【구축및수정자】 </name></respStmt>
    <item>【구축및수정작업내용】 </item>
</change>
</revisionDesc>
</teiHeader>

<text> (본문)
<body>
    <p> ... </p>
</body>
</text>

```

</tei.2>

■ 구어(대화, 일부 준구어) 헤더

```
<!DOCTYPE tei.2 SYSTEM "c:\sgml\dtd\tei2.dtd" (
  <!ENTITY % TEI.corpus "INCLUDE">
  <!ENTITY % TEI.extensions.ent SYSTEM "sejong1.ent">
  <!ENTITY % TEI.extensions.dtd SYSTEM "sejong1.dtd">
]>
```

<tei.2>

<teiHeader>

<fileDesc>

<titleStmt>

<title>【자료제목】 </title>

<author>【저자】 </author>

<sponsor>대한민국 문화관광부</sponsor>

<respStmt><resp>【자료구축내용】 </resp>

<name>【기관명】 </name>

</respStmt>

</titleStmt>

<extent>【어절수】 어절</extent>

<publicationStmt>

<distributor>국립국어연구원</distributor>

<idno>【파일명】 </idno>

<availability><p>배포 불가</p></availability>

</publicationStmt>

<sourceDesc>

<p>【원전정보】 </p>

</sourceDesc>

</fileDesc>

<encodingDesc>

<projectDesc><p>21세기 세종계획 3차년도 말뭉치 구축</p></projectDesc>

<samplingDecl><p>【구축방법】 </p></samplingDecl>

<editorialDecl><p>21세기 세종계획 말뭉치 문헌 입력 지침에 따름</p></editorialDecl>

```

</encodingDesc>
<profileDesc>
  <creation><date>【텍스트생성일】</date></creation>
  <langUsage>
    <language id=KO usage=99>한국어, 표준어</language>
  </langUsage>
  <particDesc>
    <person id=P1 sex=【성별(M/F)】 age=【연령】><p>【대화참여자】</p></person>
    <person id=P2 sex= age=><p></p></person>
    <person id=P3 sex= age=><p></p></person>
  </particDesc>
  <settingDesc>
    <p>【대화장면정보】</p>
  </settingDesc>
  <textClass>
    <catRef scheme='SJ21' target='【분류코드】'>【텍스트종류】</catRef>
  </textClass>
</profileDesc>
<revisionDesc>
  <change><date>【전자텍스트구축및수정일자】</date>

  <respStmt><resp>【지적책임성격】</resp><name>【구축및수정자】</name></respStmt>
    <item>【구축및수정작업내용】</item>
  </change>
</revisionDesc>
</teiHeader>

<text>(본문)
  <u who=P1> ... </u>
</text>

</tei.2>

```

7.4. 장르별 예

■ 문어 헤더

□ 책-정보

```
<!DOCTYPE tei.2 SYSTEM "c:\Wsgml\Wtdtd\Wtei2.dtd" [
    <!ENTITY % TEI.corpus "INCLUDE">
    <!ENTITY % TEI.extensions.ent SYSTEM "sejong1.ent">
    <!ENTITY % TEI.extensions.dtd SYSTEM "sejong1.dtd">
]>
<tei.2>
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>내 머리로 생각하는 역사이야기, 전자파일</title>
        <author>유시민</author>
        <sponsor>대한민국 문화관광부</sponsor>
        <respStmt><resp>문헌 입력, 표준화, 헤더 붙임</resp>
          <name>고려대학교 민족문화연구소</name>
        </respStmt>
      </titleStmt>
      <extent>30310 어절</extent>
      <publicationStmt>
        <distributor>국립국어연구원</distributor>
        <idno>2BGXXX25.hwp</idno>
        <availability><p>배포 불가</p></availability>
      </publicationStmt>
      <notesStmt>
        <note><p>이 텍스트는 ....</p></note>
      </notesStmt>
      <sourceDesc>
        <bibl><author>유시민</author>
          <title>내머리로 생각하는 역사이야기 </title>
          <pubPlace>서울</pubPlace>
          <publisher>한샘출판사</publisher>
          <date>1994</date>
```

```

    </bibl>
  </sourceDesc>
</fileDesc>
<encodingDesc>
  <projectDesc><p>21세기 세종계획 3차년도 말뭉치 구축</p>
  </projectDesc>
  <samplingDecl><p>CTS 파일을 PC 파일로 변환. 본문 전체 입력. 주석 제외함.</p>
  </samplingDecl>
  <editorialDecl><p>21세기 세종계획 말뭉치 문헌 입력 지침에 따름</p>
  </editorialDecl>
</encodingDesc>
<profileDesc>
  <creation><date>1994</date></creation>
  <langUsage>
    <language id="KO" usage="99">한국어, 표준어</language>
  </langUsage>
  <textClass>
    <catRef scheme="SJ21" target="M1355">책: 인문, 역사</catRef>
  </textClass>
</profileDesc>
<revisionDesc>
  <change><date>1994/05</date>
    <respStmt><resp>프로젝트 책임자</resp><name>김흥규</name></respStmt>
    <item>CTS 파일에서 디스크 파일로</item>
  </change>
  <change><date>1994/06/24</date>

<respStmt><resp>연구원</resp><name>권순희</name><name>차준경</name></respStmt>
  <item>헤더 붙임, 마킹</item>
</change>
  <change><date>1994/07</date>
    <respStmt><resp></resp><name>강범모</name></respStmt>
    <item>P3 Header로 수정</item>
  </change>
  <change><date>1999/04</date>
    <respStmt><resp></resp><name>박병선</name></respStmt>
    <item>세종 21 프로젝트 마크업 및 헤더로 수정</item>
  </change>

```

</revisionDesc>

</teiHeader>

</tei.2>

□ 뉴스

```
<!DOCTYPE tei.2 SYSTEM "c:\Wsgml\Wdtd\Wtei2.dtd" [
    <!ENTITY % TEI.corpus "INCLUDE">
    <!ENTITY % TEI.extensions.ent SYSTEM "sejong1.ent">
    <!ENTITY % TEI.extensions.dtd SYSTEM "sejong1.dtd">
]>
<tei.2>
<teiHeader>
<fileDesc>
    <titleStmt>
        <title>동아일보 종합(90), 전자파일</title>
        <author>동아일보사</author>
        <sponsor>대한민국 문화관광부</sponsor>
        <respStmt><resp>전자/표준화, 헤더붙임</resp>
            <name>고려대학교 민족문화연구원</name>
        </respStmt>
    </titleStmt>
    <editionStmt>
        <edition><date>1994</date>CTS파일에서 추출하여 전자파일화</edition>
    </editionStmt>
    <extent>58734 어절</extent>
    <publicationStmt>
        <distributor>국립국어연구원</distributor>
        <idno>D90-One.hwp</idno>
        <availability><p>배포 불가</p></availability>
    </publicationStmt>
    <notesStmt>
        <note>매월 5, 15, 25일 기사를 추출</note>
    </notesStmt>
    <sourceDesc>
        <bibl>
```

```

    <author>동아일보사</author>
    <title>동아일보 종합(90)</title>
        <pubPlace>서울</pubPlace>
        <publisher>동아일보사</publisher>
        <date>1990</date>
    </bibl>
</sourceDesc>
</fileDesc>
<encodingDesc>
    <projectDesc><p>21세기 세종계획 3차년도 말뭉치 구축</p>
    </projectDesc>
    <samplingDecl><p>CTS 파일을 PC 파일로 변환</p>
    </samplingDecl>
    <editorialDecl><p>21세기 세종계획 말뭉치 문헌 입력 지침에 따름</p>
    </editorialDecl>
</encodingDesc>
<profileDesc>
    <creation><date>1990</date></creation>
    <langUsage>
        <language id="KO" usage="99">한국어, 표준어</language>
    </langUsage>
    <textClass>
        <catRef scheme="SJ21" target="M1111">신문: 종합</catRef>
    </textClass>
</profileDesc>
<revisionDesc>
    <change>
        <date>1994</date><respStmt><resp></resp><name>한국언론연구원</name></respStmt>
        <item>CTS 파일을 전자파일로 전환</item>
    </change>
    <change>
        <date>1995</date><respStmt><resp>프로젝트
책임자</resp><name>김흥규</name></respStmt>
        <item>기사를 분류하고 TEI 마킹</item>
    </change>
    <change>
        <date>1996</date><respStmt><resp>연구원</resp><name>박봉래</name></respStmt>
        <item>띄어쓰기 교정기로 띄어쓰기 부분 교정</item>

```

```

    </change>
    <change>
      <date>1996/07</date><respStmt><resp>연구원</resp><name>이동석</name></respStmt>
      <item>TEI P3 불임</item>
    </change>
    <change><date>1999/04</date>
      <respStmt><resp></resp><name>박병선</name></respStmt>
      <item>세종 21 프로젝트 마크업 및 헤더로 수정</item>
    </change>
  </revisionDesc>
</teiHeader>

</tei.2>

```

□ 소 설

```

<!DOCTYPE tei.2 SYSTEM "c:\Wsgml\Wdtd\Wtei2.dtd" [
  <!ENTITY % TEI.corpus "INCLUDE">
  <!ENTITY % TEI.extensions.ent SYSTEM "sejong1.ent">
  <!ENTITY % TEI.extensions.dtd SYSTEM "sejong1.dtd">
]>
<tei.2>
<teiHeader>
<fileDesc>
  <titleStmt>
    <title>토지 16, 전자파일</title>
    <author>박경리</author>
    <sponsor>대한민국 문화관광부</sponsor>
    <respStmt><resp>문헌 입력, 표준화, 헤더 불임</resp>
      <name>고려대학교 민족문화연구소</name>
    </respStmt>
  </titleStmt>
  <extent>67800 어절</extent>
  <publicationStmt>
    <distributor>국립국어연구원</distributor>
    <idno>toji16.hwp</idno>
  <availability><p>배포불가</p></availability>

```



```

    </publicationStmt>
  <notesStmt>
    <note><p>이 텍스트는...</p></note>
  </notesStmt>
  <sourceDesc>
    <bibl>
      <author>박경리</author>
      <title>토지16</title>
      <pubPlace>서울</pubPlace>
      <publisher>술</publisher>
      <date>1994</date>
    </bibl>
  </sourceDesc>
</fileDesc>
<encodingDesc>
  <projectDesc><p>21세기 세종계획 3차년도 말뭉치 구축</p>
</projectDesc>
  <samplingDecl><p>CTS파일에서 디스크파일로 변환</p>
</samplingDecl>
  <editorialDecl><p>21세기 세종계획 말뭉치 문헌 입력 지침에 따름</p>
</editorialDecl>
</encodingDesc>
<profileDesc>
  <creation><date>1994</date></creation>
  <langUsage>
    <language id=KO usage=99>한국어</language>
  </langUsage>
  <textClass>
    <catRef scheme='SJ21' target='M1331'>책: 상상적 텍스트, 소설/일반</catRef>
  </textClass>
</profileDesc>
<revisionDesc>
  <change>
    <date>1994</date><respStmt><resp>프로젝트 책임자</resp><name>김흥규</name></respStmt>
    <item>CTS파일을 디스크 파일로</item>
  </change>
  <change>
    <date>1994/11</date><respStmt><resp>연구원</resp><name>오명옥</name></respStmt>

```

```

<item>TEI-P3 헤더불임</item>
</change>
<change><date>1999/04</date>
    <respStmt><resp></resp><name>박병선</name></respStmt>
    <item>세종 21 프로젝트 마크업 및 헤더로 수정</item>
</change>
</revisionDesc>
</teiHeader>

```

```
</tei.2>
```

■ 국어 헤더

□ 방송 전사

```

<!DOCTYPE tei.2 SYSTEM "c:\Wsgml\Wdtd\Wtei2.dtd" [
    <!ENTITY % TEI.corpus "INCLUDE">
    <!ENTITY % TEI.extensions.ent SYSTEM "sejong1.ent">
    <!ENTITY % TEI.extensions.dtd SYSTEM "sejong1.dtd">
]>
<tei.2>
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>김한길과 사람들 95/1/7, 전자전사자료</title>
        <author>김한길, 대화자</author>
        <sponsor>대한민국 문화관광부</sponsor>
        <respStmt><resp>전자/표준화</resp>
          <name>고려대 민족문화연구소</name>
        </respStmt>
      </titleStmt>
      <extent>25000 어절</extent>
      <publicationStmt>
        <distributor>국립국어연구원</distributor>
        <idno>GADK0124.hwp</idno>
        <availability><p>배포 불가</p></availability>
      </publicationStmt>
      <sourceDesc>
        <p>원전 없음, MBC "김한길과 사람들 (1995/1/7 일요일) 방송녹화하여 전사함.</p>
      </sourceDesc>

```

```

</fileDesc>
<encodingDesc>
  <projectDesc><p>21세기 세종계획 3차년도 말뭉치 구축</p></projectDesc>
  <samplingDecl><p>방송 녹화하여 전사</p></samplingDecl>
  <editorialDecl><p>21세기 세종계획 말뭉치 문헌 입력 지침에 따름</p></editorialDecl>
</encodingDesc>
<profileDesc>
  <creation><date>1995</date></creation>
  <langUsage>
    <language id=KO usage=99>한국어, 표준어</language>
  </langUsage>
  <particDesc>
    <person id=P1 sex=M age=50s><p>김한길, 50대 남자 사회자</p></person>
    <person id=P2 sex=F age=27><p>채시라, 27세 여자 텔런트</p></person>
    <person id=P3 sex=M age=20s><p>홍명보, 20대 남자 축구선수</p></person>
  </particDesc>
  <settingDesc>
    <p>1995년 1월 7일 일요일에 MBC에서 생방송으로 방영한 김한길과 사람들을 녹화함.
    전반부는 김한길과 채시라, 후반부는 김한길과 홍명보의 대화임.</p>
  </settingDesc>
  <textClass>
    <catRef scheme='SJ21' target='M2693'>국어 TV 녹화 전사</catRef>
  </textClass>
</profileDesc>
<revisionDesc>
  <date>1995</date><respStmt><resp>프로젝트 책임자</resp><name>김흥규</name></respStmt>
<change>
  <change><date>1995-2</date><respStmt><resp>입력자</resp>
    <name>고려대 학부생</name></respStmt>
    <item>녹화 및 전사</item>
  </change>

<change><date>1995-4-15</date><respStmt><resp></resp><name>강범모</name></respStmt>
  <item>헤더붙임</item>
</change>
<change><date>1999/04</date>
  <respStmt><resp></resp><name>박병선</name></respStmt>
  <item>세종 21 프로젝트 마크업 및 헤더로 수정</item>
</change>

```

</revisionDesc>

</teiHeader>

</tei.2>

□ 희곡

```
<!DOCTYPE tei.2 SYSTEM "c:\Wsgml\Wtd\Wtei2.dtd" [  
    <!ENTITY % TEI.corpus "INCLUDE">  
    <!ENTITY % TEI.extensions.ent SYSTEM "sejong1.ent">  
    <!ENTITY % TEI.extensions.dtd SYSTEM "sejong1.dtd">  
>  
<tei.2>
```

<teiHeader>

<fileDesc>

<titleStmt>

<title>까치교의 우화 (연극대본), 전자판</title>

<author>김상열</author>

<sponsor>대한민국 문화관광부</sponsor>

<respStmt><resp>전자/표준화</resp>

<name>고려대 민족문화연구소</name>

</respStmt>

</titleStmt>

<extent>8772 어절</extent>

<publicationStmt>

<distributor>국립국어연구원</distributor>

<idno>Ccachi.hwp</idno>

<availability><p>배포 불가</p></availability>

</publicationStmt>

<notesStmt>

<note></note>

</notesStmt>

<sourceDesc>

<p>김상열, 까치교의 우화</p>

```

</sourceDesc>
</fileDesc>
<encodingDesc>
  <projectDesc><p>21세기 세종계획 3차년도 말뭉치 구축</p></projectDesc>
  <samplingDecl><p>연극 대본을 입력함</p></samplingDecl>
  <editorialDecl><p>21세기 세종계획 말뭉치 문헌 입력 지침에 따름</p></editorialDecl>
</encodingDesc>
<profileDesc>
  <creation><date>1995</date></creation>
  <langUsage>
    <language id=KO usage=99>한국어, 표준어</language>
  </langUsage>
  <textClass>
    <catRef scheme='SJ21' target='M2506'>기타 비출판물: 총류, 연극 (희곡)</catRef>
  </textClass>
</profileDesc>
<revisionDesc>
  <change><date>1994</date><respStmt><resp></resp><name>문예진흥원</name></respStmt>
    <item>전산입력</item>
  </change>

  <change><date>1995-4-21</date><respStmt><resp>연구원</resp><name>오명옥</name></respSt
mt>
    <item>헤더붙임</item>
  </change>
  <change><date>1999/04</date>
    <respStmt><resp></resp><name>박병선</name></respStmt>
    <item>세종 21 프로젝트 마크업 및 헤더로 수정</item>
  </change>
</revisionDesc>
</teiHeader>

</tei.2>

```

8. 본문 태그

- 1) 본문은 <body>를 반드시 포함해야 하고, <front>, <back>을 임의로 포함할 수 있다. <front>는 앞표지 부분, 목차 등, 그리고 <back>은 부록, 색인, 뒷표지 등의 내

용이므로, 역사문헌 자료 등 특별한 목적이 있지 않는 한 입력하지 않는 것을 원칙으로 한다.

- 2) 합성적 텍스트의 경우 여러 개의 집단으로 구성될 수 있으며, 각 집단은 다시 여러 개의 집단 혹은 텍스트로 구성될 수 있다. 이를테면, 단편 소설집의 개별 작품을 각각 텍스트로 보고 이들을 한 집단으로 묶을 수 있는데, <group> 태그를 이용한다.

일반적 문헌의 본문 형태는 다음과 같다.

① 일반적 형태³⁾

```
<text>
  <body>
    본문 내용...
  </body>
</text>
```

② <front>, <back>을 포함한 형태

```
<text>
  <front>
    앞표지, 목차 등의 머리말 부분
  </front>
  <body>
    본문 내용...
  </body>
  <back>
    부록, 색인, 뒷표지 등의 꼬리말 부분
  </back>
</text>
```

- ③ 합성적 텍스트의 예. <group>으로 표기하며, 각 텍스트와 전체 합성적 텍스트 모두 <front>, <back>을 포함한 형태이다.

```
<text>                                     합성적 텍스트
  <front>
  </front>
  <group>
    <text>                                   첫 번째 텍스트
```

3) Indentation은 설명의 시각적 효과를 높이기 위해서 한 것이며, 실제 작업시는 하지 않는다.

```

                                <front>
                                </front>
                                <body>
                                ...
                                </body>
                                <back>
                                </back>
        </text>
        <text>                두 번째 텍스트
                                <front>
                                </front>
                                <body>
                                ...
                                </body>
                                <back>
                                </back>
    </text>
    ...
    <back>
    </back>
    </group>
</text>

```

3) 문헌 입력 텍스트에 다음 태그는 반드시 표시했다.

<head> : 제목, 부제목, 표 제목 등 모든 종류의 제목

<p> : 산문에서의 문단

<I>, <lg> : 운문에서의 시행과 시행군(연 등), 시행은 산문 내에 포함되어 나타날 수도 있다.

4) <head>는 <p> 속에 나타날 수 없고, <p>는 <head> 속에 나타날 수 없다. 보통의 산문 텍스트 본문은 다음과 같은 모양이다.

```

<text>
<boby>
<head>.....</head>
<p>....
</p>
<p>...
</p>

```

```
...
<head>.....</head>
<p>...
</p>
...
</boby>
</text>
```

- 5) 운문의 <l>은 반드시 <lg>에 포함되어야 하며, <lg>는 <poem>에 포함되어야 한다. 그리고 <p>가 <lg>를 포함할 수 없다. 시의 제목은 <lg> 속에 포함될 수도 있고 <lg> 밖에 즉, <poem>에 포함될 수도 있다.

【예】

```
<poem>
<lg>
<l>쫓아오든 햇빛인데</l>
<l>지금 교회당 꼭대기</l>
<l>십자가에 걸리었습니다.</l>
</lg>
...
</poem>
```

- 6) 희곡 등 공연 텍스트에는 다음과 같은 태그들이 필요하다.

<sp> 대사 부분, 대사자의 이름과 지문을 포함할 수 있다. 대사 자체는 <p>로 표시된다.

<speaker> 대사자의 이름 부분, 반드시 <sp> 내에 있어야 한다.

<stage> 무대 지시 등 지문. <sp> 내 혹은 밖에(즉, <body> 안에) 있을 수 있다.

【예】

```
<sp><speaker>노라</speaker><stage>고개를 흔든다</stage>
<p>당신들은 결코 저를 사랑하지 않았어요. 다만 저를 재미로 귀여워했을 뿐이에요.</p>
</sp>
```

- 7) 문헌 전사 중에 판독 불가능하여 입력할 수 없는 경우, 혹은 그림 부분 등 생략 사실을 명시적으로 밝혀야 할 경우, <gap> 태그를 사용하고 reason 속성으로 그 이유를 밝힌다.

【예】

```
<gap reason='판독 불가'>
<gap reason='그림 부분'>
```

- 8) 전사 구어의 경우, 화자를 다음과 같은 형식으로 밝혀 주어야 한다. 여기서 xx는 헤더에서 정의된 대화자의 ID이다. 전사 지침은 앞에 제시한 바와 같다.

```
<u who=xx>
```

구어 전사 자료의 본문 부분의 일반적 모양은 다음과 같다

```
<text>
<u who=xx>.....</u>
<u who=yy>.....</u>
<u who=xx>.....</u>
<u who=yy>.....</u>
.....
</text>
```

- 9) 문헌 자료의 경우 다음 태그들을 주의적으로 사용할 수 있다.

<q> : 본문 내의 인용 부분을 표시한다.

<pb>: 원전의 페이지 구분이 중요한 문헌에서의 구분 위치 표시. 페이지숫자를 표기할 필요가 있을 때에는 num 속성을 이용한다.

【예】 <pb n='12'>

<lb>: 운문에서 시행의 인쇄상의 구분 위치가 중요할 경우, 줄바꿈 위치를 표시했다.

<note>: 각주 등 주석 내용을 입력할 경우 사용했다.

<div>: 장, 절 등 본문의 구획을 명시적으로 표시하려고 할 때에 사용했다.

예)

■ 본문

□ 문어 본문 예

```

<!DOCTYPE tei.2 SYSTEM "c:\wsgml\Wtd\Wtei2.dtd" [
    <ENTITY % TEI.corpus "INCLUDE">
    <ENTITY % TEI.extensions.ent SYSTEM "sejong1.ent">
    <ENTITY % TEI.extensions.dtd SYSTEM "sejong1.dtd">
]>
<tei.2>

<teiHeader>
...생략...
</teiHeader>

<text>

<group>

<text>
<body>
<head>(1) 설명의 의의</head>
<p>우리는 일상 생활에서 다음과 같은 질문과 접하는 경우가 있다.</p>
<p>하늘은 왜 파랗게 보일까?</p>
<p>표준어를 어떻게 정하지?</p>
<p>청람 여고가 설립된 역사적 배경에 관해 말씀해 주시겠습니까?</p>
<p>환율을 인상하면 수입한 원료를 가공, 수출하는 회사는 어떻게 될까요?</p>
<p>이들 각 질문은, 묻고자 하는 바를 바꾸지 않고, 다음과 같이 ○○는 무엇인가? 의 형태로 바꾸어 말할 수 있다.</p>
<p>하늘이 파랗게 보이는 이유는 무엇일까?</p>
<p>표준어 사정의 원칙은 무엇이지?</p>
<p>청람 여고 설립의 역사적 배경은 무엇입니까?</p>
<p>환율 인상이 가공 무역 회사에 미치는 영향은 무엇입니까?</p>
<p>위의 질문들은 공통적으로 표준어 사정의 원칙은, 교양있는 사람들이 두루 쓰는 현대 서울말을 표준어로 정한다는 것이다. 와 같이, ○○는 ○○○이다. 식의 대답을 기대한다. 다시 말해, 설명을 해 달라는 요청이다.</p>
<p>설명이란, 위의 질문에 답하는 경우와 같이, 듣는 이가 잘 모르고 있는 사실이나 사물, 현상, 사건에 대하여 알 수 있도록 풀어서 말하는 것이다. 단어의 의미, 용어의 정의, 과학적 원리, 사건의 경위, 사회적·경제적 현상의 원인과 결과에 관한 설명 등으로부터 사물이나 경치 혹은 인물에 관한 묘사, 자기 자신의 견해나 입장에 대한 해명 등에 이르기까지 모두 설명의 범위 안에 든다. 또, 저가 공 차는 애들 중에 머리가 짧은 애가 철웅이고, 안경을 쓴 애가 재혁이야. 와 같은 지정도 설명에 속한다.</p>
</body>
</text>

```

```

<text>
<body>
<head>3</head>
<p>개화기를 맞아 전통극이 급속히 쇠퇴하는 과정에서 대중적인 판소리가 창극이라는 또 다른 음악극
양식을 파생시켰지만, 일본 세력과 함께 들어온 신파극으로 해서 변두리로 밀려났다. 일본 정서를
담고 있는 감상적 신파극도 3·1 운동 이후에는 지식 청년들에 의해 수용된 서구 근대극 형태에 또
다시 밀려났다. 그런데 전통극은 자연 발생적인 형태이기 때문에 구비 문학적 성격을 띠므로 작가가
따로 없었으나, 신극 이후에는 극작가들이 등장하여 희곡을 발표하기 시작했다. 1912년 조중환이
병자 삼인(病者三人)이라는 최초의 희곡을 쓴 이래, 수많은 극작가들이 등장하여 오늘날까지 수천
편의 작품을 발표했다. 신극사상 대표적인 극작가로서 박승희, 김우진, 유치진, 함세덕, 김진수,
오영진, 차범석, 윤대성, 윤조병, 오택석 등을 꼽을 수 있고, 이들 외에도 우수한 희곡을 발표한
극작가도 적지 않다. 이들은 모두가 사실주의를 기초로 하여 식민지 시대와 그 이후의 암울했던
현실을 진솔하게 표현한 것이 특징이다. 가령, 유치진의 토막(土幕) 이라든지, 함세덕의 산허구리,
오영진의 맹 진사 댁 경사, 차범석의 산불, 윤조병의 휘파람새, 오택석의 산수유 등이 그러한 예가 될
수 있을 것이다.</p>
</body>
</text>

</group>

</text>

</tei.2>

```

□ 준구어 본문 예(희곡)

```

<!DOCTYPE tei.2 SYSTEM "c:\Wsgml\Wdtd\Wtei2.dtd" [
    <ENTITY % TEI.corpus "INCLUDE">
    <ENTITY % TEI.extensions.ent SYSTEM "sejong1.ent">
    <ENTITY % TEI.extensions.dtd SYSTEM "sejong1.dtd">
]>
<tei.2>
<teiHeader>
...생략...
</teiHeader>
<text>
<front>
<head>제목: [신춘문에 당선 희곡] 꿈 잠 꿈꿈 잠....</head>

```

```

<head>주유미</head>
<castlist>
<head>등장인물</head>
<castitem><role>A (30대 초반쯤의 젊은이)</role></castitem>
<castitem><role>B (50대 중반쯤의 신사)</role></castitem>
</castlist>
<head>-꿈-</head>
<set>
<p>무대는 칠혹 같은 어둠. 포레의 음악 [꿈꾸고 난 후에]가 배경음악 (바브라 스트라이샌드의 노래)으로 깔린다. 음악이 끝나갈 때쯤 불륨이 서서히 잦아지면서 두 인물 A와 B에만 탑조명이 들어온다. A는 세미캐주얼 차림에 바바리를 입고있고 좀 피곤하고 초췌해 보인다. B는 그보다는 좀 표정에 여유가 있어 보이는 중년의 신사. 언밸런스한 차림이지만 뭔가 독특한 자기 개성이 드러나는 의상이 자유로운 분위기를 연출하고 있다. 두 사람은 가까이 앉아 있지만 둘 다 객석 쪽을 바라보고 있다. 조명이 들어옴과 동시에 A의 대사가 시작된다.</p>
</set>
</front>
<body>
<sp><speaker>A</speaker><stage>(꿈에서 덜 깬 듯 몽롱한 표정으로)</stage><p>참, 이상한 일이오.</p></sp>
<sp><speaker>B</speaker><p>뭐가요?</p></sp>
<sp><speaker>A</speaker><p>어젯밤 꿈.</p></sp>
<sp><speaker>B</speaker><p>한번 얘기해보세요.</p></sp>
<sp><speaker>A</speaker><p>난 긴 배를 노 저어 가고 있었어요. 아라비안나이트에 나오는 왕자가 신는 신발처럼 코끝이 구부러져 올라간 멋진 배였어요.</p></sp>
<sp><speaker>B</speaker><stage>(여전히 시큰둥하게, 호기심을 감추고서)</stage><p>어딜 가고 있었는데요?</p></sp>
<sp><speaker>A</speaker><p>그건 나도 몰라요. 그저 노를 저어 앞으로 앞으로만 가고 있었으니까요. 노를 젓는 느낌이 어찌나 부드러운지 하나도 힘이 들지 않았어요. 깃털처럼 부드럽고 살같이 곧게 내 배는 가고 있었어요.</p></sp>
<sp><speaker>B</speaker><p>파도가 없었나 보죠.</p></sp>
<sp><speaker>A</speaker><p>파도라고요?</p><stage>(황당하다는 듯 웃는다)</stage><p>그런 게 있을 리가 있나요.</p></sp>
<sp><speaker>B</speaker><p>그렇게 잔잔한 바다였나요?</p></sp>
<sp><speaker>A</speaker><p>바다?</p></sp>
</body>
</text>
</tei.2>

```

9. 태그 목록

9.1. 헤더

1) 공통적인 필수 태그

① <fileDesc>	컴퓨터 파일에 관한 전자 서지 사항(제목, 배포기관)과 원전에 대한 서지사항(제목, 저자, 출판사, 판 등)
<titleStmt>	저작 제목과 저자 정보
<title>	제목
<author>	저자
<sponsor>	스폰서
<respStmt>	지적 내용 책임 정보
<resp>	지적 책임 성격
<name>	고유 명사
<extent>	크기
<publicationStmt>	발행, 배포 정보
<distributor>	배포 책임 개인/기관
<idno>	파일 이름
<availability>	사용/배포 제약, 저작권
<sourceDesc>	원 텍스트의 서지 정보
<bibl>	원전 서지 사항
<pubPlace>	출판지
<publisher>	출판사
<date>	출판일
② <encodingDesc>	전자화와 관련된 방법 명시
<ProjectDesc>	전자 텍스트 구축 목적
<samplingDecl>	텍스트 선택 기준과 방법
<editorialDecl>	편집상의 원리와 실행 원칙
③ <profileDesc>	텍스트의 유형에 대한 정보 (구어 전사의 경우, 담화 상황에 대한 정보 포함)
<creation>	텍스트 생산 정보
<langUsage>	텍스트에 나타나는 언어
<textClass>	표준분류체계에 의한 분류 정보
<catRef>	분류체계에서의 범주
④ <revisionDesc>	전자 텍스트를 입력하고 수정하는 과정에 대한 정보
<change>	텍스트 수정 정보

<date>	날짜
<name>	수정자
<item>	각 항목

2) 텍스트 종류에 따른 필수 태그

- 구어

<particDesc>	대화 참여자에 대한 정보
<person>	대화 참여자
<settingDesc>	대화 장면에 대한 정보

3) 수의적 태그

<notesStmt>	부가 정보
<note>	주석

9.2. 본문

1) 공통적인 필수 태그

<text>	텍스트
<body>	본문
<head>	제목, 부제목, 표 제목 등 모든 종류의 제목
<p>	산문에서의 문단

2) 텍스트 종류에 따른 필수 태그

- 합성적 텍스트

<group>	텍스트들의 집단을 표시
---------	--------------

- 운문

<poem>	운문
<l>	운문에서의 시행
<lg>	운문에서의 시행군(연 등)
◦ 희곡	
<sp>	희곡의 대사 부분
<speaker>	대사자의 이름 부분
<stage>	무대 지시 등 지문
◦ 전사 구어	
<u who=xx>	대화. 화자를 밝혀 주며, xx는 헤더에서 정의된 대화자의 ID

3) 수의적 태그

<front>	앞표지 부분, 목차 등 표시
<back>	부록, 색인, 뒷표지 등 표시
<q>	본문 내의 인용 부분 표시
<pb>	원전의 페이지 구분이 중요한 문헌에서의 구분 위치 표시
<lb>	운문에서 시행의 인쇄상의 구분 위치가 중요할 경우, 줄바꿈 위치 표시
<note>	각주 등 주석 내용 표시
<div>	장, 절 등 본문의 구획을 명시적으로 표시

<pause> : 한동안(보통 3초 이상) 말이 없는 경우 표시.

<vocal desc= > : 웃음, 기침, 하품, 재채기 등을 기술.

<writing>, </writing> : 라디오나 tv 방송 강의 전사 자료에 많이 나오는 자막이나 제목을 표시할 때 사용.

<event desc= > : 전화벨 소리라든지 기타 음성 아닌 소리를 표시.

<note>, </note> : 전사자가 특별한 주석을 달 필요가 있을 때 사용.

<scnum></scnum> : scene number 표시

<actnum></actnum> : 대본의 장 표시

<castlist></castlist> : 등장인물 표시

<castitem></castitem> : 등장인물 소개

<actor></actor> : 배우 소개

<role></role> : 극중 역할 표시

10. 준구어 구축 작업 세부 지침

10.1 수정, 보완

■ 말줄임표의 문제.

1. 말줄임표는 텍스트의 성격상 말을 하지 않는다는 의미와 아무런 행동을 하지 않는다는 두 가지 의미로 해석될 수 있는데 이의 구분이 쉽지 않다.

▶ 결정사항 : '...'은 두 가지로 구분된다.

첫째, 행동이나 무대를 나타내는 말줄임표의 경우 - 보통 stage 안에 나타낸다.

둘째, 대사를 나타내는 말줄임표의 경우 - 이것은 말을 하지 않는다는 뜻이므로 일종의 발화행위로 보아 p 안에 나타낸다.

2. 말줄임표가 길게 나타나는 경우가 있는데 어디까지 표시할 지 결정하기 어렵다.

결정사항) 말줄임표(...)는 전각기호(3406) ... (한번) 또는 (두번)로 통일한다. 둘의 구분 기준은 말줄임표가 5개 이하이면 ...로, 5개 이상이면 (5개일 때는 작업자의 편의로 결정할 수 있다.)

■ 줄표의 문제

텍스트의 성격상 줄표(-)가 많이 나오는데 이들의 성격이 같지 않아 이에 대한 규정이 필요하였다.

▶ 결정사항 : 짧은 줄표(-)는 다음과 같은 의미가 있다.

- 말줄임표 대신일 때 : 말줄임표의 기준을 적용한다.
- 소리를 길게 끄는 것을 나타내는 경우
- 단음절의 소리가 길어지는 경우 : 긴 줄표(340a)로 표시한다.

예) 씹-어 → 씹—어

- 동일한 음절의 반복을 나타내는 경우
: 긴 줄표 하나(340a)로 표시하거나 말줄임표(3406)로 표시한다.

예) 하하하--- → 하하하—/ 하하하...

- 휴지(pause)를 표시하는 경우 : 긴 줄표로 표시한다.

예) 가만있자- → 가만있자—

- 새로운 항목을 표시하거나 '에서-까지'를 표시하는 경우 : 긴 줄표로 표시한다.

예) - 지나가 앉았다가 일어난다.

- 지나, 뛰어나간다.
- — 지나가 앉았다가 일어난다.
- 지나, 뛰어나간다.

예) 3.17-3.19 → 3.17—3.19

■ 자막의 문제

자막은 대사도, 지문도 아닌 성격을 가지고 있어 처리하기가 곤란하였다.

- ▶ 결정사항 : 자막은 <stage>로 처리한다.

■ 노래의 문제

노래는 운율을 가진 시적 어구 형태로 나타나지만, 등장인물의 대사라는 점에서 처리가 곤란하였다.

- ▶ 결정사항 : 노래는 다음과 같이 처리한다.

예) 아이들 노래 부른다

노래 : 백두산 뺨어내려 반도 삼천리.

→ <stage>아이들 노래 부른다</stage>

<sp><speaker>?</speaker>

<head>노래</head><l>백두산 뺨어내려 반도 삼천리</l></sp>

■ 불분명한 발화자의 표시 문제

발화자가 불분명할 경우 처리 기준이 없어 곤란하였다.

- ▶ 결정사항 : 발화자 표시는

(소리) => <speaker>?</speaker><stage>(소리)</stage><p>.....</p>

[석규의 소리]..... => <speaker>[석규의 소리]</speaker><p>.....</p>

로 하였다. 즉, 발화자가 분명하게 표시되지 않은 경우는 발화자 자리에 ? 표시를 하기로, 그리고 분명하게 표시된 경우는 발화자 태그를 붙이기로 하였다.

■ ‘끝, 막’ 등의 표시 문제

텍스트의 내용과 직접 상관이 없는 표시인, ‘끝, 막’ 등은 <head>로 처리하였다.

10.2. 마킹 예

【태그가 붙은 텍스트의 예】

<드라마극본>

<text>

<body>

<head> “아의정 납치소동” (9/7)</head>

<scnum>#1. 의상실 (ENG)</scnum>

<stage>의정, 대걸레질하는데 철수가 들어오자</stage>

<sp><speaker>의정</speaker><p>어서오세요</p><stage>(하다가)</stage><p>어머! 철수야! 여기 웬일이야!</p></sp>

<sp><speaker>철수</speaker><p>볼썽 찾아와서 놀랬지? 잠깐만!</p><stage>(돌아보며 손짓)</stage></sp>

<stage>입구에서 정은 들어오며 인사하고</stage>

<sp><speaker>철수</speaker><stage>(정은 손잡으며)</stage><p>내 여자친구! 신고하러 왔어!</p></sp>

<sp><speaker>의정</speaker><p>어머! 정말? 반가와요!</p></sp>

<sp><speaker>정은</speaker><p>의정씨 얘기 많이 들었어요! 전 최정은이에요!</p></sp>

<sp><speaker>철수</speaker><p>퇴근할 때 됐지? 같이 나가서 저녁 먹자!</p></sp>

<sp><speaker>의정</speaker><p>잠깐만! 금방 밀고 올게!</p><stage>(대걸레 들고 가려는데)</stage></sp>

<stage>E 전화벨소리</stage>

<sp><speaker>의정</speaker><stage>(전화 받으며)</stage><p>의상실입니다! 어! 현?</p></sp>

- 중간 생략 -

<회곡>

<text>

<front>

<head>실험극장 소극장 공연작품</head>

<head>윤심덕의 사(死)의 찬미(讚美)</head>

<p>윤대성/작</p>

<p>실험극장 제 회 공연</p>

<castlist>

<head>등장인물</head>

<castitem><role>윤심덕</role></castitem>

<castitem><role>김우진</role></castitem>

<castitem><role>홍난파</role></castitem>
 <castitem><role>요시다경부</role></castitem>
 <castitem><role>후미꼬</role></castitem>
 <castitem><role>조명희</role></castitem>
 <castitem><role>임석경관 (극중 배우와 중역)</role></castitem>
 <castitem><role>김우진의 부인 (후미꼬와 중역)</role></castitem>
 </castlist>
 <set><p>(막 오르기 전이다. 《다뉴브강 첼로 또는 다른 방식의 연주소리가 들린다》</p>
 <p>《0000》 〈 〉 </p>
 <p>《자막 당시 윤심덕의 사의찬미 가사 중 일부》</p>
 <lg><l>“광막한 황야에 달리는 인생아</l>
 <l>너희 가는 곳 그 어디이냐?</l>
 <l>쓸쓸 세상 험악한 고해에</l>
 <l>너는 무엇을 찾으려 가느냐?</l></lg>
 <p>《0000000000</p>
 <p>000000000</p>
 <p>000000000</p>
 <p>00000000” 〉 〈 〉 </p>
 <p>음악이 똑 그치면서 소란스러운 종이소리, 발자국 소리, 파도소리, 커튼 앞에 프로젝터 불빛이 비추며 다음과 같은 자막이 나타난다.</p>
 <p>“현해탄 격방 중에 청춘남녀의 정사</p>
 <p>-극작가와 음악가가 한떨기 꽃이 되어 세상시비 던져두고 끝없는 물나라로-” </p>
 <p>동아일보 1926년 8월5일</p>
 <p>“미성의 주인 윤심덕 양 청년문사 김우진 투신 정사” </p>
 <p>조선일보 1926년 8월 5일</p>
 <p>자막이 지워지면서 막이 오른다. 홍난파 혼자 침울한 표정으로 피아노 앞에 앉아 작곡에 열중하고 있다. 곡목은 “옛 동산에 올라” 멜로디 일부 밖에서 노크소리 들린다. 홍난파 그것도 듣지 못한 채 음악에 열중하고 있다.</p>
 <p>“계십니까? 아무도 없어요” 하는 목소리와 함께 요시다 경부 들어온다. 홍난파 그제사 놀라 본다.)</p>
 </set>
 </front>
 <body>
 <sp><speaker>[홍]</speaker><p>누구요?</p></sp>
 <sp><speaker>[요시다]</speaker><p>홍난파 선생?</p></sp>
 <sp><speaker>[홍]</speaker><p>에 그렇습니다.</p></sp>
 <sp><speaker>[요시다]</speaker><p>일하시는데 방해가 돼서 죄송합니다.</p></sp>
 <sp><speaker>[홍]</speaker><p>누구십니까?</p></sp>
 <sp><speaker>[요시다]</speaker><p>경성 경찰서에 요시다 경부요.</p></sp>
 <sp><speaker>[홍]</speaker><p>무슨 일로?</p></sp>

<sp><speaker>[요시다]</speaker><p>몇 가지 알고 싶은 게 있어 왔습니다.</p><stage>(악보를 본다.)</stage><p>작곡중이시군요</p><stage>(속으로 멜로디를 읊어본다.)</stage><p>아주 정서적인 멜로디군요.</p></sp>

<sp><speaker>[홍]</speaker><stage>(악보를 뺏는다.)</stage><p>아직 완성한 것 아니오.</p></sp>

<sp><speaker>[요시다]</speaker><p>선생의 봉선화는 아주 훌륭한 가곡이죠. 나도 음악에 대해선 좀 압니다. 중학 때 밴드부에서 트럼펫을 불었지요.</p></sp>

<sp><speaker>[홍]</speaker><p>제게 뭘 알고 싶어 오신걸니까?</p></sp>

<sp><speaker>[요시다]</speaker><p>김우진과 친구시라고요.</p></sp>

- 중간 생략 -

<시나리오>

<text>

<front>

<head>접속</head>

<p>STAFF</p>

<p>감독: 장윤현</p>

<p>각본: 조명주/ 장윤현</p>

<p>촬영: 김성복</p>

<p>조명: 임재영</p>

<p>미술: 신보경</p>

<p>편집: 박곡지</p>

<castlist>

<head>CAST</head>

<castitem><role>동현 역:</role><actor>한석규</actor></castitem>

<castitem><role>수현 역:</role><actor>전도연</actor></castitem>

<castitem><role>은희 역:</role><actor>추상미</actor></castitem>

<castitem><role>태호 역:</role><actor>박용수</actor></castitem>

<castitem><role>기철 역:</role><actor>김태우</actor></castitem>

<castitem><role>희진 역:</role><actor>강민아</actor></castitem>

</castlist>

</front>

<body>

<scnum>S# 1 - 영화관 전경 (밤)</scnum>

<stage>평일 밤 한산한 극장 앞. 비가 내리고 있다. 극장 문을 조용히 열고 나오는 수현. 비를 보고 난감한 표정을 짓는다. 영화가 끝났는지 쏟아져 나오는 관객들. 사람들을 피해 옆으로 비켜서는 수현. 우산을 꺼내 쓰고 가는 사람들. 여자에게 웃웃을 베풀어 씩씩주는 남자. 막 담배에 불을 붙이는 동현이 나타난다. 두 사람 출입문에 나란히 서지만 한 번도 눈이 마주치지 않는다. 그대로 담배를 던지며

수현을 지나 거리로 나가는 동현. 빗줄기를 보던 수현도 결심한 듯 핸드백을 머리에 올리고 뛰기 시작한다. 두 사람이 빠져나간 공간으로 빗줄기가 뿌옇게 더욱 짙어진다. 암전 위에 MAIN TITLE.</stage>

<scnum>S# 2 - 타이틀 몽타주 (밤)</scnum>

<stage><p>(스튜디오) 컴퓨터 모니터 화면에 접속되었다는 안내문이 뜬다. 음악이 시작된다. ON AIR에 불이 들어와 있고 릴 테잎을 거는 동현의 손. 프린트 되어 나오는 신청 메세지들. 돌아가는 CD플레이어. 음향 계측판의 불빛들.</p>

<p>(거리) 비 오는 도심의 심야. 한산한 거리와 물을 가르며 달리는 차들. 광고들이 지나가는 전광판들.</p>

<p>(스튜디오) 원고를 읽어보며 표시를 하는 진행자. 프린터에서 막 뿜은 신청 메세지들을 읽으면서 부스 안으로 들어가는 은희. 엔지니어에게 뷰스티를 설명하고 있는 동현. PC에 올라오는 신청 메세지들.</p>

<p>(도시 풍경) 동현의 프로가 비 오는 도심의 이곳 저곳에서 라디오를 통해 들려온다. 편의점, 주유소, 신문보급소, 정차해 있는 택시 안.</p>

<p>(스튜디오) 부스 안에서 신청 메세지들과 큐 시트를 뒤적이며 보던 진행자.</p></stage>

<sp><speaker>진행자</speaker><stage>(토크벨로)</stage><p>비도 오는데 뭐 썩! 한 음악 없을까요?</p></sp>

- 중간 생략 -

【준구어 예2】

<teiHeader>

...

<castlist>

<head>등장인물</head>

<castitem><role>A (30대 초반쯤의 젊은이)</role></castitem>

<castitem><role>B (50대 중반쯤의 신사)</role></castitem>

</castlist>

...

</teiHeader>

<head>-꿈-</head>

<set>

<p>무대는 철록 같은 어둠. 포레의 음악 [꿈꾸고 난 후에]가 배경음악 (바브라 스트라이샌드의 노래)으로 깔린다. 음악이 끝나갈 때쯤 불륨이 서서히 잦아지면서 두 인물 A와 B에만 탑조명이 들어온다. A는 세미캐주얼 차림에 바바리를 입고있고 좀 피곤하고 초췌해 보인다. B는 그보다는 좀 표정에 여유가 있어 보이는 중년의 신사. 언밸런스한 차림이지만 뭔가 독특한 자기 개성이 드러나는 의상이 자유로운 분위기를 연출하고 있다. 두 사람은 가까이 앉아 있지만 둘 다 객석 쪽을 바라보고 있다. 조명이 들어옴과 동시에 A의 대사가 시작된다.</p>

</set>

</front>

<body>

<sp><speaker>A</speaker><stage>(꿈에서 덜 깬 듯 몽롱한 표정으로)</stage><p>참, 이상한
일이오.</p></sp>

<sp><speaker>B</speaker><p>뭐가요?</p></sp>

<sp><speaker>A</speaker><p>어젯밤 꿈.</p></sp>

<sp><speaker>B</speaker><p>한번 얘기해보세요.</p></sp>

<sp><speaker>A</speaker><p>난 긴 배를 노 저어 가고 있었어요. 아라비안나이트에 나오는 왕자가
신는 신발처럼 코끝이 구부러져 올라간 멋진 배였어요.</p></sp>

<sp><speaker>B</speaker><stage>(여전히 시큰둥하게, 호기심을 감추고서)</stage><p>어딜 가고
있었는데요?</p></sp>

<sp><speaker>A</speaker><p>그건 나도 몰라요. 그저 노를 저어 앞으로 앞으로만 가고
있었으니까요. 노를 젓는 느낌이 어찌나 부드러운지 하나도 힘이 들지 않았어요. 깃털처럼 부드럽고
살같이 곧게 내 배는 가고 있었어요.</p></sp>

<sp><speaker>B</speaker><p>파도가 없었나 보죠.</p></sp>

<sp><speaker>A</speaker><p>파도라고요?</p><stage>(황당하다는 듯 웃는다)</stage><p>그런 게
있을 리가 있나요.</p></sp>

<sp><speaker>B</speaker><p>그렇게 잔잔한 바다였나요?</p></sp>

<sp><speaker>A</speaker><p>바다?</p></sp>

<sp><speaker>B</speaker><p>예, 바다요!</p></sp>

<sp><speaker>A</speaker><p>아니요. 바다가 아니었어요.</p></sp>

<sp><speaker>B</speaker><p>그럼, 강이요?</p></sp>

<sp><speaker>A</speaker><stage>(고개를 저으며)</stage><p>아니오.</p></sp>

<sp><speaker>B</speaker><stage>(이제서야 답을 찾았다는 듯 확신을 갖고)</stage><p>그럼
호수였겠구만!</p></sp>

<sp><speaker>A</speaker><p>아니오.</p><stage>(다시 꿈속을 헤매듯 몽롱한 시선으로 먼 곳을
바라보며)</stage><p>나는 끝없는 은빛 모래사장을 노 저어 가고 있었어요.</p></sp>

<sp><speaker>B</speaker><stage>(놀란 표정으로)</stage><p>모래사장을요? 말도
안돼요.</p></sp>

<sp><speaker>A</speaker><p>그러니까 제가 이상한 일이라고 했잖아요.</p></sp>

<sp><speaker>B</speaker><p>당신 혼자서?</p></sp>

<sp><speaker>A</speaker><p>그래요. 난 혼자였어요.</p></sp>

<sp><speaker>B</speaker><p>아무도 없었나요?</p></sp>

<sp><speaker>A</speaker><p>어디 말씀이세요?</p></sp>

<sp><speaker>B</speaker><p>모래사장 말이오.</p></sp>

<sp><speaker>A</speaker><p>웬걸요. 사람들이 있었죠.</p></sp>

<sp><speaker>B</speaker><p>그럼 사람들한테 물어보지 그랬어요. 거기가 어디냐고.</p></sp>

<sp><speaker>A</speaker><p>아니, 그럴 수가 없었어요.</p></sp>

<sp><speaker>B</speaker><p>왜요?</p></sp>

<sp><speaker>A</speaker><p>사람들이 스쳐 지나갔지만 내 배가 너무 빨라서 난 그들의 얼굴을 볼

수도 없었고 그래서 내가 어디 있는지 어딜 가고 있는지 물어볼 수가 없었어요.</p></sp>
 <sp><speaker>B</speaker><p>바보같네요. 배를 멈추고 물어보면 될걸.</p></sp>
 <sp><speaker>A</speaker><p>하지만 배를 멈출 수가 없었어요. 분명히 내가 노를 젓고 있었지만
 내 배는 마치 리모컨에 의해 작동되는 장난감처럼 앞으로 앞으로 계속 가고 있는 거예요.</p></sp>
 <sp><speaker>B</speaker><p>그래도 한번 멈추어볼 노력은 했어야죠.</p></sp>
 <sp><speaker>A</speaker><p>아니, 그러고 싶지가 않았어요.</p></sp>
 <sp><speaker>B</speaker><p>왜요?</p></sp>
 <sp><speaker>A</speaker><p>기분이 너무 좋았거든요.</p></sp>
 <sp><speaker>B</speaker><stage>(의아스런 표정으로, 그러다 심각한 듯 표정을
 바꾸며)</stage><p>젊은 양반 이거 아주 나쁜 사람이구만!</p></sp>
 <sp><speaker>A</speaker><p>네? 뭐라고 하셨죠?</p></sp>
 <sp><speaker>B</speaker><p>젊은 양반이 나쁜 사람이라고 했소.</p></sp>
 <sp><speaker>A</speaker><stage>(갑작스런 꾸지람에 영문을 몰라하며)</stage><p>제가 뭐 잘못된
 거라도 있나요?</p></sp>
 <sp><speaker>B</speaker><p>그럼, 있다마다요. 기분이 좋다고 그냥 막 가도 되는 거요? 어디가
 어딘지도 모르고. 그러다 못 돌아오면 어떻게 하려고.</p></sp>

〈부록 2〉 2004년도 세종계획 기초 말뭉치 목록

파일명	분류 코드	분류명	제목	저자	출판사	출판 년도	어절수
7BA03A01	M1111	신문-정치	조선일보 2003년 기사: 정치 해설	조선일보사	조선일보사	2003	13,777
7BA03A02	M1111	신문-정치	조선일보 2003년 기사: 종합	조선일보사	조선일보사	2003	106,211
7BA03A03	M1112	신문-사회	조선일보 2003년 기사: 지역	조선일보사	조선일보사	2003	13,209
7BA03A04	M1112	신문-사회	조선일보 2003년 기사: 사회	조선일보사	조선일보사	2003	58,267
7BA03A05	M1113	신문-경제	조선일보 2003년 기사: 경제	조선일보사	조선일보사	2003	112,726
7BA03A06	M1114	신문-문화	조선일보 2003년 기사: 문화	조선일보사	조선일보사	2003	84,160
7BA03A07	M1115	신문-생활, 과학	조선일보 2003년 기사: 생활 여성	조선일보사	조선일보사	2003	14,387
7BA03A08	M1115	신문-생활, 과학	조선일보 2003년 기사: 정보통신 과학	조선일보사	조선일보사	2003	14,606
7BA03A09	M1116	신문-국제, 외신	조선일보 2003년 기사: 국제 외신	조선일보사	조선일보사	2003	47,477
7BA03A10	M1117	신문-취미, 연예,스포츠	조선일보 2003년 기사: 스포츠	조선일보사	조선일보사	2003	43,889
7BA03A11	M1110	신문-기타	조선일보 2003년 기사: 특집	조선일보사	조선일보사	2003	45,353
7BA03A12	M1110	신문-기타	조선일보 2003년 기사: 오피니언 인물	조선일보사	조선일보사	2003	126,235
7BA03B01	M1111	신문-정치	동아일보 2003년 기사: 종합	동아일보사	동아일보사	2003	197,248
7BA03B02	M1112	신문-사회	동아일보 2003년 기사: 사회	동아일보사	동아일보사	2003	103,100
7BA03B03	M1113	신문-경제	동아일보 2003년 기사: 경제	동아일보사	동아일보사	2003	166,529
7BA03B04	M1114	신문-문화	동아일보 2003년 기사: 문화	동아일보사	동아일보사	2003	90,851
7BA03B05	M1115	신문-생활, 과학	동아일보 2003년 기사: 정보통신 과학	동아일보사	동아일보사	2003	62,705
7BA03B06	M1116	신문-국제, 외신	동아일보 2003년 기사: 국제 외신	동아일보사	동아일보사	2003	51,035
7BA03B07	M1117	신문-취미, 연예,스포츠	동아일보 2003년 기사: 스포츠	동아일보사	동아일보사	2003	61,768

파일명	분류코드	분류명	제목	저자	출판사	출판년도	어절수
7BA03E01	M1111	신문-정치	한겨레신문 2003년 기사: 종합	한겨레신문사	한겨레신문사	2003	182,204
7BA03E02	M1112	신문-사회	한겨레신문 2003년 기사: 사회	한겨레신문사	한겨레신문사	2003	68,458
7BA03E03	M1113	신문-경제	한겨레신문 2003년 기사: 경제	한겨레신문사	한겨레신문사	2003	148,521
7BA03E04	M1114	신문-문화	한겨레신문 2003년 기사: 매체	한겨레신문사	한겨레신문사	2003	6,445
7BA03E05	M1114	신문-문화	한겨레신문 2003년 기사: 문화	한겨레신문사	한겨레신문사	2003	134,267
7BA03E06	M1115	신문-생활, 과학	한겨레신문 2003년 기사: 정보통신 과학	한겨레신문사	한겨레신문사	2003	33,131
7BA03E07	M1115	신문-생활, 과학	한겨레신문 2003년 기사: 생활 여성	한겨레신문사	한겨레신문사	2003	51,076
7BA03E08	M1116	신문-국제, 외신	한겨레신문 2003년 기사: 국제 외신	한겨레신문사	한겨레신문사	2003	50,833
7BA03E09	M1117	신문-취미, 연예,스포츠	한겨레신문 2003년 기사: 방송 연예	한겨레신문사	한겨레신문사	2003	10,776
7BA03E10	M1117	신문-취미, 연예,스포츠	한겨레신문 2003년 기사: 스포츠	한겨레신문사	한겨레신문사	2003	49,806
7BB03B01	M1260	잡지-사회	주간조선 1735호	조선일보사	조선일보사	2003	26,948
7BB03B02	M1260	잡지-사회	주간조선 1736호	조선일보사	조선일보사	2003	25,527
7BB03B03	M1260	잡지-사회	주간조선 1737호	조선일보사	조선일보사	2003	29,584
7BB03B04	M1260	잡지-사회	주간조선 1738호	조선일보사	조선일보사	2003	29,423
7BB03B05	M1260	잡지-사회	주간조선 1739호	조선일보사	조선일보사	2003	22,188
7BB03B06	M1260	잡지-사회	주간조선 1740호	조선일보사	조선일보사	2003	32,730
7BB03B07	M1260	잡지-사회	주간조선 1741호	조선일보사	조선일보사	2003	28,462
7BB03B08	M1260	잡지-사회	주간조선 1742호	조선일보사	조선일보사	2003	27,541
7BB03B09	M1260	잡지-사회	주간조선 1743호	조선일보사	조선일보사	2003	25,064
7BB03B10	M1260	잡지-사회	주간조선 1744호	조선일보사	조선일보사	2003	27,211
7BB03B11	M1260	잡지-사회	주간조선 1745호	조선일보사	조선일보사	2003	25,414

파일명	분류 코드	분류명	제목	저자	출판사	출판 년도	어절수
7BB03B12	M1260	잡지-사회	주간조선 1746호	조선일보사	조선일보사	2003	27,349
7BB03B13	M1260	잡지-사회	주간조선 1747호	조선일보사	조선일보사	2003	30,458
7BB03B14	M1260	잡지-사회	주간조선 1748호	조선일보사	조선일보사	2003	25,968
7BB03B15	M1260	잡지-사회	주간조선 1751호	조선일보사	조선일보사	2003	25,128
7BB03B16	M1260	잡지-사회	주간조선 1752호	조선일보사	조선일보사	2003	27,570
7BB03B17	M1260	잡지-사회	주간조선 1753호	조선일보사	조선일보사	2003	27,329
7BB03B18	M1260	잡지-사회	주간조선 1756호	조선일보사	조선일보사	2003	28,410
7BB03B19	M1260	잡지-사회	주간조선 1757호	조선일보사	조선일보사	2003	25,167
7BB03B20	M1260	잡지-사회	주간조선 1758호	조선일보사	조선일보사	2003	31,300
7BB03B21	M1260	잡지-사회	주간조선 1762호	조선일보사	조선일보사	2003	27,478
7BB03B22	M1260	잡지-사회	주간조선 1763호	조선일보사	조선일보사	2003	29,215
7BB03B23	M1260	잡지-사회	주간조선 1764호	조선일보사	조선일보사	2003	28,361
7BB03B24	M1260	잡지-사회	주간조선 1765호	조선일보사	조선일보사	2003	27,716
7BB03B25	M1260	잡지-사회	주간조선 1766호	조선일보사	조선일보사	2003	30,853
7BB03B26	M1260	잡지-사회	주간조선 1767호	조선일보사	조선일보사	2003	29,143
7BB03B27	M1260	잡지-사회	주간조선 1768호	조선일보사	조선일보사	2003	24,540
7BB03B28	M1260	잡지-사회	주간조선 1769호	조선일보사	조선일보사	2003	24,493
7BB03B29	M1260	잡지-사회	주간조선 1770호	조선일보사	조선일보사	2003	34,259
7BE03001	M1321	책-교육	종이 먹는 개 로로	김주현	(주)문학사상사	2003	7,189
7BE04002	M1331	책-상상	꼭두의 사랑	이명훈	(주)문학사상사	2004	40,816
7BE04003	M1331	책-상상	날마다 축제	강영숙	(주)창비	2004	38,912

파일명	분류코드	분류명	제목	저자	출판사	출판년도	어절수
7BE03004	M1331	책-상상	랍스터를 먹는 시간	방현석	(주)창비	2003	49,840
7BE03005	M1331	책-상상	사람의 향기	송기원	(주)창비	2003	42,021
7BE03006	M1331	책-상상	소라단 가는 길	윤홍길	(주)창비	2003	51,240
7BE03007	M1331	책-상상	이상한 나라에서 온 스파이	최인석	(주)창비	2003	75,197
7BG03001	M1341	책-체험	다영이의 이슬람 여행: 세계사에서 숨은그림 찾기	정다영	(주)창비	2003	22,789
7BG03002	M1341	책-체험	유재현의 역사문화기행: 메콩의 슬픈 그림자, 인도차이나	유재현	(주)창비	2003	51,505
7BG03003	M1341	책-체험	현대인이 잃어버린 것들	이어령	(주)문학사상사	2003	42,377
7BH03001	M1352	책-인문	언어, 컴퓨터, 코퍼스 언어학	강범모	고려대학교 출판부	2003	50,172
7BH00002	M1352	책-인문	한국어 은유연구	박영순	고려대학교 출판부	2000	52,205
7BH04003	M1352	책-인문	한국어 의미론	박영순	고려대학교 출판부	2004	58,004
7BH01004	M1353	책-인문	메멘토 모리, 죽음을 기억하라	김열규	궁리출판	2001	51,539
7BH03005	M1353	책-인문	신화 속의 한국정신	이어령	문학사상사	2003	35,019
7BH03006	M1353	책-인문	왜 사냐면, ...웃지요	김열규	궁리출판	2003	44,984
7BH03007	M1358	책-인문	이공계대학 연구경쟁력 살리기	송충한	고려대학교 출판부	2003	23,580
7BH04008	M1362	책-사회	중국의 새로운 사회주의 탐색	이희옥	(주)창비	2004	36,162
7BH02009	M1365	책-사회	고객과 경쟁하라	박낙원	(주)북21	2002	40,773
7BH01010	M1365	책-사회	벤처산업의 미래전략	김병균	(주)북21	2001	48,451
7BH98011	M1365	책-사회	빅뱅경영	이재규	21세기북스	1998	41,874
7BH04012	M1365	책-사회	서른다섯, 행복한 도전자들	김광주	(주)북이십일	2004	29,182
7BH04013	M1365	책-사회	세계 100대 IT리더 김홍기가 들려주는 디지털 인재의 조건	김홍기	(주)북21	2004	23,660

파일명	분류코드	분류명	제목	저자	출판사	출판년도	어절수
7BH99014	M1365	책-사회	이재규 교수의 재미있는 기업이야기	이재규	21세기북스	1999	41,699
7BH04015	M1365	책-사회	자녀의 금융재능을 쑥쑥 키워주는 생활경제교실	박상섭	(주)북21 ulpaso	2004	12,794
7BH02016	M1365	책-사회	조관일의 고객죽이기	조관일	(주)북21	2002	18,323
7BH03017	M1369	책-사회	여성시대에는 남자도 화장을 한다	최재천	궁리출판	2003	26,291
7BH04018	M1371	책-자연	유전자가 세상을 바꾼다: 인간배아복제, 유전영질전환에 관한 논쟁	김훈기	궁리출판	2004	30,552
7BH02019	M1371	책-자연	하리하라의 생물학 카페	이은희	궁리출판	2002	35,383
7BH03020	M1374	책-자연	한국 식품학 입문	이철호 · 권 태완	고려대학교 출판부	2003	87,265
7BH00021	M1377	책-자연	비타민과 환경인자	우순자	고려대학교 출판부	2000	80,380
7BH00022	M1377	책-자연	의료체계와 법	이상돈	고려대학교 출판부	2000	54,411
7BH03023	M1377	책-자연	임상신경학-각론	이대희	고려대학교 출판부	2003	114,008
7BH03024	M1377	책-자연	임상신경학-총론	이대희	고려대학교 출판부	2003	91,032
7BH00025	M1377	책-자연	호스피스 · 완화의 학	최운선	고려대학교 출판부	2000	31,177
7BH03026	M1381	책-예술론	고려후기 사대부문학 연구	박성규	고려대학교 출판부	2003	62,027
7BH02027	M1381	책-예술론	발견으로서의 소설기법	송하춘	고려대학교 출판부	2002	56,600
7BH04028	M1381	책-예술론	소설 창작 강의	전상국	(주)문학사상사	2004	71,829
7BH00029	M1381	책-예술론	소설발견 송하춘 교수의 창작교실 2	송하춘	고려대학교 출판부	2000	28,224
7BH02030	M1381	책-예술론	소설발견 송하춘 교수의 창작교실 3	송하춘	고려대학교 출판부	2002	29,788
7BH02031	M1381	책-예술론	소설발견 송하춘 교수의 창작교실 4	송하춘	고려대학교 출판부	2002	26,026
7BH99032	M1381	책-예술론	소설발견1: 송하춘 교수의 창작교실	송하춘	고려대학교 출판부	1999	32,307
7BH03033	M1381	책-예술론	소설발견5: 송하춘 교수의 창작교실	송하춘	고려대학교 출판부	2003	28,720

파일명	분류 코드	분류명	제목	저자	출판사	출판 년도	어절수
7BH02034	M1381	책-예술론	인터넷시대의 시 창작론	최동호	고려대학교 출판부	2002	24,703
7BH02035	M1381	책-예술론	한국 고전문학과 비평의 성찰	김홍규	고려대학교 출판부	2002	59,374
7BH02036	M1381	책-예술론	한국고전소설사연 구	장효현	고려대학교 출판부	2002	127,581
7BH03037	M1386	책-예술론	세상에서 가장 아름다운 집	서운영	궁리출판	2003	35,295
7BH00038	M1386	책-예술론	한국건축사	주남철	고려대학교 출판부	2000	70,053
7BH04039	M1391	책-생활	야마토마치에서 만난 노인들	김동선	궁리출판	2004	27,326
합계							5,214,538

〈부록 3〉 2004년도 세종계획 기초 말뭉치 원전 소장 정보

소장처	국립중앙도서관
청구기호	326.162-2-32
서명/저자명	고객과 경쟁하라 / 박낙원 지음
발행사항	서울 : 북21, 2002
형태사항	329p. ; 23cm

소장처	국립중앙도서관
청구기호	810.904-4-2
서명/저자명	고려후기 사대부 문학연구 / 박성규 지음
발행사항	서울 : 고려대학교출판부, 2003
형태사항	iv, 381p. ; 23cm

소장처	국립중앙도서관
청구기호	813.6-4-544
서명/저자명	꼭두의 사랑 : 이명훈 장편소설 / 이명훈 지음
발행사항	서울 : 문학사상사, 2004
형태사항	244p. ; 23cm

소장처	국립중앙도서관
청구기호	813.6-4-358
서명/저자명	(날마다)축제 : 강영숙 소설집 / 강영숙 지음
발행사항	파주 : 창비, 2004
형태사항	235p. ; 23cm

소장처	국립중앙도서관
청구기호	981.802-3-1
서명/저자명	다영이의 이슬람 여행 : 세계사에서 숨은그림 찾기 / 정다영 지음
발행사항	서울 : 창작과 비평사, 2003
형태사항	234p. : 삽도 ; 21cm

소장처	국립중앙도서관
청구기호	325.04-4-47
서명/저자명	디지털 인재의 조건 / 김흥기 지음
발행사항	서울 : 북21, 2004
형태사항	237p. : 삽도 ; 21cm

소장처	국립중앙도서관
청구기호	813.6-3-1002
서명/저자명	랍스터를 먹는 시간 : 방현석 소설집 / 방현석 지음
발행사항	과주 : 창비, 2003
형태사항	332p. ; 23cm

소장처	국립중앙도서관
청구기호	114.5-1-8
서명/저자명	(한국의 죽음론)메멘토 모리, 죽음을 기억하라 / 김열규 지음
발행사항	서울 : 궁리출판, 2001
형태사항	308p. ; 23cm

소장처	국립중앙도서관
청구기호	981.402-4-1
서명/저자명	메콩의 슬픈 그림자, 인도차이나 : 유재현의 역사문화기행 / 유재현 지음
발행사항	과주 : 창비, 2003
형태사항	344p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	802.3-2-5
서명/저자명	발견으로서의 소설기법 / 송하춘 지음
발행사항	서울 : 고려대학교출판부, 2002
형태사항	vii, 330p. ; 23cm

소장처	국립중앙도서관
청구기호	324.35-1-6
서명/저자명	벤처산업의 미래전략 / 김병균 지음
발행사항	서울 : 21세기북스, 2001
형태사항	410p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	594.1-우156ㄴ
서명/저자명	비타민과 환경인자 / 우순자 지음
발행사항	서울 : 고려대학교출판부, 2000
형태사항	xiii, 394p. ; 23cm

소장처	국립중앙도서관
청구기호	325.1-이594ㄴㄴ
서명/저자명	(IMF 시대)빅뱅경영 / 이재규 지음
발행사항	서울 : 21세기북스, 1998
형태사항	323p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	813.6-3-359
서명/저자명	사람의 향기 : 송기원 연작소설 / 송기원 지음
발행사항	서울 : 창작과비평사, 2003
형태사항	275p. ; 23cm

소장처	국립중앙도서관
청구기호	320-4-14
서명/저자명	생활경제교실 / 박상섭 지음
발행사항	서울 : 북21 을파소, 2004
형태사항	178p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	325.04-4-46
서명/저자명	(서른다섯,)행복한 도전자들 / 김광주 지음
발행사항	서울 : 북이십일, 2004
형태사항	219p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	610.4-3-2
서명/저자명	(세상에서 가장 아름다운)집 / 서운영 지음
발행사항	서울 : 궁리출판, 2003
형태사항	289p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	813.6-4-50
서명/저자명	소라단 가는 길 : 윤홍길연작소설 / 윤홍길 지음
발행사항	파주 : 창비, 2003
형태사항	325p. ; 23cm

소장처	국립중앙도서관
청구기호	802.3-4-4
서명/저자명	소설 창작 강의 : 당신도 소설을 쓸 수 있다 / 전상국 지음
발행사항	서울 : 문학사상사, 2003
형태사항	404p. ; 23cm

소장처	고려대학교도서관
청구기호	3631 308 1
서명/저자명	소설발견: 송하춘 교수의 창작교실 1 / 송하춘 지음
발행사항	서울 : 고려대학교 출판부, 1999
형태사항	164 p ; 23 cm

소장처	고려대학교도서관
청구기호	3631 308 2
서명/저자명	소설발견: 송하춘 교수의 창작교실 2 / 송하춘 지음
발행사항	서울 : 고려대학교 출판부, 2000
형태사항	147 p ; 23 cm

소장처	고려대학교도서관
청구기호	3631 308 3
서명/저자명	소설발견: 송하춘 교수의 창작교실 3 / 송하춘 지음
발행사항	서울 : 고려대학교 출판부, 2002
형태사항	160 p ; 22 cm

소장처	고려대학교도서관
청구기호	808.3 1999e 4
서명/저자명	소설발견: 송하춘 교수의 창작교실 4 / 송하춘 지음
발행사항	서울 : 고려대학교 출판부, 2002
형태사항	v, 151 p ; 22 cm

소장처	고려대학교도서관
청구기호	808.3 1999e 5
서명/저자명	소설발견: 송하춘 교수의 창작교실 5 / 송하춘 지음
발행사항	서울 : 고려대학교 출판부, 2003
형태사항	iii, 158 p ; 22 cm

소장처	국립중앙도서관
청구기호	813.709-3-2
서명/저자명	신화 속의 한국정신 / 이어령 지음
발행사항	서울 : 문학사상사, 2003
형태사항	309p. ; 23cm

소장처	국립중앙도서관
청구기호	338.6-4-70
서명/저자명	야마토마치에서 만난 노인들 / 김동선 지음
발행사항	파주 : 궁리출판, 2004
형태사항	209p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	701.02-4-2
서명/저자명	언어, 컴퓨터, 코퍼스 언어학 / 강범모 지음
발행사항	서울 : 고려대학교출판부, 2003
형태사항	vi, 280p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	337.04-3-3
서명/저자명	여성시대에는 남자도 화장을 한다 / 최재천 지음
발행사항	서울 : 궁리출판, 2003
형태사항	230p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	199.1-3-100
서명/저자명	왜 사냐면, ...웃지요 : 한국인의 웃음, 그 내력과 미학 / 김열규 지음
발행사항	서울 : 궁리출판, 2003
형태사항	305p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	476.1-4-10
서명/저자명	유전자가 세상을 바꾼다 : 인간배아복제, 유전형질전환에 관한 논쟁 / 김훈기 지음
발행사항	파주 : 궁리출판, 2004
형태사항	259p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	360.1-이359○
서명/저자명	의료체계와 법 : 의료보험, 의약분업, 의료분쟁해결의 법철학적 성찰 / 이상돈 지음
발행사항	서울 : 고려대학교 출판부, 2000
형태사항	xiii, 255p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	377.7-3-14
서명/저자명	이공계대학 연구경쟁력 살리기 / 송충한 지음
발행사항	서울 : 고려대학교출판부, 2003
형태사항	vii, 157p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	813.6-3-696
서명/저자명	(이상한 나라에서 온)스파이 : 최인석 장편소설 / 최인석 지음
발행사항	파주 : 창작과비평사, 2003
형태사항	433p. ; 23cm

소장처	국립중앙도서관
청구기호	325.04-이594ㅈㄱ
서명/저자명	(이재규 교수의)재미있는 기업이야기 / 이재규 지음
발행사항	서울 : 21세기북스, 1999
형태사항	333p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	802.1-2-5
서명/저자명	인터넷시대의 시창작론 / 최동호 지음
발행사항	서울 : 고려대학교출판부, 2002
형태사항	iv, 173p. : 도판 ; 22cm

소장처	고려대학교도서관
청구기호	616.8 2003e 2
서명/저자명	임상 신경학 2 : 각론 / 이대희 [저]
발행사항	서울 : 고려대학교 출판부 , 2003
형태사항	viii, 489 p : 삽도 ; 27 cm

소장처	고려대학교도서관
청구기호	616.8 2003e 1
서명/저자명	임상 신경학 1 : 총론 / 이대희 [저]
발행사항	서울 : 고려대학교, 2003
형태사항	vii, 412 p : 삽도 ; 27 cm

소장처	국립중앙도서관
청구기호	326.161-3-1
서명/저자명	(조판일의)고객죽이기 / 조판일 지음
발행사항	서울 : 북21, 2002
형태사항	185p. : 삽도 ; 22cm

소장처	어린이도서관
청구기호	813.8-7934ㅈ
서명/저자명	종이 먹는 개 로로 / 김주현 지음 ; 이미정 그림
발행사항	서울 : 문학사상사, 2003
형태사항	141p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	340.912-4-6
서명/저자명	중국의 새로운 사회주의 탐색 / 이희옥 지음
발행사항	파주 : 창비, 2004
형태사항	315p. ; 23cm

소장처	국립중앙도서관
청구기호	472-2-3
서명/저자명	(하리하라의)생물학 카페 / 이은희 지음
발행사항	서울 : 궁리출판, 2002
형태사항	309p. : 삽도 ; 23cm

소장처	국립중앙도서관
청구기호	594-4-8
서명/저자명	한국식품학입문 / 이철호 ; 권태완 [공]지음
발행사항	서울 : 고려대학교출판부, 2003
형태사항	xi, 600p : 삽도 ; 25cm

소장처	국립중앙도서관
청구기호	610.911-주138ㅎ초
서명/저자명	한국건축사 / 주남철 지음
발행사항	서울 : 고려대학교출판부, 2000
형태사항	ix, 491p. : 삽도 ; 26cm

소장처	고려대학교도서관
청구기호	897.09 2002
서명/저자명	한국 고전문학과 비평의 성찰/ 김홍규 편저
발행사항	서울 : 고려대학교 출판부 , 2002
형태사항	355 p ; 21 cm

소장처	고려대학교도서관
청구기호	897.3309 2002g
서명/저자명	한국고전소설사연구 / 장효현
발행사항	서울 : 고려대학교 출판부 , 2002
형태사항	x, 802 p ; 24 cm

소장처	고려대학교도서관
청구기호	497.0143 2000
서명/저자명	한국어 은유연구 / 박영순 지음
발행사항	서울 : 고려대학교 출판부 , 2000
형태사항	viii, 375 p ; 23 cm

소장처	고려대학교도서관
청구기호	497.0143 1994
서명/저자명	한국어 의미론 / 박영순 지음
발행사항	서울 : 고려대학교 출판부 , 1994
형태사항	v, 295 p : 삽도 ; 23 cm

소장처	국립중앙도서관
청구기호	814.6-3-115
서명/저자명	현대인이 잃어버린 것들 / 이어령 지음
발행사항	서울 : 문학사상사, 2003
형태사항	317p. ; 23cm

소장처	고려대학교도서관
청구기호	616.029 2000
서명/저자명	호스피스·완화의학/ 최윤선 지음
발행사항	서울 : 고려대학교 출판부 , 2000
형태사항	v, 252 p : 삽도,도판 ; 23 cm

2. 형태분석 말뭉치 구축

가. 연구 배경 및 목표

1) 연구 배경

컴퓨터를 중심으로 한 과학기술이 급속히 발달하고 정보 처리에 대한 관심이 그 어느 때보다도 고조되고 있는 지금, 자연언어처리는 이미 우리가 진입하고 있는 정보화 사회의 핵심 과제로서 그 필요성이 더욱 증대되고 있다. 현재 정보 검색, 기계 번역, 음성 인식 등의 한국어 정보처리 기술이 꾸준히 연구되고 있으며 일부는 이미 실용화 단계에 접어들고 있다.

이렇듯 한국어 정보 처리에 대한 연구와 그것의 실용화에 박차를 가하고 있는 이 시점에서, 한국어 정보 처리의 근간이 되는 형태분석 말뭉치 개발은 그 중요성이 더욱 높아지고 있다. 최근의 언어 정보 처리 기술은 종래의 규칙 기반 일변도의 방법론을 탈피하여, 규칙 기반의 기법과 통계 기반의 방법론을 적절히 병행하여 상당한 성과를 올리고 있다. 현재, 한국어의 경우에도 이러한 방식을 바탕으로 하여 말뭉치 구축 작업이 활발히 이루어지고 있으며, 정밀한 언어 분석과 그에 관한 순수 이론적인 연구, 그리고 정보 처리 기술 개발을 위해서 대규모의 정확한 형태분석 말뭉치의 구축이 절실히 요구되고 있다.

형태소 분석은 각종 한국어 정보 처리 분야에서 기본이 되는 단계이지만 분석 말뭉치의 구체적인 성과가 공개된 예는 그리 많지 않다. 이는, 대규모의 자료에 품사 정보를 정확히 부여하는 작업이 매우 어렵기 때문이다. 형태소 분석은 이미 상당한 자동화율을 보이고 있다. 그러나 한국어의 특성상 불가피하게 생길 수밖에 없는 형태소 분석상의 중의성의 문제 등을 충분히 해결하기 위해서는 여전히 많은 시간과 노력을 요구하는 수작업 후처리가 필수적이다.

따라서 본 연구에서는 지난 6개 년도에 걸쳐 수행된 분석표지 연구와 1,000만 어절

규모의 형태분석 말뭉치를 바탕으로 수작업을 효과적으로 줄이는 형태소 분석 도구를 개발하는 등 작업의 효율적 처리를 위한 방안을 다각도로 모색하여 품사 정보가 부착된 형태분석 말뭉치를 새로 구축하였다.

2) 연구 목표

본 연구의 목표는 품사 표지를 부착한 200만 어절 규모의 형태분석 말뭉치를 개발하는 것이다. 1998년부터 6개년에 걸쳐 구축된 형태분석 말뭉치는 1,000만 어절 규모이다. 이번 7차년도에서는 구축된 1,000만 어절 규모의 형태분석 말뭉치를 바탕으로 200만 어절 규모의 말뭉치를 구축하였으며, 최종적으로 1,200만 어절 규모의 형태분석 말뭉치를 구축하게 된다.

나. 구축과정

1) 형태분석 말뭉치 구축 준비

가) 대상 말뭉치 선정

형태분석 말뭉치를 구축하기 위해서는 형태소 분석의 대상이 되는 말뭉치를 선정해야 한다. 7차년도에는 1998년에 구축된 연구교육용 균형 말뭉치와 2002년에 구축된 원시 말뭉치를 중심으로 200만 어절을 선정하였다.

선정된 원시 말뭉치 목록은 다음 표와 같다.

파일명	원본 파일명	어절수	제 목	저 자	출판 년도	장르 번호
7BT_001	4BB00A01.HWP	20,438	월간중앙 1월호	중앙일보 J&P	2000	M1200
7BT_002	4BB00A02.HWP	17,756	월간중앙 2월호	중앙일보 J&P	2000	M1200
7BT_003	4BB00A04.HWP	21,591	월간중앙 4월호	중앙일보 J&P	2000	M1200
7BT_004	4BB00A05.HWP	24,422	월간중앙 5월호	중앙일보 J&P	2000	M1200
7BT_005	4BB00A07.HWP	23,227	월간중앙 7월호	중앙일보 J&P	2000	M1200
7BT_006	4BB00A08.HWP	22,922	월간중앙 8월호	중앙일보 J&P	2000	M1200
7BT_007	4BB00A09.HWP	19,317	월간중앙 9월호	중앙일보 J&P	2000	M1200
7BT_008	5BA01A01.HWP	60,017	조선일보 2001년 기사: 경제	조선일보사	2001	M1113
7BT_009	5BA01A06.HWP	48,440	조선일보 2001년 기사: 사회	조선일보사	2001	M1112

파일명	원본 파일명	어절수	제 목	저 자	출판 년도	장르 번호
7BT_010	5BA01E04.HWP	52,224	한겨레신문 2001년 기사: 문화	한겨레신문사	2001	M1114
7BT_011	5BB99B01.HWP	26,155	한겨레 21: 1999/01/07	한겨레신문사	1999	M1200
7BT_012	5BB99F01.HWP	24,581	씨네 21: 1999/01/05	한겨레신문사	1999	M1200
7BT_013	5BB99F02.HWP	27,804	씨네 21: 1999/01/12	한겨레신문사	1999	M1200
7BT_014	5BB99F03.HWP	24,622	씨네 21: 1999/01/19	한겨레신문사	1999	M1200
7BT_015	5BB99F06.HWP	25,646	씨네 21: 1999/02/08	한겨레신문사	1999	M1200
7BT_016	5BE00001.HWP	54,331	냉장고	김현영	2000	M1330
7BT_017	5BE01001.HWP	55,581	가장 멀리 있는 나	윤후명	2001	M1330
7BT_018	5BE01002.HWP	57,707	구렁이들의 집	최인석	2001	M1330
7BT_019	5BE01003.HWP	39,733	당진 김씨	우애령	2001	M1330
7BT_020	5BE01004.HWP	50,755	마법성의 수호자, 나의 끼긋한 들깨	북거일	2001	M1330
7BT_021	5BE01005.HWP	57,322	미란	윤대녕	2001	M1330
7BT_022	5BE01008.HWP	36,966	아랑은 왜	김영하	2001	M1330
7BT_023	5BE01010.HWP	58,609	푸른 장미를 찾아서	김태환	2001	M1330
7BT_024	5BE01012.HWP	63,660	경성애사	이선미	2001	M1330
7BT_025	5BE01014.HWP	52,623	오디션	민해연	2001	M1330
7BT_026	5BE01015.HWP	28,822	햄릿의 여인	김지혜	2001	M1330
7BT_027	5BE02001.HWP	45,076	꽃그늘 아래	이혜경	2002	M1330
7BT_028	5BE02002.HWP	52,314	꿈꾸는 마리오네뜨	권지예	2002	M1330
7BT_029	5BE02003.HWP	48,195	멋진 한세상	공선옥	2002	M1330
7BT_030	5BE02004.HWP	47,596	푸른 수염의 첫 번째 아내	하성란	2002	M1330
7BT_031	5BE02005.HWP	51,820	황만근은 이렇게 말했다	성석제	2002	M1330
7BT_032	5BE02006.HWP	61,494	33세의 광세	김승희	2002	M1330
7BT_033	5BE02007.HWP	18,298	바다여, 아빠를 돌려줘	김재경	2002	M1330
7BT_034	5BE87001.HWP	55,584	목시의 바다	윤홍길	1987	M1330
7BT_035	5BE90001.HWP	86,551	권태	마광수	1990	M1330
7BT_036	5BE96001.HWP	56,564	사랑의 이름으로	한수산	1996	M1330
7BT_037	5BH01009.HWP	40,101	사회를 보는 논리	김찬호	2001	M1361
7BT_038	5BH01010.HWP	72,125	세계관 충돌과 한말 외교사	김용구	2001	M1355
7BT_039	5BH01011.HWP	52,046	글쓰기의 새로운 지평	이화형, 유진월	2001	M1350
7BT_040	5BH01012.HWP	40,867	대중화술	김양호	2001	M1320
7BT_041	6BG00001.HWP	64,206	참여로 여는 생태공동체: 어느 근본주의자의 환경 낚두리	박병상	2003	M1361
7BT_042	6BH00008.HWP	54,798	초민족 시대의 민족정체성	고부응	2002	M1360
7BT_043	6BH00009.HWP	40,120	글로벌 가버넌스와 NGO	주성수	2000	M1361
7BT_044	6BH00010.HWP	26,007	한국의 여성환경 운동: 그 역 사, 주체, 그리고 운동유형들	문순홍	2001	M1369
7BT_045	AC000009.HWP	18,147	이오덕의 교육일기1(3/3)	이오덕	1989	M1320
7BT_046	AE000093.HWP	23,996	광장(19-127면)	최인훈	1976	M1331
7BT_047	AH000049.HWP	15,459	여보게 저승갈 때 뭘 가지고 가 지(1/3)	석용산	1992	M1341
7BT_048	AH000050.HWP	12,920	여보게 저승갈 때 뭘 가지고 가 지(2/3)	석용산	1992	M1341
7BT_049	AH000168.HWP	28,927	역사와 민족(2/3)	함석헌	1983	M1350
7BT_050	AH000169.HWP	16,894	역사와 민족(3/3)	함석헌	1983	M1350
총계		2,025,386				

[표 1] 형태분석 말뭉치 목록

나) 장르별 구성

2004년도 형태분석 말뭉치는 다음과 같이 구성되어 있다.

분 류		구축 어절수	구축비율	비 고
신문		160,681	7.9%	
잡지		278,481	13.8%	
책-상상		1,103,607	54.4%	
책-정보	교육	18,147	0.9%	
	체험	28,379	1.5%	
	인문	250,960	12.4%	
	사회	185,131	9.1%	
합 계		2,025,386	100%	

[표 2] 2004년도 형태분석 말뭉치의 장르별 비율

다) 분석 지침 수정 및 보완

7차년도 형태분석 말뭉치는 부록에 수록된 ‘형태분석 말뭉치 구축 지침(2004-1)’을 바탕으로 구축되었다. 변화된 지침의 내용 가운데 중요한 사항 몇 가지를 다음에 제시한다.

- ‘형태분석 말뭉치 구축 지침(2004-1)’에서는 영문이나 한자, 기호 등이 어절 중간에 개입된 어절에 대한 처리를 분리형으로 수정하였다. 기존에는 이처럼 기호 삽입 등의 이유로 올바른 분석이 불가능한 경우 전체 어절을 통합하여 분석하였으나, 7차년도에는 각각 분리하여 처리하였다.

2) 형태분석 말뭉치 구축

가) 말뭉치 전처리

말뭉치 전처리는 원시 말뭉치에 자동 처리를 거쳐 품사 정보를 부착하기 위한 준비 단계라고 할 수 있다. 전처리 과정은 원시 말뭉치 가공 단계와 원시 말뭉치 오류 수정 단계로 나눌 수 있다. 원시 말뭉치 오류 수정 단계는 자동 품사 부착 시 분석 오류를 일으

키는 띄어쓰기 오류를 수정하는 것으로 자동 품사 부착 성공률을 높이기 위해 반드시 필요한 과정이라 할 수 있다.

(1) 원시 말뭉치 가공 단계

이 단계는 원시 말뭉치를 자동 태깅 프로그램에 사용하기 위한 형태로 가공하는 것이다. 원시 말뭉치에 부착되어 있는 TEI 방식의 헤더 정보와 문단 마크업 정보 등을 제거하였다.

(2) 원시 말뭉치 오류 수정 단계

이 단계는 원시 말뭉치의 오류(띄어쓰기 오류, 기호 오류 등)를 수정하여 자동 태깅 후 진행될 후처리 작업의 부담을 줄이기 위한 것이다.

나) 자동 tagger를 이용한 표지 부착

그동안 구축된 1,000만 어절 규모의 형태분석 말뭉치를 토대로 한층 개선된 태깅 방식을 사용함으로써 자동 태깅 결과의 성공률을 향상시켰다.

다) 후처리 과정

후처리는 자동 태깅에 의해 분석된 어절을 대상으로 분석의 중의성이 해소되지 않은 어절의 중의성 해소, 분석이 잘못된 어절의 오류 수정 등을 실시하는 것이다. 후처리는 수작업으로 진행된다.

(1) 중의적 어절 수정 (후처리 1단계)

자동 태깅 결과 중에서 형태론적 중의성 문제를 안고 있는 어절을 수정한다. 이 단계에서는 중의성 해결 및 형태소 결합 규칙 검사를 유형별로 처리한다.

(2) 문제 어절 수정 (후처리 2단계)

자동 태깅 결과 중에서 분석이 잘못된 문제 어절을 찾아 수정한다. 이 단계에서 후처리의 대상은 그 성격에 따라 3가지 유형으로 분류하였다.

[유형화 가능 오류] 기존 형태분석 말뭉치에서 일정하게 잘못 분석된 유형을 보이는 어절 (372,305 어절)

[유형화 불능 오류] 기존 형태분석 말뭉치에서 일정한 분석 유형 없이 잘못 분석

된 어절 (160,514 어절)

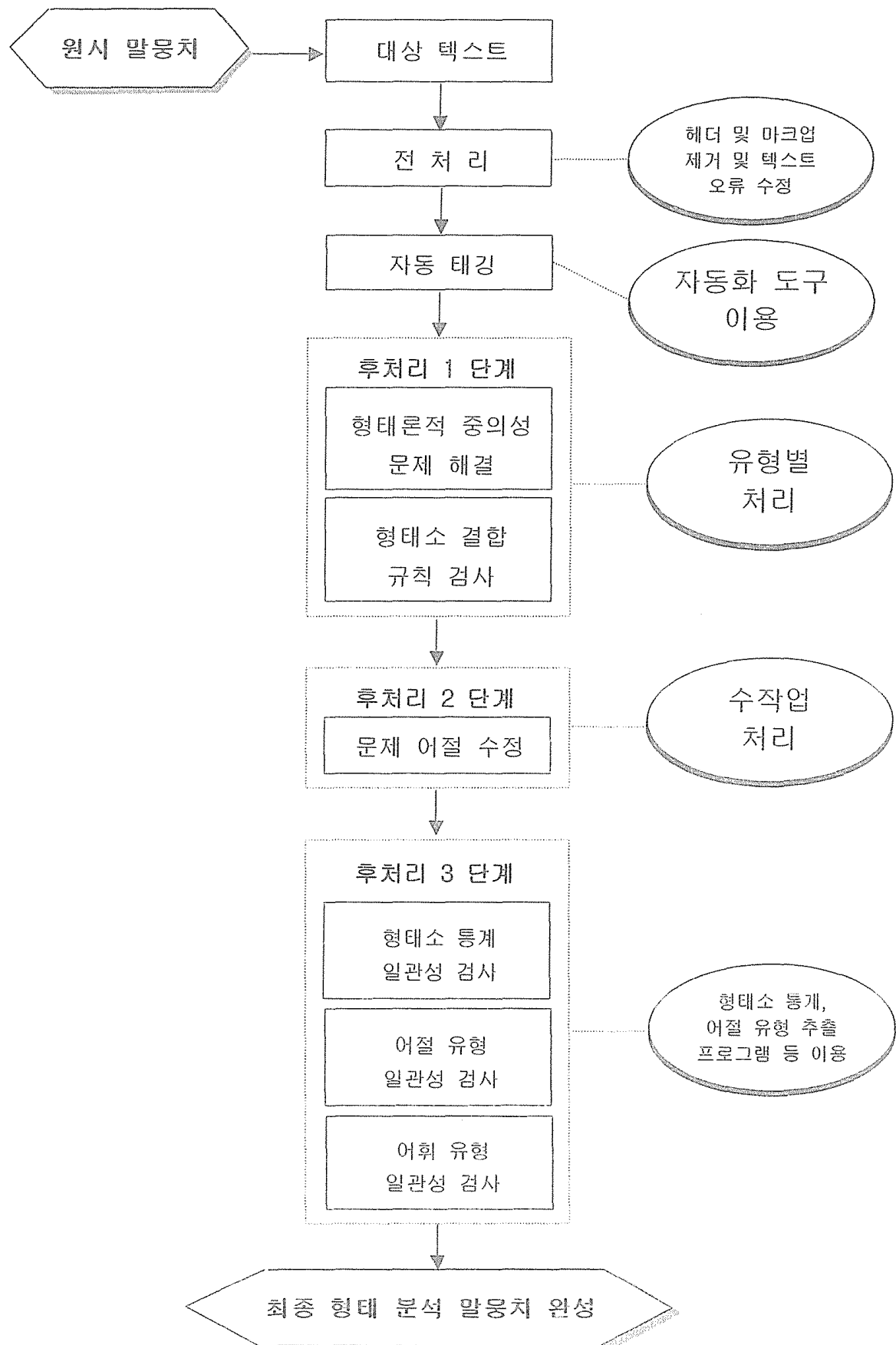
[자동 수정 불능 오류] 태거에서 처리하기 어려운 어절 유형 (38,201 어절)

(3) 통계 및 유형별 일관성 검토 (후처리 3단계)

형태소 통계 목록을 이용한 형식 오류 검사'와 '어절 유형에 따른 오류 검사' 등을 통해서, 수작업으로 진행되는 후처리 과정 가운데 생길 수 있는 작업자들 간의 분석 결과의 편차를 줄이고자 하였다.

(4) 통계 추출

최종 오류 수정을 마친 200만 어절의 형태분석 말뭉치를 토대로 각종 유용한 통계 정보를 추출한다. 이 과정에서 통계 자료를 검토함으로써 형태분석 말뭉치의 오류가 다시 드러날 가능성도 있는데, 이때 드러난 오류에 대해서도 다시 후처리를 함으로써 형태분석 말뭉치의 분석 성공률을 더욱 향상시키도록 하였다.



[그림 1] 형태분석 말뭉치 구축 과정

다. 통계 결과

200만 어절 규모의 형태분석 말뭉치에서 각 분석 표지별 항목과 빈도를 고빈도순으로 나열한 결과를 제시한다.

일반명사(NNG)	위	2070	머리	1388	학교	1042
	남자	2017	고개	1376	나라	1030
타입: [36747]	정도	1995	방	1372	내용	1022
토큰: [972652]	여성	1992	삶	1363	정신	1019
	모습	1980	운동	1359	애	1009
말	경우	1964	지역	1335	이번	1006
사람	사이	1942	시민	1304	밖	1001
때	얘기	1928	물	1289	이후	1001
일	다음	1904	민족	1287	상황	999
생각	아버지	1878	시장	1286	혼자	976
여자	길	1868	엄마	1285	이해	974
속	곳	1838	힘	1263	교육	970
자신	필요	1827	남편	1258	기억	969
집	인간	1801	이름	1253	단체	958
눈	환경	1775	소설	1235	문	958
문제	이상	1740	정치	1221	배	957
소리	날	1693	기자	1218	존재	951
세계	돈	1675	글	1196	대답	950
영화	정부	1663	기업	1191	친구	942
아이	세상	1658	목소리	1168	결과	940
시작	그때	1606	순간	1167	정책	933
손	어머니	1596	과정	1155	주장	930
얼굴	입	1592	가슴	1141	감독	928
앞	문화	1585	책	1132	인터넷	926
이야기	처음	1584	밤	1129	관련	915
몸	자리	1545	차	1110	생활	904
사회	동안	1493	술	1102	표정	900
마음	국가	1486	시대	1101	옷	898
시간	의미	1473	현실	1090	개발	897
전	후	1452	아내	1075	정보	890
사랑	관계	1441	전화	1068	땅	888
사실	점	1440	끝	1047	가치	887
안	지금	1406	활동	1045	문학	885
뒤	경제	1400	가운데	1043	오늘	883
선생	역사	1393	이유	1042	면	877

연구	877	나무	744	이제	661	학생	601
뜻	875	사업	740	전체	661	평가	599
도시	872	하늘	738	조사	657	아파트	597
옆	870	자유	735	얼마	653	잠	594
작품	868	성	732	하루	653	문명	593
결혼	862	세기	732	회장	653	개	590
사건	862	남	731	글로벌	649	인생	587
관심	861	느낌	726	달	645	아들	583
마지막	856	아침	726	표현	641	대부분	582
대학	852	작가	726	담배	640	딸	582
체제	850	자체	724	꽃	637	생산	581
결정	843	나이	723	대표	637	숨	581
상태	840	역할	721	투자	636	걱정	577
의식	828	당시	718	눈물	635	반대	572
사용	821	바람	717	요구	635	준비	570
장	818	최근	713	인사	633	행동	570
설명	812	거리	711	음악	632	냄새	569
주인공	812	발전	709	방식	628	비교	567
자연	809	기술	702	인식	626	정상	566
모양	805	인물	696	노래	625	경험	564
서구	802	국제	689	시절	625	귀	564
바다	792	웃음	685	저녁	624	국내	560
꿈	790	조약	685	입장	622	업체	560
지배	786	법	682	개인	621	노력	559
마을	784	해	682	확인	621	계획	556
사내	783	식민	681	참여	620	비	555
구조	779	상대	679	마찬가지	619	신	553
아래	775	중심	675	그날	615	생명	548
회사	773	구성	674	피	613	장면	545
기분	756	방법	673	예	609	관리	544
불	753	시	672	과거	608	손가락	542
대상	752	집단	667	가족	607	빛	541
변화	752	입술	666	발	607	컴퓨터	539
부분	746	조금	666	죽음	607	작업	537
대신	745	그림	664	금융	604	제도	533

관	533	공간	492	공동체	455	프로그램	420
다리	532	규모	490	방문	455	동네	418
성공	531	계속	489	연출	455	오후	417
행복	531	밑	489	차이	454	남성	416
대통령	529	손톱	489	미래	453	지원	416
종교	529	시선	489	언니	453	포함	415
부모	527	은행	489	행위	452	주제	414
분석	527	형성	488	마련	451	감정	413
교수	526	공부	487	조직	451	주	413
목	526	동시	487	고모	450	아무것	412
효과	526	반	485	기록	450	질서	412
발견	525	배우	485	건설	445	무대	411
전쟁	524	능력	483	발표	445	전달	411
조정	524	고향	481	개념	443	책임	411
고통	523	전통	480	주체	443	현재	407
정체	523	팔	480	회담	443	희망	403
사진	521	병	478	차원	442	불안	402
연기	521	여인	475	밥	441	강조	401
지적	520	자	474	말씀	439	도착	400
분위기	517	기관	473	예술	439	개혁	399
의사	517	약속	472	제시	438	목적	399
과학	516	사고	470	자식	437	저항	399
국민	514	비판	469	유지	436	대화	398
위기	514	사장	468	잔	436	사정	397
가버년스	513	상상	468	자료	434	성장	397
논문	513	논의	466	요즘	431	지방	397
기대	511	우리나라	465	물론	430	체결	395
태도	507	선택	463	주인	429	경찰	393
독자	505	등장	462	산업	427	지난해	391
일반	504	이용	462	이론	426	바닥	390
어둠	502	일부	461	수준	424	병원	390
해결	502	현상	460	산	422	최고	390
진행	501	주민	458	영향	422	아버	389
체계	496	공사	457	기회	420	영역	389
어깨	494	욕망	457	자동차	420	벽	387

자금	387	눈길	358	그동안	338	예정	322
서비스	386	스님	358	기존	338	이태을로기	322
예상	385	인상	358	제공	338	도움	321
상처	384	부인	357	탓	338	의자	320
여름	384	전망	357	처녀	337	권력	319
할머니	384	관객	356	머리카락	336	한편	319
의견	383	이전	356	온몸	336	세월	318
지식	382	집안	356	의지	335	기능	317
한마디	381	위험	355	차지	334	기사	317
걸음	380	인정	355	독립	333	연결	317
질문	379	작용	355	동물	333	경영	316
차례	379	추진	353	이사	333	경쟁	316
시각	378	어른	352	청년	333	동생	316
분야	377	노인	351	가게	332	촬영	316
등	376	운명	351	기본	332	학년	316
발생	374	상	349	서양	332	무역	315
택시	372	올해	349	마당	331	문서	315
물건	371	인류	349	전문가	331	부탁	315
오빠	371	진짜	349	화	331	원래	315
한참	371	유전자	348	공연	329	시기	313
신문	370	지속	347	근본	329	제목	313
회의	369	겨울	346	실패	329	맛	312
이미지	368	선	346	직원	329	제국주의	312
모두	366	게임	345	눈치	328	연대	311
어미	366	주변	345	사례	328	적극	311
형	366	음식	344	해석	327	기계	310
규정	365	나중	343	운영	326	모델	309
외국	364	처리	343	통일	326	새	309
편지	364	방향	342	언어	325	집중	309
조건	363	구체	341	판단	325	파괴	309
전문	362	장관	341	행사	325	부족	308
버스	361	단계	340	옛날	324	새벽	308
한	360	며칠	339	내일	323	자세	307
정리	359	여행	339	우주	323	주의	307
혁명	359	절	339	미소	322	노동	306

건강	305	형태	293	논리	276	전국	263
공장	303	식사	292	역	276	공식	262
기준	303	확대	292	제작	276	살	262
생태	303	약	291	보호	275	포기	262
위원회	303	부정	290	충	275	춤	261
커피	303	이날	290	투명	275	방안	260
목표	302	진실	290	거부	274	안정	260
보고	302	현장	290	거울	273	배경	259
외교	302	인기	289	소녀	273	예전	259
원인	301	새끼	288	대꾸	272	창	259
조심	301	건물	287	반면	272	텔레비전	259
지구	301	수	287	벤처	272	비밀	258
사태	300	제기	287	정의	272	전자	258
소개	300	코	287	틈	272	결	257
일상	300	한숨	287	긴장	271	나머지	257
사	299	혀	286	무릎	271	영혼	257
마누라	298	현대	286	주가	271	키	257
세대	298	출신	285	기	270	관계자	256
시도	298	사상	284	까닭	270	섬	256
주식	298	유형	283	내부	270	신경	256
기간	296	팀	283	스스로	270	해외	256
동양	296	서로	282	조작	270	통신	255
형식	296	선거	282	시인	269	방송	254
공동	295	칼	282	의심	269	심리	254
식민지	295	눈빛	281	숲	268	인수	254
전략	295	거절	280	주목	268	핵심	253
침대	295	의원	280	갈등	267	가방	252
선배	294	죄	280	교사	267	세력	252
성격	294	허리	280	소	267	작년	252
요인	294	파전	279	외국인	266	출	252
짓	294	당장	278	호텔	266	출발	252
평화	294	소식	278	짐작	265	풍경	252
가정	293	왕	277	반응	264	문장	251
불빛	293	주위	277	환자	264	이때	251
아빠	293	한쪽	277	봄	263	제일	251

표시	251	원칙	238	기획	228	공개	220
검찰	250	장소	238	만족	228	속도	220
거래	249	틀	238	부문	228	자원	220
관점	249	확보	238	완성	227	잘못	220
위원장	249	교회	237	근거	226	필자	220
자본	249	인	237	물음	226	무리	219
총선	249	화자	237	실제	226	시골	219
감사	247	부담	236	심정	226	식당	218
그룹	247	강	235	육체	226	증양	218
꿀	247	모순	235	주도	226	충격	218
싸움	247	소년	235	체험	226	측면	218
지도	247	소문	235	과제	225	표	218
홍분	247	수사	235	관능	225	요소	217
모임	246	해당	235	구별	225	유일	217
무시	246	불행	234	땀	225	졸업	217
섹스	246	공	233	비명	225	철학	217
맥주	245	광고	233	쓰레기	225	특수	217
반복	245	눈앞	233	접근	225	학문	217
상징	244	수행	233	중간	225	만화	216
의도	244	만남	232	차별	225	부부	216
이동	244	목숨	232	평소	225	청소년	216
이마	244	합의	232	한계	225	협상	216
위치	243	노릇	231	흐름	225	개선	215
직장	243	본격	231	과장	224	근대	215
도로	241	절대	231	최초	224	다음날	215
불편	241	요청	230	학자	224	완전	215
공포	240	우려	230	고민	223	창조	215
남녀	240	작성	230	대중	222	혼란	215
보장	240	증가	230	분노	222	환상	215
이익	240	할아버지	230	선정	222	가격	214
가을	239	비용	229	침묵	222	본질	214
귀신	239	상품	229	가난	221	교실	213
시험	239	안전	229	보고서	221	군	213
만일	238	언론	229	칭찬	221	여유	213
박사	238	극장	228	활용	221	감각	212

단어	212	수입	203	값	196	처지	190
기운	211	시스템	203	답론	196	도입	189
업무	211	오늘날	203	색깔	196	실장	189
형제	211	젊은이	203	짐승	196	태양	189
강화	210	구멍	202	고려	195	구분	188
민주	210	담당	202	비디오	195	서한	188
연구소	210	명예	202	사이버	195	화면	188
돌	209	이웃	202	선언	195	결심	187
폭력	209	전개	202	제외	195	사항	187
후보	209	전제	202	주먹	195	오전	187
그림자	208	제안	202	각종	194	위협	187
너머	208	화장	202	계단	194	이장	187
책상	208	경기	201	드라마	194	핵	187
결론	207	기초	201	오랫동안	194	향기	187
골목	207	당	201	정권	194	회원	187
인연	207	대형	201	종합	194	주문	186
최대	207	물질	201	소비	193	눈동자	185
한동안	207	상당	201	슬픔	193	사이트	185
계기	206	자극	201	아가씨	193	소비자	185
의료	206	군대	200	연락	193	손바닥	185
평생	206	머릿속	200	과	192	흥미	185
행정	206	인구	200	교섭	192	근무	184
화장실	206	짐	200	급	192	대책	184
극복	205	대안	199	식구	192	매력	184
근처	205	특징	199	신비	192	발달	184
동기	205	대회	198	영화제	192	사무실	184
어린이	205	마음속	198	피해	192	설정	184
엉덩이	205	적용	198	미술	191	전기	184
연극	205	추구	198	사전	191	진술	184
연상	205	파악	198	술집	191	물고기	183
출연	205	동원	197	연인	191	바탕	183
거짓말	204	번역	197	추가	191	소재	183
노예	204	시설	197	고집	190	타인	183
발	204	안개	197	영원	190	믿음	182
평등	204	연습	197	직접	190	부실	182

양	182	재산	176	합리	170	공공	162
유리	182	기쁨	175	개봉	169	다행	162
활동가	182	절차	175	부채	169	색	162
낮	181	트럭	175	분리	169	실시	162
대사	181	한국인	175	뿌리	169	중단	162
실천	181	결	174	스크린	169	통화	162
자랑	181	검토	174	용기	169	판매	162
직업	181	설득	174	고기	168	버릇	161
평균	181	증권	174	대목	168	전설	161
고객	180	미스	173	반영	168	견해	160
상상력	180	민주주의	173	여관	168	그릇	160
회복	180	보편	173	제한	168	글자	160
선물	179	수용	173	진출	168	무덤	160
신음	179	인용	173	쾌감	168	보험	160
완벽	179	작정	173	겹	167	소과	160
주부	179	특정	173	긍정	167	신고	160
참석	179	도덕	172	오염	167	종	160
공무원	178	상승	172	통합	167	하나님	160
별	178	실수	172	시나리오	166	햇빛	160
비행기	178	음성	172	장르	166	감동	159
신앙	178	이불	172	한때	166	국장	159
전환	178	인간관계	172	수단	165	기도	159
제국	178	정	172	식품	165	대응	159
창문	178	특성	172	초등학교	165	먼지	159
거실	177	순수	171	고문	164	바위	159
교장	177	이성	171	뺨	164	아저씨	159
구석	177	장치	171	여기저기	164	어사	159
불교	177	조각	171	영상	164	자전거	159
성과	177	공기	170	지지	164	정서	159
심장	177	마법	170	토론	164	개입	158
절망	177	스타일	170	옥구	163	공격	158
종이	177	아픔	170	음력	163	공법	158
구경	176	용어	170	점심	163	동의	158
움직임	176	재미	170	코스닥	163	조명	158
일종	176	통과	170	편안	163	지식인	158

진리	158	주택	153	키스	149	소주	144
채권	158	흔적	153	품	149	수도	144
형편	158	계약	152	꼬리	148	어려움	144
기반	157	놀이	152	이하	148	어젯밤	144
장애	157	되풀이	152	좌우	148	외부	144
폭	157	묘사	152	카메라	148	울음	144
공급	156	비난	152	금지	147	냉장고	143
빛	156	서술	152	노파	147	동료	143
아기	156	수출	152	상영	147	빠	143
자살	156	신부	152	시집	147	설립	143
증거	156	장난	152	언덕	147	양식	143
충돌	156	카드	152	연	147	영향력	143
허공	156	확신	152	잡지	147	일정	143
기구	155	계산	151	지	147	입구	143
당황	155	기독교	151	협력	147	자본주의	143
소득	155	다방	151	객관	146	지점	143
국회	154	대비	151	고등학교	146	건축	142
권리	154	마루	151	기여	146	계급	142
농사	154	바깥	151	나비	146	농담	142
대립	154	부여	151	도전	146	단지	142
등급	154	옥	151	언급	146	당국	142
강요	153	절반	151	외로움	146	습관	142
경계	153	피곤	151	장기	146	열차	142
군사	153	두려움	150	질	146	운전	142
권위	153	떼	150	고독	145	조성	142
닭	153	실현	150	부장	145	한잔	142
등록	153	주머니	150	살인	145	규제	141
바보	153	관념	149	설치	145	변호사	141
불만	153	바닷가	149	자국	145	보도	141
비극	153	부엌	149	각국	144	서방	141
상호	153	사과	149	내년	144	시내	141
시점	153	시행	149	덕분	144	의무	141
에너지	153	여부	149	머리칼	144	의존	141
원리	153	위원	149	명령	144	종일	141
조치	153	종류	149	복도	144	짜증	141

크리스마스	141	그리움	136	시계	133	강의	129
호장	141	수술	136	양국	133	구름	129
말투	140	실험	136	윤리	133	극	129
살림	140	염두	136	치료	133	몫	129
수익	140	영감	136	해방	133	법칙	129
이틀	140	인증	136	개성	132	변경	129
자아	140	임금	136	고정	132	비율	129
재정	140	전투	136	공원	132	용서	129
제품	140	지붕	136	비즈니스	132	주사	129
조화	140	턱	136	수업	132	추세	129
허용	140	통제	136	신화	132	가구	128
관찰	139	혐의	136	애니메이션	132	동작	128
본능	139	고백	135	중학교	132	부자	128
애인	139	낙선	135	폭발	132	이의	128
영업	139	보수	135	발전소	131	지혜	128
투쟁	139	연속	135	부사	131	창출	128
피부	139	욕심	135	상자	131	초점	128
가	138	의문	135	전부	131	경향	127
녹색	138	접촉	135	주주	131	넋	127
매각	138	종목	135	차림	131	담	127
발휘	138	확장	135	호기심	131	배제	127
순서	138	구원	134	가사	130	상대방	127
지시	138	날씨	134	구축	130	수정	127
귀국	137	논	134	네트워크	130	실질	127
마디	137	대문	134	단위	130	웹	127
면접	137	도달	134	디지털	130	중년	127
방울	137	우연	134	무게	130	후배	127
봉투	137	전반	134	범죄	130	비중	126
자산	137	중	134	상관	130	사물	126
장사	137	희생	134	소장	130	열정	126
추억	137	고생	133	예측	130	유행	126
충동	137	노동자	133	자격	130	의혹	126
확산	137	바지	133	정작	130	일자	126
곳곳	136	비닐	133	햇살	130	일치	126
규칙	136	생존	133	흙	130	제사	126

참고	126	인형	123	금	119	하	117
침	126	전형	123	농촌	119	가수	116
쿼터	126	초기	123	대변	119	계열사	116
크기	126	통치	123	똥	119	금리	116
클럽	126	풀	123	모자	119	대접	116
펀드	126	개방	122	식탁	119	발길	116
호흡	126	구입	122	열	119	선수	116
강물	125	국적	122	우산	119	수령	116
구상	125	농산물	122	이사회	119	술잔	116
모색	125	농약	122	지향	119	여학생	116
발신자	125	도구	122	호수	119	재벌	116
성적	125	도리	122	고용	118	출근	116
센터	125	문제점	122	곡	118	간직	115
압력	125	민족주의	122	극단	118	검사	115
오랜만	125	사막	122	디자인	118	계층	115
진보	125	스타	122	방해	118	곤란	115
총	125	실력	122	빵	118	기원	115
탁자	125	오줌	122	숨씨	118	논쟁	115
투입	125	장군	122	외면	118	단편	115
학습	125	제출	122	왼쪽	118	돼지	115
허락	125	지분	122	원주민	118	보유	115
구두	124	착각	122	질병	118	본래	115
도둑	124	최종	122	최소한	118	분열	115
무한	124	편	122	파업	118	악	115
발언	124	훈련	122	걸	117	안내	115
오해	124	결합	121	그중	117	예정	115
축제	124	식민주의	121	노조	117	약국	115
기색	123	연주	121	마무리	117	자존심	115
당숙	123	감	120	매일	117	개최	114
동	123	그늘	120	무력	117	결혼식	114
로비	123	서류	120	보상	117	모래	114
밀레니엄	123	연애	120	세포	117	벌레	114
사망	123	취재	120	정원	117	사원	114
소유	123	파도	120	주재	117	상실	114
요리	123	고대	119	탄생	117	숫자	114

영웅	114	독점	111	쫓	109	대체	105
예산	114	민간	111	캐릭터	109	도서관	105
우수	114	방침	111	테이블	109	동화	105
지급	114	뱀	111	효율	109	몸짓	105
집착	114	번호	111	가짜	108	무기	105
털	114	분업	111	외무성	108	세금	105
협회	114	심사	111	유리창	108	알	105
회견	114	억압	111	입학	108	여왕	105
흥행	114	예감	111	전공	108	연구원	105
경고	113	음모	111	직	108	열기	105
광장	113	점수	111	폐기물	108	증경	105
양쪽	113	편집	111	표준	108	지수	105
예외	113	글쓰기	110	결국	107	확정	105
인력	113	날개	110	관료	107	경찰서	104
지위	113	반발	110	광경	107	공유	104
후회	113	발걸음	110	보통	107	길이	104
관	112	백인	110	불꽃	107	물가	104
교통	112	신발	110	상식	107	부각	104
권태	112	양심	110	어제	107	빛깔	104
다짐	112	영화인	110	우울	107	음반	104
대출	112	오른쪽	110	강아지	106	읍내	104
붕괴	112	유혹	110	관광	106	이슈	104
승리	112	장수	110	막내	106	이혼	104
아줌마	112	진정	110	물결	106	인권	104
연합	112	침략	110	부	106	통상	104
운전사	112	포인트	110	사기	106	포장	104
접속	112	합병	110	서계	106	해체	104
집행	112	화해	110	시사	106	황금	104
초월	112	갯벌	109	유학	106	고급	103
햇바닥	112	교환	109	일본인	106	고장	103
화제	112	남쪽	109	장식	106	고전	103
흑인	112	명분	109	철	106	과일	103
개정	111	식량	109	가시	105	노출	103
구절	111	왕자	109	공정	105	뉴스	103
누나	111	일행	109	끈	105	뱃속	103

봉사	103	수상	100	고유명사(NNP)	규혁	280
부임	103	양상	100		혜자	279
상기	103	역설	100	타입: [13380]	아랑	278
손길	103	잔치	100	토큰: [90316]	정	277
신뢰	103	전생	100		안토니	266
어조	103	줄기	100	조선	명준	254
치마	103	증명	100	한국	신애	246
친절	103	직후	100	미국	리홍장	239
하락	103	특유	100	일본	독일	235
흉내	103	현관	100	김	상덕	234
가슴속	102	형사	100	서울	최	228
공항	102			중국	형민	227
상거래	102			박	연홍	224
수석	102			영국	동환	221
수집	102			완	북	218
실망	102			회수	슈펠트	213
쌀	102			여경	이강구	211
직전	102			북한	부산	210
참가	102			이	한	201
청춘	102			인도	미	199
탑	102			경지	인천	197
활성	102			프랑스	짐	192
대기	101			범수	클레오파트라	192
살해	101			유럽	유희	189
설	101			러시아	평양	182
시체	101			미니	억균	181
암	101			다비	준기	181
작전	101			금순	고종	180
조항	101			현대	들개	180
텍스트	101			영어	고	179
공통	100			남북	민이	175
균형	100			미란	동준	170
발톱	100			영주	경화	166
비평가	100			아시아	대우	166
빈곤	100			배	장	166

아프리카	163	제이슨	113	의존명사(NNB)		만큼	716
사이드	162	용규	112			바	703
이집트	161	서연	110	타입: [270]		식	689
황만근	161	은옥	110	토큰: [137493]		살	656
심	159	강화도	109			놈	618
팍스	158	정임	109	것	42472	년대	597
남가이	153	순이	108	수	15347	듯이	557
현순	153	광주	107	년	6326	편	557
부칠	150	영월	107	때문	5189	대로	519
한민족	147	왕골	105	등	4192	지	478
김대중	141	홍콩	105	거	4165	측	463
조	141	워싱턴	103	씨	3231	내	461
송	140	수희	102	월	3083	녀석	446
김홍집	137	경주	100	번	3062	셈	404
동옥	134	월간중앙	100	일	2811	달러	367
한반도	134			데	2294	리	360
민	133			듯	2214	마리	354
상구	132			중	2198	따위	349
삼성	131			뿐	1788	초	349
부처	128			명	1778	판	336
로마	127			원	1686	법	310
할리우드	126			개	1464	무렵	309
임	123			채	1427	개월	307
강	122			대	1230	척	305
이동인	121			간	1226	말	303
햄릿	119			만	1205	차	300
김범우	118			가지	1167	호	285
대만	116			쪽	1165	이	282
정옥	116			적	1113	위	278
김정일	115			터	1082	외	262
도쿄	115			출	981	세	255
현대건설	115			자	930	바람	244
대우차	114			쭈	923	장	228
황	114			시	894	나름	214
예수	113			분	887	권	196

지경	184	대명사(NP)		지	195	수사(NR)	
점	170			그거	174		
마련	166	타입: [147]		이놈	165	타입: [265]	
양	165	토큰: [94248]		저쪽	146	토큰: [11757]	
군	163			저희	141		
모양	161	나	20864	그놈	139	하나	3127
개국	159	그	17188	저기	131	만	1410
건	152	그녀	8688	저것	129	억	960
평	147	우리	6024	그대	128	둘	629
퍼센트	143	그것	5687	여러분	113	천	395
회	142	내	4080	모	108	조	369
주일	139	이	3753	그분	107	열	308
참	134	무엇	2563	그니	106	일	307
이래	113	너	2426			둘째	297
짝	108	누구	2364			백	269
조	106	자기	2349			다섯	265
주	104	저	2053			십	192
미터	103	어디	1948			첫째	192
가량	101	뭐	1885			일곱	133
		여기	1582			여섯	128
		당신	1462			셋	121
		이것	1216			수십	118
		거기	1180			삼	109
		제	529			스물	108
		네	505			셋째	104
		그곳	496			서른	102
		이곳	470			이	101
		아무	403				
		이거	333				
		언제	312				
		자네	242				
		아무개	223				
		너희	206				
		니	200				
		이쪽	196				

동사(VV)		들어가	1767	관하	943	이루어지	581
		내	1758	생기	934	움직이	572
타입: [4237]		서	1644	돌아가	924	꺼내	570
토큰: [372101]		주	1606	사	892	흐르	568
		찾	1555	얻	891	바꾸	566
하	22660	잡	1472	밝히	850	자	564
있	22026	타	1342	알리	840	배우	562
되	14896	나타나	1334	사라지	818	취하	549
보	8526	맞	1319	뜨	803	차	530
알	5661	웃	1280	넣	788	달리	524
가	5525	갓	1246	끝나	777	말	524
대하	5493	바라보	1233	이르	774	지나가	522
들	4665	남	1196	올리	750	지키	520
위하	4299	보내	1193	돌리	748	붙이	519
보이	4255	떠나	1164	오르	738	죽이	511
모르	4245	기다리	1161	그리	734	눅	508
받	4241	부르	1159	나누	733	들어서	501
오	4020	들어오	1142	넘	723	벗어나	496
나오	3719	떨어지	1135	나서	721	말	494
살	3651	내리	1134	들리	713	대	492
그러	3136	나가	1115	걸리	701	받아들이	492
쓰	3061	열	1112	걸	699	지	491
만들	2944	일어나	1109	미치	677	붙	490
가지	2333	마시	1093	놓	663	깨닫	489
나	2267	돌아오	1090	놀라	649	멈추	485
따르	2248	치	1085	좋아하	641	다가오	481
느끼	2164	향하	1051	이루	637	바라	481
든	2133	두	1035	걸	631	여기	478
만나	2132	짓	998	열리	621	팔	477
먹	2079	믿	989	울	621	잇	475
지나	1976	다니	965	세우	608	변하	474
앞	1971	빠지	964	옮기	605	던지	471
죽	1931	읽	964	읽	599	쉬	471
통하	1870	의하	962	어찌	598	드러나	470
묻	1832	입	957	떠오르	585	잇	468

끝	466	찾아오	391	내려가	320	늘어나	265
놓이	465	피우	391	갖추	317	내려다보	264
맞추	463	걸치	390	젓	317	떨리	261
전하	462	돌	390	가리	316	살리	261
버리	459	바뀌	385	내려오	315	감추	260
비하	456	담	382	내놓	310	잡히	260
쳐다보	454	살펴보	381	깨	307	매달리	259
지니	453	가르치	379	중얼거리	307	싸우	258
일으키	452	구하	379	못하	302	기르	257
낱	450	닿	377	이끌	301	기울이	257
올라가	447	벌어지	375	채우	300	정하	256
참	445	인하	373	살피	299	누르	255
원하	434	벌이	369	넘기	298	속하	255
남기	426	터지	365	취	298	쏟아지	255
허	424	홀리	361	키우	296	알아보	254
살아가	423	지르	360	달라지	291	치르	254
안	422	시키	357	막히	291	불	253
겪	421	벗	354	끄덕이	290	넘어가	252
풀	421	돌아보	345	늑	290	가리키	251
드러내	418	감	344	나아가	288	걸어가	251
당하	416	견디	344	물	286	일어서	250
피하	416	자라	343	달	284	데리	249
이어지	415	마치	342	끓	281	따지	249
삼	414	막	342	맷	279	닭	247
지내	414	끼	340	빠	278	켜	245
모이	412	흔들	335	벌	275	흔들리	245
모으	411	떠올리	334	차리	274	부리	244
찍	410	어울리	333	올라오	273	이기	244
떼	408	놀	331	가져오	272	스치	243
태어나	404	애쓰	328	찾아가	272	뵙	241
다루	402	울리	328	그치	271	이러	240
거치	401	뛰	326	띄	270	잠들	239
떨	399	내밀	322	무너지	268	밟	238
비	397	들려오	322	마르	267	헤어지	238
즐기	396	들여다보	322	빌리	266	지켜보	236

담기	232	서두르	201	별리	181	묶	164
드리	232	꿈꾸	200	젓	180	붙잡	163
때리	231	달려가	199	문히	179	줄어들	163
심	230	버티	198	속삭이	179	다녀오	162
말기	228	신	198	잃어버리	178	바치	162
고치	226	실리	198	떠	177	놓치	161
생각나	224	씻	198	추	177	빛나	160
섞이	222	물어보	197	풍기	177	시달리	160
내다보	219	살아오	197	거두	176	빠져나오	159
다가가	219	삼키	197	뚫	176	새	159
썩	219	잠기	197	머물	176	주어지	159
쓰러지	218	달하	195	비치	176	씩	158
지치	218	더불	195	찌르	176	저지르	158
퍼지	218	머무르	195	돕	175	튀	158
숨	216	뿌리	195	뛰어들	175	흠치	158
따	215	집	195	적	175	더듬	157
몰	215	커지	195	줄이	175	먹이	157
외치	215	지우	192	감싸	174	박히	157
꾸	214	퍼	192	들르	174	망설이	156
단	214	챙기	190	털	174	찾아보	156
돌아서	214	마주치	189	넘치	173	택하	156
피	214	소리치	189	닭	173	따라가	155
고르	213	앞서	189	둘러보	173	더하	154
날리	213	찾아내	189	빠져나가	173	부딪치	154
들이	213	흘러나오	188	깨어나	172	펼쳐지	154
숙이	213	높이	187	신	172	가라앉	153
밀	211	펼치	187	다물	170	날아가	153
덧붙이	208	품	187	숨기	170	터뜨리	153
늘	206	쏟	185	쓰이	170	갈히	152
블리	205	끝내	184	헤매	169	몰리	152
썩이	205	모시	184	비우	168	앞두	152
폴리	205	알아듣	184	건너	167	쳐하	152
덮	203	나타내	183	만지	165	물러나	151
빨	202	자르	183	흘러내리	165	태우	151
그만두	201	건네	182	둘러싸	164	박	150

싸	150	바르	130	질리	115	깨물	103
내려놓	149	번지	130	끓어지	114	쫓기	103
일	149	흘러가	130	도망치	114	괴롭히	102
씩	148	끼치	129	틀리	112	늑	102
묵	147	주고받	129	끓기	111	달아오르	102
잊어버리	147	튀어나오	129	노려보	111	미루	102
말리	146	달래	128	늘어지	111	부풀	102
생겨나	146	떠돌	128	모자라	111	살아남	102
가져가	145	솟	128	팔리	111	꼭	101
가하	144	움켜쥐	128	깨지	110	두려워하	101
갈리	143	흘어지	128	꾸미	110	맛보	101
틀	143	기	127	두르	110	안기	101
기대	142	달아나	126	떨어뜨리	110	끼우	100
밀리	141	돌아다니	126	뜯	110	날	100
짚	140	벗기	125	사로잡히	110	내버리	100
계시	139	비추	125	입히	110		
꺼지	139	올려다보	125	꺾	108		
빨	139	옆드리	124	놀리	108		
오가	138	짜	124	엎	108		
꽃	137	값	123	올려놓	108		
얹	137	갈	123	주저앉	108		
내쉬	136	행하	123	찢	108		
넘어서	136	걸어오	122	합치	108		
싫어하	136	끄	122	해보	108		
내뱉	135	토하	122	내세우	107		
배	134	두드리	121	줄	107		
집어넣	133	쓸	121	권하	106		
늘어놓	132	건드리	119	달려오	106		
뒤지	132	당기	119	없애	106		
쓰다듬	132	맞서	119	채	106		
활	132	끓이	118	메우	105		
누리	131	떠들	118	새기	105		
도와주	131	빌	117	드나들	104		
접어들	131	붓	115	엮히	104		
매	130	씹	115	치우	104		

형용사(VA)		적	403	의롭	204	착하	106
		빠르	402	곱	193	뒤늦	105
타입: [915]		뜨겁	397	슬프	192	묘하	105
토큰: [85568]		어둡	397	저렴	183	거세	104
		늦	388	질	183	맛있	104
없	14979	붉	387	낱	180	싸	103
갈	7726	무섭	367	날카롭	179	못하	102
그렇	6317	낮	354	멋지	178	비싸	100
크	2991	고맙	352	편하	176		
중	2966	짧	347	놀랍	171		
많	2902	밝	337	환하	166		
어떻	2320	부드럽	335	더럽	161		
이렇	1856	가볍	327	약하	160		
새롭	1305	낯설	326	안타깝	158		
길	1198	넓	309	출	156		
어렵	1089	지나치	305	드물	154		
작	976	팬찮	301	부끄럽	152		
높	959	심하	285	곤	145		
다르	924	똑같	278	빨강	142		
쉽	886	무겁	269	가늘	135		
아름답	839	엄청나	264	흔하	135		
어리	729	푸르	261	뛰어나	133		
멀	704	틀림없	260	조그맣	133		
젊	692	바쁘	256	귀엽	126		
깊	672	맑	249	반갑	123		
힘들	617	거칠	238	뻘하	123		
가깝	602	차갑	234	କୁ	120		
하얗	500	낮	233	급하	118		
싫	486	재미있	233	노랑	118		
검	481	옳	226	귀찮	117		
아프	475	수많	215	아무렇	117		
커다랗	426	예쁘	214	우습	115		
강하	420	좁	214	기쁘	114		
나쁘	416	즐겁	211	그립	107		
희	407	두렵	210	상관없	107		

보조용언(VX)		긍정지정사(VCP)		부정지정사(VCN)		관형사(MM)	
타입: [69]		타입: [3]		타입: [2]		타입: [142]	
토큰: [108103]		토큰: [91365]		토큰: [9987]		토큰: [62645]	
있	27842	이	91362	아니	9966	그	16169
하	17255	히	2	아이	21	이	7381
않	15508	아	1			한	7112
지	8080					그런	3553
보	6112					두	3428
주	6045					다른	2944
못하	4106					어떤	2355
싶	3416					이런	2264
오	3119					어느	1966
가	2829					모든	1934
버리	2601					무슨	1730
말	2065					몇	1640
놓	1939					저	942
내	1650					여러	918
나	1160					세	803
나가	923					아무	742
두	812					전	644
대	810					첫	512
달	453					새	504
들	445					각	365
먹	161					단	301
드리	158					아무런	297
가지	130					네	270
계시	119					오랜	231
						온	220
						충	216
						별	212
						약	206
						옛	184
						온갖	183

맨	176	일반부사(MAG)		곧	909	달리	456
한두	160			물론	855	벌써	454
을	158	타입: [3198]		거의	851	아마	430
웬	120	토큰: [136462]		그냥	824	참	425
저런	113			및	823	어쩌면	420
현	108	더	4175	결국	808	다만	413
몇몇	101	또	3370	그대로	753	어찌	398
이런저런	100	안	3336	늘	741	천천히	390
		다시	3146	꼭	733	가끔	380
		다	2969	전혀	704	잘못	369
		잘	2655	오히려	700	과연	368
		없이	2005	계속	684	빨리	366
		왜	1875	서로	684	멀리	361
		좀	1793	즉	677	너무나	355
		함께	1704	잠시	639	완전히	349
		가장	1673	매우	626	도대체	347
		바로	1618	역시	591	겨우	346
		이제	1551	그저	580	그리	346
		아직	1525	조금	580	자주	340
		못	1512	오래	571	별로	336
		너무	1359	아무리	556	문득	335
		이미	1335	그만	540	잠깐	331
		지금	1275	훨씬	540	마침내	327
		모두	1267	제대로	527	단지	324
		마치	1087	언제나	518	반드시	322
		정말	1078	직접	517	언젠가	313
		아주	1072	사실	502	비로소	309
		같이	1049	여전히	497	일단	309
		많이	1027	현재	485	혹시	307
		특히	984	우선	479	실제로	302
		또한	970	점점	478	자꾸	300
		더욱	944	스스로	471	분명히	292
		먼저	943	결코	469	새로	288
		얼마나	918	혹은	464	언제	288
		갑자기	912	가득	463	꽤	284

아니	279	내내	208	덜	152	다시금	122
보다	278	어느새	203	다소	151	특별히	122
가까이	273	그만큼	201	홀로	150	어쩐지	121
열심히	269	무척	200	그제야	149	이어	121
약간	268	분명	198	보통	149	막상	119
때로	266	아마도	194	예컨대	146	몰래	119
게다가	265	도저히	193	거듭	145	소위	119
가만히	263	딱	193	솔직히	145	그야말로	118
또는	261	만약	191	또다시	142	위낙	118
막	255	조용히	191	마구	142	유난히	118
항상	253	아까	187	어제	142	감히	116
차라리	252	굳이	186	제발	141	깜짝	116
그토록	250	참으로	186	한껏	141	재빨리	115
미리	248	상당히	181	단순히	139	채	115
아예	247	영원히	181	오늘	138	그다지	114
얼른	244	각각	179	살짝	134	무조건	113
주로	243	왜냐하면	179	일부러	134	차츰	113
어쨌든	240	도무지	177	텅	133	불쑥	111
아무래도	238	일찍	174	나란히	132	팬히	110
오직	238	자세히	173	대개	131	말없이	109
이른바	236	새삼	171	벌떡	131	대단히	107
깊이	234	웬지	169	우연히	131	불과	107
매일	232	끊임없이	166	방금	130	어쨌서	107
충분히	232	마침	166	곧바로	128	이따금	107
마주	231	하여금	166	한꺼번에	128	대체로	106
잔뜩	230	은통	165	이내	126	높이	105
몹시	229	제법	164	서서히	125	다행히	104
적어도	229	끝내	163	실은	125	간신히	103
금방	227	꽤	162	어차피	125	꼼짝	103
어서	227	더구나	162	이리저리	125	미처	102
비록	218	당연히	161	정확히	125	이윽고	102
흔히	216	절대로	158	한편	125	가령	101
드디어	215	심지어	157	그나마	123	좀처럼	101
정말로	214	확실히	156	이리	123	널리	100
따로	213	아무튼	154	거꾸로	122		

접속부사(MAJ)	감탄사(IC)	주격조사(JKS)	보격조사(JKC)
타입: [42] 토큰: [17689]	타입: [448] 토큰: [6856]	타입: [6] 토큰: [123360]	타입: [2] 토큰: [11389]
그리고 4101	그래 725	이 72618	이 7324
그러나 3981	아 640	가 50300	가 4065
그런데 1839	아니 626	께서 266	
하지만 1805	뭐 461	서 133	
그래서 1545	네 340		
따라서 775	응 223		
역시 702	야 184		
그럼 442	자 168		
또는 438	글쎄 148		
그러니까 357	참 143		
그러면 357	아아 136		
그리하여 354	어디 132		
그렇지만 282	그럼 124		
그러므로 256	예 118		
하긴 156			

관형격조사(JKG)		목적격조사(JKO)		부사격조사(JKB)		호격조사(JKV)	
타입: [3]		타입: [4]		타입: [71]		타입: [7]	
토큰: [97218]		토큰: [163423]		토큰: [189582]		토큰: [735]	
의	97187	을	100885	에	77844	아	363
		를	59524	으로	27148	야	232
		르	3013	에서	26152		
				로	21451		
				에게	8819		
				처럼	5762		
				과	5372		
				와	4229		
				보다	3406		
				서	1727		
				로서	1043		
				한테	973		
				로부터	749		
				으로써	667		
				으로서	666		
				같이	548		
				으로부터	432		
				에게서	427		
				하고	414		
				만큼	413		
				에서부터	215		
				랑	168		
				게	144		
				로써	102		

인용격조사(JKQ)		보조사(JX)		접속조사(JC)		선어말어미(EP)	
타입: [4]		타입: [66]		타입: [16]		타입: [28]	
토큰: [2962]		토큰: [185508]		토큰: [24445]		토큰: [130873]	
고	1388	는	65648	과	12051	었	68376
라고	670	은	47589	와	8134	았	51291
이라고	525	도	34835	이나	2283	겠	5626
하고	379	ㄴ	8386	나	1756	시	3889
		만	8196			았었	694
		까지	5858			으시	412
		부터	2909			었었	284
		나	1535			졌	180
		마다	1273				
		밖에	1220				
		이나	1132				
		야	747				
		대로	688				
		이란	686				
		조차	654				
		뿐	611				
		다	572				
		요	541				
		두	428				
		마저	426				
		란	403				
		다가	221				
		이야	176				
		이야말로	155				

종결어미(EF)		라	301	지만	117	연결어미(EC)	
		구나	299	대	111		
타입: [1004]		소	291	ㄴ데요	106	타입: [917]	
토큰: [180244]		ㄴ데	290	니까요	105	토큰: [343828]	
		습니까	281	란다	105		
다	110076	군요	279	라니	104	고	66309
ㄴ다	16811	읍니다	274	느냐	100	어	43905
어	4229	거든	265			아	37923
어요	3452	군	239			게	27777
야	3181	잖아요	239			지	21292
지	2912	으니까	236			면	9497
아	2421	거든요	227			아서	8874
비니다	2400	아라	227			며	8156
습니다	2265	는데	215			지만	6535
르까	1950	나요	214			다	6533
는다	1796	르까요	193			면서	6046
ㄴ가	1614	라고	192			어서	4842
조	1310	다고	186			다고	4749
을까	1218	네요	184			라고	4453
아요	1186	더라	182			는데	4006
어요	1117	더군요	171			라	3897
지요	849	단다	170			아야	3725
는가	842	리라	166			어야	3194
자	825	유	165			자	3169
니	614	ㄴ가요	150			니	2775
고	590	게	142			ㄴ가	2751
오	570	다니	142			다가	2582
냐	535	은가	140			는지	2529
나	520	시유	139			ㄴ지	2380
네	471	여	136			으며	2343
비니까	402	어라	134			아도	2197
잖아	368	라구	132			으면	2187
던가	350	구요	130			라도	1780
니까	326	지유	125			도록	1649
요	312	으리라	122			ㄴ데	1644

ㄴ다고	1633	여	326	르지라도	118	명사형어미(ETN)
어도	1594	냐	321	으러	118	
다면	1593	자고	314	다며	106	타입: [9]
ㄴ다	1473	으므로	313	느라고	103	토큰: [23863]
거나	1406	려면	310	ㄴ들	101	
으나	1327	어요	286	노라고	101	기 18323
듯	1109	느라	280			ㅁ 4209
려고	1087	자마자	275			음 1152
ㄴ다면	1020	어다	259			라기 100
더니	996	건	227			
든	940	냐고	225			
나	928	으려고	220			
으니	909	므로	218			
르지	879	라며	205			
라면	825	으니까	201			
으면서	817	은데	195			
르까	764	는다	184			
려	759	다시피	184			
니까	706	구나	180			
듯이	690	습니다	178			
더라도	667	느냐고	171			
곤	613	어야만	165			
는가	613	는다고	161			
구	596	아야만	159			
러	591	아라	158			
요	576	으려	157			
고서	494	는다면	156			
야	461	다니	156			
느냐	432	라서	155			
르수록	406	어라	151			
을지	404	아요	147			
자면	395	네	146			
든지	369	믄	142			
고자	364	비니다	137			
을까	354	더라면	131			

관형형어미(ETM)		접두사(XPN)		명사·과생접미사(XSN)		쟁이	112
						기	111
타입: [73]		타입: [33]		타입: [54]		치	109
토큰: [269780]		토큰: [7463]		토큰: [72386]			
ㄴ	102933	제	1078	들	32569		
는	69241	불	991	적	18842		
르	34579	대	713	성	3417		
은	21021	무	697	화	2518		
던	11759	비	630	님	2335		
을	10301	재	523	네	1192		
다는	6579	신	304	씩	1096		
라는	5849	한	301	권	1003		
ㄴ다는	3104	초	259	여	792		
란	1174	반	228	계	776		
려는	1027	소	164	째	774		
는다는	378	탈	163	상	603		
자는	358	최	148	론	573		
단	323	맨	145	꾼	561		
ㄴ단	138	헛	143	생	514		
으려는	112	피	141	질	512		
냐는	101	부	132	제	391		
		고	115	용	334		
				형	323		
				층	269		
				짜리	266		
				를	259		
				덜	257		
				별	236		
				을	177		
				시	166		
				끼리	159		
				분	157		
				당	152		
				가	148		

동사파생집마시(XSV)		형용사파생집마시(XSA)		어근(XR)			
						깨끗	188
						익숙	188
타입: [6]		타입: [8]		타입: [1265]		평범	187
토큰: [103101]		토큰: [45572]		토큰: [34582]		시원	174
						동일	172
하	83738	하	42064	가능	1544	적절	169
되	16403	스럽	2073	중요	1187	희미	169
시키	2424	롭	855	이러	933	바람직	168
당하	339	답	301	그러	593	간단	165
혀	196	혀	170	이상	588	조용	160
				다양	551	답답	154
				비롯	441	어떠	150
				분명	435	어색	146
				비슷	427	독특	145
				심각	360	영똥	140
				특별	327	위대	139
				불구	321	대단	135
				미안	320	뚜렷	131
				주요	312	불쌍	130
				단순	311	강렬	124
				불과	305	정당	120
				당연	268	일정	119
				확실	261	무관	117
				거대	247	소중	116
				따뜻	244	억울	115
				유명	238	철저	115
				정확	237	풍부	114
				화려	237	끔찍	112
				복잡	230	달콤	112
				강력	222	치열	112
				진지	218	신기	111
				궁금	199	적당	111
				충분	193	무수	108
				훌륭	193	지루	102
				진정	192	사소	100

좌송	100	기호(SF)	기호(SP)	기호(SS)
		타입: [3]	타입: [10]	타입: [37]
		토큰: [198634]	토큰: [69558]	토큰: [115020]
		.	,	"
		182756	62584	46218
		?	·	'
		12759	4320	21095
		!	/)
		3119	1607	17405
			:	(
			818	15986
			;	-
			207	3017
]
				1925
				[
				1922
				>
				994
				<
				989
				『
				881
				』
				880
				「
				787
				」
				785
				—
				741
				、
				250
				—
				225
				-
				219
				—
				136

기호(SE)		기호(SO)		기호(SW)		외국어(SL)	
타입: [7]		타입: [9]		타입: [93]		타입: [3794]	
토큰: [5384]		토큰: [783]		토큰: [4412]		토큰: [14633]	
.....	4591	~	544	%	1837	com	568
...	675	~	168	@	704	chosun	501
				=	374	NGO	496
				◆	197	UN	345
				→	179	p	305
				&	121	Q	252
						A	229
						TV	226
						kr	204
						co	190
						IMF	173
						C	172
						K	165
						hani	163
						LG	137
						a	102

한자(SH)		숫자(SN)		000	241
				29	240
타입: [4780]		타입: [1504]		60	226
토큰: [6223]		토큰: [40999]		80	223
				24	213
美	21	1	3343	98	208
北	14	3	3072	1998	200
大	13	2	2860	90	191
性	13	4	1682	70	186
道	12	5	1592	0	176
義	12	10	1527	31	176
黃河	12	6	1178	99	173
家	11	2000	1134	02	172
氣	11	7	1003	23	171
反	11	8	949	22	170
神	11	20	930	27	170
新	10	30	841	26	169
銀	10	12	761	28	166
對	9	9	739	1995	158
朴漢濟	9	11	598	1997	149
富	9	45	520	200	138
寺	9	21	472	500	133
禪	9	15	465	97	125
姓	9	50	419	1990	121
靈	9	25	372	35	116
喜囚	9	100	342	05	112
射精	8	1999	341	1980	112
陽	8	13	333	33	104
陰	8	19	329		
前	8	16	325		
曹操	8	40	310		
楚	8	14	307		
胎	8	18	277		
		01	256		
		17	250		

참고문헌

- 김태훈. 2001. 「한국어 말모듬 구축 방안 연구」 고려대학교 석사학위 논문.
- 김홍규, 강범모. 2000. 「한국어 형태소 및 어휘 사용 빈도의 분석 1」 서울 : 고려대학교 민족문화연구원.
- 김홍규 외. 1998-2003. 「21세기 세종계획 연구보고서」 문화관광부.
- 임홍빈 외. 2002. 「국어 문법 현상의 계량적 연구」 서울: 고려대학교 민족문화연구원.
- 조진현. 2000. 「형태소 분석 말뭉치 구축을 위한 태깅 방안의 연구와 개발」 고려대학교 석사학위 논문.

〈부록〉 형태분석 말뭉치 구축 지침

(2004 version 1.0)

I. 기본 원칙

1. 분석 대상

형태분석은 하나의 어절을 분석 대상으로 한다.

2. 분석 원리

본 분석은 형태소 차원이 아닌 “형태” 차원의 분석이므로 이형태를 최대한 반영한다.

3. 분석 원칙

형태분석은 분석 대상인 원시 말뭉치를 가급적 훼손하지 않는다.

II. 어절 분석 표지

1. 이 어절 분석표지는(이하 세종 형태 표지) 21세기 세종계획 국어기초자료구축 분과의 ‘형태분석 말뭉치(morpheme tagged corpus)’를 구축하기 위해 마련된 것이다.

2. 이 분석 표지는 1998년도 21세기 세종계획의 세부과제 중 하나였던 “어절분석 표지의 표준안”과 한국전자통신연구원(ETRI, 1999)에서 마련한 “태그세트 표준안”을 고려하여 설계되었다. 또한 고려대학교 자연어처리연구실에서 자체적으로 개발한 표준안도 참고하였다.

3. 이 세종 형태 표지는 단계적인 분석을 할 수 있도록 고려되었다.

대분류	소분류	세분류
(1) 체언	명사NN	일반명사NNG 고유명사NNP 의존명사NNB
	대명사NP	대명사NP
	수사NR	수사NR
(2) 용언	동사VV	동사VV
	형용사VA	형용사VA
	보조용언VX	보조용언VX
	지정사VC	긍정지정사VCP 부정지정사VCN
(3) 수식언	관형사MM	
	부사MA	일반부사MAG 접속부사MAJ
(4) 독립언	감탄사IC	감탄사IC
(5) 관계언	격조사JK	주격조사JKS
		보격조사JKC
		관형격조사JKG
		목적격조사JKO
		부사격조사JKB
		호격조사JKV
		인용격조사JKQ
	보조사JX	보조사JX
(6) 의존형태	접속조사JC	접속조사JC
	어미E	선어말어미EP
		종결어미EF
		연결어미EC
		명사형전성어미ETN
		관형형전성어미ETM
	접두사XP	체언접두사XPN
	접미사 XS	명사파생접미사XSN
		동사파생접미사XSV
		형용사파생접미사XSA
	어근XR	어근XR
(7) 기호	마침표, 물음표, 느낌표	SF
	쉼표, 가운뎃점, 콜론, 빗금	SP
	따옴표, 괄호표, 줄표	SS
	줄임표	SE
	불임표(물결, 숨김, 빠짐)	SO
	외국어	SL
	한자	SH
	기타 기호(논리 수학기호, 화폐 기호) 등)	SW
	명사추정범주	NF
	용언추정범주	NV
	숫자	SN
	분석불능범주	NA

Ⅲ. 표지별 분류 기준 및 세부 지침

1. 체언

체언은 명사, 대명사, 수사를 포괄하는 대범주로서, 조사와 결합하거나 그 자체로 다른 체언이나 용언과 어울려 하나의 문장성분이 될 수 있다.

1.1. 명사(NN)

명사는 사물의 이름을 나타내는 품사이다. 본 표지에서는 명사를 일반명사, 고유명사, 의존명사로 세분한다.

1.1.1. 일반명사(NNG)

사물의 이름을 나타내는 단어로서 고유명사와 의존명사를 제외한 모든 명사가 이에 포함된다.

※ 명사상당어의 태깅

- ① 동사의 활용형이 따옴표 없이 문장 속에서 명사처럼 기능하는 경우에도 원래 품사대로 분석한다.

예) 어디 가느냐가 그의 물음이었다. [가/VV+느냐/EC+가/JKS]

- ② 따옴표를 가진 성분이나 요소도 명사처럼 기능할 수 있으나, 원래 품사대로 분석한다.

예) 그것은 “는”이 아니라 “를”이다. [“/SS+는/JX+”/SS+0/JKC]

- ③ 부사 뒤에 격조사가 쓰이는 것도 의미론적인 따옴의 효과에 의하여 부사가 명사적인 용법을 가지는 것이므로 분석은 ‘부사’로 한다.

예) 가족을 멀리에 보냈다. [멀리/MAG+0/JKB]

- ④ 고사성어는 모두 붙여 명사로 분석한다.

예) 백척간두(百尺竿頭) [백척간두/NNG+(/SS+百尺竿頭/SH+)/SS]

단, 고사성어가 두 어절 이상으로 분리되어 있을 경우엔 분리된 채로 분석한다.

예) 백척간두(百尺竿頭) [백척간두/NNG+(/SS+百尺竿頭/SH+)/SS]

백척 간두(百尺竿頭) [백척/NNG] [간두/NNG+(/SS+百尺竿頭/SH+)/SS]

- ⑤ 일음절 한자어가 독립된 단위로 사용될 경우 이들을 모두 명사로 분석한다.

예 서울초등학교 줄 [서울초등학교/NNP+줄/NNG]

- ⑥ 외국어를 음차한 경우에도 명사로 분석한다.

예 아이 러브 유 [아이/NNG]

1.1.2. 고유명사(NNP)

고유 명사는 특정한 사물에 붙여진 이름으로, 기본적으로 최하의어에 속하는 대상을 서로 변별하기 위하여 붙인 이름이다. 고유 명사는 원칙적으로 지시 대상만 가질 뿐 의미 내용은 가지지 않는다.

※ 고유명사의 분석 기준

고유명사의 분석 기준은 작업자에 따라 애매하게 해석될 소지가 있다. 우리는 다음과 같은 경우에 한해서만 고유명사로 태깅한다.

[1] 인명

- ① 성이나 이름 혹은 호와 함께 쓰인 ‘씨(氏), 공(公), 군(君), 양(嬢), 응(翁)’ 등은 띄어 쓰기에 관계없이 각기 다른 분석 대상이 된다.

예 남수/NNP+군/NNB, 김/NNP+씨/NNB, 최치원/NNP+응/NNB, 케네디/NNP+씨/NNB

- ② 성과 이름, 호가 함께 쓰이면 하나의 단위로 분석한다.

예 김철수/NNP, 이태백/NNP

- ③ 성과 호칭어, 관직명이 함께 쓰이면 ‘고유명사+일반명사’로 분석한다.

예 정/NNP+교장/NNG, 최/NNP+선생/NNG

- ④ ‘씨, 군’ 등과 달리 ‘가(哥)’는 접미사이므로, ‘김가(金哥), 이가(李哥)’는 파생어이다.

예 김/NNP+가/XSN

[2] 지명

- ① 내륙, 바다, 강, 산, 산맥, 호수, 섬, 만, 계곡, 늪, 주 등의 이름

예 카스피해/NNP, 템즈강/NNP, 태백산맥/NNP, 미시시피호/NNP, 네바다주/NNP

- ② 도(道), 시(市), 읍(邑), 면(面), 리(里), 군(郡), 구(區), 동(洞) 등의 이름은 그 구역의 종류를 나타내는 말과 함께 전체가 고유명사가 된다.

예 서울특별시/NNP, 성북구/NNP, 강진군/NNP, 인천동/NNP

- ③ "시, 군, 구, 통, 반" 등 주소와 관련된 명칭이 독립적으로 쓰일 경우 이들을 각각 일반명사로 분석한다.

예 군을 시로 승격하기란 쉽지 않다.

[군/NNG+을/JKO]

예 종암 2동

[2/SN+동/NNG]

[3] 향로, 해로 등의 이름

예 남극향로/NNP, 북태평양향로/NNP

[4] 종족명

특정 종족의 이름은 고유명사로 분석한다.

예 알타이족/NNP, 피그마족/NNP, 돌궐족/NNP, 한족/NNP

[5] 국가명 또는 왕조명

예 대한민국/NNP, 조선/NNP, 5공화국/NNP

- ① 다른 형태가 붙어 국가나 왕조의 존립 기간을 나타내는 경우 일반명사로 분석한다.

예 대한제국기/NNG, 조선조/NNG

- ② '남한'과 '북한'을 의미하는 '남, 북, 남북'은 모두 일반명사와 고유명사를 구별한다.

남한을 뜻하는 '남'과 북한을 뜻하는 '북'을 고유명사로 분석한다.

예 남/NNP+과/JKB 북/NNP+의/JKG 의견/NNG 차이/NNG

예 남북/NNP 적십자회담/NNG

예 북/NNP+[]/NNP 회담/NNG

예 대/NNB+남/NNP, 대/NNB+북/NNP

- ③ 어떤 국가의 국민을 나타내는 '국가+인'은 통합하여 일반명사로 분석한다.

예 이집트인/NNG, 이제르바이잔인/NNG, 이스라엘인/NNG, 조선인/NNG

- ④ 어떤 국가의 군대를 나타내는 '국가+군'은 통합하여 일반명사로 분석한다.

예 미군/NNG, 북한군/NNG, 영국군/NNG, 일본군/NNG

[6] 건축물이나 시설물 혹은 구조물의 이름

- ① 도로, 항만, 철도, 전철, 지하철 및 그 명칭과 함께 쓰이는 부대 시설은 그 종류를

나타내는 말과 함께 전체가 고유명사가 된다.

예 부산항/NNP, 대전역/NNP, 서울지하철/NNP, 테헤란로/NNP

- ② 빌딩, 교회, 도서관, 학교, 박물관, 궁, 기도원, 절, 성, 아파트, 체육관, 병원 등의 이름은 그 종류를 나타내는 말과 함께 전체가 고유명사가 된다.

예 서울역사/NNP, 세종문화회관/NNP, 개나리유치원/NNP, 고려대학교/NNP

예 국립중앙박물관/NNP, 국립민속박물관/NNP, 구텐베르그박물관/NNP

- ③ 호텔, 백화점, 음식점, 극장, 병원 등의 이름은 그 종류를 나타내는 말과 함께 전체가 고유명사가 된다.

예 신라호텔/NNP, 미도파백화점/NNP, 동궁음식장/NNP, 명보극장/NNP, 고대병원/NNP

[7] 회사, 학교, 정당, 기관이나 단체의 이름

- ① 특정 회사나 학교, 정당 등의 이름은 고유명사로 분석한다.

예 삼성/NNP, 고려대학교/NNP, 한나리당/NNP, 자유민주주의연합/NNP

단, 특정 회사의 상품명은 고유명사가 아닌 일반명사로 취급한다.

예 초코하임/NNG, 한메디자인/NNP

- ② 정부기관의 명칭은 모두 일반명사로 처리한다. 그러나 거기에 인명, 지명 등의 고유명사가 포함된 경우 그 통합형을 고유명사로 처리한다.

예 헌법/NNG+제판소/NNG, 대/XPN+법원/NNG, 고등/NNG+법원/NNG, 지정/NNG+경제원/NNG

예 서울고등법원/NNP, 서울시경찰서/NNP, 서대문구치소/NNP

- ③ 특정 기관이나 단체, 연구소 등의 경우에는 분석하는 것을 원칙으로 한다. 그러나 거기에 인명, 지명 등의 고유명이나 ‘전국’, ‘국제’, ‘세계’등이 포함되면 그 통합형을 고유명사로 처리한다.

예 대한축구협회/NNP, 전국은행협회/NNP, 한국전자통신연구소/NNP,

예 생활/NNG+체육/NNG+연구소/NNG, 입주자/NNG+대표자/NNG+협의회/NNG

- ④ 약어나 준말의 처리

고유명사가 축약된 형태(준말)로 쓰일 경우 본디말과 함께 준말도 인정하여 축약된 형태 그대로를 고유명사로 분석한다. 그리고 일반명사로 분석하는 기관명의 약자는 일반명사로 분석한다.

예 육사/NNP, 고대/NNP, 5공/NNP, 자민련/NNP, 서울고법/NNP

예 정보통신위/NNG

[8] 창작물, 작품, 도서명

예 삼국사기/NNP, 손자병법/NNP

[9] 언어명

예 한국어/NNP, 일본어/NNP, 영어/NNP, 일타이아/NNP, 네덜란드어/NNP

단, ‘국가+말’은 통합하여 일반명사로 분석한다.

예 한국말/NNG, 프랑스말/NNG, 일본말/NNG, 조선말/NNG

1.1.3. 의존명사(NNB)

의존명사는 자립해서 쓰일 수 없는 명사로, 수식 성분을 반드시 동반해야 한다. 의존명사는 비단위성 의존명사와 단위성 의존명사로 나눌 수 있으나, 본 분석에서는 이를 세분해서 표지를 달지는 않는다. 한편, 단위명사 중 자립해서 쓰일 수 있는 것은 의존명사에서 제외하였다.

- ① ‘것’과 그것의 구어형 ‘거’의 분석 지침은 다음과 같다. ‘거’의 형태를 그대로 인정하여 분석한다. 그러나 다른 형태와의 결합에서 ‘거’의 형태가 유지되지 않는다면 그 때에는 ‘것’으로 복원하여 분석한다.

예 공부할 거를 준비해 왔니?	[거/NNB+를/JKO]
공부할 걸 가져왔니?	[것/NNB+ㄹ/JKO]
연습할 건 있니?	[것/NNB+ㄴ/JX]
먹을 게 모자르다	[것/NNB + 0/JKS]

- ② “연대, 연도, 연차”는 “년대, 년도, 년차”와 달리 모두 일반명사로 태깅한다.

- ③ “월, 연, 일, 주, 달러, 원” 등은 본래 의존명사이지만, 독립되어 쓰일 경우 모두 일반명사의 자격을 가지므로 일반명사로 태깅해야 한다.

예 나는 월 30만원을 받는다.	[월/NNG]
예 달러의 가치는	[달러/NNG]

- ④ 일반명사가 단위적인 용법으로 쓰인 경우에는 의존명사가 아니므로, 주의한다.

예 사람, 시간, 그릇, ...	
한 사람이 교실로 들어왔다.	[사람/NNG+0/JKS]
지장면 한 그릇만 주세요.	[그릇/NNG+만/JX]

- ⑤ “해, 달, 동, 서, 남, 북, 전, 후, 좌, 우” 등은 모두 일반명사로 태깅한다.
단, “전”은 관형사로도 쓰이니 주의해야 한다.

- ⑥ 길이, 무게, 수효, 시간 따위의 수량을 수치로 나타내는 단위들 중 “미터, 그램, 리터” 등은 의존명사(NNB)로, 외국어로 된 “m, g, l” 등은 기호(SW)로 분석한다.

- ⑦ 명사와 기호, 숫자가 통합되어 있는 경우 중 기호, 숫자와 나머지 부분을 분리할 수 없는 경우에는 이들을 하나의 명사로 간주하여 전체를 명사로 분석한다.

관왕	그는 첫 2관왕의 영광을 얻었다.	[2관왕/NNG+의/JKG]
구체	10구체는 8구체의 발전된 형태입니다.	[10구체/NNG+는/JX]
대조	그분은 나의 5대조 할아버지이다.	[5대조/NNG]
호기	화력 발전소 2호기가 가동되었다.	[2호기/NNG+가/JKS]
호선	지하철 3호선으로 나와라.	[3호선/NNG+으로/JKB]
호점	여기가 우리 식당 1호점이다.	[1호점/NNG+이/VCP+다/EF+./SF]
뱅크C유		[뱅크C유/NNG]

- ⑧ 의존명사의 전체 목록은 표준국어대사전의 그것과 일치하므로, 이를 토대로 태깅한다.

1.2 대명사(NP)

대명사는 그 자체로는 자신의 본유적 지시물을 가지지 않은 채, 다만 사람이나 사물 등 어떤 대상을 간접적으로 지시하는 품사이다. 여기서는 다음의 형태만을 대명사로 설정한다. 단, 동일한 대명사가 방언이나 고어의 이형태를 가진 경우에는 이들도 대명사로 같이 분석한다. 대명사의 전체 목록은 표준국어대사전의 그것과 같다.

[1] 1인칭 대명사

나, 내, 우리, 저, 제, 저희

[2] 2인칭 대명사

그대, 귀관(貴官), 너, 네, 너희, 당신, 닥, 어르신, 여러분, 지네, 형씨

[3] 기타 대명사

이아, 이분, 그이, 그분, 저이, 저분, 아무, 아무개, 누구, 무엇, 뭐, 어디, 언제, 자기, 게, 제, 애

이것 저것 그것 이거 저거 그거 여기 저기 거기 이곳 그곳 저곳 어디 모(某), 모모(某某), 거시기

- ① ‘모(某)’는 관형사와 대명사로 분석될 수 있으므로 주의를 요한다.

예 모 기업체	[모/MM]
김 모씨	[모/NP+씨/NNB]

- ② ‘모모(某某)’도 위와 같이 분석될 수 있다.

예 모모가 말했다	[모모/NP+가/JKS]
모모 기관의 조사를 마쳤다	[모모/MM]

- ③ ‘어르신네’는 ‘어르신/NP+네/XSN’으로 분석한다.

- ④ ‘이것, 그것, 저것; 이거, 그거, 저거’는 분석하지 않고 대명사로 인정한다. ‘~거’의 경우, 다른 형태와의 결합에서 ‘~거’의 형태가 유지되지 않는다면 그 때에도 ‘~것’으로 복원한다.

예 난 저거를 먹을래.	[저거/NP+를/JKO]
나는 여태 그걸 믿어 왔단다.	[그것/NP+ ㄹ /JKO]

- ⑤ 아래와 같이 원형을 밝힐 수 있는 대명사는 원형대로 분석한다.

내 이제부터는 내 명령을 따라라	[나/NP+의/JKG]
내게 내게 전자우편으로 알려 디오	[나/NP+에게/JKB]
네게 어제 네게 보낸 선물이 잘못 되었다	[너/NP+에게/JKB]
제게 문제가 있다면 제게 말씀해 주세요	[저/NP+에게/JKB]
누가 누가 전화를 하는 지 보고해라	[누구/NP+가/JKS]
뉘 뉘 집 얘기가 이렇게 울고 있는 거야	[누구/NP+의/JKG]
뭐가 도대체 뭐가 문제라는 거야?	[뭐/NP+가/JKS]

- ⑥ ‘뭐’는 ‘무엇’과 대등할 정도로 자주 사용되므로 그 형태 자체를 인정해 준다. 다만, 다음과 같이 조사와 축약되었을 경우에만 원형으로 복원해 준다.

예 앞으로 우리가 뭘 하지는 얘기이나?	[무엇/NP+ ㄹ /JKO]
-----------------------	-----------------

- ⑦ ‘제’의 경우, ‘제/NP+가/JKS’를 제외하고는 모두 ‘저/NP+의/JKG’로 분석한다.

예 제가 갈 것입니다.	[제/NP+가/JKS]
철수는 제 잘못을 안다.	[저/NP+의/JKG]
제 무게를 못 견디다.	[저/NP+의/JKG]

제 값을 하다.

[자/NP+의/JKG]

- ⑧ “이때, 그때, 저때”는 통합해 “이때/NNG, 그때/NNG,저때/NNG”로 분석한다.
- ⑨ “이번, 그번, 저번”은 통합해 “이번/NNG, 그번/NNG, 저번/NNG”로 분석한다.

1.3. 수사(NR)

수사는 사물의 수량이나 차례를 나타내는 품사를 말한다.

예 양수사

- 하나, 둘, 셋, 넷, 다섯, 여섯, 일곱, 여덟, 아홉, 열, 스물, 서른, 마흔, 쉰, 예순, 일흔, 여든, 아흔, 백
- 한둘, 두서넛, 서넛, 너댓, 네다섯, 네댓, 대여섯, 예닐곱, 일여덟, 일고여덟, 열두서넛, 열대여섯, 열일고여덟, 스물두서넛
- 일, 이, 삼, 사, 오, 육, 칠, 팔, 구, 십, 백, 천, 만, 억, 조
- 기십, 기백, 기천,
- 수십, 수백, 수천, 수만, 수억, 수십만, 수백만, 수천만, ...

예 서수사

- 첫째, 둘째, 셋째, 넷째, ..., 열째, 열한째, ..., 스물한째, ..., 아흔아홉째, 백째, 백한째, ...

- ① 복수의 수사가 한 어절 내에 나타날 때에는 전체를 통합해서 분석한다.

예 백만오천삼십사

[백만오천삼십사/NR]

- ② ‘여남은, 스무남은’ 등은 명사로 취급하는 일도 있으나 수사로 본다.

예 친구 여남은이 왔다.

[여남은/NR]

- ③ ‘기십(幾十), 기백(幾百), 기천(幾千)’은 수사로 보고, ‘몇 십, 몇 백, 몇 천, 몇 억’ 등은 ‘수 관형사 + 수사(단위수)’로 본다.

예 운동장에는 기백 기천의 관중이 모였다 [기백/NR]

예 하루에만 몇천 명의 인파가 몰렸다 [몇/MM+천/NR]

- ④ ‘하루, 이틀, 사흘, 나흘’ 등은 날짜의 이름으로 수사가 아니라 명사이다.

예 하루가 지나고 또 하루가 지나도 [하루/NNG+가/JKS]

- ⑤ ‘수십, 수백, 수천’ 등은 ‘수사’로 분석한다.

예 수백 [수백/NR]

- ⑥ ‘하나’는 표준국어대사전에 그 품사가 명사와 수사로 되어 있지만 이 분석에서는 모두 수사로 태깅한다.

예 광에 가서 물건 하나만 가져오렴. [하나/NR+만/JX]

우리는 하나로 뭉쳤다. [하나/NR+로/JKB]

- ⑦ 때로 수사와 수관형사의 구별이 애매한 경우가 있다. 이 분석에서는 임흥빈(1998)의 견해에 따라, 다음과 같은 특이한 형식을 가진 예만을 수관형사로 취급하고, 그 밖의 것들은 모두 수사로 분석한다. (관형사 목록 참조)

예 한, 한두, 한두어, 두, 두어, 두세, 두세너, 세, 석, 서, 세너, 네, 너, 녀, ...

- ⑧ ‘제일, 제이’ 등은 접두사 ‘제-’와 수사의 결합으로 분석한다.

예 제일, 제이, 제삼, 제사, 제오, ..., 제구십구, 제백, ...

[제/XPN+일/NR], [제/XPN+0/NR], ...

2. 용언

용언은 동사, 형용사, 지정사를 가리킨다. 용언 범주에서는 분석 대상이 본용언일 경우에만 동사와 형용사로 구분하여 표시하고, 보조용언의 경우에는 모두 ‘VX’라는 하나의 표지만을 준다. 또한 학교 문법에서 서술격조사로 다루는 ‘이다’는 조사의 범주에 넣지 않고 ‘지정사’라는 용언의 하위범주에 넣기로 한다. 지정사는 다시 긍정 지정사(VCP)와 부정 지정사(VCN)로 세분된다.

2.1. 동사(VV)

동사는 사물의 움직임이나 작용을 나타내는 용언을 말한다. 동사는 일반적으로 목적어의 필요성 여부에 따라 자동사, 타동사로 나누기도 하지만, 본 분석에서는 그것을 위한 별도의 표지를 세분하지 않고 모두 ‘VV’로 표시한다. 동사의 목록은 표준국어대사전의 그것을 따른다.

2.2. 형용사(VA)

형용사는 사물의 성질이나 상태를 나타내는 용언을 가리킨다. 형용사의 목록은 표준국어대사전의 그것을 따른다.

2.3. 보조용언(VX)

이 분석에서는 보조용언을 보조동사와 보조형용사로 하위 구분하지 않는다. 여기서는 다음에 제시된 용언들이 보조 용언으로 분석될 수 있는 후보가 된다. 이 목록은 표준국어대사전에 기대어 추출된 것이다. 물론, 아래의 형태가 항상 보조용언이 되는 것은 아니다. 보조용언이나 본용언이나의 최종 판정은 문맥 정보를 참조할 수밖에 없다.

가다	세월이 흘러 가는 대로 정처없이 떠도는 나그네	[가/VX+는/ETM]
가지다	일을 그렇게 해 가지고는 기일을 맞출 수 없다	[가지/VX+고/EC+는/JX]
계시다	손님께서 와 계십니다.	[계시/VX+버니다/EF+./SF]
나기다	정책을 추진해 나가는 과정에서 문제가 생겼다	[나기/VX+는/ETM]
나다	아침에 깨어 나 보니 그가 없어졌다	[나/VX+아/EC]
내다	힘들겠지만 잘 견뎌 내야 한다	[내/VX+아/EC]
놓다	약속을 잡아 놓고 출장을 기다니	[놓/VX+고/EC]
달다	이번 시험 문제의 정답을 알려 나오.	[달/VX+오/EF+./SF]
대다	자꾸 줄라 대는 통에 그만 허락해 주고 말았다	[대/VX+는/ETM]
두다	남겨 둔 돈도 이제 바닥이 났다	[두/VX+ㄴ/ETM]
드리다	염려는 끼쳐 드려 송구하옵니다	[드리/VX+아/EC]
들다	도무지 내 말은 믿으려 들지 않는다	[들/VX+지/EC]
말다	어렵더라도 희망을 잃지 말아야 한다	[말/VX+아/EC]
먹다	나는 오늘도 수업을 빼 먹었다.	[먹/VX+었/EP+다/EF+./SF]
못하다	그 참상을 치마 보지는 못할 것이다	[보/VX+지/EC+는/JX]
버리다	음식이 다 타 버렸다.	[버리/VX+었/EP+다/EF+./SF]
보다	이제는 새벽이 오는가 보다.	[보/VX+다/EF+./SF]
빠지다	썩어 빠진 생선을 사오다니	[빠지/VX+ㄴ/ETM]

싶다	사람 소리가 나는 듯 싶다.	[싶/VX+ㄷ/EF+./SF]
쌓다	꼬치꼬치 물어 쌓는 통에 정신이 없었다	[쌓/VX+는/ETM]
아니하다	일이 순리대로 풀리지 아니했다.	[아니하/VX+았/EP+ㄷ/EF+./SF]
않다	시간이 지나도 기차는 오지 않았다.	[않/VX+았/EP+ㄷ/EF+./SF]
오다	고향을 떠나 온 지 10년이 지났다	[오/VX+ㄴ/ETM]
있다	그녀는 검정 옷을 입고 있었다.	[있/VX+았/EP+ㄷ/EF+./SF]
주다	아버지는 아기에게 동화책을 읽어 주었다.	[주/VX+았/EP+ㄷ/EF+./SF]
지다	한 번 넘어 진 아이는 일어나는 법을 안다.	[지/VX+ㄴ/ETM]
치우다	다섯 명이 10인분의 식사를 먹어 치웠다.	[치우/VV+았/EP+ㄷ/EF+./SF]
터지다	끓인 지 오래 되어서 라면이 불어 터졌다.	[터지/VX+았/EP+ㄷ/EF+./SF]
히다	나귀를 쉬게 하는 것이 좋겠다.	[히/VX+는/ETM]

① 보조용언 태깅 원칙은 다음과 같다.

- 보조용언의 후보는 위의 표에 속한 동사에만 한정된다.
- 보조용언 앞에는 반드시 다른 용언이 위치해 있어야 한다.
- 보조용언이 동시에 두 개 이상이 연결되어 나타날 수도 있다.

② 다음과 같은 어절은 보조용언으로 취급되기도 하나, 여기서는 ‘의존명사+접사’로 분석한다. 이들 앞에는 항상 관형어가 온다는 분포적인 특성을 중시한 것이다.

예 양하다/채하다/척하다/듯하다/법하다/뻔하다 [양/NNB+하/XSA+ㄷ/EF+./SF]

③ ‘버릇하다’의 경우에는 선행 성분으로 관형형이 오는 것은 아니지만, 일반명사 ‘버릇’과 크게 구별되지 않으므로 ‘버릇’은 명사로 분석한다.

예 지꾸 물어 버릇하다. [버릇/NNG+하/XSV+ㄷ/EF+./SF]

2.4. 지정사(VC)

지정사는 학교 문법의 서술격 조사에 대응되는 것인데, 용언과 같이 활용한다는 특성을 중시한 술어이다. 여기서는 학교 문법의 ‘이다’를 긍정 지정사로, ‘아니다’를 부정 지정사로 하위 구분한다. 일반적으로 ‘아니다’는 형용사로 다루어지기도 하나, 여기서는 ‘아니다’가 ‘이다’의 부정형이라는 점을 중시하여 ‘부정지정사’로 다룬다.

예 철수는 매우 우수한 학생이다. [학생/NNG+O/VCP+ㄷ/EF+./SF]

예 철수는 모범적인 학생이 아니다. [아니/VCN+ㄷ/EF+./SF]

◦ 다음과 같은 경우에는 지정사 ‘이/VCP’를 복원해야 한다.

① 체언에 어미가 직접 연결된 경우

예 철수는 훌륭한 교시다. [교사/NNG+0/VCP+ㄷ/EF+./SF]

② 조사에 어미가 직접 연결된 경우

예 우리가 그를 본 것은 서울에서다. [서울/NNP+에서/JKB+0/VCP+ㄷ/EF+./SF]

③ ‘~였다’의 경우

예 그 당시 나는 아이였다. [아이/NNG+0/VCP+였/EP+ㄷ/EF+./SF]

④ 어미 '-라고, -라는, -라도, -라며, -라면서, -라서

예 나는 그에게 절교라고 말했다. [절교/NNG+0/VCP+라고/EC]

예 나는 친구라는 말이 좋다. [친구/NNG+0/VCP+라는/ETM]

예 거지라도 존중해 주어야 한다. [거지/NNG+0/VCP+라도/EC]

예 그는 최고라며 나를 추켜 주었다. [최고/NNG+0/VCP+라며/EC]

예 그는 실수하면서 얼버무렸다. [실수/NNG+0/VCP+라면서/EC]

예 너는 부지라서 우릴 이해하지 못할 것이다. [부지/NNG+0/VCP+라서/EC]

⑤ 인용문 뒤에 오는 “~며” 는 지정사를 복원하지 않는다.

예 얼마나 친절해나?”며 [친절/NNG + 하/XSA + 나/EF + ?/SF + "SS + 며/EC]

3. 수식언

3.1. 관형사(MM)

관형사는 체언 앞에서 그것을 꾸미는 품사를 말한다. 관형사는 지시관형사, 수관형사, 성상관형사로 세분될 수 있는데, 본 분석에서는 이를 세분하여 분석하지 않는다. 국어의 관형사는 매우 제한된 목록만을 가지고 있다. 관형사의 목록은 표준국어대사전을 따랐다.

각(各) 각 가정	[각/MM]
갖은 갖은 고생	[갖은/MM]
고 고 녀석	[고/MM]
고(校) 고 ○ ○ ○ 선생	[고/MM]
고얀 고얀 놈	[고얀/MM]
그 그 사람	[그/MM]
그까짓 그까짓 일로	[그까짓/MM]

그깐	그깐	놓	[그깐/MM]
그깃	그깃	비행기 표 한 장	[그깃/MM]
그딤	그딤	소리	[그딤/MM]
그런	그런	사람	[그런/MM]
그런저런	그런저런	얘기 끝에	[그런저런/MM]
근(近)	근	한 시간이 지나서	[근/MM]
간긴	간긴	세월	[간긴/MM]
까짓	까짓	고생쯤	[까짓/MM]
너	콩 너	말	[너/MM]
넉	금 넉	냥	[넉/MM]
네까짓	네까짓	것	[네까짓/MM]
넌	넌	것	[넌/MM]
네	연필 네	자루	[네/MM]
다른	다른	사람	[다른/MM]
단(單)	단	한 발의 총탄	[단/MM]
당(當)	당	열차는 출발합니다.	[당/MM]
댓	서른	댓	[댓/MM]
동(同)	동	회사에서 근무함.	[동/MM]
두	두	개	[두/MM]
두서너	두서너	개	[두서너/MM]
두세	두세	개	[두세/MM]
두어	사과 두어	개	[두어/MM]
딴	딴	생각	[딴/MM]
만(滿)	만	하루동안 끝냈다.	[만/MM]
매(每)	매	경기마다	[매/MM]
맨	맨	꼭대기	[맨/MM]
먼먼	먼먼	옛날	[먼먼/MM]
몇	몇	사람	[몇/MM]
몇몇	몇몇	사람	[몇몇/MM]
모(某)	모	소식통에 의하면	[모/MM]
모든	모든	사람	[모든/MM]
몹쓸	몹쓸	것	[몹쓸/MM]
무슨	무슨	일	[무슨/MM]
뭇	뭇	백성	[뭇/MM]
뭔	뭔	소리냐?	[뭔/MM]
벼리벌	벼리벌	생각	[벼리벌/MM]
벌(別)	벌	부담 없이	[벌/MM]
벌벌(別別)	벌벌	생각	[벌벌/MM]
벌의벌(別-別)	벌의벌	생각	[벌의벌/MM]
본(本)	본	법정에서는	[본/MM]
새	새	학기	[새/MM]
서너	서너	개	[서너/MM]
서	서	말	[서/MM]

석	석 냥	[석/MM]
성(聖)	성 베드로	[성/MM]
세	세 개	[세/MM]
수	수 미터의 깊이	[수/MM]
순(純)	순 살코기	[순/MM]
스무	스무 살	[스무/MM]
아무런	아무런 계획도 없다.	[아무런/MM]
아무	아무 연필	[아무/MM]
약(約)	약 두 시간	[약/MM]
양대(兩大)	양대 산맥	[양대/MM]
양(兩)	양 무릎	[양/MM]
어느	어느 학교	[어느/MM]
어떤	어떤 사람	[어떤/MM]
어인	어인 까닭으로	[어인/MM]
여느	여느 때와 달리	[여느/MM]
여러	여러 사람	[여러/MM]
연(延)	연 10만 명이 동원되다.	[연/MM]
옛	옛 추억	[옛/MM]
오랜	오랜 세월	[오랜/MM]
오른	오른 다리	[오른/MM]
오만(五萬)	오만 잠동새니	[오만/MM]
온갖	온갖 시련	[온갖/MM]
온	온 집안	[온/MM]
올	올 10월	[올/MM]
외딴	외딴 마을	[외딴/MM]
왼	왼 손목	[왼/MM]
요	요 근방	[요/MM]
웬	웬 영문인지 모르다.	[웬/MM]
이까짓	이까짓 일을 해보아	[이까짓/MM]
이갓	이갓 것	[이갓/MM]
이내	외로운 이내 신세	[이내/MM]
이런	이런 사람	[이런/MM]
이런저런	이런저런 이야기	[이런저런/MM]
이	이 사람	[이/MM]
일대(一六)	일대 혼란에 빠지다.	[일대/MM]
장장(長長)	장장 80리 길을 걸어서	[장장/MM]
저런	저런 사람	[저런/MM]
저	저 사람	[저/MM]
전(全)	전 국민	[전/MM]
전(前)	전 국회의원	[전/MM]
전전(前前)	전전 직장은 줄었다.	[전전/MM]
제까짓	제까짓 것	[제까짓/MM]
제간	제간 놈	[제간/MM]

제(諸) 제 문제	[제/MM]
조 조 녀석은	[조/MM]
주(主) 주 고객	[주/MM]
첫 첫 월급	[첫/MM]
총(總) 총 16면	[총/MM]
타(他) 타 업소와는 다르다.	[타/MM]
한두 한두 개	[한두/MM]
한두어 한두어 개	[한두어/MM]
한 한 개	[한/MM]
하튼 하튼 일	[하튼/MM]
헌 헌 구두	[헌/MM]
현(現) 현 정권	[현/MM]

- ① 위에 제시한 관형사들은 때로 문맥에 따라 다른 품사로 분석될 가능성이 있으니 문맥을 잘 살펴서 분석해야 한다.

- 관형사, 명사 통용

예 올 예산이 다 바닥이 났다.	[올/MM]
올 들어 물가가 많이 올랐다.	[올/NNG]

- 관형사, 부사 통용

예 단 세 명에서 그 일을 꾸몄다.	[단/MM]
단, 그 일은 해서는 안 된다.	[단/MAJ]

- 관형사, 명사, 부사 통용

예 이내 마음을 어찌 알리오.	[이내/MM]
아침 들판에 이내가 끼었다.	[이내/NNG]
그는 이내 떠나갔다.	[이내/MAG]

- ② 수사가 명사를 단독으로 수식하는 경우 그것을 관형사로 분석하기 쉬우나, ‘수’를 나타내는 말 가운데서 앞서 언급한 수관형사를 제외하고는 수사는 오로지 수사로만 분석한다. 즉, 수사의 관형사적 쓰임을 인정하지 않는 것이다. 따라서 다음과 같이 ‘다섯’은 모든 환경에서 중의성 없이 ‘수사’로만 분석된다.

예 다섯이 먹기에 충분하다	[다섯/NR+O/JKS]
예 다섯 명이 앉아 있었다	[다섯/NR]

- ③ 뒤에 ‘-적(的)’을 가진 말은 조사와 결합하는 것과 관계없이 모두 명사로 태깅한다.

예 명사의 부사적인 용법	[부사/NNG+적/XSN+O/VCP+L/ETM]
예 명사의 부사적 용법	[부사/NNG+적/XSN]

또한 ‘-적(的)’과 결합하는 형태가 격조사와 결합하지 못하는 특성을 가지는 경우에는 일반명사가 아니라 ‘어근’으로 태깅한다.(어근에 대한 자세한 세부 지침은 3.6.4 어근 참

조)

예 희기적	[희기/XR+적/XSN]
노골적	[노골/XR+적/XSN]
합리적	[합리/XR+적/XSN]

3.2. 부사(MA)

부사는 주로 용언을 꾸며서 그 뜻을 더 세밀하고 분명하게 해 주는 품사를 말한다. 부사는 크게 성분 부사와 문장 부사로 나뉘고, 문장 부사는 다시 양태 부사와 접속 부사로 나뉜다. 여기서는 접속부사와 일반부사로만 나누기로 한다.

3.2.1. 접속부사(MAJ)

접속부사는 다음의 목록(표준국어대사전)만을 대상으로 한다.

그래서	그러나	그러니까	그러면	그러므로
그런데	그렇지만	그리고	그리하여	단
따라서	써	연이나	연중에	연즉
이리하여	하건만	하기는	하기야	하물며
하지만				

① 접속부사는 종종 용언의 활용형으로도 쓰일 수 있으므로 주의한다.

예 그래서 마지막에는 조심하라고 했지?	[그래서/MAJ]
영희가 그래서 결석을 했구나.	[그렇/VV + 어서/EC]

② 문두의 '그리고나서' 분석의 경우

- 두 어절로 나뉜 경우

예 그리고	[그리하/VV+고/EC]
나서	[나/VX+서/EC]

- 단일 어절로 된 경우

예 그리고나서	[그리하/VV+고/EC+나/VX+서/EC]
---------	-------------------------

3.2.2. 일반부사(MAG)

일반부사는 종종 일반명사와 동일형태를 띠고 있어 구분이 애매한 경우가 있다. 이들은 뒤에 조사가 결합하느냐의 여부와, 문맥에서 후행 명사를 수식하느냐의 여부에 따라 부사와 명사로 분석될 수 있다.

예 너의 진짜 속셈이 무엇인지 말해 보라.	[진짜/NNG]
그 수학 문제는 진짜 어려웠다.	[진짜/MAG]
예 지금이 공부하기 딱 좋은 때이다.	[지금/NNG+O/JKS]
나는 지금 막 집에 도착했다.	[지금/MAG]

- ① 부사적인 용법을 가졌음에도 불구하고 일반부사가 아닌 일반명사로만 표준국어대사전에 등재되어 있는 단어는 오로지 일반명사로만 분석한다.

예 구석구석, 무작정, 여기저기, 오랫동안, 이곳저곳, 정작, 좌우간, 처음, 최근, 한때

- ② 일반부사로 분석하기 쉬운 활용상의 불완전동사인 ‘덜달아, 더붙어’는 모두 동사로 옮겨 분석해야 함에 주의한다.

예 너는 덜달아 왜 난리나?	[덜달/VV+O/EC]
우리 함께 더붙어 살아가자.	[더붙/VV+O/EC]

- ③ ‘명사+없이’는 원칙적으로 ‘일반명사+없이/MAG’로 태깅하지만, 아래와 같이 하나의 단어로 굳어진 것은 굳이 분석하지 않고 ‘없이’ 통합형 자체를 하나의 일반부사로 분석한다. 통합형의 목록은 표준국어대사전을 참고한다.

예 관계없이, 그지없이, 꾸밈없이, 끊임없이, 난데없이, 남김없이, 등

4. 독립언

4.1. 감탄사(IC)

감탄사는 화자의 부름이나 느낌, 놀람이나 대답을 직접적으로 나타내는 품사를 말한다. 다음과 같은 어절이 감탄사가 된다.

예 그림(요), 아호, 어머, 앓, 아, 예, 그래(요), 아니(요), 글썄, 참, 아이구, 외아, 오호, 세상에

- ① 사람이 입으로 직접 내는 소리를 대상으로 하되, 흉내를 내는 의도가 없는 것과 본능적인 놀람이나 느낌을 나타내는 것을 대상으로 한다. 또한 감탄사와 혼동되는 부사로서 음성상징어류의 부사어가 있는데, 이는 감탄사가 아닌 일반부사로 분석한다.

예 아호! 드디어 정상이다.	[아호/IC+!/SF]
쿨럭쿨럭 기침을 했다.	[쿨럭쿨럭/MAG]

- ② 동물의 울음소리 등은 감탄사가 아니라 일반부사로 분석한다.

예) 검둥이는 멍멍 짖으며 수풀 속으로 뛰어들어갔다. [멍멍/MAG]

- ③ 욕이나 욕설을 나타내는 말은 전체를 감탄사로 분석한다.

예) 빌어먹을! [빌어먹을/IC+!/SF]

- ④ 한 어절이 비정상적으로 늘어나거나 다른 기호가 개입되었을 경우 분석불능 범주 (NA)로 분석한다.

예) 그러아염/NA, 으~아~아/NA

5. 관계언

조사는 주로 체언과 결합하여 다른 말과의 문법적 관계를 나타내거나, 특별한 뜻을 더해 주는 품사를 말한다. 조사는 크게 격조사, 보조사, 접속조사로 나눈다. 한국어는 조사가 중첩하는 경우가 많은데, 이러한 경우 조사의 결합형은 분리해서 분석함을 원칙으로 한다.

예) 부산에서도 대형 사고가 있었다. [부산/NNP+에서/JKB+도/JX]

그녀와의 약속이 갑자기 잡혔다. [그녀/NP+와/JKB+의/JKG]

5.1. 격조사(JK)

이는 체언과 다른 성분 간의 일정한 문법 관계를 나타내는 조사이다.

5.1.1. 주격조사(JKS)

선행 체언으로 하여금 주어가 되게 하는 조사이다.

이	책이 보인다.	[책/NNG+이/JKS]
께서	선생님께서 오신다.	[선생/NNG+님/XSN+께서/JKS]
에서	정부에서 시행한 정책이다.	[정부/NNG+에서/JKS]
(이)서	둘이서 그 일을 꾸몄다고?	[둘/NR+이서/JKS]
	혼자서 그 일을 꾸몄다고?	[혼자/NNG+서/JKS]
께서	부대장님께서 오서	[부대장/NNG+님/XSN+께서/JKS]

깨옵서 황제깨옵서 드니다신다.

[황제/NNG+깨옵서/JKS]

5.1.2. 보격조사(JKC)

선행 체언으로 하여금 서술어 ‘되다, 아니다’의 보어가 되게 하는 조사이다.

이가 얼음이 물이 되었다
철수는 범인이 아니다

[물/NNG+0/JKC]

[범인/NNG+0/JKC]

5.1.3. 목적격조사(JKO)

선행 체언으로 하여금 목적어가 되게 하는 조사이다.

ㄹ/을/를 너는 바람소리를 들었다.

[바람/NNG+소리/NNG+를/JKO]

5.1.4. 관형격조사(JKG)

선행 체언으로 하여금 관형어가 되게 하는 조사이다.

의 나의 친구는 너 해니뿐이다.

[나/NP+의/JKG]

5.1.5. 부사격조사(JKB)

선행 체언으로 하여금 부사어가 되게 하는 조사이다.

(으)로 망치로 못을 박아야지.
(으)로서 장관으로서 책임을 다해야 한다.
(으)로써 돌로써 지붕을 만든다고?
같이 바보같이 웃고 다닌다.
더러 니더러 이것도 하라고 한다.
랑 너랑 많이 닮았다.
(으)로부터 TV로부터 받는 영향력이 너무 크다.
마냥 기영이마냥 놀 수만은 없다.
미따나 네 말미따나 나도 그래야 한다.
만큼 눈물만큼 콧물도 흐른다니까.
보고 영자보고 놀자고 좀 해라.
보다 직관보다는 논리가 동원되어야 한다.
에 나는 너에 대해 아무것도 모른다.

[망치/NNG+로/JKB]

[장관/NNG+으로서/JKB]

[돌/NNG+로써/JKB]

[바보/NNG+같이/JKB]

[나/NP+더러/JKB]

[너/NNP+랑/JKB]

[TV/SL+로부터/JKB]

[기영이/NNP+마냥/JKB]

[말/NNG+미따나/JKB]

[눈물/NNG+만큼/JKB]

[영자/NNP+보고/JKB]

[직관/NNG+보다/JKB+는/JX]

[나/NP+에/JKB]

에게	너에게 말하기 싫다.	[니/NP+에게/JKB]
에게서	나는 철수에게서 그 말을 들었다.	[철수/NNP+에게서/JKB]
에서	집에서 학교까지 너무 멀다.	[집/NNG+에서/JKB]
에서부터	연구소에서부터 가게까지는 너무 멀다.	[연구소/NNG+에서부터/JKB]
와/과	경미와 함께 다닌다면,	[경미/NNP+와/JKB]
처럼	사람처럼 행동하는 동물이 있다.	[사람/NNG+처럼/JKB]
하고	그 일하고 관련된 사람은 아무도 없다.	[일/NNG+하고/JKB]
한테	그 일은 경비한테 부탁해라	[경비/NNG+한테/JKB]

5.1.6. 호격조사(JKV)

주로 사람을 가리키는 체언 뒤에 연결되어 그것으로 하여금 부름의 대상이 되게 하는 조사이다.

아	호동아! 이제 그만 일어나거라	[호동/NNP+아/JKV+!/SF]
아	철수아! 밥 먹어라	[철수/NNP+아/JKV+!/SF]
여	주여, 우리에게 힘을 주소서	[주/NNG+여/JKV]
(아)시여	신이시여! 우리를 저버리지 마소서	[신/NNG+(아)시여/JKV+!/SS]

- 호격조사와 어말어미는 구분해서 태깅해야 한다.

예) 저기 오는 것이 철수아. [철수/NNP+아/VCP+0/EF+./SF]

5.1.7. 인용격조사(JKQ)

인용문이나 인용구를, 동사에 대한 부사적 성분으로 도입하는 조사이다.

고	그는 "이제 가도 좋다"고 말했다	[좋/VA+다/EF+/"SS+고/JKQ]
(아)라고	문제가 심각하대라고 보고했다	[심각/XR+하/XSA+다/EF+/"SS+라고/JKQ]
하고	영수는 "이제 가자"하고 말문을 닫았다	[가/VV+자/EF+/"SS+하고/JKQ]

- 인용격조사는 연결어미와 구별하기 어려운 경우가 있으므로 주의한다.

- 예) 가. 철수는 자기가 학생이라고 말했다
- ① [학생/NNG+0라고/JKQ]
 - ② [학생/NNG+0/VCP+라고/EC]
- 나. 철수는 "다음 주에 놀러 가도 좋다"고 말하였다
- ① [좋/VA+다/EF+/"SS+고/JKQ]
 - ② [좋/VA+다/EF+/"SS+고/EC]

인용격조사는 형태만으로 확인할 수 없고 발화 상황까지 고려해야 하는 복잡한 표지이다. 게다가 인용격조사로 인정되는 형태인 ‘라고, 하고’ 등은 원래 용언의 활용형에 불과하다. 하지만 인용격조사를 설정하지 않을 경우에는 인용부호가 들어간 (나)와 같은 어절의 처리가 어색해진다. 따라서 우리는 인용격조사를 설정하되, 그 쓰임은 인용부호(“, ’,), [예], },], >, ...)가 있는 경우로만 제한하기로 한다. 즉, (가)의 경우에는 ②의 분석을, (나)의 경우에는 ①의 분석을 취한다. 물론 (나)의 어절에서 인용부호가 빠진 경우에는 어미로 분석한다.

다. 철수는 영희가 좋다고 말했다 [좋/VA+다고/EC]

5.2. 접속조사(JC)

두 단어를 같은 자격으로 이어 주는 구실을 하는 조사를 말한다.

와	그 아주머니는 딸기와 사과를 샀다	[딸기/NNG+와/JC]
과	그 기계는 사람과 컴퓨터를 구별하지 못한다	[사람/NNG+과/JC]
나	사과나 배는 모두 몸에 좋은 과일이다	[사과/NNG+나/JC]
랑	머루랑 다래랑 먹으며 청산에 살고 싶어라	[머루/NNG+랑/JC]
하고	이번 준비물로 칼하고 연필을 샀다	[칼/NNG+하고/JC]

① ‘함께 함’의 뜻을 나타내는 접속조사는 부사격 조사와 형태상 동일하므로 주의할 필요가 있다.

예	철수와 영희가 왔다	[철수/NNP+와/JC]
	철수와 같이 놀았다	[철수/NNP+와/JKB]

② 표준국어대사전에 조사로 등재된 ‘하며’는 조사로 인정하지 않고 ‘하/VV+며/EC’로 분석한다.

5.3. 보조사(JX)

체언이나 부사 또는 용언의 연결 어미나 종결 어미의 뒤에 쓰여 특별한 뜻을 더해 주는 조사를 말한다.

그러	먹습니다그러	[먹/VV+습니다/EC+그러/JX]
까지(까정/까장)	너까지 나에게 이럴 줄이야	[너/NP+까지/JX]
깨나	너도 사람깨나 울렸겠구나	[사람/NNG+깨나/JX]
(아)나	너나 가라!	[너/NP+나/JX]
(아)나마	빵이나마 먹어라	[빵/NNG+아나마/JX]
ㄴ/은/는	이 종이는 어제 사 온 것이다	[종이/NNG+는/JX]
ㄴ 커녕/은커녕/는커녕	돈은커녕 먹을 쌀도 없다	[돈/NNG+은커녕/JX]
다	그 물건을 거기다 놓아라.	[거기/NP+다/JX]
	그 물건을 거기에다 놓아라.	[거기/NP+에/JKB+다/JX]
다가	책상을 어디다가 둘까요?	[어디/NP+다가/JX]
	집에다가 놓아 두어라.	[집/NNG+에/JKB+다가/JX]
대로(대루)	너는 너대로 살아라.	[너/NP+대로/JX]
따라	오늘따라 택시도 안 잡힌다.	[나/NP+따라/JX]
도/두	강아지도 주인은 알아본다	[강아지/NNG+도/JX]
(아)란	코알라란 호주에 사는 초식동물이다	[코알라/NNG+란/JX]
(아)ㄹ 랑	그 일에 대해선 걱정일랑 하지 말아라	[걱정/NNG+ㄹ 랑/JX]
마다	사람마다 독특한 개성이 있다	[사람/NNG+마다/JX]
마저	장미마저 시들고 말았다	[장미/NNG+마저/JX]
만	인간은 빵만으로 살 수 없다	[빵/NNG+만/JX+으로/JKB]
밖에	그래 봐야 죽기밖에 더 하라	[죽/VV+기/ETN+밖에/JX]
부터/부텀	우선 노인부터 태워라	[노인/NNG+부터/JX]
뿐	가진 건 고작 집 한 채뿐.	[채/NNB+뿐/JX]
서건	국물이나 총각김치서건 한 그릇만 주시오	[총각김치/NNG+서건/JX]
(아)사	내사 그걸 이미 했지	[내/NP+사/JX]
(아)야	그가 인간성이야 그만이지	[인간성/NNG+아야/JX]
(아)아말로	사과아말로 가을의 과일이다	[사과/NNG+아말로/JX]
요	나는요 그림을요 예쁘게 그림니다.	[나/NP+는/JX+요/JX]
조차	이젠 집조차 빼앗기는구나	[집/NNG+조차/JX]
치고	값싼 물건치고 쓸 만하다	[물건/NNG+치고/JX]

■ 보조사 분석 기준

보조사는 ‘이다’의 활용어미와 구분하기 어려운 경우가 있다. 흔히 보조사로 간주되던 몇몇 형태들은 연결어미와 의미상의 차이가 없으며, 분포상으로도 구별되지 않기 때문에 이런 대상들은 보조사로 분석하지 않는다. 즉, 다음의 두 기준을 근거로 보조사의 목록을 확정하였다.

[기준 1] 대상 형태가 용언의 어미로 사용되는가.

[기준 2] 대상 형태가 체언에 후행할 때 서술어의 자격을 가지고 사용되는가.

◦[기준 1, 2]에 부합하는 다음의 형태들은 모두 ‘연결어미’로 분석한다.

(이)ㄴ 들, (이)ㄴ 족, (이)든, (이)든지, (이)라도, (이)라서, (이)라야

• [기준 1, 2]에 부합하지 않는 다음의 형태들은 ‘보조사’가 된다.

(이)나마, (이)야, (이)≡랑, (이)야말로, (이)란

• [기준 1]에 부합하지 않으나, [기준 2]에는 부합하는 형태는 ‘중의성’을 가진다.

(이)나, (이)요

• 다음의 형태는 서술격조사 ‘이다’의 활용형과는 관계가 없으므로 모두 보조사가 된다.

까지, 깨나, 는(은/ㄴ), 대로, 도, 따라, 마다, 마저, 만, 말고, 밖에, 부터, 뿐, 조차, 치고, ㄴ커녕

① 다음의 형태들은 분석 결과에 중의성이 생기게 되므로, 이들을 분석할 때는 특히 주의해야 한다.

(이)란 코알라란 동물은 호주에 주로 서식한다	[코알라/NNG+O/VCP+란/ETM]
코알라란 매우 귀여운 동물이다	[코알라/NNG+란/JX]
(이)나 밥이나 빵을 먹도록 해라	[밥/NNG+O 나/JC]
그가 비록 열심히 하나 능력은 부족하다	[하/VV+나/EC]
어제 내가 술을 마셨나?	[마시/VV+였/EP+나/EF+?/SF]
(이)야 철수야 그 일을 할 수 있지	[철수/NNP+O/JX]
내가 좋아하는 것은 철수야.	[철수/NNP+O/VCP+O/EF+./SF]
철수야! 부르는 소리	[철수/NNP+O/JKV]
(이)요 밥을 먹다가요	[먹/VV+다가/EC+요/JX]
밥이요 빵이요	[밥/NNG+O/VCP+요/EC]

② ‘중결어미+요(보조사)’는 중결어미로 통합하여 분석한다.

예 마음이 예쁜 사람은 좋아하는 사람이 많으니까요. [많/VV+으니까요/EF+./SF]

③ ‘비중결어미+요(보조사)’는 통합하지 않고 각각 분석해 준다.

예 제가 몸이 좀 아파서요 지각을 했어요. [아프/VV+아서/EC+요/JX]

내가요, 왜요 [나/NP+가/JKS+요/JX], [왜/MAG+요/JX]

④ ‘말고’는 용언 ‘말다’의 활용형으로 처리한다.(표준국어대사전의 목록을 반영하여)

예 돈말고 지혜가 필요하다 [돈/NNG+말/VV+고/EC]

6. 의존형태

6.1. 어미

6.1.1. 선어말어미(EP)

용언이 활용할 때, 어간과 어말 어미 사이에 나타나는 것으로 높임법이나 시제, 양태를 나타내는 문법적인 요소이다. 선어말어미의 목록은 연구자에 따라 다를 수 있으나 이 분석에서는 아래의 것만을 선어말어미로 인정한다.

-겠	그 일은 내일 처리하겠다.	[처리/NNG+하/XSV+겠/EP+ㄷ/EF+./SF]
-(으)사	선생님께서 손수 만드신	[만들/VV+시/EP+ㄴ/ETM]
-옵	어머님께 선물을 받치옵고	[받치/VV+옵/EP+고/EC]
-았/었	우리가 먹었던 음식에 문제가 있다	[먹/VV+았/EP+던/ETM]
-았었/었었	거기는 우리가 전에 갔었던 곳이야	[가/VV+았었/EP+던/ETM]

① 선어말어미가 한 음절로 통합된 경우에는 각각 분리해서 분석한다.

-셨	그 일은 어머니께서 하셨다.	[하/VV+시/EP+았/EP+ㄷ/EF+./SF]
----	-----------------	----------------------------

② 다음의 선어말어미는 그 어간이 생략되었을 경우에 어간을 복원해 준다.

-겠	이것은 그대로 두어야겠다.	[두/VV+어아/EC+하/VX+겠/EP+ㄷ/EF+./SF]
-았/었	철수가 그것을 가져오랬다.	[가져오/VV+리/EC+하/VV+았/EP+ㄷ/EF+./SF]
-사	선생님께서 가지시오.	[가/VV+지/EC+하/VV+시/EP+오/EF+./SF]

③ 위의 선어말어미가 포함되지 않은 어미 형태는 그대로 연결어미로 분석한다.

-랴까-, -대야-, -래야-

6.1.2. 종결 어미(EF)

용언의 어간이나 선어말 어미 뒤에 연결되어 용언의 형식을 완성시키는 어미로서 한 문장을 끝맺는 역할을 하는 어미이다. 그러나 종결어미가 문장의 종결에만 사용되는 것은 아니고, 문맥에 따라서는 연결 어미로 쓰이기도 한다. 이 분석에서는 “SF” 앞에서만 종결 어미를 인정한다.

-게	그만한 돈이 있으면 좋게.	[좋/VA+게/EF+./SF]
-구나	넌 정말 멋지구냐!	[멋지/VA+구나/EF+!/SF]
-ㄴ가	이것이 무엇인가?	[무엇/NP+0/VCP+ㄴ가/EF+?/SF]
-ㄴ걸	이제 시작인걸.	[시작/NNG+0/VCP+ㄴ걸/EF+./SF]

-ㄴ다 이런 말도 안 된다.	[되/VV+ㄴ다/EF+./SF]
-나 제네 그리로 가냐?	[가/VV+나/EF+?/SF]
-냐 키가 얼마나 크냐?	[크/VA+냐/EF+./SF]
-네 정말 큰일 났네!	[나/VV+았/EP+네/EF+!/SF]
-느냐 그것보다 이것이 낫느냐?	[낫/VA+느냐/EF+?/SF]
-는걸 그는 벌써 갔는걸.	[가/VV+았/EP+는걸/EF+./SF]
-는구나 앞이 잘 안 보이는구나.	[보이/VV+는구나/EF+./SF]
-구려 당신도 가시겠구려.	[가/VV+시/EP+겠/EP+구려/EF+./SF]
-는구려 잘도 먹는구려.	[먹/VV+는구려/EF+./SF]
-는구먼 공부를 잘하는구먼.	[잘하/VV+는구먼/EF+./SF]
-는다 아이가 글을 잘 읽는다.	[읽/VV+는다/EF+./SF]
-는도다 짐이 조서를 내리는도다.	[내리/VV+는도다/EF+./SF]
-다 그게 사실이다.	[사실/NNG+이/VCP+다/EF+./SF]
-ㄹ게 그렇게 할게.	[하/VV+ㄹ게/EF+./SF]
-버니까 이제야 옵니까?	[오/VV+버니까/EF+?/SF]
-버니다 이렇게 합니다.	[하/VV+버니다/EF+./SF]
-세 제대로 좀 하세.	[하/VV+세/EF+./SF]
-습니까 그래도 되겠습니까?	[되/VV+겠/EP+습니까/EF+?/SF]
-습니다 정말 재미있습니다.	[재미/NNG+있/VV+습니다/EF+./SF]
-버시다다시 만납시다.	[만나/VV+버시다/EF+./SF]
-버시오서둘러 주십시오.	[주/VX+시/EP+버시오/EF+./SF]
-아라/어라 웃기지 말아라.	[말/VX+아라/EF+./SF]
-으냐 물이 얼마나 깊은냐?	[깊/VA+으냐/EF+?/SF]
-은가 그것이 좋은가?	[좋/VA+은가/EF+?/SF]
-자 밥 먹으러 가자!	[가/VV+자/EF+!/SF]
-오/으오/소 물이 깨끗하오.	[깨끗/XR+하/XSA+오/EF+./SF]
-버디다/습디다 참 좋은 곳입니다.	[곳/NNB+이/VCP+버디다/EF+./SF]
-거든 나는 이것이 좋거든!	[좋/VA+거든/EF+!/SF]
-ㄴ 걸/은걸 힘이 꽤 센걸.	[세/VA+ㄴ 걸/EF+./SF]
-ㄹ 걸/을걸 모른다고 할걸.	[하/VV+ㄹ 걸/EF+./SF]
-ㄹ까 이제 밥을 할까?	[하/VV+ㄹ까/EF+?/SF]
-다오 그가 가지고 있다오.	[있/VV+다오/EF+./SF]
-다네 일을 망쳤다네	[망치/VV+았/EP+다네/EF+./SF]
-다구 돈이 많다구?	[많/VA+다구/EF+?/SF]
-다니까 돈이 없대니까!	[없/VA+다니까/EF+!/SF]
-냐고/느냐고 그가 누구냐고?	[누구/NP+이/VCP+냐고/EF+?/SF]
-도다 꽃이 아름답도다.	[아름답/VA+도다/EF+./SF]
-대니 무엇을 읽었대니?	[읽/VV+았/EP+대니/EF+?/SF]
-는가 그가 집에 있는가?	[있/VV+는가/EF+?/SF]
-버디까/습디까 보기에 좋습디까?	[좋/VA+습디까/EF+?/SF]
-다면서 술은 싫다면서?	[싫/VA+다면서/EF+?/SF]
-대나 그도 가겠대나.	[가/VV+겠/EP+대나/EF+./SF]
-렴/으렴 맘대로 해 보렴.	[보/VX+렴/EF+./SF]

-려무나 책이나 읽으려무나.	[읽/VV+으려무나/EF+./SF]
-라니까 그 사람이 아니라니까.	[아니/VCN+라니까/EF+./SF]
-자 잠이나 자자.	[자/VV+자/EF+./SF]
-세 일이나 하세.	[하/VV+세/EF+./SF]
-자꾸나 약속을 좀 미루자꾸나.	[미루/VV+자꾸나/EF+./SF]
-자니까 그만 따지자니까.	[따지/VV+자니까/EF+./SF]
-지 그가 언제 오지?	[오/VV+지/EF+./SF]
-아/아/여 밥 먹어!	[먹/VV+아/EF+!/SF]
-ㅁ 세/음세 그날 꼭 음세.	[오/VV+ㅁ 세/EF+./SF]
-단다 애들이 다쳤단다.	[다치/VV+있/EP+단다/EF+./SF]
-니 그게 없니?	[없/VA+니/EF+?/SF]

① ‘중결어미+요’는 통합해서 중결어미로 분석한다.

예 말씀대로 했는걸요. [하/VV+있/EP+는걸요/EF+./SF]

② ‘-세요’는 다음과 같이 선어말어미까지 분석한다.

예 어서 출근하세요. [출근/NNG+하/XSV+시/EP+어요/EF+./SF]

③ ‘-죠’는 축약형을 그대로 태깅한다.

예 어서 출근하죠. [출근/NNG+하/XSV+죠/EF+./SF]

6.1.3 연결 어미(EC)

용언의 어간이나 선어말 어미 뒤에 연결되어 용언의 형식을 완성시키는 어미로서 문장을 종결시키지 못하고 뒤에 오는 절을 연결시켜 주는 어미를 말한다.

-거나 누가 오거나 알은 체 할 것 없다.	[오/VV+거나/EC]
-거니 비가 오겠거니 생각했다.	[오/VV+겠/EP+거니/EC]
-건대 내가 보건대, 네 말이 옳다.	[보/VV+건대/EC]
-고 일을 하고 밥을 먹자.	[하/VV+고/EC]
-곤 숙제한 것도 빌려가곤 한다.	[빌리/VV+아/EC+가/VX+곤/EC]
-느니 앉아서 걱정하느니 나가서 하겠다	[걱정/NNG+하/XSV+느니/EC]
-든지 외모가 어떠하든지 무슨 상관인가?	[어떠하/VA+든지/EC]
-며 노래하며 춤을 춘다.	[노래/NNG+하/XSV+며/EC]
-면서 푸르면서 검은 물빛	[푸르/VA+면서/EC]
-자 오지마자 당했다.	[오/VV+자/EC]
-락 오르락 내리락	[오르/VV+락/EC]
-거든 가거든 말해라.	[가/VV+거든/EC]
-거늘 이미 늦었거늘 어찌 빨리 가는가?	[늦/VV+있/EP+거늘/EC]
-건마는 말렸건마는 아직도 축축하다.	[말리/VV+있/EP+건마는/EC]

-고자 병을 낫고자 몸부림쳤다.
 -기에 실수했기에 응서해 주었다.
 -ㄴ 다손/다손 밑다손 치다리도 구박하지 말자.
 -ㄴ 다기에 앞으로 잘 한다기에 승낙했다.
 -ㄴ 데/는데 예쁜데 미워한다.
 -ㄴ 들/는들 간다 한들 아주 길까?
 -ㄴ 즉 배가 고프즉 속이 쓰리다.
 ㄴ 지라/는지라 눈이 온지라 길이 미끄럽다.
 -나 눈이 오나 비가 오나 같다.
 -네니 멀리 보이네니 넓은 들아로다.
 -니마 맛이 좋지 못하나마 많이 드십시오.
 -노니 묻노니, 포부가 무엇이나?
 -는다기에 빵을 먹는다기에 주었다.
 -니 멍하니 서있다.
 -니까 웃기니까 좋다.
 -다가 자랑하다가 망신당했다.
 -다기에 그녀가 예쁘다기에 보러 왔소.
 -대도 시간이 있대도 만나 주질 않는다.
 -더라도 가더라도 꼭 돌아와라.
 -던들 잔작 알았던들 방법을 취했지.
 -도록 미치도록 알했다.
 -되 싸우되 꼭 지도록 해라.
 -ㄹ 뽀뽀러 비가 올뽀뽀러 바람도 분다.
 -ㄹ 수록 갈수록 태산이다.
 -ㄹ 지 비가 얼마나 올지 천둥이 다 친다.
 -ㄹ 지리도 이길지리도 명예롭지는 않다.
 -ㄹ 지연정 죽을지연정 그 일은 못하겠다.
 -라고 바보라고 생각한다.
 -람시고 그는 반장이람시고 행패만 부린다.
 -러 청소하러 가자.
 -려 웃으려 한다.
 -려니와 비용도 문제려니와 일꾼도 문제다.
 -려마는 보면 반가우려마는 볼 수가 없네.
 -면 지옥이 존재하면 만원일 것이다.
 -므로 비가 오므로 가지 않겠다.
 -아도/어도 임만 봐도 모르겠다.
 -아서/어서 땀을 흘려서 꺾을 잡았다.
 -ايا 이 일은 잘해야 한다.
 -으나 밥을 먹으나 마니이다.
 -으니마 맛은 없으니마 많이 드세요.
 -음직 먹음직한 음식
 -지마는 비가 오지마는 가야 한다.
 -게 개를 굶게 하지 마라.

[낫/VV+고자/EC]
 [실수/NNG+하/XSV+있/EP+기에/EC]
 [밑/VA+다손/EC]
 [하/VV+ㄴ 다기에/EC]
 [예쁘/VA+ㄴ 데/EC]
 [하/VX+ㄴ 들/EC]
 [고프/VA+ㄴ 즉/EC]
 [오/VV+ㄴ 지라/EC]
 [오/VV+ㄴ 나/EC]
 [보아/VV+네니/EC]
 [못하/VX+니마/EC]
 [묻/VV+노니/EC+/,SP]
 [먹/VV+는다기에/EC]
 [멍하/VV+니/EC]
 [웃기/VV+니까/EC]
 [자랑/NNG+하/XSV+다가/EC]
 [예쁘/VA+다기에/EC]
 [있/VV+대도/EC]
 [가/VV+더라도/EC]
 [알/VV+있/EP+던들/EC]
 [미치/VV+도록/EC]
 [싸우/VV+되/EC]
 [오/VV+ㄹ 뽀뽀러/EC]
 [가/VV+ㄹ 수록/EC]
 [오/VV+ㄹ 지/EC]
 [이기/VV+ㄹ 지리도/EC]
 [죽/VA+ㄹ 지연정/EC]
 [바보/NNG+0/VCP+라고/EC]
 [반장/NNG+0/VCP+람시고/EC]
 [청소/NNG+하/XSA+러/EC]
 [웃/VV+으려/EC]
 [문제/NNG+0/VCP+려니와/EC]
 [반갑/VA+려마는/EC]
 [존재/NNG+하/XSA+면/EC]
 [오/VV+므로/EC]
 [보/VV+아도/EC]
 [놓/VV+아서/EC]
 [잘하/VV+아야/EC]
 [먹/VV+으나/EC]
 [없/VA+으니/EC]
 [먹/VV+음직/EC+하/VX+ㄴ /ETM]
 [오/VV+지마는/EC]
 [굶/VV+게/EC]

-고	일단 먹고 보자.	[먹/VV+고/EC]
-아/어	입을 막아 버렸다.	[막/VV+아/EC]
-지	우기지 못해 버렸다.	[우기/VV+지/EC]

- ① 어미에 따라서는 분석의 중의성이 생길 수 있으므로 문맥 확인을 통해 형태분석을 결정한다.

- '-는데'는 '는데/EC'와 '는데/ETM+데/NNB'로 중의성을 갖는다.

예	너는 내가 왔는데 기쁘지도 않니?	[오/VV+왔/EP+는데/EC]
	내가 지금 있는데가 어디지?	[있/VV+는/ETM+데/NNB+기/JKS]

- '-는데'도 '는데/EC'와 '는데/ETM+지/NNB'로 중의성을 갖는다.

예	다들 만족하는지 아무런 불평이 없다.	[만족/NNG+하/XSV+는지/EC]
	너를 만난지도 꽤 오래구나.	[만나/VV+는/ETM+지/NNB+도/JX]

- ② '-음직'은 “음직/EC”로 분석한다. 그러나 ‘바람직’은 그 자체가 하나의 어근이므로 더 이상 분석할 수 없다는 것에 유의한다.

예	그것 참 먹음직스럽다.	[먹/VV+음직/EC+스럽/XSA+다/EF+./SF]
예	그것은 매우 바람직한 일이다.	[바람직/XR+하/XSA+는/ETM]

- ③ ‘-라카/-라케-’의 경우 ‘라고+하’형은 ‘라고/EC+카/VV’로 분석한다. 또한 ‘려고+하’형은 ‘려/EC+카/VX’로 분석한다.

예	그리고 병원 가는 데도 무료라카데	[무료/NNG+아/VCP+라고/EC+카/VV+다/EF+./SF]
	또 피리 만들라케나?	[만들/VV+라고/EC+카/VX+나/EF+./SF]

6.1.4. 명사형 전성 어미(ETN)

한 문장의 성격을 임시로 바꾸어 다른 문장 속에서 명사적인 역할을 하게 하는 어미를 말한다.

-기	그 일은 정말 중요하기 때문이다	[중요/XR+하/XSA+기/ETN]
-마/-음	장사는 신용을 얻음이 제일이다	[얻/VV+음/XSN+아/JKS]

- ① 불규칙 용언 어간에 명사형 전성 어미가 붙어 있을 경우 ‘-음’이 아닌 ‘-기’으로 분석한다.

예	김철수 지음	[짓/VV+기/ETN]
---	--------	--------------

- ② “음, 기”가 붙은 말이 단순히 명사형이냐 아니면 굳어진 명사이냐 하는 것은 물론

문맥에 따라 결정되어야 하지만 먼저 그것이 “사전”에 등재되어 있느냐의 여부를 살펴보아야 한다.

예 책을 읽기가 어렵다	[읽/VV+기/ETN+기/JKS]
읽기 교육이 문제가 된다	[읽기/NNG]

6.1.5. 관형사형 전성 어미(ETM)

용언의 성격을 임시로 바꾸어 다른 문장 속에서 관형사적인 역할을 하게 하는 어미이다.

-ㄴ/은	어제 먹은 빵에 이상이 있었다	[먹/VV+은/ETM]
-는	읽어버린 물건을 찾는 일은 요즘은 어렵다	[찾/VV+는/ETM]
-던	이제까지 미루던 일을 오늘 해치웠다	[미루/VV+던/ETM]
-ㄹ/을	나에게는 아직 처리할 일이 있다	[처리/NNG+하/XSV+ㄹ/ETM]
-런	어제런 듯하다.	[어제/NNG+O/VCP+런/ETM]

① 불규칙 용언 어간에 관형사형 전성 어미가 있을 경우 ‘-은, -을’이 아닌 ‘-ㄴ, -ㄹ’로 분석한다.

예 그녀의 고운 얼굴	[곱/VA+ㄴ/ETM]
그녀는 매우 아름다울 것이다.	[아름답/VA+ㄹ/ETM]

② 연결 어미에 이어서 전성 어미가 올 경우 통합해서 전성어미로 처리한다.

예 어느 쪽에 더 비중을 두느냐는 것이	[두/VV+느냐는/ETM]
-----------------------	----------------

6.2. 접두사(XP)

접두사는 명사와 수사에 결합하는 접사류를 묶어서 체언접두사만을 설정하기로 한다.

6.2.1. 체언 접두사(XPN)

명사 접두사에는 한자어계 접두사와 고유어계 접두사가 있는데, 그 목록의 풍부함에 비해 대개가 생산성이 그리 높지 않다. 일단 여기서는 비교적 생산성이 높다고 인정되는 접두사와, 접두사를 분리했을 경우 단일한 표제어로 등재될 수 있는 경우에 한해서 접두사 분석을 하기로 한다.

가(假)

가건물

고(高)	고물가
과(過)	과보호
구(舊)	구소련
날	날음식
노(老)	노부부
대(大)	대선배
말	말아들
맨	맨몸
무(無)	무의식
미(未)	미완성
반(反)	반독재
범(凡)	범세계
부(不)	부도덕
불(不)	불합리
비(非)	비논리
생(生)	생김치
소(小)	소강당
신(新)	신정당
왕(王)	왕족발
재(再)	재충전
저(低)	저임금
제(第)	제13차
준(準)	준전시
초(超)	초만원
최(最)	최고급
친(親)	친러시아
탈(脫)	탈냉전시대
폐(廢)	폐광산
풋	풋살구
피(彼)	피고소인
한	한가운데
헛	헛고생

6.3. 접미사(XS)

파생 접미사에는 어기의 품사를 바꾸는 것과 그렇지 않은 것이 있는데, 이들을 별도로 구별하여 표지를 부여하지는 않는다.

6.3.1. 명사파생접미사(XSN)

명사파생접미사는 명사나 다른 어근에 후행하여 그것이 명사의 기능을 수행할 수 있

도록 만들어 주는 의존 형태이다. 그러나 명사파생접미사는 연구자에 따라 그 목록이 다르며, 실제로도 구분이 애매한 경우가 많다. 본 분석에서는 접미사의 생산성과 접미사를 제외한 형태의 독립성을 기준으로 다음과 같이 목록을 마련하였다.

가(價) 매매가
 가(哥) 김가
 경(頃) 두 시경
 계(系) 몽고계
 계(界) 교육계
 광(狂) 메모광
 권(圈) 운동권
 권(權) 참정권
 당(當) 한 사람당
 대(臺) 역대
 덕(宅) 청주덕
 론(論) 비평론
 별(別) 가구별
 여(餘) 삼십여
 류(類) 자연류
 률, 율(率) 경쟁률
 리(裡) 비밀리
 분(分) 분량 일인분
 분(分) 3분의
 산(産) 중국산
 상(上) 역사상
 생1(生)갑자생
 생2(生)간습생
 성(性) 인간성
 시(視) 영웅시
 용(用) 전쟁용
 적(的) 사상적
 형(型) 기본형
 형(形) 도시형
 제(制) 봉건제
 층(層) 선수층

치(值) 보름치
 풍(風) 북고풍
 화(化) 도구화
 기 기름기
 깨 10분께
 꿀 십 원꿀
 까리 전우까리
 곤 노름곤
 네 동이네
 님 선생님
 들 우리들
 들이 1ℓ들이
 배기 열 살배기
 뽕 조끼뽕
 씹 만원씹
 장이 간판장이
 쟁이 심술쟁이
 질 서방질
 짜리 백 원짜리
 쟈1 이틀쟁
 쟈2 옹기쟁
 치레 인사치레
 투성이 먼지투성이

- ① 명사파생접미사인 ‘-들’은 그 분포가 매우 다양하여 일부에서는 이를 보조사와 접미사로 나누어 분석하기도 한다. 그러나, 본 분석에서는 이들을 모두 명사파생접미사로 처리한다. ‘먹고들’의 ‘-들’도 선행성분이 어미이긴 하나, 일치하는 대상은 선행하는 명사로 해석할 수도 있기 때문이다.

예 사람들이 우리 집에 왔다. [사람/NNG+들/XSN]
 그들은 밥을 먹고들 싶었다. [먹/VV+고/EC+들/XSN]

- ② ‘-님’은 다음과 같이 세 가지의 분석 중의성을 가지므로 주의해서 분석한다.

- ‘임’의 의미로 쓰인 경우: 보통명사

예 님과 이별하다. [님/NNG+과/JKB]

- 사람의 ‘이름’이나 ‘성’ 뒤에서 쓰인 경우: 의존명사

예 김철수님께서 오셨습니다. [김철수/NNP+님/NNB+께서/JKS]

- 그 밖의 경우: 명사파생접미사

예 과장님이 부르십니다. [과장/NNP+님/XSN+0/JKS]

6.3.2. 동사파생접미사(XSV)

동사파생접미사는 어기 또는 어근에 붙어서 그것을 동사로 만들어 주는 기능을 갖는 접미사이다. 여기서는 그러한 접미사 중 현재 생산성을 가지고 쓰이는 것만을 인정하여 분석한다.

당하	아군이 공격당하는 데에는 이유가 있다.	[공격/NNG+당하/XSV+는/ETM]
되	아침식사가 이미 준비되어 있었다.	[준비/NNG+되/XSV+0/EC]
시키	오늘 강아지를 운동시키려고 공원에 나갔다.	[운동/NNG+시키/XSV+려고/EC]
하	외국에서 공부하는 일이 쉬운 것은 아니다.	[공부/NNG+하/XSV+는/ETM]

6.3.3. 형용사파생접미사(XSA)

형용사파생접미사는 어기나 어근에 붙어서 그것을 형용사로 파생시키는 접미사이다. 여기서는 그러한 접미사 중 현재 생산성을 가지고 쓰이는 것만을 인정한다.

답	사람이 사람답게 행동해야 사람이지	[사람/NNG+답/XSA+게/EC]
되	자식된 도리로 어떻게 그런 짓을..	[자식/NNG+되/XSA+ㄴ/ETM]
롭	어려운 일일수록 슬기롭게 대처하라	[슬기/NNG+롭/XSA+게/EC]
스럽	그녀의 사랑스러운 표정을 보거라	[사랑/NNG+스럽/XSA+ㄴ/ETM]
하	멍청한 표정을 짓지 말아라	[멍청/XR+하/XSA+ㄴ/ETM]

◦ 표준국어대사전에 따라, 기존에 접미사로 분석하던 ‘만하’의 지침을 변경하여, ‘만’을 보조사로, ‘하’를 동사로 분석한다. ‘만하’는 ‘만/NNB+하/XSA’로 분석되는 경우도 있으므로 주의해야 한다.

예 그 일을 처리하는 데 철수만한 인제가 없다 [철수/NNP+만/JX+하/VV+ㄴ/ETM]
 예 이 음식은 먹을 만하다. [먹/VV+을/ETM] [만/NNB+하/XSA+ㄴ/EF+/SF]

6.4. 어근(XR)

국어에는 하나의 단어가 조사에 의해 분리되는 현상이 있다. 즉, 파생된 용언에서 보

조사 등의 삽입에 의해 어근과 접사가 분리되는 현상이 있다. 지금까지 알려진 형태소 분석 표지에서는 이들에 대해 적절한 표지를 제공하지 못한 채 어근을 명사 등의 범주 속에 포함시켜 왔다. 그러나, 세종 표준안에서는 어근 단위에 대한 표지를 설정하였으므로 이들에 어근이라는 표지를 할당할 수 있다.

예 땀도 하다.

[땀/XR+도/JX] [ᄒ/VV+ㄷ/EF+./SF]

■ 어근의 분석 지침

- ① ‘하다’가 결합한 어휘 중 ‘하다’에 선행하는 음절이 2음절 이상일 경우에는 이들은 일단 어근으로 분석될 수 있는 후보가 된다. 어근이 보조사 등에 의해 분리되지 않은 경우에도 물론 어근으로 분리해서 분석해 준다.

예 명칭은 하다.

[명칭/XR+은/JX] [ᄒ/VA+ㄷ/EF+./SF]

명칭하다.

[명칭/XR+ᄒ/XSA+ㄷ/EF+./SF]

- ② ‘하다’가 결합한 어휘 중 ‘하다’에 선행하는 음절이 1음절일 경우에는 어근 분리 현상이 매우 제한적이므로 이 경우에는 통합형으로 분석한다.

예 취하다.

[취ᄒ/VV+ㄷ/EF+./SF]

- ③ 물론, ‘하다’에 선행하는 음절이 독립된 품사로 인정되는 경우에는 해당 품사로 분석한다.

예 일하다.

[일/NNG+ᄒ/XSV+ㄷ/EF+./SF]

- ④ ‘양’, ‘체’, ‘척’, ‘만’, ‘뻔’, ‘듯’ 등의 의존명사와 ‘하다’가 결합한 성분도 분리해서 분석한다.

예 양하다.

[양/NNB+ᄒ/XSV+ㄷ/EF+./SF]

체하다.

[체/NNB+ᄒ/XSV+ㄷ/EF+./SF]

- ⑤ 일반적인 명사의 용법을 보이지 못하는 한자어 어근의 경우도 명사가 아니라 어근에 포함된다.

예 적합하다.

[적합/XR+ᄒ/XSA+ㄷ/EF+./SF]

■ 어근 목록(표준국어대사전과 일치하지 않음)

가능 기쁜 기미득 기무잡잡 기쁜 가지런 기혹 각다분 각별 간악 감감 감승 갑갑 강격 강열 강은 기름 갈쪽 거대 거든 거룩 거무뎡뎡 거무레 거무스레 거무스름 거무죽죽 거무칙칙 거무튀튀 거무티티 거뭇 건건쫘쫘 건심 건전 걸직 곁쪽 곁쪼 곁성드뭇 결국 고가누룩 고리타분 고만고만 고이

고스낙 골돌 골똥 골번 과대 과도 팔팔 광대 쾌센 괴상망칙 괴팍 구가 구리척척 구부정 구저분 구
주레 구중중 구지레 구차 군잡 군건 그들먹 그든 그털싸 그림직 그림 그만그만 그옥 극명 극밀 극
박 글박 글적 기우듬 기이 긴가민가 깊숙 까끌 까리 까마득 까마아득 까무잡잡 까무죽죽 까칠 깎듯
간간 깔끄레 감감 감직 감썸 감깁 감곰 깨끗 깨긋 깨름 깨굴 깨림 깨림직 깨림척 깨병 깨분 꺼칠 켄렁
켄썸지근 겹겹 꼬로스럼 꼬로스름 꼬부장 꼬불탕 꼬끔 꼬끗 꼬궁땅꼬 꼬죄죄 꾸깃 구부정 꾸적꾸적
꾸준 꼬느름 꼬무레 끈끈 끈적 끔직 끔썸 끔썸끔썸 니긋 니들 니른 니름니른 나약 나즈막 나지막
날렵 날씬 날썸지근 날감 낱딱 내만 냉냉 너끈 너더분 너저분 너절 널널 널죽 널직 널썸 널데데 넓
죽 넓직 넓넌 녀석노루푸레 노르께 노르끼리 노르스름 노리께 노리끼 노리끼리 노릿노릿 녹작지근
녹진 뇌리끼리 누르끼 누르스름 누척지근 눅눅 녹진 뉘엿 느리치분 느리터분 느릿 느물 느슨 느즈
막 느지막 느직 녹실 늘썸지근 늘썰 늘편 늑수구레 늑수그레 능엄 니끼 다각 다복 다소곳 다양 다
이내믹 다정 단단 단순 단출 단출 달썸 달썸지근 달착지근 달콤 달콤세콤 담지 답답 당당 대담 대
대 더풀 덜큰 덩덤 데데 델리케이트 도도록 도통 도통 독랄 독득 돌방 동그마 동그스럼 동그스름
둥글 둥글납작 동일 되직 두두룩 두런 두루몽수리 두루몽술 두루몽실 두리몽실 두리번 두툼 동그스
럼 동그스름 둥글납작 둥글넓적 둥글납직 뒤숭숭 드리마틱 든든 들락 등성긋 등직 띠끈 띠끔 띠땃
띠땃무레 띠끈 땃땃 땡땡 땡땡막 떠들석 떠르르 떠름 떼떠름 떼땃 땡글 또렷 뚜릿뚜릿 뚱뚱 뜨끈
뜨끔 뜨뜻미지근 뜨뜻 뜨뜻미지근 뜨아 뜨악 뜨직 뜻뜻 땡땡 로맨틱 리드미컬 마땅 마득 만무 만발
말썸 망척 매끈 매카 매케 매케 매콤 매름 맨들맨들 맵살 맵쌀 맵자 맹상 맹송맹송 맹송맹송 머뭇
머쓱 머춤 멀뚱 멀뚱 말썸 말썰 말썰막 말썸 멍청 명량 물렁 못마땅 뚱뚱뚱뚱 무궁 무궁광대 무궁
무궁 무끼럼 무던 무덥덥 무득 무뚱 무뚱뚱 무뚱뚱 무료 무룩 무모 무지근 무추름 무해 목직 물렁
물커덩 뚱글케케 뚱뚱 뚱글 뚱툭 마끈덩 미니멀 미세 미안 미옥 미지근 미칠 미흠 민둥 민송 입광
밍밍 비듯 비람직 발그대대 발그스름 발칙 배다 번드레 번잡 범용 변변 보들녹진 보송 보유품 볼
그죽죽 부당 부듯 부산 부음 부유품 부음 분명 분주 불과 불그레 불그스럼 불그스레 불그스름 불
그죽죽 불리 불순 불쌍 불가 불과 붉으스름 붉으죽죽 비긋 비뿔 비리치근 비스듬 비줄 뻗땃 뻗근
뻗덩 뻗젓 뻗뻗 뻗로통 뻗그죽죽 뻗로통 뻗까뻗썸 뻗닥 뻗뚜름 시근 살뿔 살멧 살벌 상냥 새춤 새
치름 새침 새콤 새콤달콤 새콤텃텃 생뚱 생생 서글 서글서글 서늘 서먹 서먹서먹 섣뚱섣뚱 선량 선
명 설면설면 섭섭 세련 섹시 센세이셔널 소도록 소들 소백 속상 쇄쇄 쇄양배양 쇼킹 수꿀 수더분
수두룩 수부룩 수성수성 순진 스산 슬깃 시금털털 시글시글 시들 시무룩 시뻘장 시뻘 시시껄렁 시
시곱작 시원섭섭 시척지근 시커메티티 시큰둥 시름 시통 시푸르스름 식상 신바막썸 신평넌 실랄 심
드렁 삼승삼승 싸리 싸아 쌀벌 싹싹래 싹싹 씨늘 썰렁 쓰리쓰리 씹스름 씹쌀 씨그동 씨니컬 씨무룩
씨식 씨언 아둔 아득 아령척 아리 아리몽송 아리송 아심삼 아수쿠러 아스므레 아슴 아쌀 아이러니
컬 아이러닉 아이로니컬 아찔 알뜰 알랑 알썸지근 앙증 앙큼 애돌애돌 애매모호 애틀 아시시 아트
막 악악 알삭 알쌍 알알 알팍 알포름 앓상 어긋동 어긋똥 어두컴컴 어둑스근 어둑신 어둑침침 어둑
컴컴 어더 어더어더 어떨떨 어린 어렵풋 어리더덤 어리둥절 어리벙벙 어리석 어리숙 어리송 아마두
저 아마어마 어머어마 어벙 아수룩 아수룩 아수신 아삭 아웅 어정쩡 어지간 어정쩡 아찔 얼근 얼떨
떨 얼명영성 얼쑤덜쑤 얼썸 얼큰 엇비슷 엇비슷 엉뚱 엉설 엉성 엉큼 예로틱 엔간 엔간 여들 여들
여들 여럿 여유민만 여차저차 여툼 열열 예쁘장 오련 오봇 오사 바사 오중중 와일드 욕살 욕자 왕
성 요리저러 요리조러 요상 우복 우악살 원기왕성 웅간 웅만 위세당당 유니크 유력 유머러스 유명
유명짜 유용 유익 으늑 으리으리 으스 으스름 으쓱 으슬 으젓 은근 음훈 응큼 의아 이러이러 이러
저러 이룩 이쁘장 이상 이상망칙 이악 이지랑 이지러 익숙 일관 일정 자그마 자깁 자우룩 자지분
자잘 자잘구레 잔인 잔조름 짙룩 잠자무룩 저러저러 저속 적대 적합 전무 점직 정갈 정당 정확 조
그마 조봇 조초 조출 주도 주요 중대 중요 즐편 지긋지긋 지리 지저분 지질 질썸 질척 질척 질퍽
질퍽 짐짐 짜릿짜릿막 짹막 짹달막 짹막 짹잘 짹조름 짹째름 짹잘 짹째레 짹째 짜릿 짹째 쪼그마
쪼자 쪼깃 쪼그마 쪼그마 쪼그똥 쪼뻗둥 쪼뻗드 쪼뻗드드 쪼뻗뚱 쪼뻗둥 쪼둑 쪼뽀뽀 쪼뽀 쪼뽀 쪼뽀

직작 침담 침흑 처절 천천 철저 초과 초롱 촉박 촌촌 추레 추레 출무성 총일 총직 취악 촛촛 치렁
 치열 친밀 침통 가랑 칼칼 감감 겹겹 겹겹무레 겹겹스레 곱곱 괴괴 굴굴 퀴퀴 퀴퀴칙칙 귀귀
 그낙 큼지막 큼직 큼큼 타이트 탐탁 터분 터뜨 텃수룩 텃텃 통째 두미 툇툇 특별 특이 튼튼 툃툃
 툃툃걸직 파근 파르스름 파름 파릇 편온 평퍼짐 포근 포르쭈쭈 포실 푸군 푸근 푸더분 푸르덩덩 푸
 르무레 푸르스럼 푸르스레 푸르스름 푸릿 푸짐 푸더분 하구 하아무레 한미 할랑 함초롱 함추름 해
 괴망측 해꼬무레 해곰 해낙낙 해납작 해말썽 해반드르르 헬썽 헬썽 항긋 허룩 허름 허망 허수룩 허
 수름 허술 허스키 허악 허어말끔 허어말썽 허우룩 허천 현썰 현칠 혈령 혈직 협수룩 행뎡 현격 현
 란 호끈 호동그레 호젓 호줄근 혼재 혼탁 훌기분 훌쭉 흥건 회금 황망 황폐 황홀 행뎡그렁 회뿌옴
 후덕지근 후덕지근 후련 후줄근 후줄근 후출 후텃지근 훌렁 훌쭉 휘어 흰썰 흰칠 행뎡그렁 행뎡그
 레 휘번 휘형 횡동그레 흉악 흉축 흉치 흉폭 흐끄무레 흐드분 흐리멍덩 흐리멍텅 흐리터분 흐릿 흐
 뭇 흐뭇 혼곤 흡족 흥건 희끄무레 희곰 희뚝마룩 희뚝 희부 희부옴 희부옴 희봄 희뿌옴 희뿌옴 희
 뽕 희어말끔 희읍스름 하쭈구리

7. 기타

7.1. 기호

영문이나 한자, 기호 등이 어절 중간에 개입하여 올바른 분석이 불가능한 경우에는 각각의 요소를 분리하여 분석한다.

예) 마이크로소프트(microsoft)사	[마이크로소프트/NNP + (/SS + microsoft/SL +)/SS
+ 사/NNG]	
농·수산물	[농·수산물/NNG]
초·중·고	[초/NNG + ·/SP + 중/NNG + ·/SP + 고/NNG]

단, 모두 분리하였을 때 표지를 줄 수 없는 불완전한 형태가 생기는 경우에는 예외적으로 전체 어절을 통합하여 처리하기로 한다.

충분(充分) \bar{O} | [충분(充分) \bar{O} /MAG]

7.2. 주말

준말은, 그것이 본딤말과 대등하게 사용되고 분석결과가 동일한 어절 단위를 형성할 경우에 한해서만 복원한다. 그러나 다음에서처럼, 본딤말로 복원할 경우 어절 수에 변화가 생길 뿐 아니라 본딤말로 복원하는 정도가 일관성을 띠지 않게 되는 경우는 굳이 복원하지 않는다. 그러나, 이러한 원칙이 모든 경우에 일관적으로 적용될 수 있는 것은 아니다. 결국 준말의 처리는 해당 어절에 따라 임의적일 수 있다.

[예] 리는 리/ETM (○)

[라고/JKQ+하/VV+는/ETM] (×)

7.3. 분석불능범주(NA)

그 자체가 사전에 등재되어 있지도 않으면서, 축약의 정도가 심하거나 분석하기 어려운 방언형의 경우 분석불능범주로 처리한다.

예) 담배가 쪼메턴게 히마 지라서 빼나? [쪼메턴게/NA]

7.4. 합성어

합성어를 어느 차원까지 분석할 것인지를 명확히 규정하는 것이 필요하다. 그렇지 않으면 분석 결과의 일관성을 확보하기 어렵기 때문이다. 세종 형태분석 말뭉치 구축에서 합성어는 보다 정확하고 일관된 기준의 설정을 위해 다음과 같은 원칙에 의해 처리한다.

■ 합성어의 분석 지침

- ① 표준국어대사전의 표제어 목록을 최대한 반영한다. 즉, 표준국어대사전에 합성어로 등재된 단위는 세분해서 분석하지 않는다.

예) 정치권력 [정치권력/NNG]

- ② 다만, 합성어로 등재되어 있되 띄어쓰기를 허용한 합성어는 세분하여 분석하는 것을 원칙으로 한다.

예) 학생운동 [학생/NNG+운동/NNG]

- ③ 사전에 등재된 합성어가 말뭉치상에서 어절이 분리되어 띄어쓰기가 되어있는 경우 어절을 분리하여 분석한다.

예) 정치 권력 [정치/NNG] [권력/NNG]

예) 민간 단체 [민간/NNG] [단체/NNG]

☞ 의미 합성성이 준수되는 복합 형태는 다음과 같은 원칙에 따라 분석한다.

- 복합 형태의 뒤쪽으로부터, 복합 형태를 구성하고 있는 일부분이 사전에 등재되어 있고, 그것을 제외한 나머지 또한 독립성이 있다면 분석한다.

예) 이등품 [0/NR+등품/NNG]

■ 1음절어 처리

○ 대상 : 본 분석 지침에서 의존명사나 명사파생접사로 설정하지 않은 1음절어
 ○ 지침 : 표준국어대사전에 명사나 의존명사로 등재되어 있고 독립되어 사용될 수 있는 1음절어는 나누어 분석하는 것을 원칙으로 한다. 독립되어 사용될 수 없는 1음절 명사나 의존명사는 붙여서 분석하되 그 목록을 별도로 유지한다. 다만, 표준 국어대사전에서 하나의 1음절어가 명사와 접사, 또는 의존명사와 접사로 동시에 등재될 수 있다. 이에 대한 세부적인 원칙은 다음과 같다.

- 표준국어대사전에 명사나 의존명사로만 등재되고 접미사로 등재되지 않은 경우 나누어 분석한다. 단, 의존명사로만 등재된 경우 본 지침의 의존명사 목록을 참고하여 분석한다.

예 농약병은

[농약/NNG+병/NNG+은/JX]

예 근로기준법을 (법률)

[근로/NNG+기준/NNG+법/NNG+을/JKO]

- 표준국어대사전에 명사나 의존명사로 등재되어 있다 하더라도 결합되는 단어가 비자립성을 띄는 경우나 판단이 어려운 경우 붙여서 분석한다.

예 오솔길이

[오솔길/NNG+0/JKS]

- 표준국어대사전에 명사와 접사 또는 의존명사와 접사로 모두 등재되어 있는 경우 한 어절 내에서는 붙여서 분석하고 어절이 분리되어 있는 경우에는 명사나 의존명사로 분석한다. 단, 분석 지침의 의존명사 목록을 참조해서 분석한다.

예 비람회장

[비람회장/NNG]

예 대화의 장이 마련되었다.

[장/NNG+0/JKS]

예 조리법을 (방법, 방식)

[조리법/NNG+을/JKO]

예 간장 만드는 법은

[법/NNB+은/JX]

- 표준국어대사전에 접미사로만 등재된 경우에는 붙여서 분석한다.

예 포장지

[포장지/NNG]

예 조직책

[조직책/NNG]

cf. 역사책

[역사/NNG+책/NNG]

- 표준국어대사전에 등재되어 있지 않고 본 지침의 의존명사나 접사 지침에도 없는 1음절어의 경우 붙여서 분석한다.

예) 사실혼

[사실혼/NNG]

☞ 표준국어대사전에 명사로 등재되어 있지만 독립적으로 사용되기 어렵기 때문에 붙여서 분석한 1음절 명사 목록

- 감
 - 1) 일부 명사 뒤에 붙어 ‘자격을 갖춘 사람’의 뜻을 나타내는 말.
예) 신랑감/며느릿감/사윗감/장군감.
 - 2) 일부 명사 뒤에 붙어 대상이 되는 도구, 사물, 사람, 재료의 뜻을 나타내는 말.
예) 구경감/놀림감/떨감/양념감/안줏감/웃음감/장난감.
- 고(庫)
 - 1) 곳간(庫間)
예) 생선취급고.
- 교(敎)
 - 1) 종교.
예) 개신교/유태교
- 구(區)
 - 1) 숫자를 나타내는 말 뒤에 쓰여 넓은 범위의 것을 일정한 기준에 따라 몇으로 나눈 하나하나의 구획.
예) 상업구
 - 2) 일부 명사 뒤에 붙어 ‘법령 집행을 위하여 정한 구획’의 뜻을 나타내는 말.
예) 선거구/투표구.
- 구(球)
 - 1) 공처럼 둥글게 생긴 물체. 또는 그런 모양.
 - 2) 한 점에서 같은 거리에 있는 모든 점으로 이루어진 입체. ≒공1
예) 결정구/광속구/사사구/승부구/유인구
- 난(亂)
 - 1) 분쟁, 재해 따위로 세상이 소란하고 질서가 어지러워진 상태.
 - 2) 작은 소동을 비유적으로 이르는 말.
예) 구직난/기술난
- 대(隊)
 - 1) 소대, 중대, 대대 따위의 편제(編制) 부대.
 - 2) 편성된 대열.
예) 구급대/방위대/봉사대
- 대(臺)
 - 1) 흙이나 돌 따위로 높이 쌓아 올려 사방을 바라볼 수 있게 만든 곳.
 - 2) 물건을 떠받치거나 올려놓기 위한 받침이 되는 기구를 통틀어 이르는 말.
예) 검색대/전지대/경매대/망월대/분수대
- 도(度)
 - 1) 어떠한 정도나 한도.
예) 강직도/공현도/기여도

- 동(棟)
 - 1) 규모가 큰 집채.
 - 예 강의동/연구동/학사동
- 량(量)
 - 1) 한자어 명사 뒤에 붙어 분량이나 수량의 뜻을 나타내는 말.
 - 예 가사량/노동량/작업량
- 례(例)
 - 1) 일부 명사 뒤에 붙어 ‘본보기’의 뜻을 나타내는 말.
 - 예 인용례/판결례
- 명(名)
 - 1) ‘이름’의 뜻을 나타내는 말.
 - 예 작품명/저자명
- 모(母)
 - 1) 문어체에서, ‘어머니’를 이르는 말
 - 예 맹금모/악덕모
- 벽(癖)
 - 1) 고치기 어렵게 굳어 버린 버릇.
 - 예 청결벽
- 변(邊)
 - 1) 물체나 장소 따위의 가장자리.
 - 예 강변
- 보(譜)
 - 1) 사물의 계통과 순서를 세워, 그것을 체계적으로 기록한 것.
 - 예 족보/악보/청년보
- 불(佛)
 - 1) ‘부처’의 뜻을 나타내는 말.
 - 예 삼존불
- 사(死)
 - 1) 죽음
 - 예 파로사, ※사-(死) 예 사백신
- 상(上)
 - 1) 물건의 위나 위쪽.
 - 예 천판상
- 상(相)
 - 1) 볼 수 있고, 알 수 있는 모습.
 - 예 구체상/궁핍상
- 상(像)
 - 1) ‘모범’, ‘본보기’의 뜻을 나타내는 말.
 - 예 어머니상/교사상
- 좌(座)
 - 1) 앉을 자리나 지위.
 - 예 주교좌
- 쪽

- 1) 쪼개진 물건의 한 부분.

예 금쪽

• 환(換)

- 1) 멀리 있는 채권자에게 현금 대신에 어음, 수표, 증서 따위를 보내어 결제하는 방식.

예 어음환

《주의1》

복합 형태로 된 나무이름은 분석할 수 없는 ‘느티나무, 너도밤나무’ 등을 고려하여 그 통합형 전체를 하나의 명사로 분석한다.

예 사과나무/NNG, 감나무/NNG, 느티나무/NNG, 너도밤나무/NNG

《주의2》

‘꽃’이 다른 말과 복합 형태를 이룰 경우, 선행어가 식물이나 계절 공간을 의미하는 단어일 경우에 분석하고, 그렇지 않을 경우엔 통합형 전체를 하나의 단어로 분석한다.

예 개나리/NNG+꽃/NNG, 봄/NNG+꽃/NNG, 들/NNG+꽃/NNG

분꽃/NNG, 니팔꽃/NNG, 접시꽃/NNG

《주의3》

‘색’이나 ‘빛’이 다른 단어와 복합 형태를 이룰 경우, 그것의 선행어가 ‘색깔’을 나타내는 말이면 통합형으로 처리하고 그렇지 않은 경우엔 분석한다.

예 노란색/NNG, 바나나/NNG+색/NNG

노란빛/NNG, 바나나/NNG+빛/NNG

빈 면

3. 형태이미분석 말뭉치 구축

가. 연구의 목적과 필요성

1) 연구 목적

이 연구의 목적은 21세기 세종계획 제2단계 2차년도(2002)에 구축된 250만 어절 규모의 형태분석 말뭉치를 대상으로 품사 단계의 동음이의어 중의성을 해결한 분석 말뭉치를 구축하는 것이다. 이를 위해 지난 3년에 걸쳐 550만 어절 규모의 형태이미분석 말뭉치를 개발한 경험을 바탕으로 분석 방법론 및 도구를 보완하여 분석 말뭉치를 구축하였다.

2) 연구의 필요성

형태분석 말뭉치는 어절을 대상으로 형태 단위의 중의성을 해소한 것이지만, 의미의 중의성을 해결한 것은 아니기 때문에 기계번역과 같은 고급 자연 언어 처리나 정밀한 언어학적 연구에 이용하기에는 여러 가지 한계를 갖고 있다. 형태이미분석 말뭉치는 구축에 있어 시간과 노력이 많이 들지만, 각종 자연 언어 처리와 언어 연구 분야에서 활용 가능성이 매우 높으므로 해외 각국에서 개발되고 있으며, 국내에서도 연세대학교와 국립국어연구원 등에서 구축이 이루어졌는데, 연세대학교의 의미 주석 말뭉치는 연세 한국어 사전을 참조로 의미 주석을 시도한 것이며, 국립국어연구원의 의미 주석 말뭉치는 표준국어대사전을 기초로 한 것이다.

본 연구에서 구축한 형태이미분석 말뭉치는 형태소 분석 말뭉치를 기반으로, 이 단계에서 해결하지 못한 동음(동철)이의어의 중의성을 해결한 것으로, 단어 의미 중의성을 해결한(word sense disambiguation) 말뭉치의 전 단계에 해당한다. 단어 의미를 분석한 말뭉치의 구축에는 많은 시간과 노력, 그리고 정교한 방법론이 요구되며, 동형어 수준의 중의성 해결이 전제된다. 따라서 본 연구를 통해 구축된 분석 말뭉치는 보다 정밀한 언어학적 연구와 자연 언어 처리 기술 개발에 사용될 수 있는 중요한 언어 자원이 된다.

나. 연구 내용

1) 개요

21세기 세종계획 제2단계 2차년도(2002)에 구축된 250만 어절 규모의 형태분석 말뭉치를 대상으로, 품사 차원에서의 중의성을 해소한 형태의미분석 말뭉치를 구축한다. 이를 위해 각종 분석 도구를 이용하여 형태분석 말뭉치에 표지를 부착한다. 기능어를 배제한 내용어가 주 분석 대상이 되고, 국립국어연구원의 <표준국어대사전>의 표제어 기술을 기준으로 하여 동형어 즉, 동음이의어의 구별에 중점을 둔다.

이같은	이/NP + 같/VA + 은/ETM
평가	평가__03/NNG
결과에	결과__02/NNG + 에/JKB
대한	대하__02/VV + ㄴ/ETM
대학	대학__01/NNG
사회의	사회__07/NNG + 의/JKG
관심은	관심__01/NNG + 은/JX
다만	다만/MAJ
그에	그/NP + 에/JKB
따른	따르__01/VV + ㄴ/ETM
재정적	재정__05/NNG + 적/XSN
지원	지원__02/NNG
때문만은	때문/NNB + 만/JX + 은/JX
아니다.	아니/VCN + 다/EF + /SF

[그림 1] 형태의미분석 말뭉치 분석 예

2) 분석 지침 수정

전년도에 사용되었던 분석 지침을 검토하고 일부 내용을 조정한다. 형태분석 말뭉치의 분석 결과 중 동음이의어에 <표준국어대사전>의 어께번호를 부착하는 방식으로 의미를 구별한다. 자세한 내용은 <부록: 형태이미분석 말뭉치 구축 지침>에 수록되어 있다.

3) 작업 대상 선정

가) 대상 자료

2002년도에 구축된 250만 어절 규모의 형태분석 말뭉치를 기준으로 하며, 파일명은 형태분석 말뭉치와 동일하게 하되, 파일명 내의 T(morph-Tagged)를 S(Sense-tagged)로 바꾼다. 말뭉치의 목록은 다음 [표 1]과 같다.

파일명	어절수	제 목	작 가	출판년도	장르기호
5BS_B01	67,372	월간조선 158호	월간조선 편집국	1993	M1260
5BS_B02	10,093	난장이가 쏘아올린 작은 공	조세희	1978	M1331
5BS_B03	18,649	아이를 잘 만드는 여자	김영희	1992	M1341
5BS_B04	12,464	여보게 저승갈때 뭘 가지고 가지	석용산	1992	M1341
5BS_B05	28,840	역사와 민족	함석헌	1983	M1350
5BS_B06	32,795	제3의 물결	앨빈 토플러	1994	M1360
5BS_B07	22,761	조선일보 사설(91)	조선일보사	1991	M1118
5BS_B08	63,477	조선일보 종합(91)	조선일보사	1991	M1111
5BS_B09	30,220	동아일보 외신(91)	동아일보사 편집부	1991	M1116
5BS_B10	19,796	한겨레신문 경제(91)	한겨레신문사	1991	M1113
5BS_B11	36,979	한겨레신문 문화(92)	한겨레신문사	1992	M1114
5BS_B12	72,803	한겨레신문 생활(92)	한겨레신문사	1992	M1115
5BS_B13	47,907	조선일보 생활(93)	조선일보사	1993	M1115

파일명	어절수	제 목	작 가	출판년도	장르기호
5BS_B14	28,089	조선일보 외신(93)	조선일보사	1993	M1112
5BS_B15	34,719	한겨레신문 칼럼(93)	한겨레신문사	1993	M1119
5BS_B16	29,372	뉴스피플	경향신문사 편집부	1993	M1200
5BS_B17	46,890	주간경향	주간경향 편집부	1994	M1290
5BS_B18	54,198	우리교육 초등용 94년 11월호	우리교육 편집부	1994	M1258
5BS_B19	56,858	객주5	김주영	1996	M1331
5BS_B20	64,590	연꽃과 진흙	김성동	1993	M1331
5BS_B21	29,038	만세전	염상섭	1987	M1331
5BS_B22	81,151	잠자는 갈매기	엄홍섭 외 14인	1995	M1331
5BS_B23	47,988	통도사 가는 길	조성기	1992	M1331
5BS_B24	23,857	바닷가 아이들	권정생	1988	M1321
5BS_B25	35,134	이 땅덩이와 밥상	천규석	1993	M1341
5BS_B26	36,321	발해를 찾아서	송기호	1993	M1343
5BS_B27	49,575	길은 길을따라 끝이 없고: 한샘 미네르바문고1	김홍규, 윤구병 역음	1993	M1350
5BS_B28	46,475	초월에서 포월로	김진석	1994	M1353
5BS_B29	52,358	대중 문화의 겉과 속	강준만	1994	M1368
5BS_B30	56,169	인간을 위하여 미래를 위하여	김광식	1995	M1362
5BS_B31	43,572	중년기 건강클리닉	이상종	1994	M1377
5BS_B32	25,130	CD-ROM으로 보는 이규태 코너	이규태	1993	M1977
5BS_B33	60,403	한국일보 96-02 사회	한국일보사	1996	M1112
5BS_B34	26,108	한국일보 96-02 경제	한국일보사	1996	M1113
5BS_B35	31,826	중앙일보 96-09 과학	중앙일보사	1996	M1115
5BS_B36	19,853	중앙일보 96-06 사설	중앙일보사	1996	M1118
5BS_B37	29,947	금융 97년 3월호	금융연합회	1997	M1265
5BS_B38	90,195	라벨르 9월호	중앙일보사	1998	M1290
5BS_B39	64,659	보이지 않는 나라	김지용	1993	M1331

파일명	어절수	제 목	작 가	출판년도	장르기호
5BS_B40	64,320	헐리우드키드의 생애	안정효	1992	M1331
5BS_B41	61,956	한평 구휼의 안식	채희운	1993	M1331
5BS_B42	56,491	완전한 만남	김하기	1990	M1331
5BS_B43	21,678	아기참새 찌꾸	곽재구	1992	M1321
5BS_B44	36,084	피리부는 소년	이주홍	1991	M1321
5BS_B45	18,893	열두 컷의 낡은 필름	이영호	1991	M1321
5BS_B46	41,330	동양철학 에세이	김교빈, 이현구	1993	M1353
5BS_B47	31,004	가을에 만난 사람	박완서	1990	M1341
5BS_B48	42,157	결혼과 성	김동길, 유안진 외	1992	M1343
5BS_B49	41,553	가장 작은 것으로부터의 사랑	윤금초	1992	M1341
5BS_B50	79,054	한국언론의 좌표	이효성	1996	M1368
5BS_B51	59,321	강좌, 민족문학	고은 외	1990	M1381
5BS_B52	46,430	9시 뉴스를 기다리며	신은경	1992	M1343
5BS_B53	60,963	광야의 끝에서	이신범	1991	M1343
5BS_B54	44,604	안녕하십니까 MBC 전국패트를 봉두완입니다. (봉두완칼럼)	봉두완	1991	M1343
5BS_B55	38,071	영원의 뱃길(함석헌 전집 19)	함석헌	1985	M1300
5BS_B56	75,253	정치와 언어	최정호	1977	M1300
5BS_B57	24,099	돈의 여행	이슬기	1989	M1300
5BS_B58	18,718	고맙다 논리아	위기철	1992	M1322
5BS_B59	38,181	세계문학문학선(1)	오헨리 외	1974	M1331
합 계	2,528,791				

[표 1] 형태이미분석 말뭉치 목록

나) 대상 범주

형태이미분석 말뭉치의 분석 대상 어휘는 조사, 어미 등의 기능어(function words)를 제외한 내용어(content words)를 중심으로 선정하였다. 기능어의 경우 동일 품사 내에서 동음이의어가 나타나는 경우가 거의 없기 때문이다. 형태분석 말뭉치의 분석표지 중 작업

대상 범주는 다음과 같은데, 내용어 중 어근(XR)은 중의성을 갖지 않는 것으로 나타났다.

일반명사(NNG)

의존명사(NNB)

동사(VV)

형용사(VA)

관형사(MM)

일반부사(MAG)

250만 어절 형태분석 말뭉치의 품사별 통계 결과와 표준국어대사전 중 동음이의어에 해당하는 표제어 목록을 비교하여 분포를 확인하였다. (표2 참조) 품사별 형태분석 통계에 나타나는 어휘 중 동음이의어 목록에 등재되어 있지 않은 단어는 품사 단위 내에서는 하나의 의미로만 사용되는 것이므로, 의미분석 표지를 부착할 필요가 없다. 그리고 작업 대상은 동일 품사 내의 동음이의어로만 한정된다. 본 말뭉치는 일차적으로 형태 분석된 말뭉치를 기반으로 하므로, 전 단계에서 서로 다른 품사간의 중의성은 해소되기 때문이다.

품 사	단 어 종 류	출 현 빈 도
일반명사 (NNG)	12,166	793,857
동 사 (VV)	565	162,259
일반부사 (MAG)	397	11,736
의존명사 (NNB)	56	64,348
형 용 사 (VA)	43	4,528
관 형 사 (MM)	6	1,634
계	13,233	1,038,362

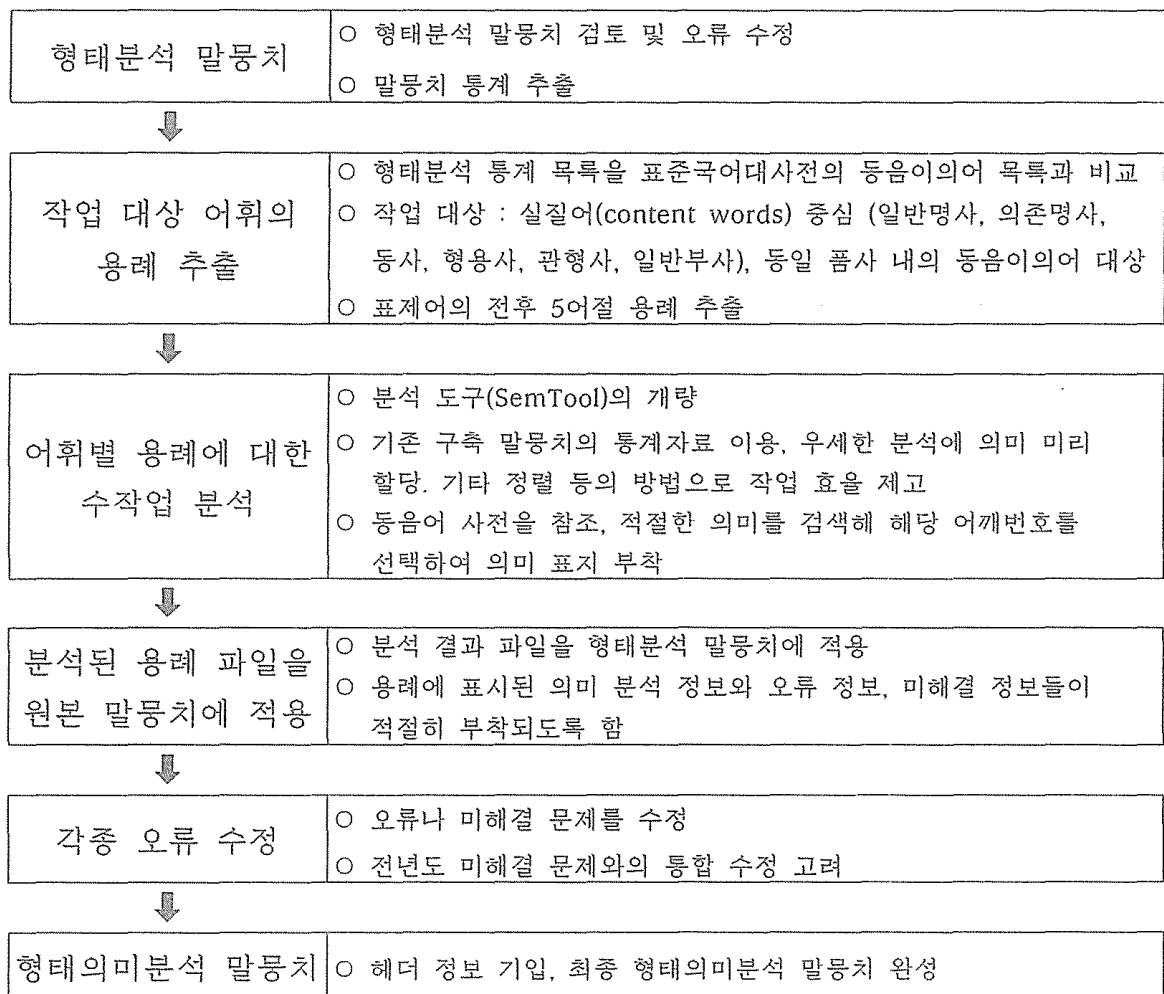
[표 2] 대상 어휘의 분포

4) 분석 도구 개발

형태이미분석 말뭉치를 효율적으로 구축하기 위해서는 적절한 도구의 개발이 필요하다. 작업 대상 어휘에 대한 용례 추출 및 정렬 프로그램, 추출된 용례에 사전 정보를 참조하여 의미 태그를 부착할 수 있는 프로그램 등을 수정, 보완하였다.

5) 말뭉치 구축 절차

형태이미분석 말뭉치 구축의 구체적인 절차는 다음과 같다.



[그림 2] 형태이미분석 말뭉치 구축 절차

가) 형태분석 말뭉치의 통계 추출

작업대상 어휘를 선택하기 위한 첫 번째 단계로, 형태분석 말뭉치의 형태소별(NNG, NNB, VV, VA, MM, MAG) 통계를 추출한다. 작업 전에 형태분석 말뭉치를 다시한번 검토하여 오류가 있는 경우 수정한다.

나) 동음이의어 사전과 비교

가)에서 추출된 형태소의 목록을 표준국어대사전의 동음이의어 목록과 비교하여 형태 의미분석의 대상이 되는 어휘 목록을 추출한다.

다) 작업대상 어휘의 용례 추출

형태분석 말뭉치에서 작업 대상 어휘의 용례(concordance)를 추출한다. 표제어를 기준으로 전후 다섯 어절씩을 추출한 것인데, 해당 어절의 말뭉치 내 위치 정보도 포함하고 있다. 일반적인 경우에는 전후 다섯 어절만으로 해당 어휘의 의미를 판별하는 데 충분하나, 그렇지 못한 경우에는 다음 단계에서 별도의 표시를 한 후, 원본 말뭉치의 문맥을 충분히 검토하여 의미를 판단한다.

라) 분석 도구를 이용한 분석

분석 도구를 이용하여 대상 용례에 대한 의미를 입력한다. 분석 도구는 동음어 사전과 용례 파일을 동시에 보여 주며, 사전 목록의 적절한 의미를 검색해 해당 어께번호를 선택하면 자동적으로 단어에 의미 표지가 입력된다. 이와 같은 방법으로 약 백만 개 정도의 어휘 의미를 직접 검토하였다. 분석도구를 이용하기 전 단계에서 기존 구축된 550만 어절 규모의 형태의미분석 말뭉치의 통계 자료를 이용하여, 해당 어휘의 어께번호가 한 가지로만 나오는 경우가 일정한 비율 이상일 때는 미리 그 번호를 입력하여 작업의 편의와 효율을 도모하였다. (올해 구축 작업에서는 기존 분석의 95% 이상이 동일한 경우에 미리 표시했다) 또한, 해당 용례를 표제어 중심으로 정렬하는 것도 같은 문맥들이 모이기 때문에 작업에 도움이 된다.

마) 원본 말뭉치에 결과 적용

분석 결과 파일을 형태분석 말뭉치에 적용한다. 용례에 표시된 의미 분석 정보와 오류 정보, 미해결 정보들이 적절히 부착되도록 한다.

바) 오류 수정

분석 지침에 따라 분석시 표시한 오류, 말뭉치 형식상의 오류, 기타 여러 가지 문제점을 각종 도구를 이용하여 수정한다. 수정이 끝난 자료에는 헤더 정보를 기입하여 최종적으로 형태의미분석 말뭉치를 완성한다.

다. 구축결과의 통계

250만 어절 규모의 형태의미분석 말뭉치의 작업 대상 품사의 표지별 항목과 빈도를 고빈도순으로 나열한 결과를 제시한다.

일반명사(NNG)	기자__05 2299	애기 1658	책__01 1273
	전__08 2265	영화__01 1640	오늘 1261
말__01 14761	시대__02 2218	처음 1634	머리__01 1254
사람 13906	얼굴 2208	학생 1606	당시__02 1252
때__01 8978	다음__01 2174	결과__02 1580	개발 1239
일__01 7884	생활 2157	주장__03 1570	뜻 1231
생각__01 6930	어머니 2097	가운데 1567	요구__03 1231
문제__06 5317	관계__05 2084	의미 1567	기업__01 1228
사회__07 4410	운동__02 2060	국가__01 1565	입 1227
속__01 3975	길__01 2019	관련 1561	결정__01 1223
경우__03 3878	손__01 2008	물__01 1555	연구__03 1219
소리__01 3643	대학__01 2007	조사__30 1550	상황__02 1211
집__01 3457	후__08 2001	모습__01 1541	친구__02 1208
자신__01 3412	위__01 1951	정보__06 1526	과정__03 1207
시간__04 3204	아버지 1940	내용__02 1524	발전__01 1200
언론 3143	동안__01 1928	선생__01 1522	대중__02 1197
앞 3077	역사__04 1927	교육 1518	밖 1194
뒤__01 3012	돈__01 1908	사용__04 1502	은행__02 1193
사실__04 2975	마음__01 1905	이름 1450	이후__02 1181
아이__01 2958	이상__05 1895	이번__01 1444	상태__01 1138
정치__03 2906	이야기 1876	힘__01 1435	대표 1134
필요 2879	여성__01 1865	학교 1434	이용__01 1132
시작__01 2805	안__01 1852	국민 1421	방__07 1120
정도__11 2719	여자__02 1840	세상__01 1398	밤__01 1119
정부__08 2717	방송__01 1820	문학__01 1396	일반__02 1117
나라 2699	사건__01 1813	현실__02 1392	그때 1116
눈__01 2660	지금__03 1804	자유__03 1386	교수__06 1104
문화__01 2619	사이__01 1786	이유__04 1378	환경__02 1101
경제__04 2611	날__01 1762	활동__02 1339	민족 1096
세계__02 2599	지역__03 1741	정신__12 1337	삶 1092
신문__10 2521	몸__01 1736	시장__04 1324	관심__01 1090
점__10 2463	자리__01 1701	변화 1291	정책__02 1084
인간__01 2398	방법 1677	제도__01 1286	발표__01 1083
대통령 2384	보도__04 1671	결혼 1277	모양__02 1082
곳__01 2336	사랑__01 1663	계획__01 1275	설명 1077

끝	1061	의식__03	931	미래__02	844	능력__02	754
글	1053	장관__02	924	발견__01	843	광고__02	752
대상__11	1050	전체__01	911	지방__05	843	책임	752
경찰__04	1046	마련	908	가슴__01	843	원인__02	749
산업	1044	하늘	907	할머니	838	표현	745
이해__06	1042	비판__01	893	가족__01	837	사업__04	744
존재	1041	마을__01	892	회의__04	836	기관__11	743
땅	1037	의견__01	890	도시__03	835	동시__02	742
기술__01	1037	해__01	887	형사__02	827	바람__01	736
선거__04	1031	남__01	886	효과__01	821	강조__02	735
자체__02	1030	정권__04	883	아내__01	815	작업__01	731
부분__01	1030	프로그램	882	참여	808	대답	730
의원__05	1020	작품__01	881	반대__03	805	아래__01	729
체제__02	1015	확인__02	881	통일__02	803	판단	728
단체__02	1009	말씀	880	얼마	796	유지__09	728
대부분	998	엄마	879	개혁	795	세력	727
법	996	노인__01	877	문__05	795	과학	726
지적__05	986	방식__01	873	고개__01	794	아들	721
국제__02	983	준비	872	목소리	788	발생	712
전쟁	974	계속__04	871	예__08	787	밑__01	710
생명	969	시__13	871	뉴스	785	형태	708
기능__03	969	하루__01	870	행동	784	지배__01	707
생산	968	현상__04	869	중심__01	783	성__07	706
자연__01	968	전화__07	869	혼자__01	781	기사__10	704
과거__03	968	평화__02	866	진행__02	781	영향__04	701
개인__02	957	옆	864	시민	781	건강__03	700
이날	954	술__01	863	행위	779	아저씨	698
반장__08	949	국내__02	860	형__01	774	구성__07	697
역할	949	경쟁	859	평가__03	774	소년	693
입장__04	946	회사__04	859	상품__03	773	관리__04	688
남자__02	946	일부__02	855	해결__02	768	시절__01	688
달__05	945	아침	850	산__01	764	민주주의	685
노력__01	942	조직	849	마찬가지	762	단계__03	684
이제__01	938	권력__02	848	규모	762	구조__08	682
남편__01	935	민주__02	847	대신__03	758	선택	675

나이__01 675	수준 629	분석__02 589	오늘날 547
발__01 673	추진__02 629	현재__02 588	잠__01 544
옷 671	이론__01 629	조금__01 588	컴퓨터 544
핵 671	소비자 629	처리__02 587	인류__01 544
바다 670	교사__09 625	언어__01 586	죽음__01 542
실시__03 669	면__05 623	원칙 585	성장__01 542
불__01 668	인식 623	최고__02 584	소설__03 540
방향__01 665	상대__04 621	담배 583	자금__08 538
어린이 661	태도__03 621	논의__02 583	그림__01 537
포함__02 661	금융 620	순간__03 583	인기__01 537
운영__03 661	공부__01 620	개념 578	기회__03 536
위원회 659	차이 618	주인__01 576	논리 535
마지막 658	노래__01 611	부모__01 575	제품__02 534
차원__01 658	등장__01 606	자식__01 574	값 532
목적__03 657	통신__01 606	청년 572	분위기 531
지난해 654	집단 605	구체__02 567	조건__02 530
아기__01 654	수__26 605	의사__12 565	분야 530
회담 654	할아버지 604	맛__01 565	성공__01 529
꿈__01 647	현대 604	참새__01 563	민중 526
전국__03 646	기본 603	자료__03 562	저녁 524
시설__03 645	공동__02 601	가정__06 561	적극 524
사상__15 644	업체 599	차레__01 561	영역__03 523
환자__03 642	텔레비전 598	비교__01 559	건설 522
병원__02 641	인정__08 597	병__04 558	활용 521
배__02 641	세기__03 596	외국__02 557	드라마 518
가치__06 641	경험 596	제공__04 557	공개__02 515
성격__02 640	제시__02 596	물결 555	어른 513
밥__01 639	걱정 596	각종 553	공정__02 511
기대__03 638	주민 594	음식 552	눈물__01 511
회장__07 638	나무__01 594	형성__01 551	예정__02 511
전통__06 636	공장__02 594	기록__02 551	대책__03 511
기간__07 635	확대__02 594	치료 549	투자__02 510
전문__08 634	세대__02 593	물건 549	오후__02 510
지원__02 633	사태__06 592	매체 549	노동__03 507
기준 630	모두__01 590	판매 547	혁명 506

신경__04 505	재산 472	총리__01 444	질서__03 426
표정__03 504	이익__02 472	동네 443	스타 425
방안__01 502	개선__01 471	관계자 443	공공__02 425
기분__01 502	보호__01 471	기존 442	서비스 424
옛날 501	사진__07 469	도입 441	군사__04 423
제작__02 501	전문가 469	종교 440	전달__03 421
종합 500	느낌 468	기계__07 440	취재__02 420
철학 500	대회__02 468	작가__01 440	발달 419
현장__03 497	소비__05 466	애__02 439	전자__06 419
한편 497	시기__04 466	죄__03 439	빛 419
경영__02 496	팀 464	군__05 436	출신 418
귀__01 496	강화__04 464	인생__01 435	편집__03 418
인물 491	쌀 463	협상__01 435	학년 418
농민 491	채기__06 461	사내__01 434	고통 418
소__03 490	감정__06 460	검토 434	고향__02 417
커뮤니케이션 490	담당 458	피__02 434	공화국 417
딸__01 489	출연__02 458	내부__04 434	목__01 417
독립 488	이전__03 458	업무__02 434	사회주의 416
국회 486	차__06 458	해방__05 434	폭력 415
부부__03 484	예술 457	수사__18 433	정상__02 414
가격__03 484	질문 457	제한__01 432	주택 412
방문__03 483	작용__01 456	설치__02 432	사장__15 411
올해 483	당__14 456	보장__01 431	새벽__01 411
기억 482	부족__01 456	수입__02 431	검찰__02 409
하나님 482	규정__04 456	시위__04 430	편지__02 408
사고__12 481	반__07 455	우주__02 430	위치__01 408
조치__04 481	예상__02 454	실천__01 430	전망__03 408
자세__02 480	아파트 452	거부__02 430	부담__01 407
개방__04 480	그동안 451	위험 430	추구__01 407
수단__01 479	투쟁 451	근거 429	위기__01 406
공간__05 476	도움 450	문명__03 428	측면 405
한__06 474	시청자 448	냄새 427	공격__02 405
시험__03 474	차지__01 448	이때 427	진실__02 405
요즘 472	주변__04 446	합의__01 426	동물 404
학자__01 472	건물__03 444	여론__02 426	집중__02 404

증가__01 404	반면__02 388	확보__01 362	고기__01 345
행복__02 404	근대__03 387	방침__02 362	아빠 345
남성__01 403	부정__09 386	장소__05 361	박사__01 345
지식__02 402	도덕 386	선__14 361	요청 344
주위__02 402	우려 385	나머지 360	만일__01 344
소유__03 402	윤리 384	중앙 360	코__01 343
한마디 401	영감__01 384	속도__01 360	주의__07 343
목표 400	구경__01 383	여인__01 359	식품__01 342
꽃__01 400	가을__01 383	대화__06 359	졸업 342
물질__02 399	약__07 383	최근 359	서양 341
사정__07 398	싸움 381	종류__02 358	음악__01 341
소개__02 398	요소__04 380	시인__10 357	목사__05 340
집안__01 398	여부__01 379	장면__04 357	희망 340
부인__01 398	자동차 379	우유__02 356	어깨__01 339
피해__01 397	신__09 378	배__09 355	양__20 339
연결__01 397	해석__04 378	혐의 355	갈등 338
벽__06 396	통합 377	제일__04 354	지구__04 338
젊은이 395	층__02 375	규제 354	이웃 338
해당__05 395	노동자 375	등__01 354	행정__01 337
그날 394	통제__02 375	눈치 353	여름__01 336
모임__01 394	숨__01 373	전개__02 352	지도자 336
변호사 393	내년 371	의회__02 352	재판__06 335
안전__03 393	까닭 370	커피 352	요인__03 334
체계__03 393	적용 370	구속__02 352	감독__02 333
반응 392	한참 369	사고__14 352	제목__02 332
안정__01 392	다리__01 368	행사__01 351	분단__03 331
나중__01 392	파악__01 368	줄__01 349	소식__04 331
약속 392	거리__01 367	겨울 349	모델 331
여행__02 392	주체__02 367	걸음 349	평균 330
거래__02 392	경향__02 366	한쪽 348	도착__01 330
정리__09 391	창조__03 366	전제__08 347	인사__03 329
직원__03 391	그룹__01 366	저널리즘 347	보통 329
자본주의 390	새끼__02 365	공익__01 346	불안__01 328
직업 389	인__02 365	참석 346	노릇 328
공동체 389	곳__01 365	스스로 345	오전__02 327

연구소 327	허용 313	제외__02 300	개__03 290
포기__02 326	당국__02 313	수행__02 300	자본__02 289
농업 324	잡지 312	가난__01 299	긴장 289
남녀 324	실현 312	인구__01 299	지지__06 289
짓__01 324	소금__01 311	바닥__01 299	검사__03 289
근본 323	연극 311	인사__01 298	비디오 288
선언__01 323	주제__04 311	당뇨병 298	자치__06 287
독자__04 322	완전 311	오염 298	이동__03 287
시행__01 322	팔__01 310	주먹 298	장__25 286
출발 322	내일 310	허리__01 298	최초 286
거리__08 322	실제__02 310	불만 297	크기 286
재벌__02 322	전략__03 309	회원 296	주인공 286
해외 322	발행__01 309	공급__02 296	신청__01 286
기구__15 322	흐름 308	비난 296	기초__06 286
비용__03 322	정당__07 308	원래__01 296	무당__01 285
마당 321	특징 307	장군__04 296	배__01 285
공식__01 321	위원 307	압력 296	손가락 285
버스__02 320	체험 307	자랑 296	주목__03 285
계장__03 320	뿌리 307	어둠 296	명령__01 284
수술__05 319	의지__06 307	인사__02 295	야당 284
접근 319	연기__10 306	과제__04 295	탓 284
파괴 318	웃음 305	아낙 295	농산물 283
청소년 318	소문__02 305	재료__01 294	고객__04 283
식사__03 318	축구 305	한국인 294	농약 283
지시__02 317	보고__03 304	외교__01 294	흥미 282
고려__01 317	교회__02 304	최대 294	수용__06 282
후보__04 316	특정 303	무역__02 293	극장 282
자율 315	전환__03 303	가입 293	비__01 281
형식__01 315	촬영 302	시도__07 292	시청__03 279
의도__02 315	사항__02 301	방송사 292	상당__04 279
주부__03 315	이외__01 301	한계 291	농촌 279
토론__01 315	지향__02 300	등록__01 291	백성 278
이미지 314	지도__09 300	급__04 291	궁정 277
구입__03 314	실패__02 300	봄 291	고문__04 277
언론인 313	극복__01 300	자극__01 291	문제점 277

상징	277	정상__11	267	독재	257	비율__02	248
혼란	276	가게	267	피부__02	256	대형__04	248
며칠	276	상상__07	267	동료	256	위협	247
사례__05	275	충격__02	267	결론__02	256	실험	247
필자__02	275	균형	267	부문__06	256	계급__02	246
돌__02	275	부장__07	266	반영__01	256	예산__02	246
대응__02	275	일종__03	265	동양__03	255	살__01	246
회복	274	금지__04	265	형편__01	255	특성__01	245
카드	273	개최	265	시스템	254	지속__01	245
피고인	273	의심__03	265	연락__02	254	공해__01	244
특수__02	273	눈__04	265	증상__01	254	발언__02	244
감각__02	273	배경__01	264	부동산	253	살인	244
방송국	273	종이__01	264	초기__04	253	행사__02	243
주식__03	273	의사__02	264	세월__02	253	지급__01	243
상처__02	273	협력	264	위원장	253	구분__06	243
억압	272	의존	263	제출__02	252	놀이__01	243
무시__04	272	어려움	263	잔__03	252	장치__07	242
동원__02	272	짐작	263	새__03	252	과__04	242
고민	272	직장__05	263	호랑이	252	심장__02	241
참가__01	271	미디어	263	학원__02	251	분리__04	241
화면__05	271	본래	262	미군	251	질__08	241
민간	271	시선__03	262	설립__02	251	접촉	240
만족__01	271	조작__02	262	교통__01	251	땀__01	240
장기__10	271	계기__04	261	선배	251	발휘	240
간부__05	270	주도__02	261	현지__03	251	유일__06	240
인터넷	270	꿀__01	261	의무__01	251	이데올로기	240
감시__02	270	누나	260	라디오	251	단위__02	240
합리__01	270	확산	260	눈길__01	251	마디__01	239
틈__01	270	고생	260	칼__01	250	대립__03	239
농사__01	269	자녀__01	259	동생__01	250	아무것	239
왕__04	269	구멍	258	이성__08	250	이하__02	239
본질__02	268	운명__01	258	비밀	250	노조__02	239
연방__09	268	견해__02	258	조정__09	249	취급	239
추가__02	268	아주머니	257	중간__01	249	외부__02	239
도로__07	267	희견	257	여유	249	핵심	238

군인	238	재미__01	228	성명__10	220	법원__01	215
의학__02	238	계산__01	228	이상__09	220	원장__07	215
수익__01	238	근무	227	신체__02	220	협의__01	215
절대__05	238	평등	227	평소	220	위반__03	215
바탕	238	신고__01	227	상업__02	220	규칙__02	214
사망__04	237	중단__04	227	시골	220	수출__03	214
총장__01	237	공연__02	226	검사__02	220	고등학교	214
객관__03	237	심리__01	226	봉사__03	220	발	214
인권__01	236	서류__02	226	강제__01	220	의료__02	214
신문사	236	기독교	226	지위__04	220	키__01	214
유통__04	236	언론사	226	상호__04	220	진출__02	213
서로__01	235	자__14	226	완성__01	220	선정__07	213
어제__01	234	무기__05	225	권리	219	총선	213
장사__01	234	서방__02	225	별	219	한숨__02	213
차별	234	보험	225	개정__02	219	홍보__01	213
시각__04	233	연예인	225	조심__02	219	사나이	212
모순	233	항의__01	225	배치__03	219	범죄	212
사찰__06	233	손__02	224	물론__01	219	무릎	212
인터뷰	233	무대__06	224	색__03	219	흙	212
화학__01	232	움직임	224	만화__10	219	고양이	211
예방__02	232	복지__10	224	채택	219	사물__10	211
부탁	232	선물__03	223	가구__03	218	상승__01	211
절차__02	231	귀국__02	223	탄생	218	의장__12	211
잘못	231	수도__09	223	소녀	218	정치인	211
소득	231	신분__02	222	뼈	218	오랫동안	211
증거	231	반복__01	222	관점__02	218	공사__02	211
확정__02	231	낮	222	퍼센트	218	용기__02	211
사무실	230	강__01	222	공산주의	217	섭취	211
생존	230	배우__01	221	집권__01	217	진단__02	211
질환	230	구석__01	221	탤런트	217	물가__02	210
시__06	230	입시__04	221	연기자	216	화__06	210
협회	229	훈련	220	통과	216	영향력	210
안__04	228	증인__01	220	임금__03	216	불교	209
협정	228	반__11	220	직접	216	근로자	209
학문__02	228	혈압	220	불편__01	216	식구__01	209

결	209	결합	201	육체__03	197	도전__04	189
관리__01	209	수용자__02		부인__02	196	의자__03	189
순수__03	208	201		지식인	196	작전__01	189
사냥	208	원리__02	201	소속__01	196	유적__08	189
진화__05	208	폭__06	201	자원__04	196	방영__02	188
에너지	208	감옥__02	201	과일__01	196	판사__01	188
숲__01	207	예외	201	침묵	196	양심__02	188
처지	207	명예	201	비닐	196	바위__01	188
조절__02	207	구별__02	200	증언	196	분노	188
백화점	207	재정__05	200	수요__06	195	이틀__01	188
실장	206	작정__01	200	인상__03	195	신앙	188
대출__03	206	본부__03	200	떡__01	195	국장__01	187
대비__10	206	반성__01	200	입술	195	공통	187
신탁__02	205	총재__03	200	카메라	195	주__24	187
기__13	205	수업__04	199	헌법	194	유행__02	186
군대__03	205	조성__04	199	실명__05	194	사전__13	186
그릇	205	연출__02	199	폐지__02	194	언급	186
번역	205	이념	199	고위__04	193	냉전	186
성과__01	205	평생	199	의문__02	193	최종	186
바깥	205	책상__01	199	근처	193	신용__01	186
작년	204	일상__04	199	일제__02	193	부여__04	185
법률	204	국민학교	199	메시지	193	울음	185
기반__01	204	좌우__01	198	시점__02	193	살림__01	185
공무원	204	길이	198	미소__05	192	허가__01	185
수석__02	204	논쟁	198	오류__03	192	형제__01	185
목숨	204	도구__10	198	영상__01	192	조화__07	185
증년__01	204	역__06	198	연합__03	192	앵커	185
금리__01	203	해소__03	198	신뢰__02	191	책임자	184
희생	203	말__05	198	주의__02	191	석방	184
실정__04	202	추세__03	197	무덤	191	직후	184
진보__02	202	토지__02	197	정기__07	191	일치__01	184
연습__03	202	범위	197	교실	191	제안__02	184
저항	201	유권자	197	불법__01	191	인상__06	184
계약	201	시내__03	197	상식__06	190	대변인	184
부정__02	201	비극	197	욕구	189	결국	184

과장__07 184	자살__01 179	대량__01 175	대문__03 170
이마__01 184	총__03 179	상대방__02 175	조상__07 170
작성__01 183	심정__01 179	일자__05 175	방해__01 170
수입__01 183	요금__01 179	세금__01 174	호소__01 170
통치__03 183	천__01 179	관념__02 174	공기__06 170
거짓말 183	양식__04 178	지혜__02 174	간섭 170
보관__01 183	연기__09 178	비중 174	현금__04 169
유학__04 183	증세__01 178	유형__07 174	채소 169
활성 183	서구 178	차량__01 174	이식__09 169
효율 183	성질 178	습관 174	연애__05 169
공포__08 182	임신__02 178	투표__01 174	일방__01 169
남쪽 182	완화__02 178	교환__01 173	법정__01 169
극__05 182	향상__01 178	가스 173	프로__04 169
매스 182	채널 177	도__10 173	인력__01 169
강요__01 182	날씨__01 177	고집__02 173	늪은이 169
설정__02 182	상관__03 177	범인__02 173	제자__01 168
경찰서 182	환상__02 177	병__05 173	의혹 168
성적__03 182	암__08 177	소장__08 172	토대 168
컵 182	시체__02 177	입학 172	이상__12 168
피로__02 181	예측 177	불행 172	부근__03 168
방지__04 181	숫자 176	계집 172	분쟁__03 168
골목__01 181	고추__01 176	고급__02 172	경고__04 168
악화__01 181	구축__03 176	제재__01 172	홍분 168
교류__01 181	외무__01 176	염려 172	조각__01 168
춤__01 181	안보__02 176	정의__03 172	선고__04 168
소화__06 181	사흘 176	기__21 171	축소__02 168
보고서 181	되풀이__01 176	양국__01 171	보급__01 168
긴급 181	애정__02 176	양쪽 171	일본인 168
임금__01 181	기름__01 175	매력 171	글자 168
무리__08 181	기획__01 175	개입 171	기여__04 167
정서__06 180	호텔 175	기차__01 171	한때 167
정치권력 180	영양__05 175	신비__02 171	박물관 167
이래__03 180	집승 175	조기__05 171	최소한 167
제의__04 179	비행기 175	발굴 170	집회 167
임명__01 179	인용__02 175	서방__01 170	식물__02 167

스포츠 167	연상__13 162	요리__05 157	본능 154
계층 166	상__25 162	너머 157	실태__02 154
개별 166	왜곡 162	핵무기 157	붕괴 153
접__05 166	미안__01 162	판결 157	당초__03 153
동의__02 166	제거__03 162	진술__02 156	첨단__01 153
진리 166	실수__01 161	콩 156	구름__01 153
경쟁력 165	스님 161	파견__01 156	결 153
표시__01 165	가수__11 161	스트레스 156	성숙 153
전반__03 165	온몸 161	독점__02 156	비리__08 153
대처__04 165	전기__15 161	조합__01 156	여기저기 153
떼__01 165	연령__01 160	철수__02 156	뇌__03 153
센터__02 165	귀신__01 160	개편__01 156	처녀 153
보편__01 165	알 160	교포__02 156	번호 153
집행 165	이자__05 160	기금__02 156	간__08 153
정착 165	전면__01 160	비상__01 156	생리__03 153
필연__01 165	승리 160	인쇄 155	용서__01 153
제정__01 164	우선__07 160	모색__03 155	안개 152
한글 164	승인__02 159	일반인 155	사안__07 152
문장__02 164	양파 159	기도__03 155	확신__01 152
배제__03 164	공업__01 159	생물__01 155	부엌 152
위인__02 163	상자__10 158	그림자 155	깍질 152
중소기업 163	적응__02 158	전부__05 155	섬__03 152
촉진__01 163	수능 158	성__08 155	동무__01 151
절반 163	도둑__01 158	비료 155	처벌 151
외국인 163	선전__03 158	호르몬 155	디자인 151
부처__04 163	후보자 158	사무__05 155	성경__03 151
고정__06 163	대안__05 157	중__09 155	화살__01 151
억제 163	요즈음 157	영원__02 154	게임 151
구역__04 163	몫 157	추정__02 154	혈액 151
장애__02 163	다음날 157	해체__03 154	수립__02 151
다수 163	행수__02 157	도리__10 154	연장__05 151
여당__01 163	재단__01 157	권위 154	지난달 151
관료 162	연설__02 157	조__13 154	기소__03 151
선진__03 162	비서관 157	징역 154	감동__02 151
성폭력 162	위주 157	보수__06 154	논문 151

마늘	150	선장__06	146	밥상	143	일행__01	140
짐__01	150	바지__01	146	출판__02	142	피해자	140
거짓	150	주머니	146	초점__03	142	음성__02	140
논란	150	욕망	146	대사__04	142	정__20	140
저축__03	150	날개__01	145	교장__03	142	곳곳	140
감사__08	150	그중	145	한인__04	142	중학교	140
마무리	150	출입	145	부위__04	142	짜귀	140
예전__01	149	전공__05	145	지__x28	142	장악__02	140
성인__01	149	꼬리__01	145	완벽__01	142	버릇__01	139
굿당	149	처분	145	진상__03	142	진짜	139
일선	149	엘리트__01	145	화제__07	142	시계__01	139
법인__02	149	유물__04	145	여관__03	142	새__01	139
앞	149	인연__03	144	슬픔	142	시중__03	139
소수__08	149	소송__01	144	원고__03	142	증대	139
국어	149	자동__01	144	이혼__03	141	사방__03	139
가치관	149	관찰__01	144	전투	141	도달__01	139
하오__03	148	손바닥	144	발길	141	감안__02	139
과세__04	148	북쪽	144	손해	141	최선__02	139
지면__03	148	서명__03	144	기타__01	141	교도소	139
풍경__01	148	불빛	144	입구__02	141	만남	139
단__06	148	증명	144	강의	141	생성	139
고대__04	148	식민지	144	일정__03	141	모자__08	139
훼손	148	눈앞	143	확장	141	인격	138
폴__02	148	신랑__02	143	주최	141	시조__07	138
표준__01	148	개설__02	143	점심	141	무장__06	138
재야	148	열__07	143	절__01	141	교과서	138
안내__01	147	요법__03	143	동지__05	141	고유__03	138
자격__04	147	고발__02	143	예금__01	141	다짐	138
우수__11	147	여건__01	143	편리__03	141	결심__01	138
질병__02	147	의면__03	143	아동__02	141	순서	138
해제__03	147	인민__01	143	파출소	141	스타일	138
선출__02	147	초청__01	143	태양__02	140	절약	138
달걀	146	오빠	143	곤란	140	석유__01	138
침__01	146	가방__01	143	장식__05	140	덕__05	137
쇼__02	146	적자__02	143	씨__01	140	선발__03	137

연대__02 137	아줌마 134	지붕 131	무료__01 129
색깔 137	화장실 133	초월__02 131	토끼 129
답변 137	똥 133	인형__01 131	일대__03 129
눈빛__01 137	신설__03 133	대선__88 131	지대__07 129
부작용 137	침대__02 133	미사일 131	제조__07 128
부각__05 137	구상__08 133	조국__01 131	공__04 128
장래 137	답__03 133	대결__03 131	음모__04 128
부대__08 137	유도__08 133	수표__01 131	무늬 128
빵__01 136	국회의원 133	당사자 131	진전__08 128
실질 136	요원__02 133	어음__01 130	용어__02 128
지옥 136	허락 133	부패__04 130	임시__02 128
조약__02 136	쓰레기 133	식당 130	소외 128
빛 136	보상__02 133	동기__07 130	건축__01 128
지방__09 136	대체__03 132	경매__03 130	한동안 128
투명__02 136	패러디 132	폭행 130	마루__03 128
양반__03 136	표시__02 132	변동 130	품질__03 128
단어 136	시비__05 132	구실__01 130	아픔 127
관행__02 136	초래__01 132	갈매기 130	사__18 127
감소__01 136	창__09 132	절__08 130	중생__04 127
보존 136	근원 132	이내__05 130	허__01 127
협의회 135	고함__01 132	혁신 130	다행 127
흡수__02 135	과목__02 132	지정__14 130	창문 127
충돌 135	합격 132	높이 130	이치__06 127
병력__01 135	만물__02 132	약화__01 130	권한 127
심화__04 135	설득 132	혈관 129	주문__04 127
오해__02 135	오락__01 132	부총리 129	거절__01 127
분류__03 135	직__06 132	게임 129	연관__06 127
표면 134	경계__07 132	시사__02 129	모래__01 127
아가씨 134	표__04 132	금액__02 129	하우스 127
통보__02 134	후배__06 132	동화__07 129	영업 126
차림 134	설탕 132	증권__01 129	비명__02 126
경력__02 134	적__13 132	장해__02 129	간주__02 126
근육 134	현행__01 132	순환__01 129	쇠고기 126
동상__09 134	아이디어 131	위성__06 129	신사__16 126
소재__05 134	바닷가 131	법안__01 129	부인__04 126

언덕	126	영당이	123	체중	120	업계__01	118
보완	126	단지__08	123	턱__01	120	총체__02	118
본인	126	타락__02	123	침해	120	부서__12	118
승용차	126	국방부	123	개성__03	120	사학__04	118
주가__05	126	취미__04	122	축적	120	소식통	118
보유__01	126	학과__01	122	차관__01	120	정신대	118
들	126	먼지__01	122	탄압	120	반발__01	118
국방	125	착각__03	122	댁	120	통상__01	117
법칙	125	야단__01	122	실력__02	120	실행__02	117
대가__06	125	닭	122	신화__04	120	호흡	117
혜택	125	전향__03	122	분비__02	120	여야__02	117
수집__02	125	상담__01	122	일시__01	120	조항__02	117
어미__01	125	배추__01	122	기쁨	120	비유__01	117
변명__01	125	퇴직	121	생선	119	열차__02	117
확립	125	과도	121	입찰	119	명의__02	117
제약__01	125	틀__01	121	처방	119	체결	117
지하	125	인근	121	식량__03	119	계열	117
전시__04	124	갓가지	121	배우자__02	119	변경__04	117
미술	124	선비__01	121	변신__01	119	차__09	117
비타민	124	상실__05	121	본성__02	119	기법__01	117
약간	124	여학생	121	흉내	119	한정__01	117
무한	124	서부__01	121	학습	119	초원__04	117
문서	124	안심__04	121	발해사	119	당황	116
측정__01	124	무게	121	왼쪽	119	악사__04	116
편__04	124	궁극	121	각국	119	바보	116
간접	124	연립	121	계절__01	119	큰일__01	116
상오__02	124	민심	121	최후	119	전용__04	116
장비__07	124	사설__04	121	심사__08	118	수정__09	116
일요일	124	손님	121	풀이	118	응답	116
보리__01	124	연속__02	121	대꾸	118	이튿날	116
설계__02	123	쿠테타	121	뉴우스	118	맥주	116
전후__01	123	술집	121	이사__11	118	단백질	116
선진국	123	원주민	121	국무__02	118	실감	116
시장__03	123	살해__01	121	정의__05	118	택시	116
대학원	123	문자__02	120	지점__03	118	이전__08	116

곡식	116	가루__01	113	장난	112	재배__03	110
떨치	116	주말__02	113	모집	112	선수__05	110
고독__02	116	충족	113	일생__01	112	풍토	110
입증__02	116	데모	113	소프트웨어	112	공중__02	109
운용	116	의욕	113	정지__06	111	보안__01	109
구성원	116	욕__02	113	읍내	111	그물	109
요란	115	직전__02	113	자전거	111	원단__02	109
임무__01	115	통계__04	113	털	111	묘사__02	109
외무부	115	낭비	113	시나리오	111	추측	109
공항__02	115	난__06	113	중지__02	111	노출	109
포장__01	115	동맥__01	113	필수__02	111	알레르기	109
욕심	115	전후__04	113	이중__03	111	자존심	109
뇌물	115	칭찬	113	내면	111	주연__01	109
소나무	115	문의__03	113	가지__01	111	권위주의	109
이사__14	115	이달__01	113	명칭__02	111	정작	109
존경	115	상인__07	113	온도	111	골짜기__01	109
정성__11	115	배달	113	할인__01	111	독신__01	109
취소__01	115	정식__01	113	동작__03	111	친척	109
교육부	115	유발__02	113	매입	111	복도__04	109
카드	115	노화__01	113	중반__03	111	구호__02	109
추적__03	115	재봉틀	113	대목__01	111	추석__01	109
지하철	115	규명	113	장계	110	계단__04	109
안팎	114	용품__01	113	안경__03	110	전형__04	109
성립	114	해명__02	113	기운__01	110	정직__01	108
숨씨	114	흔적	112	출현	110	공중__03	108
성분__01	114	광장	112	장교__03	110	수사관	108
청소__06	114	맘__01	112	이불__01	110	겨냥	108
명분__01	114	끈	112	관객	110	거주__02	108
거론	114	한자__02	112	분양__01	110	개척	108
뒷받침	114	타협	112	나으리	110	점점	108
치즈	114	학부모	112	고백	110	국__x10	108
변혁	114	실무	112	노사__07	110	경기__05	108
최대한	114	타결	112	오른쪽	110	자각__03	108
노선	114	양보__03	112	창립__02	110	군부__06	107
맥락	114	봉투__02	112	편성__02	110	도끼__01	107

탈출__02 107	실체__02 105	입맛 103	환영__02 101
호기심 107	뱀 105	총회__02 103	국토 101
애__01 107	밀가루 105	목욕 103	파업 101
협조__01 107	단절__05 105	약물__01 103	입원__01 101
종일__01 107	인공__01 105	취업 103	민요__02 101
도중__04 107	거울__01 105	발상__03 103	동__12 101
이윤 107	창작 105	단속__01 103	내란 101
즐거움 107	대접__05 105	철도 103	향기__01 101
대사관 107	정면__01 105	못__01 103	고속__02 101
유리__05 107	무선__02 105	지점__01 103	편의__02 101
믿음 107	철거__01 105	국물 103	선포__01 101
영혼__02 107	심의__05 105	젓 102	다리__02 101
동__15 106	폭발__01 104	방학 102	당선__04 101
연예__01 106	연말__02 104	물음__01 102	화장__02 101
이__03 106	연기__05 104	시리즈 102	비결__03 101
덕분 106	후자__01 104	베일 102	복합__02 100
계통__01 106	바늘 104	덩어리 102	친절 100
위__06 106	별명__01 104	지침__02 102	계재 100
생태__02 106	우연__02 104	사법__02 102	차단 100
행정부 106	육군__02 104	한__05 102	채용__01 100
간직__01 106	공존 104	아나운서 102	웬일 100
달성 106	참기름 104	지불__02 102	점수__06 100
제국__02 106	접수__03 104	인수__07 102	탐구__02 100
별도__01 106	전날 104	다방__02 102	월급 100
통증 106	진행자 104	전신__01 102	혼__02 100
대전__07 106	연간__02 104	패션__01 102	보증 100
존중 106	제사__07 104	성향__02 102	수지__04 100
짜 105	문항 104	협박 102	성실__02 100
골수 105	국경__01 104	원료__02 102	풍요__02 100
순사__02 105	시각__03 104	병사__02 101	김치 100
신중__04 105	파__01 104	식용유 101	본사__03 100
세제__04 105	고등__02 104	절망__02 101	정말 100
지__12 105	눈썹 103	지상__01 101	장애인 100
정보부 105	중시__03 103	햇빛 101	후춧가루 100
농부__01 105	장점__02 103	인체 101	

의존명사(NNB)		내		지경	253	동사(VV)	
		자_18	823	녀석	253		
것	52092	바_03	792	범_01	248	하	36938
수_02	19172	세_13	623	점	247	있	30695
등_05	9727	분_01	610	판	236	되_01	20113
년_02	8281	식_04	607	설	228	보_01	9600
때문	6553	듯이	588	바람_01	227	대하_02	8149
일	5853	대로	583	큰술	222	위하_01	6807
씨	4444	셈	580	권	215	가	5913
명	4072	차	580	군	215	받_01	5717
월	4009	개월	556	주_26	210	알	5293
중_04	3853	달리	532	도	207	오	4845
번	3695	회	480	위	175	나오	4075
원_01	3008	말_11	475	등등	164	만들	4064
데	2983	외	465	가량	162	따르_01	3828
개	2613	지_02	437	리_02	150	살_01	3634
거	2577	대_11	430	부_15	136	모르	3505
뿐	1971	만_01	428	척_08	129	가지	2993
듯	1964	마리	416	양_02	123	보이_01	2783
가지	1947	따위	408	척_01	123	먹_02	2754
간_10	1907	평	404	양_25	120	나	2635
눔	1844	살	390	년도	119	듣_01	2586
시	1544	초_03	385	주년	118	지나	2571
쪽_05	1401	장_22	383	초_07	109	통하	2503
채_09	1121	편_04	366	푼	109	들_01	2469
년대	1071	호_14	359	검	107	않	2433
측	1042	주일	344	개국	106	그리	2208
터	1019	무렵	339	체_02	104	의하	2202
쭙	992	나름	312	년_01	100	주	2195
분_08	973	이	307			밝히	2139
적_03	943	리_06	295			쓰_03	2113
대_06	933	편_09	288			들_04	2054
만큼	875	마련	284			들어가_01	2048
줄_04	865	대_15	284			죽_01	1965
만_02	834	전	277			갖	1955

잡__01	1915	이루어지	952	견__02	650	시키	519
내__02	1868	떠나	923	모으	642	사라지	511
찾	1774	기다리	921	잇__01	631	모이	508
쓰__01	1772	열리__02	912	막__01	630	맞__02	506
앞	1703	오르	906	삼__02	623	좋아하	500
느끼__02	1698	올리__01	877	올라가	615	변하	498
만나	1693	끝나	862	버리__01	608	들어서	494
나타나	1682	넘__01	861	자__01	605	바라	482
서	1661	향하	853	냥__01	604	지내	480
나가	1596	걸리__01	852	풀	602	원하__02	474
보이__02	1545	나누	847	웁기	601	거치__01	474
일어나	1535	놓__01	834	걸__02	599	바뀌	469
부르	1526	돌아가	829	움직이	596	대__01	468
문__03	1433	입	821	어찌	596	빠지__01	463
넙	1409	당하	818	맞추__01	588	드리나	461
내리	1335	끝	817	흐르__01	586	받아들이	459
떨어지	1332	말__01	789	그리__02	584	가르치__01	459
짓__01	1306	웃	779	인하__01	583	걸치	456
관하__02	1286	돌아오	776	불	577	줄이	456
두	1266	바꾸	764	다루	564	말__03	450
알리	1266	돌리__04	763	비하	562	미치__02	448
생기	1255	팔	760	놀라	555	남기	441
얻	1251	전하	753	빠지__02	553	지나가	437
들어오	1235	지키__01	748	여기	552	취하__01	437
사	1176	바라보	744	피하	551	늘	433
읽	1168	세우__01	739	담__01	549	던지	431
남	1165	불이	725	들	546	태어나	431
열__02	1112	마시	710	내려오	540	찾아오	430
치__02	1087	울__01	704	겪	529	나아가	425
보내	1071	벌이	695	죽이__01	528	파지__01	424
믿	1062	일으키	684	갖추	525	즐기	424
다니	1049	이루__01	677	벗어나	524	쩍__02	419
이르__01	1028	맞__01	674	늘어나	523	정하__03	413
나서	1019	배우__01	672	읽	523	앞서	413
타__02	1016	들리__03	664	지니	521	꺼내	413

들이__02 411	닭__01 339	뿌리 299	마르__01 262
빼__01 409	가져오 339	뜨__05 296	밀__01 262
펴 408	떨__01 337	속하__02 295	실리__01 261
해보 408	벌어지__02 335	불리__04 293	채우__03 260
살펴보 404	비__01 332	어울리 293	띄__01 257
눅__01 402	깨닫 328	더불 292	박 257
내려가 401	차리 327	빌리 291	알아보 257
듭 401	맷 327	달리__01 290	떠__01 255
내놓 395	먹이 326	늪 284	쏟아지 255
이어지 387	달리__04 326	거두 283	끓 254
떼__01 386	쥐__01 325	멈추 283	틀리__01 254
높이 384	지__05 324	끼__03 282	이기__01 253
차__01 382	허 323	뛰__02 280	누르 252
치르 379	차르__01 323	숨__01 280	끝내 251
안 379	기울이 323	몰__01 280	앞두 250
뽐 376	살피 323	일어서 280	이러 248
자라__01 374	싹 319	당 279	심__01 248
홀리 367	떠오르 317	부리__02 278	쌓이 244
맞__03 367	드러내 317	커지 277	내밀 242
놀 366	데리 317	따__01 277	쓰__02 242
싸우 364	찾아가 316	말기 276	누리 241
피우__01 363	살아가 315	벌__02 272	감__01 240
구하__01 358	넘기 314	쌍 271	덧붙이 239
터지 358	드리__01 314	미치__01 270	올라오 238
지르__03 358	쳐다보 313	소리치 270	때리__01 236
잡히__02 357	씻 313	저지르 269	적__01 236
살리 356	불__01 311	쓰러지 269	주어지 235
마치__02 353	외치__01 310	키우 269	달__03 234
나타내 352	흔들 310	가라키__01 269	가하 231
참 345	견디 309	몰__02 268	내세우 231
잊 344	그치 309	벗 267	들려오 230
고르__01 343	넘어가__01 304	그만두 265	덮 229
이끌 341	다가오 302	쉬__03 265	달려가 228
달라지 341	기르 301	뛰__01 263	묶 226
놓이 340	고치__01 301	막히 263	몰리__01 223

다지	222	타__01	194	두르	167	끌려가	149
숨기	222	가리__03	193	새기__01	166	돌아보	148
끓이__01	218	섞이	192	걸어가	166	잃어버리	148
바르__01	218	섞	192	들르	165	빼앗	148
쓰이__03	218	애쓰	192	내다보	164	벌리__01	147
셀__01	217	붓__02	191	품__01	163	젓	147
울리__01	215	비추	191	권하	162	벗기__02	147
쏘__01	215	떨리__01	190	익히__02	162	건네	146
퍼지	214	버티	189	집__01	161	쫓기	146
둘러싸	214	찌르	188	돌아다니	161	살아오	145
챙기	213	펼치	186	간히	160	닫__02	145
폴리	213	미루	186	끼치__02	159	쉬__04	145
찾아내	213	찾아보	185	아끼	159	합치	144
감추	212	싸__01	184	볶	158	떨어뜨리	144
무너지	209	떠들__01	184	밀리__02	158	바치__01	144
담기__01	209	뛰어들	183	묻__01	158	차__02	144
젓__01	208	행하	181	짜__01	158	달하	143
줄어들	207	굽__01	181	중얼거리	157	틀	143
모시	206	지켜보	180	시달리	157	치하	142
밟	205	들여다보	180	일__01	155	뒤지__02	142
꾸미	204	박히	180	넘치	155	다스리	142
섞	202	털	179	달아나	155	날아가	141
없애	202	답	178	엎	155	켜__01	140
건너	201	빚	178	깨__01	155	깎	140
기__01	200	얹	178	두드리	155	흘어지	139
계시	200	흔들리	177	생각나	155	알아듣	138
서두르	200	이르__02	175	지우__01	155	둘러보	136
날__01	199	붙잡	174	뿔	153	파__01	135
뜨__01	199	생겨나	173	갈	152	건지	135
매달리	199	따라가	173	끓어지	152	줍__01	134
눌리	199	비치__01	172	쫓	151	헤어지	133
팔리	198	놓치	170	지치__01	150	끼우__01	133
만지	197	갈__02	170	택하	150	엎드리	133
쏟	196	피__01	167	모자라	149	물러나	132
끄덕이	195	갈리__01	167	머무르	149	값	132

흙치__02 132	반하__03 120	앞장서 111	베__02 102
넘어서 131	묻히__02 119	넓히 111	깨__02 102
숙이 131	묵이 119	가르 110	가라앉 102
터뜨리 131	가져가 119	내보내 110	지나치 102
도와주 131	잠기__02 119	뜨__04 110	펼쳐지 102
끼이__01 131	겹치__01 119	접 110	삶 101
뒤따르 131	입히 118	잡아먹 109	새__01 101
달려오 130	뒤집 118	달래__01 109	주저앉 101
줄 130	삼키 118	망하 109	꿈꾸 100
돌아서 130	해매 118	매__01 109	
더하 128	속이 118	짚 109	
빨__01 127	꽃 118	말리__03 108	
꾸__01 127	접하__01 118	깨지 108	
치__10 126	살아남 117	맞서 108	
다투 126	문__02 117	스치__01 107	
부딪치 125	싫어하 117	홀러가 107	
적히 125	다가가 117	두들기 107	
달려들 125	늘어지 116	빠져나가 107	
뜯 124	다녀오 116	태우__01 107	
추__02 124	비우__01 116	휘두르__01 107	
꺾 124	물어보 116	병들 106	
빨 124	뽑__01 115	씹 106	
내려다보 123	더듬 115	낮추 106	
접어들 123	치__05 115	잊어버리 105	
머물 123	잠들 115	업 105	
맛보 122	신 115	기대__01 104	
가꾸 121	면하__01 115	내버리 103	
빛나 121	도망치 114	담그 103	
상하__02 121	끄__01 113	뽑히__01 103	
빌__01 121	번지__01 113	익 103	
날리__02 121	잇따르 112	감싸 103	
당기__01 121	해내 112	문지르 103	
해아리 121	빼앗기 112	돌보 103	
떠올리 120	따라오 111	취하__03 103	
풍기 120	다물 111	메__02 102	

형용사(VA)	넓	411	푸르	183	차_04	105
	엄청나	369	질	178	묘하	105
없	싫	355	낯	174	어지럽	105
갈	커다랗	349	고맙_01	174	뒤늦	105
그렇	짧	342	드물	174	세_03	104
많	희	321	슬프	170	귀엽	103
크	검	306	맛있	162	거세	102
좋	어둡	303	조그맣	161	멋지	101
어떻	하얗	296	거칠	158	등글	100
이렇	밝	289	반갑	158		
새롭	예쁘	288	부끄럽	151		
다르	바쁘	287	싸_05	151		
어렵	좁	286	날카롭	150		
높	똑같	286	못하	149		
쉽	악하	281	기쁘	149		
작_01	부드럽	278	낮설	143		
적	가볍	265	굵	143		
길	뜨겁	262	급하	142		
깊	틀림없	240	괴롭	142		
짧	맑	239	춡	138		
멀	즐겁	238	더럽	135		
아름답	무겁	235	외롭	135		
어리_03	재미있	231	굳	134		
심하	북_01	225	차갑	132		
강하_02	낯_02	220	착하	130		
가깝	편하	217	어리석	130		
빠르	미싸	216	흔하	127		
힘들	팬찮	214	얇	125		
웁	올바르	203	안타깝	123		
지나치	수많	201	저렇	122		
낮	바르_03	201	두렵	121		
나쁘_01	뛰어나	199	빨강	117		
늦	느리	194	이르	112		
무섭	놀랍	190	가늘	108		
아프	곱_02	188	알맞	106		

관형사(MIM)		현	196	일반부사(MAG)		우선	805
		한두	179			전혀__01	791
그	17976	양	176	더	4497	꼭__03	728
이	12364	몇몇	159	안	3794	계속	727
한	8922	웬__01	118	잘	3039	정말	725
두	3696	저런	117	다시	2996	제대로	713
그런	3540	서너	110	함께	2648	매우	690
이런	3079	탄	109	다	2540	직접	671
다른	3036			가장__01	2413	아무리	645
어떤	2653			없이	2327	잘못	632
모든	2358			바로	2220	스스로	631
어느	2224			못	2186	마치	619
몇	1891			모두	1858	갑자기	618
여러	1887			이미	1832	훨씬	589
무슨	1512			특히	1723	빨리	570
세	968			물론	1721	사실	552
저	878			많이	1666	그냥	552
새	844			아직	1665	반드시	545
각	837			이제	1660	달리	541
첫	675			왜	1592	여전히	522
아무	596			같이	1547	벌써	510
전__08	544			좀	1449	보다	505
전__07	458			더욱	1313	결코	492
약	404			지금__03	1206	어찌	487
아무런	402			거의	1172	그만	484
네__02	350			너무	1167	주로	480
은	338			얼마나	982	일단	478
총	310			먼저	982	자주	477
단	292			곧	973	언제나	468
옛	289			서로	968	조금	462
은갓	241			그대로	955	완전히	461
오랜	237			현재	911	늘	460
을	235			최근	900	실제로	455
맨	230			결국	888	그저	450
별	228			아주	883	과연__01	446

점점	438	어서	271	단지	191	또다시	140
자꾸	431	참	271	널리	189	극히	140
흔히	426	대개__03	265	언젠가	189	덜	140
오래	420	소위	260	그야말로	184	마구	139
항상	411	어쩌면	257	내지	181	새삼	139
각각	399	당장	249	마주	180	절대	139
겨우	399	만약	249	몹시	173	똑같이	138
게다가	381	오직__01	247	대체로	171	그나마	137
별로	378	깊이	246	끝내	166	조용히	137
멀리	375	그만큼	238	거꾸로	166	정말로	137
분명히	371	꽤	235	제법	164	채	137
잠시	358	어쨌든	233	정확히__01	163	아까	136
아마	351	천천히	230	서서히	162	일부러	136
매일	344	점차	229	은통	161	절대로	135
아예	339	일찍	228	잔뜩	160	함부로	135
새로	333	상당히	228	굳이	159	철저히	134
마침내	329	얼른	225	한꺼번에	158	나란히	131
미리	327	다소	223	무조건	158	높이	130
도대체	323	잠깐	222	막__01	158	예컨대	128
너무나	323	심지어	222	당연히__01	155	도무지	126
비록	321	무척	221	확실히	154	즉시	125
따로	318	때로	216	자세히	154	특별히	124
적어도	316	하여금	215	드디어	153	내내	124
금방	308	대단히	215	이어	153	제발	123
열심히	304	가까이	214	감히	153	영원히	122
그리__02	303	보통	207	가령	152	깜짝__02	121
약간	299	도저히	205	어느새	152	꼭__02	118
가만히	298	문득	205	분명	152	간신히	117
참으로	294	마침	205	끊임없이	151	곧바로	117
비로소	291	아무래도	203	몰래	149	이내	115
충분히	290	차라리	200	살짝	146	어쨌서	115
단순히	290	거듭	196	꽤	144	벌떡	114
언제	288	혹시__01	194	위낙	143	지극히	114
가득	276	아마도	193	즉	141	아니	111
가끔	275	아울러	192	대폭	140	자칫	109

곧장	109
이윅고	109
그토록	108
도리어	108
무려	108
전부	108
저절로	104
딱__03	103
불과	103
우연히	103
비단	103
일찍이	103
고작	102
말없이	102
그다지	102
오로지	101
한층	101
더욱더	100

참고문헌

김홍규, 강범모. 2000. 「한국어 형태소 및 어휘 사용 빈도의 분석 1」 서울: 고려대학교 민족문화연구원.

김홍규 외. 1998-2003. 「21세기 세종계획 연구보고서」 문화관광부.

김한샘. 2000. “말뭉치의 주석과 활용: 의미 주석 말뭉치의 활용을 위한 기초 연구.” 「언어정보 연찬회 발표 논문집」 (연세대학교 언어정보개발연구원) 14.

신지현, 최민우, 강범모. 2001. “어휘의미분석 말뭉치 구축의 절차와 문제.” 「제11회 한글 및 한국어 정보처리 학술대회 논문집」.

국립국어연구원. 1999. 「표준국어대사전」 서울: 두산동아.

〈부록〉 형태 의미 분석 말뭉치 구축 지침 (2004 version 1.0)

1. 기본 원칙

- 형태 분석 말뭉치의 원문(원어절과 분석 결과 모두)은 가급적 수정하지 않는다. 단, 형태 분석 말뭉치에서 발견되는 오류에 대해서는 이를 별도로 기록한 후 형태 분석 말뭉치의 개선에 반영될 수 있도록 한다.
- 동음이의어 분석 말뭉치의 구축시, 형태 분석 말뭉치를 어휘 분석 말뭉치 수준으로 변환하여 구축할 수 있다. 형태 분석 말뭉치를 대상으로 품사내 동음이의어 중의성을 해결한 말뭉치를 형태 의미 분석 말뭉치라 한다면, 접사 처리 과정을 거쳐 만들어진 어휘 분석 말뭉치를 대상으로 동음이의어 중의성을 해결한 말뭉치는 단어 의미 분석 말뭉치로 부를 수 있다. 당해 연도에 구축한 형태 의미 분석 말뭉치는 이중 형태 의미 분석 말뭉치에 속한다.

2. 세부 사항

- 분석 대상 품사는 기능어(function words)를 제외한 내용어(content words)를 중심으로 선정하였다. 대상 범주는 다음과 같다. : 일반명사(NNG), 의존명사(NNB), 동사(VV), 형용사(VA), 관형사(MM), 일반부사(MAG)
- 동음이의어 목록은 국립국어연구원의 <표준국어대사전>을 기준으로 한다.
- 말뭉치 정보 부착 방법은 해당 어절의 분석 결과에 ‘_’과 두자리 어깨번호를 부착하는 방식을 사용한다. 어깨번호는 국립국어연구원 <표준국어대사전>에 나타난 표제어 번호와 일치한다.

예) 5BT_B4920255840 비비/MAG 결의했다. 천(千)의 다리, 천의 팔이 [비비 : 비비_02/MAG] 꼬인 이 매듭을. 재갈 물린
index 대상형태소 앞 5어절 어절 분석 뒤 5어절

- 형태분석 말뭉치와 <표준국어대사전>의 품사 분류가 일치하지 않아서 그 분석이 달라진 경우, 다음과 같이 어깨번호 앞에 'x' 표지를 부여한다.

예) 현지의 값싼 노동력(임금이 우리 나라의 [10분의 : 10/SN + 분_x15/NNB + 의/JKG] 1 이하라고 한다)으로 물건을 만들어

'10분의 1' 같은 표현에서의 '분'은 사전의 정의에 의하면 접미사이나, 형태분석 말뭉치에서는 의존명사로 분류하고 있다. 위의 예는 이러한 불일치를 보여 준다. 이처럼 사전과 형태분석 말뭉치 간의 차이가 나타날 수 있는 경우는 일반명사와 의존명사, 의존명사와 접미사, 관형사와 접두사 등에서도 찾아볼 수 있다.

- 사전에 등재되지 않은 어휘가 나타날 경우, 어깨번호는 '88'으로 한다.

예) 12일자 '갈'난에 보도된 '교도소 가는 [부정'의 : 부정_88/NGG + '/SS + 의 /JKG] 인태군 씨에게 전해 달라며 박찬수...

반의어인 '모정'(母情)이라는 단어는 사전에 존재하지만 위의 '부정'(父情)이라는 단어는 사전에 등재된 어휘가 아니다. 이 경우에는 어깨번호를 '00'으로 붙이고 추후 해결 방안을 모색한다.

- 형태분석 말뭉치의 분석 오류가 발견되면, 어깨번호를 '99'로 하고 이후 후처리 단계에서 형태분석 오류를 수정한다. 형태분석 오류 수정은 당해연도 형태분석 지침에 따른다.
- 좌우 5어절의 환경만으로는 동음이의어 중의성 해결이 어려운 경우, 또는 기타 오류로 인해 추후 처리가 필요한 경우에는, 용례의 맨 앞에 @를 붙이고, 최종적인 말뭉치 수정 단계에서 이를 찾아 해결한다.

빈 면

4. 구문분석 말뭉치 구축

가. 연구의 목적과 필요성

1) 연구 목적

이 연구의 목적은 21세기 세종계획의 일환으로 구축된 형태분석 말뭉치 가운데 일부인 20만여 어절을 대상으로 하여, 문장 단위 차원에서 이들 어절이 가지는 구문 정보를 분석하여 구문분석 말뭉치를 구축하는 것이다. 이를 위해 2003년도에 작성된 구문분석 지침을 수정·보완하였고 이것을 바탕으로 20만 4천 어절(총 20,798개 문장)에 달하는 구문분석 말뭉치를 구축하였다.

2) 연구의 필요성

구문분석 말뭉치 구축은 다음과 같은 의의를 갖는다.

첫째, 구축된 구문분석 말뭉치는 통계적인 구문분석기 개발의 학습 자료로 삼을 수 있다. 대규모의 말뭉치를 구축·처리하는 데에는 인력만으로는 절대 부족하므로 (반)자동적인 처리를 위하여 구문분석기를 개발하는 것이 필수적이다. 이러한 구문분석기 개발에는 구문 요소 각각의 문법 관계들을 기계 학습 방법으로 도출해야 하는데, 이때 그 기반이 되는 자료로 구문분석 말뭉치가 필요하다.

둘째, 고품질의 지식 기반 시스템 구축에 있어 중요하게 여겨지는 텍스트 마이닝(text mining) 기법의 개발을 위해서도 구문분석 말뭉치는 필수적이다. 텍스트 마이닝 기술이란, 텍스트 자체의 정보만을 뽑는 것이 아니라 그 텍스트에 숨겨져 있는 다른 정보를 이차로 가공하여 사용자에게 유용한 형태로 만들어 주는 것을 의미한다. 그리고 이를 위해서는 단어들 간의 관계를 추출하거나 구문들의 다양한 구조를 학습하여 좀더 고차원적인 지식 표현 요소를 만들어 두는 것이 필요한데, 이때에 구문분석 말뭉치가 필요하다.

이렇게 구축된 고차원의 텍스트를 활용할 수 있는 분야는 또한 광범위하다. 대표적으로 기계 번역, 정보 검색, 전문가 시스템 등이 있다. 특히 전문가 시스템은 이미 축적된 지식을 바탕으로 질의나 문제가 입력되면 적절한 해답을 제시해 주는 시스템을 말하며, 그 인터페이스 자체가 자연언어로 되어 있을 경우 문장의 구문분석이 필수불가결하다.

셋째, 대규모의 한국어 구문분석 말뭉치 구축을 통하여 한국어에 관한 실증적인 이론 언어학적 연구에 매우 큰 기여를 할 수 있다. 앞서 든 두 가지의 연구가 주로 자연 언어 처리에 일차적인 목적을 둔 것이라면 지금 말한 것은 한국어 문법에 관한 순수 국어학적인 탐색과 관련이 있다. 개화기 이후 지금까지 이루어진 현대 국어학의 연구는 주로 연구자의 직관에 의존한 것이 대다수였다. 그러나 그러한 국어 연구는 실제 언중들이 사용하고 있는 다양한 구문 패턴을 제대로 반영하지 못하는 한계를 지닌다. 따라서 한국어 문장의 문법 문제에 대한 총체 서술이나 개별 구문들에 대한 좀더 치밀한 연구를 위해서는 구문분석 말뭉치를 이용하는 것이 더욱 바람직할 것이다.

3) 관련 연구 검토

구문분석 말뭉치 구축과 관련된 기존의 연구는 대상 문장의 구문 정보를 구구조(Phrase Structure) 형식을 빌려 표시하느냐 아니면 지배-의존이라는 이진 관계 중심의 의존 구조(Dependency Structure) 형식으로 나타내느냐에 따라 크게 두 가지로 대별된다.

먼저 구 구조 형식의 구문분석에는 Penn Korean Treebank와 STEP 2000을 들 수 있다. Penn Korean Treebank는 미국 Pennsylvania 대학에서 (중세) 영어와 중국어와 함께 한국어를 대상으로 하여 만든 구문분석 말뭉치이다. 이는 54,366개 어절로 이루어진 5,078개 문장 규모인데, Chomsky 등의 지배·결속 이론을 연상하게 만들 정도로 매우 세부적인 문법 정보, 예컨대 무형 논항(PRO, pro)이나 흔적(trace)의 정보까지 명시한 것이 특징적이다.

한편, STEP 2000은 KAIST에서 주관하고 SERI 자연어 처리부에서 협력하여 구축한 것이다. 총 12,084개 문장을 대상으로 하였으며, Penn Korean Treebank가 어절 단위로 구문분석을 수행한 것과 달리 형태소를 기본 단위로 하여 구문 정보를 표시하였다.

한편, 이와 달리 의존 구조를 통하여 구문분석을 시도한 것은 서강대와 포항공대 등에서 이루어진 바 있다. 이들 연구에서는 구 구조 형식의 STEP 2000 말뭉치를 의존 구조 형식의 말뭉치로 변환하여 연구하였다. 이들 연구는 모두 어절 단위를 기초로 한 것인데, 서강대와 달리 포항공대의 연구에서는 계층구조를 도입하여 의존 관계를 더욱 명시화하였다.

나. 연구 내용

1) 분석 방법론 연구

앞 절에서 간략히 살펴본 바와 같이 기존의 구문분석 말뭉치 구축의 방법론은 구 구조 형식에 의한 것과 의존 구조 형식에 의한 것으로 대별되는데, 본 연구에서는 전자의 방식을 취한다. 그 이유는, 이미 Penn Korean Treebank에서 내세운 바대로 구 구조 형식이 순수 의존 구조 형식보다 더욱 풍부한 구문 정보를 표시해 줄 수 있기 때문이다. 즉, 구 구조 형식에서는 NP, VP 등과 같은 구문 표지나 SBJ나 OBJ와 같은 기능 표지를 나타내고, 문장 가운데 생략된 성분을 명시적으로 표시해 줄 수 있을 뿐만 아니라, 보충어와 부가어를 분명하게 구별해 줄 수 있으며, 명사구 이동에 따른 흔적 정보까지 사용할 수 있기 때문이다. Penn Korean Treebank는 이와 같은 입장에서 대상 문장에 대한 풍부한 구문 정보를 구문분석 말뭉치에 담아내려고 하였던 것이다. 한편, 구 구조 형식은 의존 구조 형식으로 변환 가능한 데 비해, 의존 구조 형식은 구 구조 형식으로 변환이 가능하지 않다는 점도 구 구조 형식을 택하게 된 주요한 근거가 된다.

그러나 본 연구에서는, 비록 Penn Korean Treebank와 같이 구 구조 형식을 구문분석의 기본 방법론으로 택하고는 있으나, 생략된 논항(PRO, pro)이나 논항 이동에 따른 흔적(trace) 정보를 위한 공범주(Empty Category)는 일체 도입하지 않는다는 점에서 Penn Korean Treebank와 구별된다. 공범주를 도입하려는 이유는 문장의 구문 정보를 명시적으로 나타내기 위함인데, 그러나 그러한 시도는 첫째, 구문분석의 자동 처리를 거의 불가능하게 만든다는 문제점을 지닌다. 둘째로 그러한 공범주의 설정 없이도 공범주 효과를 낼 수 있는 방안이 있다는 점이다. 예컨대, “S \rightarrow NP_OBJECT S”와 같은 다시 쓰기 규칙은 목적어 명사구가 문두로 이동했다는 것을 의미하는 것으로 이해할 수 있다. 따라서 이와 같은 다시 쓰기 규칙을 인정한다면 흔적을 표시하기 위한 공범주 설정은 불필요해진다. 또한 생략된 논항에 대한 정보를 표시하고자 무형 논항을 도입하려는 시도도 마찬가지로의 문제에 직면하게 된다. 더욱이 무형 논항의 설정이 주로 필수 보충어를 겨냥한 것인 만큼, 그것은 한국어 용언들의 논항구조(Argument Structure) 혹은 의미역 틀(Thematic Role Grid)에 대한 정보 베이스를 필요로 하게 되는데, 아직은 이에 관한 국내외의 표준적인 틀이 마련되어 있지 않다. 그럼에도 불구하고, 만일 무리하게 무형 논항의 표시를 강행한다면, 결국 확실한 기준 없이 연구자 혹은 작업자의 개인적인 직관이나 판단에 의거하여 동일한 동사 구문에 서로 다른 무형 논항의 표시가 이루어져 결국 균일하지 못한 구문분석 말뭉치를 산출해 낼 수밖에 없을 것이다.

결국, 본 연구는 구문분석에 있어서 구 구조 형식을 바탕으로 하되, 자동적인 자연 언어 처리에 부합할 수 있는 새로운 접근 방식을 취하게 되었다. 이와 관련된 자세한 사항은 본 보고서의 말미에 첨가되어 있는 <구문분석 말뭉치 구축 지침>을 참고하기 바란다.

2) 작업 대상 선정

가) 대상 자료

구문분석의 대상 자료는, 21세기 세종계획의 일환으로 구축된 형태분석 말뭉치 가운데 일부인 20만여 어절을 대상으로 한 총 7개 파일(총 205,138개 어절)이다. 대상 파일들 내부의 문장이 보이는 복잡성의 정도를 고려하였고, 작년까지 구축된 구문분석 말뭉치의 장르 분포를 염두에 둔 채 교육용 말뭉치로 간주할 만한 형태분석 파일들을 중점적으로 채택하였다.

이와 같은 고려에 의해 선정된 파일의 특성과 그 분석 결과를 보이면 다음과 같다.

파일명	제목	장르	어절수	문장수	1문장당 평균 어절수	비율 (어절 기준)
4BG_B18.BNK	우리교육 중등용 95년 1월호	잡지, 인문, 교육	52,295	4,673	11.19	25%
6BG_087.BNK	마이너리그	책, 상상, 소설, 일반	45,073	4,666	9.65	22%
6BG_B14.BNK	사람의 아들(2/3)	책, 상상, 소설, 일반	20,034	1,715	11.68	10%
6BG_B15.BNK	사람의 아들(3/3)	책, 상상, 소설, 일반	16,240	1,787	9.08	8%
6BG_B20.BNK	심리학개론	책, 인문, 철학, 사상	21,813	1,621	13.45	11%
6BG_B43.BNK	경향신문 96-01 생활여성	신문, 보도, 해설-사회	26,893	2,990	8.99	13%
6BG_B60.BNK	하늘에 뜬 돌도끼	책, 교육, 아동/비허구	22,790	3,346	6.81	11%
총계			205,138	20,798	9.82	100%

[표1] 구문분석 말뭉치 목록

올해 구축 분량을 포함하여 지금까지 구축된 구문분석 말뭉치는 총 35만 어절이다. 2002년에는 3만 어절, 2003년엔 12만 어절, 올해는 20만 어절을 구축하였다. 문장 수로 환산해 보면 총 33,683개 문장(총 362,518개 어절)이다.

통상 국어의 1문장당 어절 수는 10어절 정도라고 얘기하는데, 사실상 1파일 내부의 문장당 평균 어절 수를 산출한다는 것은 큰 의미를 가지지 못한다. 왜냐하면 다음의 자료에서 보듯이 실제 구문분석 파일들이 보여주는 문장당 어절 수 분포는 매우 다양하기 때문이다.

구성 어절수	해당 문장수	문장 비율	소구분	중구분	대구분
1~5	8,916	26.47%	26.47%	26.47%	55.34%
5~10	9,725	28.87%	28.87%	28.87%	
11~15	7,470	22.17%	22.17%	44.56%	
16~20	4,047	12.01%	12.01%		
21~25	1,945	5.77%	10.48%		
25~30	888	2.63%			
31~35	406	1.20%			
36~40	160	0.47%			
41~45	67	0.19%			
46~50	26	0.07%			
51~55	19	0.05%			
56~60	4	0.01%			
61~65	1	0.0029%			
66~70	2	0.0059%			
71~75	2	0.0059%			
76~80	2	0.0059%			
81~85	1	0.0029%			
86~90	0	0%			
91~95	1	0.0029%			
96~100	0	0%			
101~105	1	0.0029%			
총계	33,683	100%		100%	100%

[표2] 문장당 어절 수 분포

위의 도표는 구문분석 말뭉치에 포함된 각 문장들의 어절 수와 그 분포를 보인 것이다. 3개년 간 구축한 35만 문장의 문장당 어절 수 분포는 1개에서 102개에 이르기까지 다양한데, 10어절 이하의 문장이 전체의 55.34%를 차지하고, 그 이상의 어절 수를 가지는 문장은 전체의 44.66%를 차지한다. 1문장당 보통 10개의 어절로 구성되어 있다고 보는 통상적인 견해에 비추어 보더라도 현 구문분석 작업 대상 문장들의 복잡도는 충분히 높은 수준이라고 할 수 있다.

나) 분석 단위

본 연구에서 대상으로 삼은 언어 단위는 어절이다. 선행 연구에 따라서는 형태소를 기본 단위로 삼는 경우도 있다. 형태소를 기본 단위로 삼는 경우의 장점은 가령 국어의 조사나 어미를 독립적인 통사적 원자(Syntactic Atom)로 보는 국어학의 기본 시각을 잘 반영해 줄 수 있다는 것 등이다. 그러나 이러한 방식은 첫째, 지나치게 어절 수가 늘어나

구문분석 작업의 효율을 떨어뜨린다는 점과, 둘째 군이 형태소를 단위로 하지 않아도 어절 단위의 분석에서 이러한 점이 무시되는 것은 아니라는 점이다.

먼저 첫 번째 문제점과 관련하여 다음과 같은 실제 문장을 살펴보자.

예) 월남전의 망령이 20여 년이 지난 지금 미국 민주당 진영에서 되살아나고 있다.

[어절 기반 구문분석]의 경우

월남전/NNP + 의/JKG	1행
망령/NNG + 이/JKS	2행
20/SN + 여/XSN	3행
년/NNB + 이/JKS	4행
지나/VV + ㄴ/ETM	5행
지금/NNG	6행
미국/NNP	7행
민주당/NNP	8행
진영/NNG + 에서/JKB	9행
되살/VV + 아/EC + 나/VX + 고/EC	10행
있/VX + 다/EF + ./SF	11행

[형태소 기반 구문분석]의 경우

월남전/NNP	1행
의/JKG	2행
망령/NNG	3행
이/JKS	4행
20/SN	5행
여/XSN	6행
년/NNB	7행
이/JKS	8행
지나/VV	9행
ㄴ/ETM	10행
지금/NNG	11행
미국/NNP	12행
민주당/NNP	13행

진영/NNG	14행
에서/JKB	15행
되살/VV	16행
아/EC	17행
나/VX	18행
고/EC	19행
있/VX	20행
다/EF	21행
/SF	22행

위에서 보듯이, 임의의 문장에 대해서 형태소를 기본 단위로 한 구문분석의 경우, 어절을 기본 단위로 한 경우보다 2배나 행이 더 길어진다. 구구조이거나 의존 구조이거나 어느 형식을 취하더라도 조사나 어미가 국어학적으로 통사 원자가 된다는 점만을 부각시켜 그 분석을 시도한다면, 실제 대규모의 말뭉치 구축에 드는 시간과 비용은 더욱 커질 수밖에 없고 상대적으로 복잡도 역시 더욱 증가할 것이다. 이러한 작업은 실제로 구문분석기 프로그램이 제공하는 창(window)에서 이루어질 수밖에 없기 때문이다.

또한 굳이 이렇게 복잡한 형태소 기반의 분석 방법이 아니더라도 어절 단위 분석에서도 조사나 어미의 정보는 그것의 어기(base)의 구문 정보에 포함되어 어기가 투사하는 상위 마디(node)로 고스란히 삼투되기 마련이다. 따라서 순수 국어학적인 입장만을 고려하여 조사나 어미가 그것의 선행어들을 자매로 취하는 형상을 보여 주기 위해 형태소 기반 구문분석을 취한다는 것은 득보다는 실이 더 많은 것으로 보인다. 따라서 본 연구에서는 어절 단위를 기반으로 한 구문분석을 시도한다.

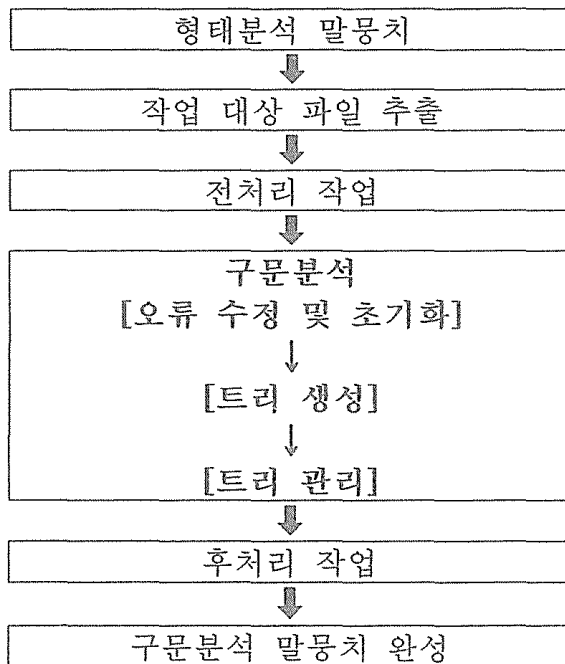
이렇듯 어절을 기본 단위로 함에 따라 구문분석의 대상도 아래와 같이 한정된다.

형태분석 범주	구문분석 범주
체언	NP(체언구)
용언	VP(용언구)
수식언	AP(부사구), DP(관형사구)
독립언	IP(감탄사구)

[표3] 구문분석의 대상

다) 말뭉치 구축 절차

구문분석 말뭉치 구축의 절차를 정리하면 다음과 같다.



[그림 1] 구문분석 말뭉치 구축 절차

• 작업 대상 파일 추출

장르를 적절히 고려하여 이미 구축된 형태분석 말뭉치로부터 대상 파일을 선정한다.

• 전처리 작업

추출된 형태분석 말뭉치의 각 파일 내부에 들어가 어절의 행들을 문장 단위로 끊고, 각 문장들에서 어절 구분이 제대로 되어 있는가를 확인하고 잘못된 것이 있으면 고쳐준다.

• 구문분석

고려대학교 자연어처리연구실에서 제작한 “구문태그부착 말뭉치 구축 도구” 프로그램을 이용하여 다음의 세부 절차를 통해 구문분석을 해 나간다.

- 오류 수정 및 초기화: 해당 어절의 구문 표지나 기능 표지의 확인 및 수정
 - 트리 생성: 구문분석 지침에 따라 문장 전체를 이분지 구조로 분석
 - 트리 관리: 분석된 개별 문장을 저장
- 후처리 작업

일관성 검사 및 최종 수정 작업, 헤더(Header) 부착

다. 연구 결과

최종 구축된 구문분석 말뭉치는 어절별 형태분석 결과와 함께 구문 표지 및 기능 표지, 그리고 구 구조 정보를 지닌다. 총 파일 수는 7개, 총 어절 수는 205,138개, 총 문장 수는 20,798개이다.

<분석 말뭉치의 예>

```

<!DOCTYPE tei.2 SYSTEM "c:\sgml\td\tei2.dtd" [
  <ENTITY % TEIcorpus "INCLUDE">
  <ENTITY % TEIextensions.ent SYSTEM "sejong1.ent">
  <ENTITY % TEIextensions.dtd SYSTEM "sejong1.dtd"> ]>
<tei.2>
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>우리교육 중등용 95년 1월호</title>
        <author>우리교육 편집부</author>
        <sponsor>대한민국 문화관광부</sponsor>
        <respStmt>
          <resp>문헌입력, 표준화, 형태소 정보 부착</resp>
          <name>고려대학교 민족문화연구원</name>
        </respStmt>
      </titleStmt>
      <extent>52295 어절, 4673문장</extent>
      <publicationStmt>
        <distributor>국립국어연구원</distributor>
        <idno>4BG_B1&BNK, 원본4BT_B1&HWP</idno>
        <availability><p>배포 불가</p></availability>
      </publicationStmt>
      <notesStmt>
        <note><p>균형말뭉치에서 선정</p></note>
      </notesStmt>
      <sourceDesc>
        <bibl>
          <author>우리교육 편집부</author>
          <title>우리교육 중등용 95년 1월호</title>
          <pubPlace>서울</pubPlace>
          <publisher>우리교육 편집부</publisher>
          <date>1995</date>
        </bibl>
      </sourceDesc>
    </fileDesc>
    <encodingDesc>
      <projectDesc><p>21세기 세종계획 4차년도 형태소 말뭉치 구축</p></projectDesc>
      <samplingDecl><p>21세기 세종계획 말뭉치 파일을 형태소 정보 부착 파일로 변환</p></samplingDecl>
      <editorialDecl><p>21세기 세종계획 형태소 분석 말뭉치 구축 지침(2000-2)에 따른 형태소 정보 부착</p></editorialDecl>
    </encodingDesc>
    <profileDesc>
      <creation>
        <date>1995</date>
      </creation>
      <langUsage>
        <language id="KO" usage="99">한국어, 표준어</language>
      </langUsage>
      <textClass>
        <catRef scheme="SJ21" target="M123">잡지 : 인문 ; 교육</catRef>
      </textClass>

```

```
</profileDesc>
<revisionDesc>
  <respStmt>
    <resp>프로젝트 책임자</resp>
    <name>김홍규</name>
  </respStmt>
  <change>
    <date>2001</date>
    <respStmt>
      <resp>연구보조원</resp>
      <name>김일환, 김의수, 조진현, 김태훈, 김혜영</name>
    </respStmt>
    <item>21세기 세종계획 형태소 분석 말뭉치 구축 지침(2000-2)에 따른 형태소 정보 검토, 수정</item>
  </change>
  <change>
    <date>2004</date>
    <respStmt>
      <resp>연구보조원</resp>
      <name>김의수, 박현아, 송경화</name>
    </respStmt>
    <item>21세기 세종계획 구문 분석 말뭉치 구축 지침(2004-1)에 따른 구문 정보 부착</item>
  </change>
</revisionDesc>
</teiHeader>
```

<text>

<body>

; 꽃과 문화

(NP (NP_CNJ 꽃/NGG + 과/JC)
(NP 문학/NGG))

; 언 땅 속에 감춰 둔 향긋한 봄내음 냉이의 향취는 씹스레하면서도 흙냄새 가운데서도 가장 좋은 내음만을 모은 듯해서 캐는 이 데쳐 무치는 이 먹는 이 모두를 즐겁게 한다.

(S	(NP_SBJ	(NP	(VP_MOD	(NP_AJT	(NP	(VP_MOD 앞/VV + ㄴ/ETM) (NP 땅/NG))	
						(NP_AJT 속/NG + 예/JKB))	
			(VP_MOD	(VP 감추/VV + 어/EC)			
				(VP_MOD 두/VX + ㄴ/ETM))			
		(NP	(VP_MOD 향긋/XR + 하/XSA + ㄴ/ETM)				
			(NP 봄내음/NG))				
	(NP_SBJ	(NP_MOD 냉이/NG + 의/JKG)					
		(NP_SBJ 향취/NG + 는/JX))					
	(VP	(VP	(VP	(NP_OBJ	(VP_MOD	(NP_AJT	(NP 흠뻑/NG)
		(VP	(NP_OBJ	(VP_MOD	(NP_AJT	(NP	(NP_AJT 가운데/NG + 서/JKB + 도
/JX))							
					(VP_MOD	(AP 가장/MAG)	
						(VP_MOD 총/VA + 음/ETM))	
				(NP_OBJ	내음/NG + 만/JX + 을/JKO))		
			(VP	(VP_MOD 모으/VV + ㄴ/ETM)			
				(VP	뒹/NNB + 하/XSA + 아서/EC))		
	(VP	(NP_OBJ	(NP_CNJ	(VP_MOD	깨/VV + 는/ETM)		
				(NP_CNJ	이/NP + ㄴ/SP))		

(NP_OBJ (NP_CNJ (VP_MOD (VP 테치/VV + 어/EC)
(VP_MOD 무치/VV + 는/ETM))
(NP_CNJ 이/NP + ,/SP))
(NP_OBJ (VP_MOD 먹/VV + 는/ETM)
(NP_OBJ 아*/NP))))
(VP (NP_OBJ 모두/NG + 를/JK)
(VP (VP 줄갑/VA + 게/EC)
(VP 하/VX + ㄴ다/EF + ,/SF))))))

; 앞보다 뿌리가 더 싹하고 냄새도 많이 나므로 뿌리까지 다 먹는데 비타민과 무기질도 많다.

(S (VP (S (S (NP_AJT 앞/NG + 보다/JKB)
(S (NP_SBJ 뿌리/NG + 가/JKS)
(VP (AP 더/MAG)
(VP 싹하/VA + 고/EC))))
(S (NP_SBJ 냄새/NG + 도/JX)
(VP (AP 많이/MAG)
(VP 나/VV + 프로/EC))))
(VP (NP_OBJ 뿌리/NG + 까지/JX)
(VP (AP 다/MAG)
(VP 먹/VV + 는데/EC))))
(S (NP_SBJ (NP_CNJ 비타민/NG + 과/JC)
(NP_SBJ 무기질/NG + 도/JX)
(VP 많/VA + 다/EF + ,/SF))))

; 묵은 김장 김치찌개에 냉이를 넣는다면 이미 묵은 김치가 아니다.

(VP (VP (NP_AJT (VP_MOD 묵/VV + 은/ETM)
(NP_AJT (NP 김장/NG)
(NP_AJT 김치찌개/NG + 예/JKB))))
(VP (NP_OBJ 냉이/NG + 를/JK)
(VP 넣/VV + 는다면/EC)))
(VP (AP 이미/MAG)
(VP (NP_CMP (VP_MOD 묵/VV + 은/ETM)
(NP_CMP 김치/NG + 가/JKC)
(VP 아니/VCN + 다/EF + ,/SF))))

; 특히, '농가월령가' 2월령에도 기록되어 있는 냉이국은 만가에서 가장 많이 먹는 봄철의 국 가운데 하나이다.

(S (NP_SBJ (VP_MOD (AP 특히/MAG + ,/SP)
(VP_MOD (NP_AJT (NP 'SS + 농가월령가/NNP + 'SS)
(NP_AJT 2/SN + 월령/NG + 예/JKB + 도/JX)
(VP_MOD (VP 기록/NG + 되/XSV + 어/EC)
(VP_MOD 있/VX + 는/ETM))))
(NP_SBJ 냉이국/NG + 은/JX))
(VNP (NP (NP (VP_MOD (NP_AJT 만가/NG + 에서/JKB)
(VP_MOD (AP (AP 가장/MAG)
(AP 많이/MAG)
(VP_MOD 먹/VV + 는/ETM))))
(NP (NP_MOD 봄철/NG + 의/JKG)
(NP 국/NG))
(NP 가운데/NG))
(VNP 하나/NR + 이/VCP + 다/EF + ,/SF))))

; 국문학자 이병주는 뽕안 싹뜨물을 받아서 모시조개 약간을 넣고 된장을 풀어서 끓이면 그 씹살한 맛이 향긋한 내음과 더불어 아취를 북돋운다고 하였고
불문학자 이봉구는 토장에다 고추장 기운을 약간 가해 모시조개를 넣고 끓인 냉이국은 해토(解土) 냄새가 풍기는데, 봄맞이국으로는 최고에 속하는 음식이라

[illegible]

(VP_MOD 속하/VV + 는/ETM))
(VNP_CMP 음식/NNG + 이/VCP + 라고/EC))
(VP 하/VV + 았/EP + 다/EF + /SF)))

; 모든 일이 마찬가지이지만 운동도 꾸준한 자기의 노력과 확신이 없으면 원하는 결과를 얻기는 힘든 것이다.
(S
(S
(NP_SBJ (DP 모든/MM)
(NP_SBJ 일/NNG + 이/JKS)
(VNP 마찬가지/NNG + 이/VCP + 지만/EC))
(S
(NP_SBJ 운동/NNG + 도/JX)
(S
(S
(NP_SBJ (NP_CNJ (VP_MOD 꾸준/XR + 하/XSA + ㄴ/ETM)
(NP_CNJ (NP_MOD 자기/NP + 의/JKG)
(NP_CNJ 노력/NNG + 과/JC))
(NP_SBJ 확신/NNG + 이/JKS))
(VP 없/VA + 으면/EC))
(S
(VP_SBJ (NP_OBJ (VP_MOD 원하/VV + 는/ETM)
(NP_OBJ 결과/NNG + 를/JKO))
(VP_SBJ 얻/VV + 기/ETN + 는/JX)
(VNP (VP_MOD 힘들/VA + ㄴ/ETM)
(VNP 것/NNB + 이/VCP + 다/EF + /SF))))))

; 이 겨울 한가지 운동이라도 시작해 건강하게 보내자.
(VP
(NP_AJT (DP 이/MM)
(NP_AJT 겨울/NNG))
(VP
(VP
(VNP_OBJ (NP 한/MM + 가지/NNB)
(VNP_OBJ 운동/NNG + 이/VCP + 라도/EC))
(VP 시작/NNG + 하/XSV + 아/EC))
(VP
(VP 건강/NNG + 하/XSA + 게/EC)
(VP 보내/VV + 자/EF + /SF))))

; 그러나 시작이 빈이라는 말과 함께 "가다가 그치면 아니간만 못하니라"라는 말도 잊지 말자.
(VP
(AP 그러나/MAG)
(VP
(AP
(NP_AJT (S_MOD (NP_SBJ 시작/NNG + 이/JKS)
(VNP_MODAL 받/NNG + 이/VCP + 라는/ETM))
(NP_AJT 말/NNG + 과/JKB))
(AP 함께/MAG))
(VP
(NP_OBJ (VNP_MODAL (VP
(L "/SS)
(VP
(VP
(VP
(VP
가/VV + 다가/EC)
그치/VV + 면/EC))
(NP_AJT 아니/MAG + 가/VV + ㄴ/ETM + 만/NNB)
못/MAG + 하/VV + 나라/EC))
(R "/SS))
(VNP_MODAL 이/VCP + 라는/ETM))
(NP_OBJ 말/NNG + 도/JX)
(VP
(VP 앞/VV + 자/EC)
(VP 말/VX + 자/EF + /SF))))))

참고문헌

- 강범모. 2001. 「범주문법: 한국어의 형태론, 통사론, 타입논리적 의미론」, 고려대학교 민족문화연구원.
- 강범모·김의수. 2004, 세종 구문분석 말뭉치를 위한 구문 분석 방법, 「코퍼스와 어휘 데이터베이스」(강범모 외 편), 월인.
- 고광주. 대등 접속문에 대한 재검토, 「한국어학」 9. 한국어학회.
- 김의수. 2002. 언어단위로서의 상당어 설정 시고, 「형태론」 4-1, 박이정 출판사.
- 김의수. 2003. 국어의 격과 의미역 연구, 고려대 박사학위논문.
- 김의수·강범모. 2002. 구문분석 말뭉치 구축을 위한 분석의 원칙, 방법, 문제, 2002년도 제14회 한글 및 한국어 정보처리 학술대회 발표논문집.
- 김의수·조진현·김태훈·박현아·송경화. 2004, 말뭉치에 나타난 어근의 분포, 「코퍼스와 어휘 데이터베이스」(강범모 외 편), 월인.
- 김흥규 외. 2001. 형태소 분석 말뭉치 구축 지침, [21세기 세종계획 국어 기초자료 구축 연구보고서], 문화관광부.
- 서정수. 1996. 「국어문법」, 한양대 출판원.
- 이관규. 1992. 「국어 대등구성 연구」, 서광학술자료사.
- 이기용 외. 1997. 구문분석기 데이터 정보개발, ms. 한국과학기술원.
- 이선웅. 2000. '의문사+(이)--ㄴ가' 구성의 부정(不定) 표현에 대하여, 「국어학」 36. 한국어학회.
- 임준호 외. 2002. 구문패턴을 이용한 반자동 구문분석 말뭉치 구축도구, 2002년도 제 14회 한글 및 한국어 정보처리 학술대회 발표논문집.

임해창 외. 2001a. 구문 태그 부착 말뭉치 구축 도구 개발, 21세기 세종계획 국어 기초자료 구축 분과 제2 워크숍 발표문.

임해창 외. 2001b. 구문 태그 부착 말뭉치 구축 도구 개발, [21세기 세종계획 국어 기초자료 구축 연구보고서], 문화관광부.

홍종선 · 강범모 · 최호철. 2001. 한국어 연어 관계 연구, 월인.

Han, Chung-hye, Na-Rae Han and Eon-Suk Ko. 2001. Bracketing Guidelines for Penn Korean TreeBank, IRCS Report 01-09, Institute for Research in Cognitive Science University of Pennsylvania.

〈부록〉 구문분석 말뭉치 구축 지침 (2004 version 1.0)

I. 기본 원칙

- 자연언어처리에서 일반적으로 고려되는 일관성 유지와 효율성 제고에 초점을 두되, 일반언어학적 관점에서도 크게 벗어나지 않도록 한다.
- 표층 구조를 중시하여 분석한다.
- 이분지 가설을 취하며 다분지를 허용하지 않는다.
- 공범주를 인정하지 않는다.
- 어절을 분석의 기본 단위로 한다.
- 보어와 부가어를 구분하되 보어의 범위를 엄격히 제한한다.
- 원칙적으로 접속과 내포를 구별하지 않으며 접속절은 모두 부사절로 분석한다.
(다만 명사구 접속만은 인정한다.)
- 하나의 주어가 모문과 내포문 모두에 관련되어 있어 구문 분석의 중의성이 발생할 경우, 모문의 주어로 우선 분석한다.

II. 구문 분석 표지

- 이 구문 분석 표지는 21세기 세종 계획 국어기초자료구축 본과의 구문분석 말뭉치 (2004년 12월말 현재까지 35만 어절, 향후 총 100만 어절 규모)를 구축하기 위해 마련된 것이다.
- 이 분석 표지는 2001년도 세종 계획의 세부 과제 중 하나였던 '구문 표지 부착 말뭉치 구축 도구 개발'에서 처음 제안되었던 것을 수정·확충한 것이다.

· 구문 표지

	범주	사례
S	문장	
Q	인용절	인용부호(“”) 안에 들어 있는 두 개 이상의 문장
NP	체언구	체언(명사, 대명사, 수사)
VP	용언구	용언(동사, 형용사, 보조용언)
VNP	긍정 지정사구	긍정 지정사 '이다'
AP	부사구	부사
DP	관형사구	관형사
IP	감탄사구	감탄사

· 기능 표지

	범주	사례
SBJ	주어	주격 체언구, 명사 전성 용언구, 명사절 (NP_SBJ, VP_SBJ, S_SBJ)
OBJ	목적어	목적격 체언구, 명사 전성 용언구, 명사절 (NP_OBJ, VP_OBJ, S_OBJ)
CMP	보어	보격 체언구, 명사 전성 용언구, 인용절 (NP_CMP, VP_CMP, S_CMP)
MOD	체언 수식어	관형격 체언구, 관형형 용언구, 관형절 (NP_MOD, VP_MOD, S_MOD)
AJT	용언 수식어	부사격 체언구, 문말어미+부사격조사 (NP_AJT VP_AJT, S_AJT)
CNJ	접속어	접속격 체언(NP_CNJ)
INT	독립어	체언 (NP_INT)

· 기타 표지

	범주	사례
PRN	삽입어구	삽입된 성분의 기능표지 위치에 (예: [NP_PRN])
X	의사 구 (pseudo phrase)	인용부호와 괄호를 제외한 나머지의 부호나, 조사, 어미가 단독으로 어절을 이룰 때 그 구문표지 위치에 표시(예: [X_CMP])
L, R	부호	인용부호나 괄호의 구문표지 위치에. 왼쪽 부호에는 L을, 오른쪽 부호에는 R을 표시
Q, U, W, Y, Z	인용절	인용부호(“”)에 이끌려 나온 두 개 이상의 인용절을 대신하여 표기되는 부호

Ⅲ. 표지별 분류 기준 및 세부 지침

1. 절의 기본 구조

1.1 문장의 기본 구성 성분

1.1.1 서술어(Predicates)

· 문장(S)의 서술어에는 VP와 S, NP, AP가 있다.

예) [S [NP_SBJ 그가] [VP 떠났다.]]

예) [S [NP_SBJ 그녀가] [S [NP_SBJ 얼굴이] [VP 예쁘다.]]]

예) [S [NP_SBJ 우리 학교가] [NP 우승.]]

예) 그는 [S [NP_SBJ 아무 생각이] [AP 없이]] 그 일을 했다.

《주의》 본 구문 분석에서 기본이 되는 원칙이 있는데, 그것을 다시쓰기 규칙(Rewrite Rule)으로 나타내면 다음과 같다.

$$S \rightarrow \{NP/VP/S\}_{SBJ} + VP$$

$$S \rightarrow \{NP/VP/S\}_{SBJ} + S$$

$S \rightarrow \{NP/VP/S\}_{\text{SBJ}} + NP$

$S \rightarrow \{NP/VP/S\}_{\text{SBJ}} + AP$

이러한 원칙은 “VP나 S 등이 주어($\{NP/VP/S\}_{\text{SBJ}}$)를 만나면 S를 이룬다”는 말로 간략히 표현할 수 있다. 이 원칙은 일견 매우 당연해 보이지만 실제의 구문 분석에서 매우 중요한 역할을 수행한다. 특히 주격 중출 구문이나 도치(어순재배치) 구문에서 그러하다.

☞ /Upenn TreeBank 상동/

· 서술어를 위해 따로 기능 표지를 마련하지 않는다. 다만 VP가 논항이나 수식어로 기능할 경우 그에 부합하는 기능 표지를 가질 수는 있다.

예) 그는 [VP [VP 일등을 하기 위해] 열심히 공부했다.]

예) 그는 [S [VP_SBJ 거기에 가기가] 싫었다.]

예) 그는 [NP_SBJ [VP_MOD 공부하는] 것이] 좋았다.

예) 그것은 [VP [VP_AJT 얼마나 쓰는가]에] 달려있다.]

1.1.2 논항(Arguments)

1) 주어(SBJ) : 주격조사를 가진 명사구, 그 조사가 생략되거나 보조사로 대체된 명사구, 그리고 이러한 명사구와 같이 기능하는 용언구와 절

《주의》 단체 주어 인정 여부

학교문법에서는 단체 주어를 인정하여 아래 예문의 ‘우리 학교에서’를 주어로 분석하기도 한다.

예) 우리 학교에서 운동회를 열었다.

그러나, 다음과 같이 “단체명+에서”가 주어인지 부사어인지 불분명한 경우가 매우 많다.

예) 우리 학교에서 일등을 하였다.

이러한 경우에는, ‘우리 학교’가 일등을 한 것인지 우리 학교에서 ‘아무개’가 일등을 한 것인지가 불분명하다.

또한, 다음의 경우에서 보이듯, 장소 부사어가 출현할 경우 단체 주어는 ‘에서’가 아닌 ‘가’로 실현되어야만 하는 제약이 있다.

예) 이번 전국체전에서 우리 학교가 일등을 하였다.

예) *이번 전국체전에서 우리 학교에서 일등을 하였다.

따라서, 본 구문분석에서는 ‘에서’를 단체 주어가 아닌 부사어로 분석하기로 한다.

2) 목적어(OBJ) : 목적격 조사를 가진 명사구나 그 조사가 생략되거나 보조사로 대체된 명사구, 그리고 이러한 명사구와 같이 기능하는 용언구와 절

예) 철수는 [NP_OBJ 밥을] 먹었다.

철수는 [NP_OBJ 밥도] 먹었다.

철수는 [NP_OBJ 밥] 먹었다.

철수는 [S_OBJ 우리가 함께 가기를] 청했다.

철수는 [VP_OBJ 가기를] 싫어했다.

《주의》 목적어/부가어의 구별

i) ‘을/를’만을 취할 수 있는 명사구는, ‘을/를’이 생략되거나 특수조사가 있어도 ‘_OBJ’로 분석해 준다.

예) 철수가 {[NP_OBJ 밥을], [NP_OBJ 밥]} 먹었다.

ii) ‘을/를’ 이외에도 어휘격 조사(‘에’, ‘로’ 등)를 허용하는 명사구나, 시간 혹은 처소를 나타내는 명사구는, ①‘을/를’을 동반하는 경우에는 ‘_OBJ’로, ②어휘격 조사를 동반하는 경우에는 ‘_AJT’로, ③조사 없이 쓰일 경우에는 ‘_AJT’로 분석해 준다. (이 경우는 특히 시간이나 거리의 명사구에서 자주 문제가 된다.)

예) 현우는 학교를 하루 쉬었다.

⇒ 현우는 학교를 [NP_AJT 하루] 쉬었다.

현우는 학교를 하루를 쉬었다.

⇒ 현우는 학교를 [NP_OBJ 하루를] 쉬었다.

현우는 아침 일찍 학교 갔다.

⇒ 현우는 아침 일찍 [NP_AJT 학교] 갔다.

현우는 아침 일찍 학교를 갔다.

⇒ 현우는 아침 일찍 [NP_OBJ 학교를] 갔다.

현우는 아침 일찍 학교에 갔다.

⇒ 현우는 아침 일찍 [NP_AJT 학교에] 갔다.

- 3) 보어(CMP) : 보격조사를 가진 명사구나 그 조사가 생략되거나 보조사로 대체된 명사구 혹은 이러한 명사구와 같이 기능하는 용언구, 그리고 인용절 보문

《주의》 보어/부가어의 구별

보어와 부가어의 구별 자체가 어렵다. 따라서 일관성 있고 간결한 구문 분석을 위해 명사구의 경우 학교문법의 처리에 따라 ‘되다, 아니다’의 앞에서 보격조사 ‘이/가’나 특수조사를 취하는 명사구만을 보어(_CMP)로 분석하기로 한다(단, ‘되다, 아니다’ 앞의 성분이 어휘격 조사(로/으로)를 가질 경우엔 부가어(_AJT)로 처리한다). 따라서 ‘되다, 아니다’를 제외한 2항의 비대격 술어가 가지는 두 번째 논항 ‘NP+이’도 보어가 아닌 주어로 분석한다. 한편, 절과 관련하여서는 인용절 보문만을 보어로 분석한다.

예) 철수가 [NP_CMP 의사가] 되었다.

철수가 [NP_AJT 의사로] 되었다.

철수가 [NP_CMP 의사만] 되었어도 좋았을텐데.

철수는 [NP_SBJ 고향이] 그림다.

철수는 [S_CMP 자신이 잘못을 하지 않았다고] 말했다.

1.1.3 부가어(Adjuncts)

· 부가어는 문장 성분 가운데 위에 정의된 서술어와 논항, 그리고 아래에 서술될 독립어를 제외한 나머지 성분들을 모두 아우른다. 따라서 순수 언어학적인 입장에서 보어로 처리될 수 있는 것들(예, 수여동사 구문에서의 여격 성분 등)도 본 구문 분석에서는 모두 부가어로 처리된다.

《주의》 관형사와 부사는 부가어의 범주에 속하며 그것의 구문 표지 자체가 이미 그것의 기능을 나타내 준다. 따라서 이들에게는 별도로 부가어의 기능 표지를 더해 주지 않는다.

예) [DP 그] 사람은 [AP 매우] 잘생겼다.

· 부가어의 종류 I : 피수식 성분의 종류에 따라

1) 체언 수식어(MOD) : NP를 수식하는 부가어로서 관형격 체언(NP_MOD)과 관형형 용언(VP_MOD), 관형절(S_MOD)을 가리킨다.

예) 철수는 [NP_MOD 그녀의] 동생을 만났다.

예) 철수는 [VP_MOD 예쁜] 여자와 사귄다.

예) 철수는 [S_MOD 내가 좋아하는] 음식을 사왔다.

예) 철수는 [VP_MOD 먹던] 음식을 옆으로 밀쳐 내었다.

《주의》 일부 부사는 NP를 직접 수식할 수도 있다.

예) [NP_SBJ [AP_MOD 바로] [NP_SBJ 그가] 그 일을 꾸민 게 틀림없어.

2) 용언 수식어(AJT) : VP나 S, 혹은 서술성 명사(Verbal Noun)를 수식하는 부가어로서 체언(NP_AJT), [문말어미 + 부사격조사]의 구나 절(VP_AJT, S_AJT) 등을 가리킨다.

예) 철수는 [NP_AJT 학교에] 갔다.

예) 그것은 [VP_AJT 얼마나 쓰는가에] 달려 있지요.

예) 날씨가 [S_AJT 우리가 수영하기에는] 좀 추웠다.

· 부가어의 종류 II : 부가의 층위에 따라

1) S-층위 부가어 : 주어 앞에 위치하면서 문장 전체를 수식하는 부가어.

예) [S [NP_AJT 밤마다] [S 철수는 공부를 열심히 한다.]]

예) [S [AP 매일] [S 철수는 공부를 열심히 한다.]]

예) [S [S_AJT 비가 왔기에] [S 우리들은 밖에 나가 놀 수가 없었다.]]

예) [S [S 비가 와서] [S 우리들은 밖에 나가 놀 수가 없었다.]]

2) VP-층위 부가어 : 주어와 VP 사이에 위치하면서 VP를 수식하는 부가어.

예) 철수는 공부를 [VP [NP_AJT 밤마다] [VP 한다.]]

예) 철수는 공부를 [VP [AP 열심히] [VP 한다.]]

예) 우리 반은 [VP [S_AJT 학생들이 매우 열심히 공부했기에] 이번에 1등을 차지하였다.]

예) 우리 반은 [VP [S 학생들이 매우 열심히 공부해서] 이번에 1등을 차지하였다.]

3) NP-충위 부가어 : ‘바로’와 같은 일부 특수한 부사의 경우도 이에 속한다.

예) [NP_SBJ [NP_MOD 그의] 고향은] 여기서 멀다.

예) [NP_SBJ [DP 그] 집은] 여기서 가깝다.

예) [NP_SBJ [AP_MOD 바로] [NP_SBJ 그가]] 그 일을 꾸민 게 틀림없어.

예) 철수는 [NP_OBJ [S_MOD 어제 그가 잡은] 고래를] 놓아주었다.

1.1.4 독립어

· 독립어는 다른 문장 성분들과 직접적인 관련이 없이 문두에 나타나는 언어단위를 가리킨다. 이러한 독립어에는 부름이나 감탄, 대답을 나타내는 어구들이 있다.

예) 철수야, 얼른 이리 와!

⇒ [VP [NP_INT 철수야,] 얼른 이리 와!]

예) 앗, 불이 났다.

⇒ [S [IP 앓,] 불이 났다!]

예) 예, 제가 가겠습니다.

⇒ [S [IP 예,] 제가 가겠습니다.]

예) "글쎄, 쓰려면 쓰지 뭐."

(VP (L "/SS))

(VP (VP (IP + 글쎄/IC + /SP)

(VP (VP (VP 쓰/VV + 려면/EC)

(VP 쓰/VV + 지/EC))

(IP 뭐/IC + /SF)))

(R + "/SS)))

1.1.5 공범주(Null elements)

· 순수 언어학이 아닌 자연언어처리의 관점에서 공범주 설정은 구문 분석의 자동화와 일관성 있는 분석에 큰 부담이 된다. 따라서 본 구문 분석에서는 공범주를 인정하지 않기로 한다.

예) 모문 : [VP 어서 가라!]

- 예) 관계절 : 그는 [VP_MOD 먹던] 사과를 나에게 내던졌다.
 예) 인용절 : 그가 나에게 [VP [VNP_CMP 바보라고] 말했다.]
 예) 접속문 : [VP [VP 오니까] [VP 갔다.]]

☞ “4. 공범주 문제” 항목 참고.

1.2 절의 기본 유형

1.2.1 단순절

· 보어를 ‘되다, 아니다’ 앞에 실현된 ‘NP_이/가’로 엄격히 제한하기 때문에, 결과적으로 불완전 타동사 구문은 단순 타동사 구문의 일종으로 처리된다.

· 기본 문형의 종류와 분석 :

1) 완전 자동사문 :

예) 아이가 걸어간다.

(S (NP_SBJ 아이/NNG + 가/JKS)
 (VP 걸어가/VV + ㄴ다/EF + ./SF))

2) 완전 타동사문 :

예) 철수가 밥을 먹었다.

(S (NP_SBJ 철수/NNP + 가/JKS)
 (VP (NP_OBJ 밥/NNG + 을/JKO)
 (VP 먹/VV + 었/EP + 다/EF + ./SF))))

3) 불완전 자동사문 :

예) 밭이 논이 되었다.

(S (NP_SBJ 밭/NNG + 이/JKS)
 (VP (NP_CMP 논/NNG + 이/JKC)
 (VP 되/VV + 었/EP + 다/EF + ./SF))))

4) 불완전 타동사문 :

예) 그들이 황무지를 논으로 만들었다.

(S (NP_SBJ 그/NP + 들/XSN + 이/JKS)
 (VP (NP_OBJ 황무지/NNG + 를/JKO)
 (VP (NP_AJT 논/NNG + 으로/JKB)
 (VP 만들/VV + 었/EP + 다/EF + ./SF))))

5) 서술어 생략 구문:

· 일반동사 구문에서의 서술어 생략

예) "응, 새가 어디서?"

(S (L "/SS)
 (S (S (IP + 응/IC + ./SP)
 (S (NP_SBJ 새/NNG + 가/JKS)
 (NP_AJT 어디/NP + 서/JKB + ?/SF)))
 (R + "/SS)))

· 계사 구문에서의 서술어 생략

예) 지나고 보면 모두가 정든 사람들뿐!

(S (VP (VP 지나/VV + 고/EC)
 (VP 보/VX + 면/EC))
 (S (NP_SBJ 모두/NNG + 가/JKS)
 (NP (VP_MOD 정들/VV + ㄴ/ETM)
 (NP 사람/NNG + 들/XSN + 뿐/JX + !/SF))))

1.2.2 비단순절

(가) 내포절

(가 -1) 논항절

· 논항절에는 ‘주어절, 목적어절, 보어절’이 있으며 이들은 각기 적절한 기능 표지들을 갖는다(S_SBJ, S_OBJ, S_CMP)

1) 주어절(S_SBJ) :

예) 가을에 학생들이 공부를 하기가 좋다.

(S (S_SBJ (NP_AJT 가을/NNG + 예/JKB)
 (S_SBJ (NP_SBJ 학생/NNG + 들/XSN + 이/JKS)
 (VP_SBJ (NP_OBJ 공부/NNG + 를/JKO)
 (VP_SBJ 하/VV + 기/ETN + 가/JKS))))
 (VP 좋/VA + 다/EF + ./SF))

2) 목적어절(S_OBJ) :

예) 철수가 이번 시험에서 자기가 일등이기를 바랬다.

(S (NP_SBJ 철수/NNP + 가/JKS)
 (VP (S_OBJ (NP_AJT (NP 이번/NNG)
 (NP_AJT 시험/NNG + 에서/JKB))
 (S_OBJ (NP_SBJ 자기/NP + 가/JKS)
 (VNP_OBJ 일등/NNG + 이/VCP + 기/ETN + 를
 /JKO))))
 (VP 바래/VV + 었/EP + 다/EF + ./SF)))

3) 보어절(S_CMP) :

예) 정탐꾼이 적들이 우리에게 다가오고 있다고 알려 주었다.

(S (NP_SBJ 정탐/NNG + 궂/XSN + 이/JKS)
 (VP (S_CMP (NP_SBJ 적/NNG + 들/XSN + 이/JKS)
 (VP_CMP (NP_AJT 우리/NP + 에게/JKB)
 (VP_CMP (VP 다가오/VV + 고/EC)
 (VP_CMP 있/VX + 다고/EC))))
 (VP (VP 알리/VV + 어/EC)
 (VP 주/VX + 었/EP + 다/EF + ./SF))))

(가 -2) 비논항절 (부가어절 ≡ 접속절, 서술절)

· 부가어절과 접속절(대등, 종속)은 원칙적으로 구별이 쉽지 않다. 따라서 그 둘을 궂

이 구별한다는 것은 구문 분석에서 일관성과 자동화의 효율을 떨어뜨리므로 모두 부가어절로 분석한다.

· 부가어절(≡ 접속절) : 이것은 피수식어에 따라 두 가지 유형으로 구분된다.

1) 체언 수식절 :

예) 철수는 [S_MOD 내가 좋아하는] 음식을 사왔다.

예) 철수는 [VP_MOD 먹던] 음식을 옆으로 밀쳐 내었다.

2) 용언 수식절 :

예) 우리 학교가 학생들이 책을 열심히 읽어 타교에 모범이 되고 있다.

(S (NP_SBJ (NP 우리/NP)
 (NP_SBJ 학교/NNG + 가/JKS))
 (VP (S (NP_SBJ 학생/NNG + 들/XSN + 이/JKS)
 (VP (NP_OBJ 책/NNG + 을/JKO)
 (VP (AP 열심히/MAG)
 (VP 읽/VV + 어/EC))))))
 (NP_AJT 타교/NNG + 에/JKB)
 (VP (NP_CMP 모범/NNG + 이/JKC)
 (VP (VP 되/VV + 고/EC)
 (VP 있/VX + 다/EF + .(SF))))))

《주의》 다만, 용언 수식절의 서술어가 {문말어미+부사격조사} 꼴로 나타나면 “_AJT”를 붙인다.

예) 그 날의 날씨는 [VP [S_AJT 우리가 소풍을 가기에] 적당하지 않았다.]

예) 그는 [VP [S 우리가 희생할 수 있도록] 도와 주었다.]

《주의》 ‘의하다’나 ‘대하다’, ‘관하다’ 등 특수한 서술어가 이루는 용언구의 경우 역시 보통의 경우처럼 분석해 준다.

예) 철수가 경찰들에 의해 붙잡혔다.

(S (NP_SBJ 철수/NNP + 가/JKS)
 (VP (VP (NP_AJT 경찰/NNG + 들/XSN + 에/JKB)

(VP 의하/VV + 아/EC))
 (VP 붙잡히/VV + 었/EP + 다/EF + ./SF)))

· 서술절 :

예) 아이가 얼굴이 예뻐다.

(S (NP_SBJ 아이/NNG + 가/JKS)
 (S (NP_SBJ 얼굴/NNG + 이/JKS)
 (VP 예쁘/VA + 었/EP + 다/EF + ./SF)))

(나) 접속절

· 대등접속절과 종속접속절은 둘 간의 변별이 쉽지 않을 뿐만 아니라, 부가어절과도 구별이 잘 되지 않는다. 따라서 이들 모두 부가어절로 취급하여 분석하기로 한다.

예) [S 철수는 [VP [VP 밥을 먹고] [VP 학교에 갔다.]]]

예) [S 철수가 [VP [VP 밥을 먹고] [VP 물을 마셨다.]]]

예) [S 나는 [VP [S 호랑이가 무서웠지만] [VP 아들을 위하여 호랑이 굴에 뛰어 들었다.]]]

예) 소나무는 푸르고 꽃은 붉었다.

(S (S (NP_SBJ 소나무/NNG + 는/JX)
 (VP 푸르/VA + 고/EC))
 (S (NP_SBJ 꽃/NNG + 은/JX)
 (VP 붉/VA + 었/EP + 다/EF + ./SF)))

2. 구의 세부 구조

2.1 명사구

2.1.1 명사 + 명사 구성

· 선행 명사와 후행 명사의 결합이 구인지 복합어인지 불분명할 경우가 많아 선행 명

사에 MOD라는 기능 표지를 선불리 주기 힘들다. 따라서 이 경우 아무런 기능 표지를 붙이지 않는다.

예) 그 일은 [NP_AJT [NP [NP 정부] [NP 수립]] [NP_AJT 이후]] 발생했다.

· 명사 연결 구성을 분석할 때에는 명사들 간의 의미를 고려하여 의미관계가 긴밀한 것들끼리 우선적으로 묶어 준다.

예) ... [NP [NP [NP 도중] [NP 하차]] [NP [NP 전도] [NP 무효]]] ...

2.1.2 명사 접속 구성

<유형>

· 1유형: “접속조사”나 “첨표”에 의한 명사구 접속(무표적인 경우 포함)

예1-1) [철수와 내가] 집에 갔다.

예1-2) [철수와 영수와 내가] 집에 갔다.

예1-3) [철수, 영수, 순이가] 집에 갔다.

예1-4) [철수 영수 순이가] 집에 갔다.

· 2유형: 접속부사(MAJ)에 의한 명사구 접속

예2-1) [송신기 {그리고/혹은/및/또는/내지} 수신기가] 고장이 났다.

예2-2) [송신기, 수신기, 그리고 중계기가] 고장이 났다.

<분석 방식>

· 분석의 기본 원칙

1) [NP1 and NP2]의 형식에서 NP1에 명사구 접속의 기능표지 “_CNJ”를 붙인다.

☞ [NP_SBJ [NP_CNJ 철수와] 영희가] ...

2) 복수의 명사구가 접속될 경우에는 맨 뒤에서부터 연결시켜 나아간다.

☞ [NP1 [NP2 [NP3 NP4]]]

· 1유형의 경우

예1-1) 철수와 내가 집에 갔다.

☞ [NP_SBJ [NP_CNJ 철수와] [NP_SBJ 내가]] ...

예1-2) 철수와 영수와 내가 집에 갔다.

☞ [[NP_CNJ 철수와] [[NP_CNJ 영수와] [NP_SBJ 내가]]] ...

예1-3) 철수, 영수, 순이가 집에 갔다.

☞ [[NP_CNJ 철수,] [[NP_CNJ 영수,] [NP_SBJ 순이가]]] ...

예1-4) 철수 영수 순이가 집에 갔다.

☞ [[NP_CNJ 철수] [[NP_CNJ 영수] [NP_SBJ 순이가]]] ...

· 2유형의 경우

예2-1) 송신기 {그리고/혹은/또는} 수신기가 고장이 났다.

☞ [NP_SBJ [NP_CNJ 송신기] [NP_SBJ [AP 그리고] [NP_SBJ 수신기가]]]

...

예2-2) 송신기, 수신기, 그리고 중계기가 고장이 났다.

☞ 형상: [송신기, [수신기, [그리고 [중계기가]]]]

☞ 명세: [NP_SBJ [NP_CNJ 송신기,] [NP_SBJ [NP_CNJ 수신기,] [NP_SBJ [AP 그리고] [NP_SBJ 중계기가]]]]

《주의》 비록 “명사+명사” 구성이라 할지라도 수식 구성이 아닌 접속 구성일 경우 위와 달리 “_CNJ”라는 기능 표지를 붙인다.

예) 나는 [NP_OBJ [NP_CNJ 오이,] [NP_OBJ [NP_CNJ 당근,] [NP_OBJ 토마토를]]] 제배한다.

예) 나는 [NP_OBJ [NP_CNJ 오이] [NP_OBJ [NP_CNJ 당근] [NP_OBJ 토마토를]]] 제배한다.

《주의》 비록 “NP+나 NP+나” 구성에서 두 번째 “NP+나”는 “NP_CNJ”로 분석하지 않는다.

예) [NP_SBJ [NP_CNJ 너나] [NP_SBJ 나나]] 모두 골칫덩어리들이야.

☞ 이 경우 두 번째 “나”는 “JX”이기 때문이다.

《주의》 Upenn TreeBank에서는 문장 중간에 들어가는 구두점이 별도의 어절을 이루

면서 분석되지만(그러나 별도의 구문 표지나 기능 표지는 가지지 않음), 본 구문 분석에서는 인접한 선행어의 일부로 분석하는 것을 원칙으로 한다.

예) Upenn TreeBank : [오이/NNG] [/SP] [당근/NNG] [/SP] [토마토/NNG + 를/JKO] 재배한다.

예) 본 구문 분석 : [[오이/NNG + ,/SP] [[당근/NNG + ,/SP] [토마토/NNG + 를/JKO]]] 재배한다.

《주의》 동일 형식의 절 접속의 경우 별도의 접속표지 없이 분석한다.

예) 속임 없이 말한다면 그때 현우는 음악이 무엇인지 또 음악가라는 것이 무엇인지 몰랐다.

⇒ 속임 없이 말한다면 그때 현우는 [S_OBJ [S_OBJ 음악이 무엇인지] [또 [S_OBJ 음악가라는 것이 무엇인지]]] 몰랐다.

2.1.3 관형절 수식 구성

· 관형절은 체언을 수식해 주는 절이다.

예) 철수는 [S_MOD 내가 어제 산] 책을 반품해 버렸다.

《주의》 절(S)이 되려면 주어가 구비되어 있어야 한다. 따라서 다른 성분이 다 잘 갖추어 있다 하더라도 주어가 생략되어 있는 것은 S가 아니라 VP인 것으로 분석한다.

⇒ [VP_MOD 어디서 많이 본] 사람이 다가왔다.

2.1.4 명사 전성 구문(동명사 구문)

· 명사 전성은 용언으로 하여금 서술의 기능뿐만 아니라 체언의 구실도 겸할 수 있도록 해 주는 기제이다. 따라서 명사 전성을 겪은 용언은 그 자체로서 다른 술어의 논항이 될 수 있으며 또한 그 자신이 다른 논항을 취하여 논항절을 형성할 수 있다.

예) 단독으로 논항 기능 수행 : 그는 [VP_SBJ 주저함] 없이 떠났다.

예) 관형어 수식의 경우 : 그는 [VP_SBJ [NP_MOD 한치의] [VP_SBJ 주저함]] 없이 떠났다.

예) 논항을 취해 논항절이 되는 경우 : 그는 [S_OBJ 철수가 떠났음을] 아직도 믿지 않고 있다.

2.1.5 (명사+명사 구성) + (명사 전성 구문)

· “-기 때문에”와 같은 구성은 2.1.1의 “명사+명사 구성”과 2.1.4의 “명사 전성 구문”의 특성을 모두 가지고 있다. 따라서 VP 단독으로 후행 명사를 수식할 경우는 2.1.1을 따라 VP에 별도의 기능 표지를 주지 않으며, VP가 주어를 가질 경우 S로 분석한다.

예) 철수는 [NP_AJT [VP 아프기] 때문에] 결근하였다.

예) 철수는 [NP_AJT [VP 어제부터 매우 아팠기] 때문에] 결근하였다.

예) 비단순 구성: [NP_AJT [S 그가 [VP 배신했기]] 때문에] 그녀가 떠났다.

2.1.6 동일 기능 명사구 반복 구조

· 동일한 문법 기능을 하는 명사구가 반복되었으나 선행 명사구에 격이 표시되어 있지 않은 경우, 후행 명사구에 준하여 기능 표지를 붙여 준다.

예) [NP_AJT [NP_AJT 어느 지방,] [NP_AJT 어느 곳에서도]] 그런 일은 볼 수 없을 것이다.

2.1.7 동격 명사구 구성

· 동격의 명사구들은 우선적으로 그 둘을 묶어 준다. 이때 첫 번째 명사구는 아무런 기능 표지를 갖지 않는다.

예) 그런데 현우 넌 어디루 이사 갔니?

⇒ 그런데 [NP_SBJ [NP 현우] [NP_SBJ 넌]] 어디루 이사 갔니?

예) 그동안 나는 철수 그 사람을 죽 지켜보았습니다.

⇒ 그동안 나는 [NP_OBJ [NP 철수] [NP_OBJ 그 사람을]] 죽 지켜보았습니다.

《주의1》 방금 살펴본 예들에서나 아래에 살펴볼 예들이나 모두 일종의 제시어들이라 할 수 있다. 그런데, 특히 다음의 예들은 독립어로 보기 쉽다.

예) 마을에서 약간 떨어진 외딴집, 그것이 현우 집 식구가 새로 든 집이었다.

이 예에서 밑줄 친 명사구 구성은 분명히 ‘그것이’와 관련된 제시어로 볼 수 있다. 그러나 이 명사구 구성 또한 앞의 두 예들과 구성 양상이 동일하므로 역시 동격 명사구로 분석해 준다.

⇒ [NP_SBJ [NP 마을에서 약간 떨어진 외딴집,] [NP_SBJ 그것이]] 현우 집 식구가 새로 든 집이었다.

《주의2》 그러나 동격 구성의 사이에 다른 요소가 끼어들면 앞의 요소를 독립어로 분석한다.

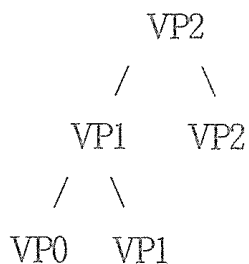
예) 솔직히 말해서, [NP_INT 이왕 당할 것] 어서 [선생님의 꾸지람이] 있었으면 했다.

2.2 동사구

2.2.1 본용언 + 보조용언 구성

· 순수 언어학적인 입장에서는 보조용언이 최종 VP를 관할하는 것이 합당하겠지만, 자연언어처리적 관점에서는 VP들을 먼저 묶어 주는 것이 더욱 효율적이므로 후자의 방식을 좇아 본용언과 보조용언을 먼저 묶는다.

예) [VP [VP 본용언 [VP 보조용언1]] [VP 보조용언2]]



예) 철수가 집에 [VP [VP [VP 가지] [VP 앓고]] [VP 있다]].

《주의》 다음과 같이 “본용언 + 보조용언” 구성에서 “본용언”에 {가/를}이 붙는 경우가 있다.

예) 나는 그때 음식을 먹어를 보았다.
그때 나는 그렇게를 하지 못했다.

철수가 밥을 이틀내 먹지를 않았다.

철수는 그 밥을 먹고를 싫었다.

그 산은 그리 높지가 않다.

열차가 움직이지가 않는다.

이 예들은 공히 “본용언 + 보조용언” 구성에서 “본용언”에 화용적 기능을 가지는 조사 {가/를}이 붙은 것으로서, 만일 “본용언 어간 + {가/를}”을 논항처럼 취급하면, 보조용언이 본용언을 자신의 논항으로 취한다는 이상한 결론에 이른다. 따라서 “본용언 어간 + {가/를}”은 그냥 “VP”로만 분석한다.

2.2.2 본용언 + 본용언 구성

· 복합된 두 용언이 합쳐져 연속동사를 이루는 구성으로서, 이 둘을 합친 전체 구성은 두 가지로 구별하여 분석해 주어야 한다.

1) 복합된 두 용언이 명사구를 수식할 경우 :

예) 우리들은 [VP_MOD [VP 먹고] [VP_MOD 마시던]] 것을 모두 그만 두었다.

예) [VP_MOD [VNP 여자라면] [VP_MOD 몸서리치던]] 그가 일주일 뒤에 결혼을 한다고 했다.

2) 접속된 두 용언이 서술어 구실을 할 경우 :

예) 그들은 휴일 내내 [VP [VP 먹고] [VP 마셨다.]]

2.2.3 의사 보조용언 구성(pseudo-auxiliary verb constructions)

· 주서술어 뒤에, 보조용언은 아니지만 서법(mood, modal)을 나타내는 언어 단위들이 올 경우 이들을 먼저 한데 묶는 것을 원칙으로 한다. 이는 앞서 “본용언”과 “보조용언”을 먼저 묶는 것과 동개의 처리 방식이다.

《주의》 이와 같은 의존 구성을 처리하는 데에는 세 가지 방식을 생각해 볼 수 있다.

예) 그는 밥을 곧 먹을 것이다.

<1안> : [VNP [S 그는 밥을 곧 먹을] 것이다.]

<2안> : [S 그는 밥을 곧 [VNP [VP_MOD 먹을] 것이다.]] => 기존의 방식

<3안> : [S 그는 [VNP [VP 밥을 곧 먹을] 것이다.] => Upenn의 방식
즉, 위의 예를 구문 분석할 때 위와 같이 세 가지 방식이 제안될 수 있다.
<1안>은 문면에 나타난 그대로 “관형절+것이다”로 분석하는 것이고, <2안>은 서술어를 먼저 묶고 나서 명제를 구성하는 논항들이나 부사어들을 묶어 나가는 방식이다. 반면 <3안>은 주어를 제외한 나머지 서술부에서 다시 ‘것이다’를 제외한 나머지 성분을 관형구로 묶고 그것이 뒤의 ‘것이다’를 수식해 준다는 분석으로서 Upenn TreeBank에서 취한 입장이다. 본 분석에서는 <2안>을 따르기로 한다. 그 이유는, 이 방식이 앞서 “본용언”과 “보조용언”을 먼저 묶는 것과 동계의 처리 방식일 뿐만 아니라, 문장을 구성하는 서술어가 본질적으로 “술어+서법/양태”으로 구성된다는 점에서, 그리고 구문 분석에서의 비중의적인(따라서 효율적인) 처리안이라는 점에서 합리적인 것으로 판단 되기 때문이다.

1) “-는 {모양/노릇/지경}이다”의 구성

예) 철수가 고향이 그리운 모양이다.

예) 우리가 매일 놀 수만은 없는 노릇이다.

예) 철수는 정말 미칠 지경이었다.

⇒ 철수가 [S 고향이 [VNP [VP_MOD 그리운] [VNP 모양이다.]]

⇒ 우리가 매일 [VNP [S_MOD [NP_SBJ [VP_MOD 놀] 수만은] 없는] 노릇이다.]

⇒ 철수는 [VNP 정말 [VNP [VP_MOD 미칠] [VNP 지경이다.]]

《주의》 단, “무렵이다, 때문이다, 셈/적이다”의 경우는 이에 해당하지 않는다.

예1) 지금은 여름이 끝나갈 무렵이다.

⇒ [S [NP_SBJ 지금은] [VNP [S_MOD 여름이 끝나갈] 무렵이다.]]

예2-1) 그것은 철수가 배고프기 때문이었다.

⇒ [NP_SBJ 그것은] [VNP [S 철수가 배고프기] 때문이었다.]

예2-2) 철수가 어제 간 때문이었다.

⇒ [VNP [S_MOD 철수가 어제 간] 때문이었다.]

예3) 우리가 이긴 셈이다.

⇒ [VNP [S_MOD 우리가 이긴] 셈/적이다.]

2) “-ㄴ + 의존명사 + 이다” 구성

예) 할 {것/뿐/터/따름}이다

예) 철수가 밥을 곧 먹을 것이다.

예) 한국이면 어디서라도 있음직한 고향일 뿐이다.

⇒ [VNP [VP_MOD 할] [VNP 것/뿐/터/따름이다.]]

⇒ 철수가 밥을 곧 [VNP [VP_MOD 먹을] [VNP 것이다.]]

⇒ [VP_MOD 한국이면 어디서라도 있음직한] [VNP [VNP_MOD 고향일] 뿐이다.]

《주의》 단, “뿐” 뒤에 계사가 오지 않을 경우에는 다음과 같이 분석한다.

예) 시간이 지날수록 정이 깊어 갈 뿐 불협화음이 생기지 않았다.

⇒ ... [NP_AJT [S_MOD 정이 깊어 갈] 뿐] ...

3) “-ㄴ 것 같다” 구성

예) 비가 곧 올 것 같다.

⇒ 비가 곧 [VP [NP [VP_MOD 올] 것] 같다.]

예) 어디선지 가까운 자리에서 아버지가 만족스레 웃음을 짓고 계시는 것 같았다.

⇒ ... 아버지가 만족스레 [VP 웃음을 [VP [NP [VP_MOD 짓고 계시는] 것] 같았다.]

4) “-기로 하다”의 구성

예) 그가 포기하기로 하였다.

예) 그가 나와 같이 가기로 하였다.

⇒ 그가 [VP [VP_AJT 포기하기로] 하였다.]

⇒ 그가 나와 같이 [VP [VP_AJT 가/VV + 기/ETN + 로/JKB] 하였다.]

☞ /Upenn TreeBank 상동/

5) “-기{도/만/는} 하다”의 구성

예) 그들은 서로 다르기도 하다.

⇒ 그들은 서로 [VP [VP 다르기도] 하다.]

⇒ 철수는 하루 종일 음악을 [VP [VP 듣기만] 하였다.]

⇒ 장남감 자동차가 [VP [VP 움직이기는] 했다.]

☞ /Upenn 상동/

6) “-기 시작하다”의 구성: 이 경우는 “시작하다”가 기동상을 나타냄.

예) 철수가 밥을 먹기 시작하였다.

⇒ 철수가 밥을 [VP [VP_OBJ 먹기] 시작하였다.]

《주의》 이와 같이 의사보조용언 구성에서 목적격 조사를 필히 붙여야 하는 경우는 다음과 같다.

~기 시작하-/ ~기 위하-/ ~ 말고/ ~ 모르다

예) 날이 [VP [VP_OBJ 어두워지기] 시작했다.]

예) [VP [VP_OBJ 먹기] 위하여] 산다.

예) [VP [NP_OBJ 너] 말/VV+고/EC] 철수 데려 와.

예) 그가 집에 [VP [VP_OBJ 같지도] 모른다.]

7) “-든지 하다” 구성

예) 철수는 빵을 먹든지 할 것이다.

예) 철수는 밥을 먹든지 빵을 먹든지 할 것이다.

⇒ 철수는 [VNP 빵을 [VNP [VP_MOD [VP 먹든지] 할] 것이다.]]

⇒ 철수는 [VNP [VP 밥을 먹든지] [VNP 빵을 [VNP [VP_MOD [VP 먹든지] 할] 것이다.]]]

8) “-어서는 안되다” 구성

예) 우리는 경계를 늦추어서는 안된다.

⇒ 우리는 경계를 [VP [VP 늦추어서는] 안된다.]

9) “-ㄴ {만/뻬}하다” 구성

예) 그 영화는 청소년들도 볼 만하다.

⇒ 그 영화는 [S 청소년들도 [VP [VP_MOD 볼] 만하다.]]

10) “-고 해서” 구성

예) 시간도 없고 해서 그는 친적집에 들르지 않았다.

⇒ [S 시간도 [VP [VP 없고] 해서]] 그는 친적집에 들르지 않았다.

☞ /Upenn TreeBank 상동/

11) “-ㄴ/ㄷ지도 모르다” 구성

예) 그가 집에 갈지도 모른다.

⇒ 그가 집에 [VP [VP_OBJ 갈지도] 모른다.]

12) “-ㄴ/ㄷ지도 모르는 일이다” 구성

예) 앞으로 어떤 벌이 내릴는지 모를 일이다.

⇒ 앞으로 어떤 벌이 [VNP [VP [VP_OBJ 내릴는지] 모를] 일이다.]

13) “-ㄴ 예정이다” 구성

예) 그가 오늘 도착할 예정이다.

⇒ 그가 오늘 [VNP [VP_MOD 도착할] 예정이다.]

14) “-ㄴ/ㄷ 것이다/냐” 구성

예) 너는 장차 무엇을 할 것이냐?

예) 그래서 내가 그렇게 한 것이다.

⇒ 너는 장차 무엇을 [VNP [VP_MOD 할] 것이냐?]

⇒ 그래서 내가 그렇게 [VNP [VP_MOD 한] 것이다.]

15) “-ㄴ/ㄷ 듯(도) 하다” 구성

예) 따스한 손길로 소리 없이 머리를 어루만져 주시는 듯도 했다.

⇒ 따스한 손길로 소리 없이 머리를 [VP [NP [VP_MOD 어루만져 주시는] 듯도] 했다.]

16) “-ㄴ 짓을(걸) 그랬다” 구성

예) 거기 나가서 동창들한테 협찬금 받아낼 걸 그랬잖아.

⇒ 거기 나가서 동창들한테 [NP_OBJ 협찬금 [VP [NP_OBJ 받아낼 걸] 그랬잖아.]]

2.2.4 부사절 수식 구성

· 부사절은 용언이나 문장 전체를 수식해 주는 기능을 한다.

예) 철수는 [VP [S 다람쥐가 나무를 오르듯] 나무를 잘 탔다.]

예) [S [S 내가 그 문제를 풀었듯이] [S 철수도 그 문제를 해결하였다.]]

2.2.5 경동사 구성(Light Verb constructions)

1) 서술성 명사(Verbal Noun)가 중첩된 경우 : 경동사 구성에서 서술성 명사가 중첩될 경우 그것을 VP로 분석해 준다.

예) 곧장 달려가 적진을 [VP [NP 습격] [VP파괴한다.]]

예) 곧장 달려가 적진을 [VP [NP_OBJ [NP 습격] [NP_OBJ 파괴를]] 했다.]]

2) 의사 경동사 구성(pseudo-light verb construction)

· ‘필요로 하다’ 구성

예) 그는 우리의 도움을 필요로 했다.

⇒ 그는 우리의 도움을 [VP [NP_AJT 필요로] 했다.]

· ‘즈음 해서’류 구성

예) 광복절 즈음 해서 독립기념관을 찾았다.

⇒ [VP [NP 광복절 즈음] 해서] 독립기념관을 찾았다.

예) 한밤쯤 해서 아버지가 돌아오셨다.

⇒ [VP [NP 한밤쯤] 해서] 아버지가 돌아오셨다.

3) 경동사 생략 구성

· 경동사가 생략된 채 서술성 명사와 그에 딸린 논항들만 나온 경우

예) 신문사의 공격적 논지에 역효과 강조

⇒ [NP [NP_AJT 신문사의 공격적 논지에] [NP [NP 역효과] [NP 강조]]]

예) 신문사의 공격적 논지에 역효과를 강조

⇒ [VP [NP_AJT 신문사의 공격적 논지에] [VP [NP_OBJ 역효과를] [NP 강조]]]

2.3 부사구

2.3.1 부사구 연결 구성

· 비록 연결하지는 않았지만 복합 조사로 알려진 “~부터 - ~까지”를 먼저 묶는다.

☞ <김석득(1992), 「우리말 형태론」, 탑출판사.> 참조.

· 해당 사례: NP에서 NP로, NP에서 NP까지, NP부터 NP까지, NP부터 NP로, NP에서 NP에로의, NP에서 NP에의

예) 그가 [NP_AJT [NP_AJT 작년부터] [NP_AJT 지금까지]] 열심히 일했다.

예) 그 전투는 [NP_AJT [NP [NP_AJT 4시에서] [NP 5시 반]] 사이에] 터졌을 것이다.

예) [NP_SBJ [NP_AJT 학교에서부터] 집까지가] 문제다.

2.3.2 논항을 취한 부사구

· ‘없이’나 ‘같이’와 같은 부사는 그 자체로 논항을 취한다. 명사의 경우도 마찬가지다. 이 경우 역시 다른 경우와 마찬가지로 처리한다. 즉, AP가 주어를 취하면 그 전체 범주는 S가 되고, AP가 목적어를 취하면 그 전체 범주는 VP가 되며, AP가 AJT를 취하면 그 전체 범주는 AP가 된다.

예) 철수는 [S [NP_SBJ 아무 생각도] [AP 없이]] 길을 나섰다.

예) 난 [AP [NP_AJT 너와] [AP 같이]] 살고 싶어.

예) [S [NP_SBJ 고등학생과 대학생은] [AP 물론]] 중학생까지 그 일에 참여했다.

예) [S [NP_SBJ 고등학생과 대학생은] [VNP 물론이고]] 중학생까지 그 일에 참여했다.

예) 철수는 [AP [NP_AJT 아버지] 몰래] 집을 빠져나갔다.

예) 누구도 [S [NP_SBJ 빠짐] 없이] 다 참여하였다.

《주의》 예컨대, 부사 ‘없이’ 앞에 동사구가 오고 그 앞에 다시 주어처럼 보이는 성분이 올 경우, 그 처리가 애매해진다. 이때 동사구와 그 앞의 주어처럼 보이는 성분을 먼저 묶는다. 단, 그 주어처럼 보이는 성분이 문장 전체의 주어 노릇을 하는 것이라면 동사구와 부사 ‘없이’를 묶어 준다.

예) O: 철수는 [S [S_SBJ 한 사람도 빠짐] 없이] 회비를 걷어 내었다.

예) X: 철수는 [S 한 사람도 [S 빠짐 없이] 회비를 걷어 내었다.

예) O: 누구도 [VP [S 주저함 없이] 그 일에 동참했다.]

예) X: [VP [S [S 누구도 주저함] 없이] 그 일에 동참했다.]

《주의》 명사구가 목적어를 취하는 것처럼 보이는 구문이 있으나, 이는 사실상 동사가 생략된 구문으로 보아야 하므로, 아래와 같이 그 전체 범주를 VP로 분석해 준다.

예) 우리는 [VP [NP_OBJ 성실과 근면을] [NP_AJT 기초로]] 열심히 공부하고 있다.

《주의》 “자기 스스로”와 같은 경우는 다음과 같이 분석한다.

예) 성인이 되면 자기/그 스스로 일을 처리해야 한다.

⇒ 성인이 되면 [AP [NP 자기/그] 스스로] 일을 처리해야 한다.

2.4 유동 양화사(floating quantifiers)

· 유동 양화사는 격 중출 구문과 상관이 있다. 이론 언어학적인 관점에서는 그 지위가 다소 모호할 수 있으나, 자연언어처리적 관점에서는 그들에 실현된 격표지의 성격을 고려하여 모두 논항으로 취급하기로 한다.

예) 주격인 경우: [S 나는 친구가 [S [NP_SBJ 두 명(이)] 있습니다.]]

예) 목적격인 경우: 나는 [VP 술을 [VP [NP_OBJ 두 병(을)] 마셨다.]]

3. 절의 세부 구조

3.1 논항절의 기능 표지 부착 시점

· 논항절의 경우 해당 전성어미를 가진 서술어로부터 기능 표지를 붙여 올라간다.

예) 철수가 [S_OBJ 이번 시험에 [S_OBJ 자기가 [VNP_OBJ 일등이기를]]] 바랬다.

· 논항절이 다시 그 내부에 서술절을 포함할 경우에도 마찬가지이다.

예) 뉴스에서 [S_CMP 강이 [S_CMP 얼음이 [VP_CMP 얼었다고]]] 보도하였다.

3.2 접속의 층위 문제

· 문접속을 인정하지 않기 때문에 절 접속의 층위 문제는 발생하지 않는다. 다만, 순수 언어학적인 입장에서 문제가 될만한 예들을 통해, 본 구문 분석의 분석 방식을 보이면 다음과 같다.

1) 주어를 공유하는 VP접속의 경우 :

예) [S 철수가 [VP [VP 밥을 먹고] [VP 물을 마셨다.]]]

2) 주어를 공유할지라도 접속 성분 중 어느 하나만이 S인 경우 :

예) [S 나는 [VP [S 호랑이가 무서웠지만] [VP 아들을 위하여 호랑이 굴에 뛰어 들었다.]]]

3.3 계사 구문

· 계사 구문에서 서술어는 ‘체언+이다’이다. 따라서 그것은 관형어 수식과 부사어 수식을 모두 허용한다. 그 각각의 경우에 있어 수식어의 기능 표지는 “_MOD”와 “_AJT”로 구별하여 분석한다.

1) 부사어 수식 :

예) 내가 붓글씨에 손방이다.

(S (NP_SBJ 내/NP + 가/JKS)

(VNP (NP_AJT 붓글씨/NNG + 에/JKB)

(VNP 손방/NNG + 이/VCP + 다/EF + ./SF)))

예) 그는 매우 부자이다.

(S (NP_SBJ 그/NP + 는/JX)

(VNP (AP 매우)

(VNP 부자/NNG + 이/VCP + 다/EF + ./SF)))

2) 관형어 수식 :

예) 너는 나의 아들이다.

(S (NP_SBJ 너/NP + 는/JX)
 (VNP (NP_MOD 나/NP + 의/JKG)
 (VNP 아들/NNG + 이/VCP + 다/EF + ./SF)))

예) 그가 착한 학생이다.

(S (NP_SBJ 그/NP + 가/JKS)
 (VNP (VP_MOD 착하/VA + ㄴ/ETM)
 (VNP 학생/NNG + 이/VCP + 다/EF + ./SF)))

3) 계사 생략 :

<단순문>에서

예) 지나고 보면 모두가 정든 사람들뿐!

(S (VP (VP 지나/VV + 고/EC)
 (VP 보/VX + 면/EC))
 (S (NP_SBJ 모두/NNG + 가/JKS)
 (NP (VP_MOD 정들/VV + ㄴ/ETM)
 (NP 사람/NNG + 들/XSN + 뿐/JX + !/SF))))

<복합문>에서

예) 나는 학생, 너는 회사원이다.

⇒ [S [S [NP_SBJ 나는] [NP 학생,]] [S 너는 회사원이다.]]

예) 1천 석 가까운 추수가 있는 외에 자작만으로도 논밭 70여 마지기를 짓는 대농가
 이니만큼, 머슴이 네댓, 계집종이 두셋, 집안은 노상 잔칫집같이 법석거리는 터이었다.

(S (VP (VP_MOD (NP_AJT (S_MOD (NP_SBJ (VP_MOD (NP_AJT (NP I/SN + 천/NR)
 (NP_AJT 석/NNB)
 (VP_MOD 가깝/VA + ㄴ/ETM))
 (NP_SBJ 추수/ANG + 가/JKS))
 (VP_MOD 있/VV + 는/ETM))
 (NP_AJT 외/ANG + 예/JKB))
 (VP_MOD (NP_AJT 자작/ANG + 만/JX + 으로/JKB + 도/JX)
 (VP_MOD (NP_OBJ (NP 논/ANG + 밭/ANG)
 (NP_OBJ (NP 70/SN + 여/XSN)
 (NP_OBJ 마지기/NNB + 를/JKO)))
 (VP_MOD 짓/VV + 는/ETM))))
 (VNP 대/XPN + 농가/ANG + 이/VCP + 니/EC + 만 큼/JKB + ./SP))

(S (S (S (NP_SBJ 미슴/NNG + 이/JKS)
(NP 네댓/NR + /SF))
(S (NP_SBJ 재깍/NNG + 중/NNG + 이/JKS)
(NP 두셋/NR + /SF)))
(S (NP_SBJ 집안/NNG + 은/JX)
(VP (AP 노상/MAG)
(VP (NP_AJT 잔칫집/NNG + 같이/JKB)
(VP (VP_MOD 법석거리/VV + 는/ETM)
(VNP 터/NNB + 이/VCP + 었/EP + 다/EF +
/SF))))))

3.4 주격 중출 구문

- 서술절 가설을 그대로 인정하여 분석한다.
- 보어를 엄격히 제한하였으므로 주격 보어를 가지는 2항 이상의 비대격 동사의 경우 모두 다중 주격 중출 구문으로 분석한다.

예) 그는 호랑이가 무섭다.

(S (NP_SBJ 그/NP + 는/JX)
(S (NP_SBJ 호랑이/NNG + 가/JKS)
(VP 무섭/VA + 다/EF + /SF)))

예) [S [NP_SBJ 철수가] [S [NP_SBJ 나이가] [S [NP_SBJ 한 살] 늘었다.]]

예) [S [NP_SBJ 철수가] [S [NP_SBJ 한 살] [S [NP_SBJ 나이가] 늘었다.]]

3.5 목적격 중출 구문

- 목적격을 가지는 NP들은 모두 목적어로 분석되며, 목적어를 포함한 [VP]는 다시 [VP]로 투사한다.

예) 나는 [VP [NP_OBJ 영화를] [VP [NP_OBJ 책을] 주었다.]]

예) 그는 [VP [NP_OBJ 영화를] [VP [NP_OBJ 뺨을] 때렸다.]]

예) 어느덧 그는 [VP [NP_OBJ 나이를] [VP [NP_OBJ 한두 살] 먹어 갔다.]]

예) 어느덧 그는 [VP [VP [NP_OBJ [NP 한두 살] 나이를] 먹어 갔다.]]

3.6 주어 인상 구문(ECM 구문)

· 이론 언어학적인 입장에서는 내포문의 주어가 주절의 목적어로 인상한 것이라고 분석하기도 한다. 공범주를 인정하지 않는 본 구문 분석에서는 흔적이나 공범주 주어 등을 복원해 주지 않는다.

· 기본 골격:

[S 모문 주어 [[S_CMP 내포문 주어 [VP_CMP 내포문 술어]] 모문 술어]]

[S 모문 주어 [NP_OBJ 모문 목적어] [[VP_CMP 내포문 술어] 모문 술어]]

[S 모문 주어 [[VP_CMP 내포문 술어] 모문 술어]]

1) 내포문이 계사일 경우 :

예) 그는 철수가 천재라고 생각한다.

⇒ 그는 [VP [S_CMP 철수가 [VNP_CMP 천재라고]] 생각한다.]

예) 그는 철수를 천재라고 생각한다.

⇒ 그는 [VP [NP_OBJ 철수를] [VP [VNP_CMP 천재라고] 생각한다.]]

예) 그는 천재라고 생각했다. (“천재”가 “그”는 아님)

⇒ 그는 [VP [VNP_CMP 천재라고] 생각한다.]

2) 내포문이 일반 용언일 경우:

예) 그는 영희가 매우 예쁘다고 생각한다.

⇒ 그는 [VP [S_CMP 영희가 매우 예쁘다고] 생각한다.]

예) 그는 영희를 매우 예쁘다고 생각한다.

⇒ 그는 [VP [NP_OBJ 영희를] [VP [VP_CMP 매우 예쁘다고] 생각한다.]]

예) 나는 잘못되었다고 생각했다.

⇒ 나는 [VP [VP_CMP 잘못되었다고] 생각한다.]

3) ‘-고 하다’의 처리

예) 그가 간다고 합니다.

⇒ 그가 [VP [VP_CMP 간다고] 합니다.]

예) 그가 학교에 간다고 합니다.

⇒ 그가 [VP [VP_CMP 학교에 간다고] 합니다.]

예) 그가 밥을 먹는다고 한다.

⇒ 그가 [VP [VP_CMP 밥을 먹는다고] 합니다.]

예) 그는 영희가 그 일을 했다고 합니다.

⇒ 그는 [VP [S_CMP 영희가 그 일을 했다고] 합니다.]

《주의》 어떤 NP가 중의성을 가져 내포문이나 모문의 주어 모두로 인식될 수 있을 경우에는 그것을 모문의 주어로 분석한다.

예) 영희가 간다고 합니다.

⇒ 영희가 [VP [VP_CMP 간다고] 합니다.]

《주의》 ‘전해지다’와 같은 경우 역시 다음과 같이 분석한다.

예) 적들이 섬멸되었다고 전해졌다.

⇒ [S_CMP 적들이 섬멸되었다고] 전해졌다.

4) 완형 보문의 처리

· 인용 부호가 없는 경우:

예) 운동을 하면 추운 기가 조금이라도 가셔질까 해서였다.

⇒ [완형보문 운동을 하면 추운 기가 조금이라도 가셔질까] 해서였다.

(VNP (S_CMP (VP (NP_OBJ 운동/NNG + 을/JKO)
(VP 하/VV + 면/EC))
(S_CMP (NP_SBJ (VP_MOD 춥/VA + ㄴ/ETM)
(NP_SBJ 기/NNG + 가/JKS))
(VP_CMP (NP_AJT 조금/NNG + 이라도/JX)
(VP_CMP 가시/VV + 어/EC + 지/VX + ㄴ까/EC))))
(VNP 하/VV + 아서/EC + 이/VCP + 었/EP + 다/EF + ㄴ/SF))

· 인용 부호가 있는 경우:

예) ‘나도 아버지가 있다’로서만이 아니라 ‘내게는 이러한 훌륭한 아버지가 계시다’ 하는 벽차오르는 자랑이었다.

.....

(S_CMP (L '/SS)
(S_CMP (S_CMP (NP_AJT + 나/NP + 예게/JKB + 는/JX)
(S_CMP (NP_SBJ (VP_MOD 이러하/VA + ㄴ/ETM)
(NP_SBJ (VP_MOD 훌륭/XR + 하/XSA + ㄴ

/ETM)

(NP_SBJ 아버지/NNG + 가/JKS)))

(VP_CMP 계시/VV + 다/EC)))

(R +'/SS))))

하/VV + 는/ETM))

.....

5) ‘-고 {야단/난리}이다’의 처리

· 이 구성은 ‘-고 야단을 하다’에 준하여 역시 보문절을 가진 복문으로 분석한다.

예) 그들은 추장이 죽었다고 야단들이다.

⇒ 그들은 [VNP [S_CMP 추장이 죽었다고] 야단들이다.]

6) 기능범주만 있고 구문범주가 없는 보문 구성

· 이 경우 보어 표지 “_CMP”는 인용격 조사(JKQ)인 ‘고’에게 붙여준다.

예) 마을 여자들도 일부러 현우 어머니를 찾아와서는 “왜 기득이 만나러 안 가보느냐?”고 궁금해했다.

⇒ ... [VP [VP_CMP [VP “ [VP 왜 기득이 만나러 안 가보느냐?”] [X_CMP 고 /JKQ]] 궁금해했다].

《주의》 인용의 “하/VV” 앞에 부사나 감탄사, 명사가 구두점을 가지며 인용 부호를 동반하더라도 결코 “_CMP”라는 기능표지를 주지 않는다. “_CMP”가 붙는 경우는, ‘되다, 아니다’ 앞의 NP이거나 보문동사 앞의 VP나 VNP, 그리고 S의 경우에만 한정된다.

· 부사의 경우

예1) 가마솔이 핑 하고 터졌다.

⇒ 가마솔이 [VP [AP 핑] 하고] 터졌다.

예2) 가마솔이 “핑” 하고 터졌다.

⇒ 가마솔이 [VP [AP “ [AP [AP 핑]]] 하고] 터졌다.

예3) 가마솔이 핑! 하고 터졌다.

⇒ 가마솔이 [VP [AP 핑!] 하고] 터졌다.

예4) 가마솔이 “핑!” 하고 터졌다.

⇒ 가마솥이 [VP [AP “ [AP 평!]”] 하고] 터졌다.

· 감탄사의 경우

예1) 아버지께서 아이고 하고 쓰러지셨다.

⇒ 아버지께서 [VP [IP 아이고/IC] 하고] 쓰러지셨다.

예2) 아버지께서 “아이고” 하고 쓰러지셨다.

⇒ 아버지께서 [VP [IP “ [IP [IP 아이고]]”] 하고] 쓰러지셨다.

예3) 아버지께서 아이고! 하고 쓰러지셨다.

⇒ 아버지께서 [VP [IP 아이고!] 하고] 쓰러지셨다.

예4) 아버지께서 “아이고!” 하고 쓰러지셨다.

⇒ 아버지께서 [VP [IP “ [IP 아이고!] ”] 하고] 쓰러지셨다.

· 명사의 경우

※문맥상 부름말일 경우에만 독립어 기능표지 “_INT”를 붙인다.

예1) 철수가 아버지 하고 불렀다.

⇒ 철수가 [VP [NP_INT 아버지] 하고] 불렀다.

예2) 철수가 아버지! 하고 불렀다.

⇒ 철수가 [VP [NP_INT 아버지!] 하고] 불렀다.

예3) 철수가 “아버지” 하고 불렀다.

⇒ 철수가 [VP [NP_INT “ [NP_INT [NP_INT 아버지] ”]] 하고] 불렀다.

예4) 철수가 “아버지!” 하고 불렀다.

⇒ 철수가 [VP [NP_INT “ [NP_INT [NP_INT 아버지!] ”]] 하고] 불렀다.

※문맥상 부름말이 아닐 경우에는 다음과 같이 “_INT”를 붙이지 않는다.

너 뭐 먹고 싶어?

예5) 철수는 “빵” 하고 대답했다.

⇒ 철수는 [VP [NP “빵”] 하고] 대답했다.

《주의》 내포되지 않은 단독 인용문에서 명사구가 나뉘는 표지를 가지고 나왔을 경우, 원래의 표지를 살려서 분석해 준다.

예1) “우리나라가”

⇒ [NP_SBJ “ [NP_SBJ [NP_SBJ 우리나라가] ”]]

예2) “하던…” => 형태분석: “/SS + 하/VV + 던/ETM + …/SE + ”/SS

⇒ [VP_MOD “ [VP_MOD [VP_MOD 하던…] ”]]

7) 보문동사의 어간이 생략되고 어미만 남아 있는 경우

· 이 경우 조사가 단독 어절을 이루는 경우와 마찬가지로 구문 표지 위치에 “X”를 붙여준다.

예) 그는 [VP [S_CMP “자식이 장원급제를 했다”] [X 며/EC]] 기뻐했다.

《주의》 보문동사라 할지라도 그 앞에 절이 아닌 명사구가 올 경우에는 본 지침의 원칙대로 “_CMP”를 붙이지 않고 그대로 놓아둔다. 이때 주의할 것은, 아래와 같이 “라고/EC”가 단독으로 올 경우 그것을 위의 “며/EC”의 경우와 평행하게 [X 라고/EC], [X 라며/EC]로 처리해 준다.

예) 그는 “가족 때문에 고생이 많다는 의미”라고/며 말했다.

⇒ 그는 [NP_CMP [NP “가족 때문에 고생이 많다는 의미”] [X 라고/JKQ]] 말했다.

⇒ 그는 [VNP [NP “가족 때문에 고생이 많다는 의미”] [X 라며/EC]] 말했다.

8) “하/VV + 고/EC”와 “하고/JKB”의 처리 문제

· 형태분석 오류로 인해 문제가 발생하였을 경우

예) 어느 아이인가 하고 술래잡기를 하였다.

⇒ 잘못된 형태분석:

어느/MM

아이/NNG + 이/VCP + ㄴ가/EC

하/VV + 고/EC

...

⇒ 형태분석 교정:

어느/MM

아이/NNG + 이/VCP + ㄴ가/EC

하고/JKB

...

⇒ 구문분석:

[VNP_AJT [VNP 어느 아이인가] [X_AJT 하고/JKB]] ...

3.7 목적어 인상 구문(Tough 구문)

- 이동한 성분이 남긴 흔적이나 생략된 일반 주어(arbitrary PRO)는 복원하지 않는다.

예) [VP_SBJ 영어를 매일 공부하기가] 쉽지 않다.

예) [S [S_SBJ 영어가] [S [VP_SBJ 매일 공부하기가] 쉽지 않다.]]

예) [S 나는 [S 영어가 [S [VP_SBJ 매일 공부하기가] 어렵다.]]]

3.8 장형 사동 구문

- 장형 사동 구문에서의 내포절은 부가어절로 처리하며, 본 분석은 공범주를 인정하지 않기 때문에 모문으로 이동한 성분의 흔적이나 모문의 주어와 구별되는 생략된 내포문 주어를 복원하지 않는다.

예) 그는 철수가 집에 가게 하였다.

⇒ 그는 [VP [S 철수가 집에 가게] 하였다.]

예) 그는 철수를 집에 가게 하였다.

⇒ 그는 [VP [NP_OBJ 철수를] [VP [VP 집에 가게] 하였다.]]

예) 그는 죽게 하였다.

⇒ 그는 [VP [VP 죽게] 하였다.]

3.9 도치 구문(어순재배치 구문)

- 이동한 성분이 남긴 흔적을 별도로 표시하지 않는다. 다만, 모든 경우에 있어 VP가 주어를 만나면 S를 이룬다는 원칙이 있기 때문에, 비록 “흔적”을 위한 공범주를 따로 설정하지 않고서도 구문의 어순재배치 정보를 표시해 줄 수 있다.

예) [S [NP_OBJ 밥을] [S [NP_SBJ 철수는] 맛있게 먹었다.]

- ☞ 즉, 목적어가 문두로 어순재배치된 위와 같은 문장의 정보는 “S → NP_OBJ S”와 같은 다시쓰기 규칙으로 포착해 줄 수 있다.

예) "뭘 하는 거냐, 야단스럽게."

(VNP (L "/SS)
 (VNP (VNP (VNP (NP_OBJ 무엇/NP + ㄹ/JKO)
 (VNP (VP_MOD 하/VV + 는/ETM)
 (VNP 거/NNB + 이/VCP + 냐/EC + ./SP)))
 (VP 야단/NNG + 스럽/XSA + 게/EC + ./SF))
 (R "/SS)))

3.10 삽입 구문(parenthetical constructions)

· 삽입 구문은 “_PRN”으로 분석하고 부가어처럼 처리한다. 삽입구문이 문장의 맨 앞에 위치할 때에는 별 문제가 없지만, 문장의 중간에 끼일 경우에 문제가 된다. 이때 삽입구문은 보통 선행 어절을 부연해 주는 기능을 하므로 선행 어절과 묶어 주고 그 둘을 관할하는 상위 마디의 구문/기능 표지는 선행 어절의 그것을 따른다.

예) (어둠 속에서) 나 좀 봐요!

(VP (NP_PRN (L (/SS)
 (NP_PRN (NP_AJT (NP + 어둠/NNG)
 (NP_AJT 속/NP + 에서/JKB))
 (R_PRN +)/SS)))
 (VP (NP_OBJ 나/NP)
 (VP (AP 좀/MAG)
 (VP 보/VV + 아요/EF + ./SF))))

예) 자, (차를 가져 오면서) 이것 좀 드세요.

(VP (IP 자/IC + ./SP)
 (VP (VP_PRN (L (/SS)
 (VP_PRN (VP (NP_OBJ + 차/NNG + 를/JKO)
 (VP (VP 가지/VV + 어/EC)
 (VP 오/VV + 면서/EC)))
 (R_PRN +)/SS)))
 (VP (NP_OBJ 이것/NP)
 (VP (AP 좀/MAG)
 (VP 들/VV + 드세요/EF + ./SF))))

예) 상부에서 지도원 (소위) 1명을 내려보냈다.

(VP (NP_AJT 상부/NNG + 에서/JKB)
 (VP (NP_OBJ (NP (NP 지도원/NNG)
 (NP_PRN (/SS + 소위/NNG +)/SS))
 (NP_OBJ 1/SN + 명/NNB + 을/JKO))
 (VP 내려보내/VV + 았/EP + 다/EF + .SF))))

《주의》 명사와 조사 사이에 삽입어구가 들어갔을 경우 아래와 같이 주의해서 분석한다.

예) 그 남자(어제 우리 집까지 따라왔던)가 또 왔다.

⇒ 먼저 한 어절로 된 “남자(어제)”를 “남자V(어제)”와 같이 두 어절로 분리한다. 그런 다음 구문분석을 가한다.

...

(NP_SBJ (NP (NP (DP 그/MM)
 (NP 남자/NNG))
 (VP_PRN (L +(/SS)
 (VP_PRN (VP_MOD (AP 어제/MAG)
 (VP_MOD (NP_AJT (NP 우리/NP)
 (NP_AJT 집/NNG + 까지/JX))
 (VP_MOD 따라오/VV + 았/EP + 던
 /ETM)))
 (R_PRN +)/SS))))
 (X_SBJ 가/JKS))

...

《주의》 다음과 같은 명시적인 부연 어구는 삽입구문으로 분석하지 않는다.

예) 만일 적들이 아군의 진지를, 예컨대 두 군데나 세 군데를, 동시에 침투해 들어오면 어떻게 하는가?

⇒ ... [NP_OBJ 아군의 진지를, [NP_OBJ 예컨대 두 군데나 세 군데를,]] ...

3.11 직접 인용문으로서의 단순문

· 직접 인용문으로서의 단순문은 인용 부호를 동반하는 것이 보통이다. 이 때 각각의 문장 부호는 모두 독립적인 교점을 부여받으며 다음과 같은 순서로 분석된다.

예) "어머니, 추워요."

[" [[어머니, 추워요.] "]]

(VP (L "/SS)

(VP (VP (NP_INT + 어머니/NNG + ./SP)

(VP 춥/VA + 어요/EF + ./SF))

(R +"/SS)))

예) "요놈의 닭!"

(NP (L "/SS)

(NP (NP (NP_MOD + 요놈/NP + 의/JKG)

(NP 닭/NNG + !/SF))

(R +"/SS)))

3.12 VP가 아닌 서술어를 가진 문장 구성

· 다음과 같이, 하나의 절이지만 그 서술어는 VP가 아닌 경우가 있다. 부사구가 서술어인 경우에는 부사의 특성상 주어와 부사가 결합한 상위 범주를 그냥 "AP"로만 분석하였다. 그러나 본질의 경우에는 그 전체 범주를 "S"로 분석해 준다.

예) [S [NP_SBJ 종수란] [NP_AJT 금년 봄에 같이 졸업을 한 한 반 아이로서,]] 나는 그와 둘도 없는 친구이다.

4. 공범주 문제

4.1 기본 입장

· 고려될 수 있는 공범주의 종류:

- 1) 이동된 성분의 '흔적(trace)'
- 2) 생략된 주어나 목적어를 위한 것(pro)
- 3) 생략된 동사나 동사구를 위한 것

4) 공운용소(null operator)에 해당하는 것

· 구문 분석에서의 입장: 이론 언어학적인 입장(예컨대, 지배·결속 이론)에서는 이들을 포착해 주기 위해 다양한 공범주(null element)들을 인정하고 그것을 표시해 주기도 하지만 이론에 따라서는 반드시 그런 것은 아니다. 자연언어처리의 관점에서 볼 때 그러한 설정은 자동적인 언어 처리에 매우 큰 부담을 준다. 따라서 쉽게 규칙화거나 예측할 수 없는 그러한 공범주는 일절 인정하지 않기로 한다.

☞ /Upenn TreeBank/에서는 4가지 경우 모두 공범주를 인정함.

4.2 관련 문제와 분석 방안

4.2.1 성분 이동에 따른 흔적(trace)

- 이를 위한 공범주 설정은 하지 않는다.
- 논항 이동시 생기는 흔적: 이미 3.6절~3.9절에서 다룬 “주어 인상 구문, 목적어 인상 구문, 장형 사동 구문, 어순재배치 구문”의 처리를 참조.
- 부가어 이동시 생기는 흔적: 부가어의 기저 위치를 정확히 가려내는 것 자체가 순수 이론적으로도 매우 어려운 문제이므로 부가어가 나타난 그 위치를 기저 위치로 보아도 무방할 것이다.

4.2.2 탈락된 주어나 목적어(pro)

- 모두 고려하지 않는다.

1) 모문의 생략된 주어 :

예) 모문 : [VP 어서 가거라!]

2) 관형절에서 피수식명사(=표제명사)가 수식절의 비주어(목적어 등)일 경우 관형절에서 문맥상 생략된 주어

예) 관형절 : 그는 [NP_OBJ [VP_MOD 먹던] 사과를] 나에게 내던졌다.

3) 보어절(인용절)에서 생략된 주어

예) 인용절 : 그가 [VP [VNP_CMP 바보라고] 말했다.]

그가 [VP [OBJ 영희를] [VP [VNP_CMP 바보라고] 말했다.]]

《주의》 ‘전해지다’와 같은 경우 따로 모문의 주어를 상정할 수 없으므로 모문의 주어 없이 보문절만을 보충어로 취하는 것으로 분석한다.

예) 적들이 섬멸되었다고 전해졌다.

⇒ [VP [S_CMP 적들이 섬멸되었다고] 전해졌다.]

4) 주어를 달리하는 접속문들에서 분명히 맥락상에서 생략되었다고 보이는 주어

예) [VP [VP 오니까] [VP 갔다.]]

· 이 밖의 어떤 경우에서도 생략된 논항을 복원하지 않는다.

《주의》 일반적인 사람이나 사물을 나타내는 주어도 복원하지 않는다.

예) 영어는 공부하기가 어렵다.

⇒ [S 영어는 [VP [VP_SBJ 공부하기가] 어렵다.]]

4.2.3 공운용소(null operator)

· 고려하지 않는다.

4.2.4 계사 구문에서의 ‘이다’ 생략

· 서술성 명사를 NP로 분석한다.

예) 나는 학생, 너는 회사원이다.

⇒ [S [S [NP_SBJ 나는] [NP 학생,]] [S 너는 회사원이다.]]

4.2.5 V 혹은 VP 생략(V or VP deletion)

· 비록 (조)동사나 동사구가 생략되었다 하더라도 그것을 복원하지 않고 표면 그대로 분석한다.

1) 동사 생략 :

- 다른 성분은 남기고 동사만 생략된 경우

예) 철수는 밥을, 나는 떡을 먹었다.

⇒ [S [S [NP_SBJ 철수는] [NP_OBJ 밥을,]] [S 나는 떡을 먹었다.]]

2) 보조용언 생략 :

- 다른 성분은 남기고 보조용언만 생략된 경우

예) 철수는 밥을 먹고 나는 떡을 먹고 있다.

⇒ [S [S [NP_SBJ 철수는] [VP [VP_OBJ 밥을] [VP 먹고]]] [S 나는 떡을 먹고 있다.]]

3) 동사구 생략 :

- 다른 성분은 남기고 동사구만 생략된 경우

예) 우리는 서울에서, 그들은 부산에서 시합을 합니다.

⇒ [S [S [NP_SBJ 우리는] [NP_AJT 서울에서,]] [S 그들은 부산에서 시합을 합니다.]]

예) 철수는 병수에게, 영희는 순이에게 책을 주었다.

⇒ [S [S [NP_SBJ 철수는] [NP_AJT 병수에게,]] [S 영희는 순이에게 책을 주었다.]]

《주의》 동사 없이 나머지 문장 성분들만이 존재할 경우, 그 전체 범주를 무엇으로 분석해 줄 지가 문제가 되는 예들이 있다. 이러한 사례는 특히 신문에서 많이 발견된다. 이 경우 아래와 같이 분석해 준다. 문장 일부의 전체 범주 설정의 문제가 제기될 수 있다. 이 경우 아래와 같이 분석해 준다.

- ① 주어(_SBJ)가 있으면 그것을 포함한 전체 범주를 S로 분석한다.

예) [S [NP_SBJ 죄수 3명이] 탈옥]

- ② 주어가 없고 목적어(_OBJ)나 보어(_CMP)가 있으면 그것을 포함한 전체 범주를 VP로 분석한다.

예) [VP [NP_OBJ 최다 국내의 관광지과 관광업체들] [NP_AJT 한 자리에서]]

예) [VP 헌법재판소에서 [S_CMP 그것은 위헌이었다고] 판결]

- ③ 주어나 목적어가 모두 없고 부사어(_AJT)가 있으면 그것을 포함한

전체범주는 NP로 분석한다.

예) [NP [NP_AJT 매일] 민족 공연 등 직접 홍보]

cf) [VP [NP_AJT 매일] 민족 공연 등을 직접 홍보]

5. 기타 구문

5.1 특정 구문

5.1.1 “-기에” 구성

· “-기에”는 일종의 수식 성분으로 처리한다.

예) 듣기에 네가 일등이라고 하더라.

예) 그에게 듣기에 네가 일등이라고 하더라.

⇒ [VP_AJT 듣기에] 네가 일등이라고 하더라.

⇒ [VP_AJT 그에게 [VP_AJT 듣기에]] 네가 일등이라고 하더라.

5.1.2 “관형절 + 것은 아니다” 구성

예) 그가 잘못된 것은 아니다.

※이 구문의 기본(기저) 구조는 다음과 같을 것으로 보인다.

: [NP_SBJ 그가 잘못된 것은] [NP_CMP 사실이/진실이/참이] 아니다.

따라서 이를 토대로 위 문장은 다음과 같이 분석하기로 한다.

⇒ [NP_SBJ [S_MOD 그가 잘못된] 것은] 아니다.]

《주의》 ‘것은 아니다’ 앞에 복문이 개재하는 경우

위와 같은 구성에 복문이 개재하는 경우가 있다.

예) 나는 독재에 반대이지만 그렇다고 해서 방종을 원하는 것은 아니다.

이와 같은 경우 모문 주어 확보 우선의 원칙에 의거하여 다음과 같이 분석한다.

예) [S [S 나는 독재에 반대이지만] [S [NP_SBJ 그렇다고 해서 방종을 원하는 것은] 아니다.]]

5.1.3 “NP + 말고” 구성

예) 그것 말고 다른 것을 가져 오렴.

⇒ [VP [NP_OBJ 그것] [VP 말고]] 다른 것을 가져 오렴.

5.1.4 “-ㄴ 수/리(가) 있다/없다” 구성

· 통합형으로 우선 분석해 준다.

예) 철수가 일어날 수 있다.

⇒ [S 철수가 [S [NP_SBJ [VP_MOD 일어날] 수] 있다/없다.]]

5.1.5 “뿐만 아니라” 구성

예) 너뿐만 아니라 나도 괴롭다.

예) 많은 사람이 다쳤을 뿐만 아니라 죽기도 하였다.

⇒ [S [VP [NP_CMP 너뿐만] 아니라] [S 나도 괴롭다.]]

⇒ 많은 사람이 [VP [VP [NP_CMP [VP_MOD 다쳤을] 뿐만] 아니라] 죽기도 하였다.]

5.1.6 “-밖에 안 되다” 구성

예) 그 사람은 그릇이 그것밖에 안 된다.

예) 공부를 열심히 하였지만 철수는 7급 공무원밖에 안 되었다.

⇒ 그 사람은 그릇이 [NP_CMP 그것밖에] 안 된다.

⇒ 공부를 열심히 하였지만 철수는 [NP_CMP 7급 공무원밖에] 안 되었다.

5.1.7 “-만 해도” 구성

예) 그것만 해도 큰 성공이다.

⇒ [VNP [VP [NP_OBJ 그것만] 해도] 큰 성공이다.]

5.1.8 “얼마나 되다” 구성

예) 그 돈을 다 모으면 얼마나 되니?

예) 그 돈을 다 모으면 얼마가 되니?

⇒ 그 돈을 다 모으면 [AP 얼마나] 되니?

⇒ 그 돈을 다 모으면 [NP_CMP 얼마가] 되니?

5.1.9 “~는 말이다” 구성

예) 밥을 먹었던 말이나?

⇒ 밥을 [VNP 먹었던 말이나?]

5.1.10 “~는 {길/참}이다” 구성

예) 나는 외갓집에 심부름을 갔다 오는 길이야.

예) 그는 막 손가락을 놓으려던 참이었다.

⇒ 나는 외갓집에 심부름을 [VNP [VP_MOD 갔다 오는] 길이야.]

⇒ 그는 막 손가락을 [VNP 놓으려던 참이었다.]

5.2 부호 처리 문제

5.2.1 인용 부호와 괄호

· 인용 부호와 괄호는 공히 두 개의 기호가 한 쌍을 이룬다. 이때 여는 기호와 닫는 기호를 구별하여 여는 기호에는 L을, 닫는 기호에는 R을 써 준다. L과 R은 구문표지의 자리에 써 준다.

예) 철수가 “영희는 나빠!”라고 말했다.

(S (NP_SBJ 철수/NNP + 가/JKS)
 (VP (S_CMP (S (L “/SS)
 (S (S (NP_SBJ + 영희/NNP + 는/JX)
 (VP 나빠/VA + 아/EF + !/SF))
 (R + “/SS)))
 (X_CMP 라고/JKQ))
 (VP 말하/VV + 았/EP + 다/EF + ./SF)))

예) 영희는 연필(개당 1천원짜리)을 몇 자루를 샀다.

(S (NP_SBJ 영희/NNP + 는/JX)
 (VP (NP_OBJ (NP (NP 연필/NNG)
 (NP_PRN (L (/SS)
 (NP_PRN (NP (NP + 개당/NNG)
 (NP 1/SR + 천원/NNG + 짜
 리/XSN))
 (R_PRN +)/SS))))
 (X_OBJ 을/JKO))
 (VP (NP_OBJ (DP 몇/MM)

(NP_OBJ 자루/NING + 를/JKO))
(VP 사/VV + 았/EP + 다/EF + ./SF)))

• 인용 부호와 괄호가 동시에 나올 경우는 다음과 같이 처리한다.

예) 가출 동기의 1위가 '가정 불화'(14%)였다.

⇒ 어절 나누기의 예(구문분석의 과정에서 다음과 같이 어절을 재분리해 준다)

<가공 전: 형태분석 말뭉치의 예>

가출

동기의

1위가

‘가정

불화'(14%)였다.

<가공 후: 구문분석 말뭉치 구축에서>

가출

동기의

1위가

‘가정

물화'

(14%)였다.

<구문분석 사례>

; 가출 동기 1위가 '가정 불화' (14%)였다.

(S (NP_SBJ (NP (NP 가출/NNG) (NP 동기/NNG)) (NP_SBJ 1/SN + 위/NNB + 가/JKS)) (VNP (NP (L 'SS) (NP (NP (NP + 가정/NNG) (NP 불화/NNG)) (R + 'SS))) (VNP (/SS + 14/SN + %/SW +)/SS + 이/VCP + 었/EP + 다/EF + /SF)))

5.2.2 문장 종결기호 및 기타 기호

· 문장이나 구의 끝에 나오는 종결 부호나 기타 부호 등은 모두 바로 앞의 어절의 일부로서 따로 독립 어절을 이루지 않는다.

예) 이라크 사태 급박//

(NP (NP (NP 이라크/NNP)
(NP 사태/NNG))
(NP 급박/NNG + ///SP))

예) 김 대통령 방북.

(NP (NP (NP 김/NNP)
(NP 대통령/NNG))
(NP 방북/NNG + ./SF))

예) 철수는 집에 있니?

(S (NP_SBJ 철수/NNP + 는/JX)
(VP (NP_AJT 집/NNG + 예/JKB)
(VP 있/VV + 니/EF + ?/SF)))

· 독립어절 분리의 원칙: 인용 부호와 괄호 모두 한 쌍의 부호들 가운데 왼쪽의 부호(여는 부호)가 독립 어절로 나타날 경우, 오른쪽 부호(닫는 부호) 역시 독립 어절로 분석해야 한다.

예) 그가 ‘오딧세이(로 드 이세이)’를 내놓았다.

=> 어절 분리:

그가 그/NP + 가/JKS

‘ /SS

오딧세이 오딧세이/NNP

(/SS

로 로/NNG

드	드/NNG
이세이	이세이/NNG
))/SS
')/SS
를	를/JKO
내놓았다. 내놓/VV + 았/테 + 다/EF + ./SF	

<구문 분석 결과>

; 그가 '오딧세이(로 드 이세이)'를 내놓았다.

(S
(NP_{SBJ} 그)NP + 가/는)
(VP NP_{CJ}) (NP (L /SS)
(NP (NP (NP + 오닷?)\NP)
(NP_{HVN} (L /SS)
(NP_{HVN} (NP (N P
(NP 는)\NG)
(NP 드)\NG)
새이)\NG))
(R_{HVN} -)SS)))
(X_{CJ} 를)\(O)
(VP 나\W + 앞테 + 다EF - /SE)))
(R = /SS)))

5.3 인용부호를 동반한 두 개 이상의 인용절 처리 문제

- 인용부호가 있고 그 안에 두 개 이상의 문장이 들어 있을 경우, 모문과 내포절을 모두 같은 차원으로 놓고 분석을 하게 되면 문장의 길이가 매우 길어지고 복잡성이 너무 많이 증가하여 구문 분석이 어려워지게 된다.
- 이러한 문제점을 해결하기 위해 길어진 인용절과 모문을 다른 차원으로 놓아 분석하는 방법을 취하기로 한다.

5.3.1 독립된 인용절의 경우

- 두 개 이상의 문장들을 인용부호만이 감싸고 있을 뿐, 여타의 모문 성분들이 보이지 않는 경우를 말한다.

예) “영희는 좋은 여자야. 그녀와 결혼하고 싶어!”

· 분석 방법

i) 구문분석 이전의 전처리 단계에서 독립된 인용절을 색출해 낸다.

※참고: 전처리 단계에서의 자료 형태는 형태분석 말뭉치와 다를 바가 없다. 다만 문장별로 한 줄씩 떨어져 있을 뿐이다.

실례)

“영희는 “/SS + 영희/NNP + 는/JX

좋은 좋/VA + 은ETM

여자야. 여자/NNG + 이/VCP + 야/EF + ./SF

그녀와 그녀/NNG + 와/JKB

결혼하고 결혼/NNG + 하/XSV + 고/EC

싶어!” 싶/VX + 어/EF + !/SF + ”/SS

ii) 인용부호만을 모아 두고 그 내부의 문장들을 오려서 그 아래에 붙여 놓는다.

실례)

“” “/SS”/SS

영희는 영희/NNP + 는/JX

좋은 좋/VA + 은ETM

여자야. 여자/NNG + 이/VCP + 야/EF + ./SF

그녀와 그녀/NNG + 와/JKB

결혼하고 결혼/NNG + 하/XSV + 고/EC

싶어! 싶/VX + 어/EF + !/SF

iii) 인용부호 사이에 Q=를 넣고 Q= 오른쪽에 인용절의 개수만큼의 숫자를 써넣는다.

실례)

“Q=2” “/SS + Q=2/Q + ”/SS

영희는 영희/NNP + 는/JX

좋은 좋/VA + 은ETM

여자야. 여자/NNG + 이/VCP + 야/EF + ./SF

그녀와 그녀/NNG + 와/JKB

결혼하고 결혼/NNG + 하/XSV + 고/EC

싫어! 싫/VX + 어/EF + !/SF

iv) 인용절 각각의 맨처음 어절에 순서대로 Q1, Q2 등의 순서를 나타내는 문자열을 써넣는다.

※참고: 여기까지가 전처리 단계에서 해 주어야 할 작업이다.

실례)

“Q=2” “/SS + Q=2/Q + ”/SS

Q1영희는 영희/NNP + 는/JX

좋은 좋/VA + 은ETM

여자야. 여자/NNG + 이/VCP + 야/EF + ./SF

Q2그녀와 그녀/NNG + 와/JKB

결혼하고 결혼/NNG + 하/XSV + 고/EC

싫어! 싫/VX + 어/EF + !/SF

v) 구문분석 최종 결과

; “Q=2”

(Q “/SS + Q=2/Q + ”/SS)

; Q1영희는 좋은 여자야.

(S;Q1 (NP_SBJ 영희/NNP + 는/JX)
 (VNP (VP_MOD 좋/VA + 은ETM)
 (VNP 여자/NNG + 이/VCP + 야/EF + ./SF)))

; Q2그녀와 결혼하고 싶어!

(VP;Q2 (NP_AJT 그녀/NNG + 와/JKB)
 (VP (VP 결혼/NNG + 하/XSV + 고/EC)
 (VP 싶/VX + 어/EF + !/SF)))

5.3.2 내포된 인용절의 경우

☞ [5.3.1]의 경우와 동일한 방식으로 처리한다.

예) 보훈병원 진영수 박사(체육학)는 "물속 운동은 관절에 부담을 주지 않아 재활치료에서도 물 속 걷기 - 달리기 등이 자주 활용된다. 자기 몸무게를 못 이겨 무릎이나 허리부상을 당하기 쉬운 뚱뚱한 사람이나 환자들의 운동으로도 추천할 만하다"고 말한다.

<형태분석 말뭉치의 출력 결과>

보훈병원 보훈병원/NNP

진영수 진영수/NNP

박사(체육학)는 박사/NNG + (/SS + 체육학/NNG +)/SS + 는/JX

"물속 "/SS + 물/NNG + 속/NNG

운동은 운동/NNG + 은/JX

관절에 관절/NNG + 에/JKB

부담을 부담/NNG + 을/JKO

주지 주/VV + 지/EC

않아 앓/VX + 아/EC

재활치료에서도 재활/NNG + 치료/NNG + 에서/JKB + 도/JX

물 물/NNG

속 속/NNG

걷기 걷/VV + 기/ETN

- -/SS
 달리기 달리/VV + 기/ETN
 등이 등/NNB + 이/JKS
 자주 자주/MAG
 활용된다. 활용/NNG + 되/XSV + ㄴ다/EF + ./SF
 자기 자기/NP
 몸무게를 몸무게/NNG + 를/JKO
 못 못/MAG
 이겨 이기/VV + 어/EC
 무릎이나 무릎/NNG + 이나/JC
 허리부상을 허리/NNG + 부상/NNG + 을/JKO
 당하기 당하/VV + 기/ETN
 쉬운 쉽/VA + ㄴ/ETM
 뚱뚱한 뚱뚱/XR + 하/XSA + ㄴ/ETM
 사람이나 사람/NNG + 이나/JC
 환자들의 환자/NNG + 들/XSN + 의/JKG
 운동으로도 운동/NNG + 으로/JKB + 도/JX
 추천할 추천/NNG + 하/XSV + ㄴ/ETM
 만하다"고 만/NNB + 하/XSA + 다/EC + "/SS + 고/JKQ
 말한다. 말/NNG + 하/XSV + ㄴ다/EF + ./SF

<인용절 분리 과정을 거친 상태>

보훈병원 보훈병원/NNP
 진영수 진영수/NNP
 박사(체육학)는 박사/NNG + (/SS + 체육학/NNG +)/SS + 는/JX
 "Q=2"고 "/SS + Q=2/Q + "/SS + 고/JKQ
 말한다. 말/NNG + 하/XSV + ㄴ다/EF + ./SF

Q1물속 물/NNG + 속/NNG
 운동은 운동/NNG + 은/JX
 관절에 관절/NNG + 에/JKB
 부담을 부담/NNG + 을/JKO

주지 주/VV + 지/EC
 않아 앓/VX + 아/EC
 재활치료에서도 재활/NNG + 치료/NNG + 에서/JKB + 도/JX
 물 물/NNG
 속 속/NNG
 걷기 걷/VV + 기/ETN
 - -/SS
 달리기 달리/VV + 기/ETN
 등이 등/NNB + 이/JKS
 자주 자주/MAG
 활용된다. 활용/NNG + 되/XSV + ㄴ다/EF + ./SF

Q2자기 자기/NP
 몸무게를 몸무게/NNG + 를/JKO
 못 못/MAG
 이겨 이기/VV + 어/EC
 무릎이나 무릎/NNG + 이나/JC
 허리부상을 허리/NNG + 부상/NNG + 을/JKO
 당하기 당하/VV + 기/ETN
 쉬운 쉽/VA + ㄴ/ETM
 똥똥한 똥똥/XR + 하/XSA + ㄴ/ETM
 사람이나 사람/NNG + 이나/JC
 환자들의 환자/NNG + 들/XSN + 의/JKG
 운동으로도 운동/NNG + 으로/JKB + 도/JX
 추천할 추천/NNG + 하/XSV + ㄴ/ETM
 만하다 만/NNB + 하/XSA + 다/EC

<구문 분석된 결과>

; 보훈병원 진영수 박사(체육학)는 "Q=2"고 말한다.

(S (NP_SBJ (NP 보훈병원/NNP) (NP_SBJ (NP 진영수/NNP) (NP_SBJ 박사/NNG + (/SS + 체육학/NNG +)/SS + 는/JX))) (VP (Q_CMP "/SS + Q=2/Q + "/SS + 고/JKQ)

(VP 말/NNG + 하/XSV + ㄴ다/EF + ./SF)))

; Q1물속 운동은 관절에 부담을 주지 않아 재활치료에서도 물 속 걷기 - 달리기 등이 자주 활용된다.

(S;Q1 (S (NP_SBJ (NP 물/NNG + 속/NNG)
(NP_SBJ 운동/NNG + 은/JX)
(VP (NP_AJT 관절/NNG + 에/JKB)
(VP (NP_OBJ 부담/NNG + 을/JKO)
(VP (VP 주/VV + 지/EC)
(VP 않/VX + 아/EC))))))
(S (NP_AJT 재활/NNG + 치료/NNG + 에서/JKB + 도/JX)
(S (NP_SBJ (VP (NP_AJT (NP 물/NNG)
(NP_AJT 속/NNG)
(VP (VP (VP 걷/VV + 기/ETN)
(X -/SS)
(VP 달리/VV + 기/ETN)))
(NP_SBJ 등/NNB + 이/JKS))
(VP (AP 자주/MAG)
(VP 활용/NNG + 되/XSV + ㄴ다/EF + ./SF))))))

; Q2자기 몸무게를 못 이겨 무릎이나 허리부상을 당하기 쉬운 뚱뚱한 사람이나 환자들의 운동으로도 추천할 만하다

(VP;Q2 (NP_AJT (NP_MOD (NP_CNJ (S_MOD (VP (NP_OBJ (NP 자기/NP)
(NP_OBJ 몸무게/NNG + 를/JKO))
(VP (AP 못/MAG)
(VP 이기/VV + 어/EC)))
(S_MOD (NP_OBJ (NP_CNJ 무릎/NNG + 아나/JC)
(NP_OBJ 허리/NNG + 부상/NNG
+ 을/JKO))
(S_MOD (VP_SBJ 당하/VV + 기/ETN)
(VP_MOD 쉽/VA + ㄴ/ETM))))
(NP_CNJ (VP_MOD 뚱뚱/XR + 하/XSA + ㄴ/ETM)
(NP_CNJ 사람/NNG + 아나/JC))
(NP_MOD 환자/NNG + 들/XSN + 의/JKG))
(NP_AJT 운동/NNG + 으로/JKB + 도/JX))
(VP (VP_MOD 추천/NNG + 하/XSV + ㄴ/ETM)
(VP 만/NNB + 하/XSA + 다/EC)))

5.3.3 두 개 이상의 내포된 인용절의 경우

☞ [5.3.1]의 경우와 동일한 방식으로 처리한다. 다만, 두 개 이상의 인용절을 구별하기 위해 Q 이외에도 U, W, Y, Z를 사용한다.

예) '급제하고야 만다는 무거운 마음은 심신을 다치게 하는 일이다. 다친 마음에서 좋은 글이 나올 수는 없다.' <고암선생가훈(顧菴先生家訓)> 한 것을 고쳐 '이기지 않으면 안 된다는 강박감은 바로 심신을 다치게 하는 일이다. 다친 몸으로 잘 싸운다는 것은 거짓말이다.'

<형태분석 말뭉치의 출력 결과>

'급제하고야 '/SS + 급제/NNG + 하/XSV + 고/EC + 야/JX
만다는 말/VV + ㄴ다는/ETM
무거운 무겁/VA + ㄴ/ETM
마음은 마음/NNG + 은/JX
심신을 심신/NNG + 을/JKO
다치게 다치/VV + 게/EC
하는 하/VX + 는/ETM
일이다. 일/NNG + 이/VCP + 다/EF + ./SF
다친 다치/VV + ㄴ/ETM
마음에서 마음/NNG + 에서/JKB
좋은 좋/VA + 은/ETM
글이 글/NNG + 이/JKS
나올 나오/VV + ㄴ/ETM
수는 수/NNB + 는/JX
없다.' 없/VA + 다/EF + ./SF + '/SS
<고암선생가훈(顧菴先生家訓)> </SS + 고암/NNP + 선생/NNG + 가훈/NNG +
(/SS + 顧菴先生家訓/SH +)/SS + >/SS
한 하/VV + ㄴ/ETM
것을 것/NNB + 을/JKO
고쳐 고치/VV + 어/EC
'이기지 '/SS + 이기/VV + 지/EC
않으면 않/VX + 으면/EC
안 안/MAG
된다는 되/VV + ㄴ다는/ETM
강박감은 강박감/NNG + 은/JX
바로 바로/MAG

심신을 심신/NNG + 을/JKO
 다치게 다치/VV + 게/EC
 하는 하/VX + 는/ETM
 일이다. 일/NNG + 이/VCP + 다/EF + ./SF
 다친 다치/VV + ㄴ/ETM
 몸으로 몸/NNG + 으로/JKB
 잘 잘/MAG
 싸운다는 싸우/VV + ㄴ다는/ETM
 것은 것/NNB + 은/JX
 거짓말이다.' 거짓말/NNG + 이/VCP + 다/EF + ./SF + '/SS

<인용절 분리 과정을 거친 상태>

'Q=2' '/SS + Q=2/Q + '/SS

<고암선생가훈(顧菴先生家訓)> </SS + 고암/NNP + 선생/NNG + 가훈/NNG +
 (/SS + 顧菴先生家訓/SH +)/SS + >/SS

한 하/VV + ㄴ/ETM

것을 것/NNB + 을/JKO

고쳐 고치/VV + 어/EC

'U=2' '/SS + U=2/Q + '/SS

Q1급제하고야 급제/NNG + 하/XSV + 고/EC + 야/JX

만다는 말/VV + ㄴ다는/ETM

무거운 무겁/VA + ㄴ/ETM

마음은 마음/NNG + 은/JX

심신을 심신/NNG + 을/JKO

다치게 다치/VV + 게/EC

하는 하/VX + 는/ETM

일이다. 일/NNG + 이/VCP + 다/EF + ./SF

Q2다친 다치/VV + ㄴ/ETM

마음에서 마음/NNG + 에서/JKB

좋은 좋/VA + 은/ETM

글이 글/NNG + 이/JKS
 나올 나오/VV + ㄴ/ETM
 수는 수/NNB + 는/JX
 없다. 없/VA + 다/EF + ./SF

U1이기지 이기/VV + 지/EC
 앞으면 앞/VX + 으면/EC
 안 안/MAG
 된다는 되/VV + ㄴ다는/ETM
 강박감은 강박감/NNG + 은/JX
 바로 바로/MAG
 심신을 심신/NNG + 을/JKO
 다치게 다치/VV + 게/EC
 하는 하/VX + 는/ETM
 일이다. 일/NNG + 이/VCP + 다/EF + ./SF

U2다친 다치/VV + ㄴ/ETM
 몸으로 몸/NNG + 으로/JKB
 잘 잘/MAG
 싸운다는 싸우/VV + ㄴ다는/ETM
 것은 것/NNB + 은/JX
 거짓말이다. 거짓말/NNG + 이/VCP + 다/EF + ./SF

<구문 분석된 결과>

; 'Q=2' <고암선생가훈(顧菴先生家訓)> 한 것을 고쳐 'U=2'

(VP (VP (NP_OBJ (VP_MOD (Q (Q '/SS + Q=2/Q + '/SS)
 (NP_PRN <SS + 고암/NNP + 선생/NNG + 가훈/NNG + /SS + 顧菴先生家訓
 /SHI - /SS + >/SS))
 (VP_MOD 하/VV + ㄴ/ETM))
 (NP_OBJ 것/NNB + 을/JKO))
 (VP 고치/VV + 어/EC))
 (Q '/SS + U=2/Q + '/SS))

; Q1급제하고야 만다는 무거운 마음은 심신을 다치게 하는 일이다.

(S:Q1 (NP_SBJ (VP_MOD (VP 급제/NNG + 하/XSV + 고/EC + 야/JX)

(VP_MOD 말/VV + ㄴ다는/ETM))
 (NP_SBJ (VP_MOD 무겁/VA + ㄴ/ETM)
 (NP_SBJ 마음/NNG + 은/JX)))
 (VNP (VP_MOD (NP_OBJ 심신/NNG + 을/JKO)
 (VP_MOD (VP 다치/VV + 게/EC)
 (VP_MOD 하/VX + 는/ETM))))
 (VNP 일/NNG + 이/VCP + 다/EF + /SF)))

; Q2다친 마음에서 좋은 글이 나올 수는 없다.

(S;Q2 (NP_AJT (VP_MOD 다치/VV + ㄴ/ETM)
 (NP_AJT 마음/NNG + 에서/JKB))
 (S (NP_SBJ (VP_MOD 좋/VA + 은/ETM)
 (NP_SBJ 글/NNG + 이/JKS))
 (S (NP_SBJ (VP_MOD 나오/VV + ㄴ/ETM)
 (NP_SBJ 수/NNB + 는/JX))
 (VP 없/VA + 다/EF + /SF))))

; U1이기지 않으면 안 된다는 강박감은 바로 심신을 다치게 하는 일이다.

(S;U1 (NP_SBJ (VP_MOD (VP (VP 이기/VV + 지/EC)
 (VP 앓/VX + 으면/EC))
 (VP_MOD (AP 안/MAG)
 (VP_MOD 되/VV + ㄴ다는/ETM)))
 (NP_SBJ 강박감/NNG + 은/JX))
 (VNP (AP 바로/MAG)
 (VNP (VP_MOD (NP_OBJ 심신/NNG + 을/JKO)
 (VP_MOD (VP 다치/VV + 게/EC)
 (VP_MOD 하/VX + 는/ETM))))
 (VNP 일/NNG + 이/VCP + 다/EF + /SF))))

; U2다친 몸으로 잘 싸운다는 것은 거짓말이다.

(S;U2 (NP_SBJ (VP_MOD (NP_AJT (VP_MOD 다치/VV + ㄴ/ETM)
 (NP_AJT 몸/NNG + 으로/JKB))
 (VP_MOD (AP 잘/MAG)
 (VP_MOD 싸우/VV + ㄴ다는/ETM)))
 (NP_SBJ 것/NNB + 은/JX))
 (VNP 거짓말/NNG + 이/VCP + 다/EF + /SF)))

5.3.4 내포된 인용절 속에 또 다른 인용절이 내포된 경우

☞ 앞서의 경우와 동일하게 처리한다. 다만, 서로 다른 층위의 인용절을 구별하기 위

해 Q 이외에 U, W, Y, Z를 사용한다. 아래 예를 통해 특히 전처리 주석 다는 방식에 유의하라.

예) [좌석 앞줄에서 듣고 있던 그 학생이 강단 앞으로 달려나가더니 목사를 손가락질 하며 소리쳤습니다. <집어쳐! 말씀이 우리에게 무얼 줄 수 있단 말이야? 기껏 너 따 위에게 이용되어 당연히 우리 입에 들어가게 되어 있는 빵조차 가로채 갈 뿐이야!> 그러자 목사는 강단 위에서 <사탄아, 물러가라!>고 외치더군요.]

<형태분석 말뭉치의 출력 결과>

[좌석 [/SS + 좌석/NNG
 앞줄에서 앞/NNG + 줄/NNG + 에서/JKB
 듣고 듣/VV + 고/EC
 있던 있/VX + 던/ETM
 그 그/MM
 학생이 학생/NNG + 이/JKS
 강단 강단/NNG
 앞으로 앞/NNG + 으로/JKB
 달려나가더니 달려나가/VV + 더니/EC
 목사를 목사/NNG + 를/JKO
 손가락질하며 손가락/NNG + 질/XSN + 하/XSV + 며/EC
 소리쳤습니다. 소리치/VV + 었/EP + 습니다/EF + ./SF

<집어쳐!</SS + 집어치/VV + 어/EF + !/SF

말씀이 말씀/NNG + 이/JKS
 우리에게 우리/NP + 에게/JKB
 무얼 무엇/NP + 르/JKO
 줄 주/VV + 르/ETM
 수 수/NNB
 있단 있/VV + 단/ETM
 말이야? 말/NNG + 이/VCP + 야/EF + ?/SF

기껏 기껏/MAG
 너 너/NP
 따위에게 따위/NNB + 에게/JKB
 이용되어 이용/NNG + 되/XSV + 어/EC
 당연히 당연히/MAG
 우리 우리/NP
 입에 입/NNG + 에/JKB
 들어가게 들어가/VV + 게/EC
 되어 되/VX + 어/EC
 있는 있/VX + 는/ETM
 빵조차 빵/NNG + 조차/JX
 가로채 가로채/VV + 어/EC
 갈 가/VV + ㄹ/ETM
 뿐이야!> 뿐/NNB + 이/VCP + 야/EF + !/SF + >/SS

그러자 그러자/MAJ
 목사는 목사/NNG + 는/JX
 강단 강단/NNG
 위에서 위/NNG + 에서/JKB
 <사탄아, </SS + 사탄/NNG + 아/JKV + ,/SP
 물러가라!>고 물러가/VV + 아라/EF + !/SF + >/SS + 고/JKQ
 외치더군요.] 외치/VV + 더군요/EF + ./SF +]/SS

<인용절 분리 과정을 거친 상태>

[Q=3] [/SS + Q=3/Q +]/SS

Q1좌석 좌석/NNG
 앞줄에서 앞/NNG + 줄/NNG + 에서/JKB
 듣고 듣/VV + 고/EC
 있던 있/VX + 던/ETM
 그 그/MM
 학생이 학생/NNG + 이/JKS

강단 강단/NNG

앞으로 앞/NNG + 으로/JKB

달려나가더니 달려나가/VV + 더니/EC

목사를 목사/NNG + 를/JKO

손가락질하며 손가락/NNG + 질/XSN + 하/XSV + 며/EC

소리쳤습니다. 소리치/VV + 었/EP + 습니다/EF + ./SF

Q2<U=3> </SS + U=3/Q + >/SS

U1집어쳐! 집어치/VV + 어/EF + !/SF

U2말씀이 말씀/NNG + 이/JKS

우리에게 우리/NP + 에게/JKB

무얼 무엇/NP + ㄹ/JKO

줄 주/VV + ㄹ/ETM

수 수/NNB

있단 있/VV + 단/ETM

말이야? 말/NNG + 이/VCP + 야/EF + ?/SF

U3기껏 기껏/MAG

너 너/NP

따위에게 따위/NNB + 에게/JKB

이용되어 이용/NNG + 되/XSV + 어/EC

당연히 당연히/MAG

우리 우리/NP

입에 입/NNG + 에/JKB

들어가게 들어가/VV + 게/EC

되어 되/VX + 어/EC

있는 있/VX + 는/ETM

빵조차 빵/NNG + 조차/JX

가로채 가로채/VV + 어/EC

갈 가/VV + ㄹ/ETM

뿐이야! 뿐/NNB + 이/VCP + 야/EF + !/SF

Q3그러자그러자/MAJ

목사는 목사/NNG + 는/JX

강단 강단/NNG

위에서 위/NNG + 에서/JKB

<사탄아, </SS + 사탄/NNG + 아/JKV + ,/SP

물러가라!>고 물러가/VV + 아라/EF + !/SF + >/SS + 고/JKQ

외치더군요. 외치/VV + 더군요/EF + ./SF

<구문 분석된 결과>

; Q=3]

(Q [/SS + Q=3/Q + !/SS)

; Q1좌석 앞줄에서 듣고 있던 그 학생이 강단 앞으로 달려나가더니 목사를 손가락질하며 소리쳤습니다.

(S:Q1 (NP_SBJ (VP_MOD (NP_AJT (NP 좌석/NNG
(NP_AJT 앞/NNG + 줄/NNG + 에서/JKB))
(VP_MOD (VP 듣/VV + 고/EC)
(VP_MOD 있/VX + 던/ETM)))
(NP_SBJ (DP 그/MM)
(NP_SBJ 학생/NNG + 이/JKS)))
(VP (VP (NP_AJT (NP 강단/NNG
(NP_AJT 앞/NNG + 으로/JKB))
(VP 달려나가/VV + 더니/EC))
(VP (VP (NP_OBJ 목사/NNG + 를/JKO)
(VP 손가락/NNG + 질/XSN + 하/XSV + 며/EC))
(VP 소리치/VV + 었/EP + 습니다/EF + ./SF))))))

; Q2<U=3>

(Q:Q2 </SS + U=3/Q + >/SS)

; U1잡어쳐!

(VP:U1 잡어쳐/VV + 어/EF + !/SF)

; U2말씀이 우리에게 무얼 줄 수 있단 말이야?

(S:U2 (NP_SBJ 말씀/NNG + 이/JKS)
(VP (NP_AJT 우리/NP + 에게/JKB)
(VP (NP_OBJ 무엇/NP + ㄹ/JKO)
(VNP (S_MOD (NP_SBJ (VP_MOD 주/VV + ㄹ/ETM)
(NP_SBJ 수/NNB))
(VP_MOD 있/VV + 단/ETM))
(VNP 말/NNG + 이/VCP + 야/EF + ?/SF))))))

; U3기껏 너 따위에게 이용되어 당연히 우리 입에 들어가게 되어 있는 빵조차 가로채 갈 뿐이야!

(VNP;U3 (VP (AP 기껏/MAG
(VP (NP_AJT (NP 너/NP)
(NP_AJT 따위/NNB + 예/EC))
(VP 이용/NNG + 되/XSV + 어/EC)))
(VNP (NP_OBJ (VP_MOD (AP 당연히/MAG)
(VP_MOD (NP_AJT (NP 우리/NP)
(NP_AJT 입/NNG + 예/EC))
(VP_MOD (VP (VP 들어가/VV + 게/EC)
(VP 되/VX + 어/EC))
(VP_MOD 있/VX + 는/ETM))))
(NP_OBJ 땀/NNG + 조차/JX)
(VNP (VP_MOD (VP 가로채/VV + 어/EC)
(VP_MOD 가/VV + ㄹ/ETM))
(VNP 뵈/NNB + 이/VCP + 아/EF + !/SF))))

; Q3그러자 목사는 강단 위에서 <사탄아, 물러가라!>고 외치더군요.

(S;Q3 (AP 그러자/MAJ)
(S (NP_SBJ 목사/NNG + 는/JX)
(VP (NP_AJT (NP 강단/NNG)
(NP_AJT 위/NNG + 에서/JKB))
(VP (VP_CMP (VP (L </SS)
(VP (NP_INT 사탄/NNG +
(VP 물러가/VV + 아
(VP 외치/VV + 더군요/EF + !/SF))))
(R >/SS)))
(X_CMP 고/JKQ))
(VP 외치/VV + 더군요/EF + !/SF))))

5.4 구문 표지 없이 기능 표지만 부착되는 경우

· 인용 부호 다음에 조사들이 오는 경우가 이에 해당한다.

예) ‘나도 아버지가 있다’로서만이 아니라 ‘내게는 이러한 훌륭한 아버지가 계시다’ 하
는 벽차오르는 자랑이었다

⇒ 형태분석 상태

‘나도 ’/SS + 나/NP + 도/JX
아버지가 아버지/NNG + 가/JKS
있다’로서만이 있/VV + 다/EC + ’/SS + 로서/JKB + 만/JX + 이/JKC
아니라 아니/VCN + 라/EC

.....

⇒ 구문 분석 도구에 읽힌 모습

구문	기능	태그
'/SS	I	I
나/NP + 도/JX	NP	NP_SBJ
아버지/NNG + 가/JKS	NP	NP_SBJ
있/VV + 다/EC	VP	VP
'/SS	R	R
로써/JKB + 만/JX + 이/JKC	CMP	CMP
아니/VCN + 라/EC	VP	VP
'/SS	I	I
나/NP + 내게/JKB + 는/AD	NP	NP_SBJ
이러한/VA + 는/ETM	VP	VP_MOD
훌륭한/VA + 한/XSA + 는/ETM	VP	VP_MOD
아버지/NNG + 가/JKS	NP	NP_SBJ
계시/VV + 다/EC	VP	VP
'/SS	R	R
아니/VV + 는/ETM	VP	VP_MOD
법치오관/VV + 는/ETM	VP	VP_MOD
자랑/NNG + 이/VCP + 한/FP + 다/EF + 는/SF	VP	VP

현재 문장: '나도 아버지가 있다'로서만이 아니라 '내게는 이러한 훌륭한 아버지가 계시다' 하는 법치오르는 자랑이었다.

입력 파일 정보: 현재 문장 번호: 29, 전체 문장 개수: 29, 입력 파일 이름: 주장중인 문장1.A

⇒ 구문분석의 실제

· 구문 표지 없이 기능 표지만을 가지는 경우 구문 표지의 빈 자리에 “X”를 넣어주며, 그것은 그 위의 묶기 단계에서 선행 어절의 구문 표지로 대체된다.

예) '나도 아버지가 있다'로서만이 아니라 '내게는 이러한 훌륭한 아버지가 계시다' 하는 법치오르는 자랑이었다.

.....

(VP (S_CMP (S (L '/SS)

(S (S (NP_SBJ + 나/NP + 도/JX)

(S (NP_SBJ 아버지/NNG + 가/JKS)

(VP 있/VV + 다/EC)))

(R +'/SS)))

(X_CMP 로써/JKB + 만/JX + 이/JKC))

(VP 아니/VCN + 라/EC))

.....

· 조사가 떨어져 있을 뿐만이 아니라, 선행 명사구가 다시 수식을 받는 경우의 분석 순서 문제.

⇒ [선행 명사구 + 조사의 결합체]가 먼저 결합되고 난 뒤, 그 전체가 수식어의 수식을 받는 것으로 분석한다.

예) 세계적 조각가인 알렉산더 칼더 조각전행사의 하나로 미술관측이 마련한 '어린이를 위한 모빌실습'의 현장이다.

⇒ ... [NP_MOD [S_MOD 미술관측이 마련한] [X_MOD [S '어린이를 위한 모빌실습'] 의/JKG]] ...

⇒

(VP (NP_MOD (S_MOD (NP_AJT (NP_MOD (VP_MOD (NP 세계/NNG + 적/XSN)
(VNP_MOD 조각가/NNG + 이/VCP + ㄴ/ETM))
(NP_MOD (NP (NP 알렉산더/NNP)
(NP 칼더/NNP))
(NP_MOD 조각전/NNG + 행사/NNG + 의/JKG)))
(NP_AJT 하나/NR + 로/JKB))
(S_MOD (NP_SBJ 미술관/NNG + 측/NNB + 이/JKS)
(VP_MOD 마련/NNG + 하/XSV + ㄴ/ETM)))
(NP_MOD (NP (L 'SS)
(S (NP (VP_MOD (NP_OBJ + 어린이/NNG + 를
(VP_MOD 위하/VV + ㄴ/ETM))
(NP 모빌/NNG + 실습/NNG))
(R +'/SS)))
(X_MOD 의/JKG)))
(VNP 현장/NNG + 이/VCP + 다/EF + /SF))

5.5 “명사+계사”가 논항처럼 기능하는 경우

· 이 경우 구문 표지를 VNP로 하고 기능표지를 명시해 준다.

예) 형제간이라도 있었으면 좀 나왔으련만, ...

=> 형태분석의 경우:

형제간이라도 형제/NNG + 간/NNB + 이/VCP + 라도/EC

있었으면 있/VV + 었/EP + 으면/EC

좀 좀/MAG

나왔으련만, 낮/VA + 왔/EP + 으련만/EC + ,/SP

=> 구문분석의 경우:

(VP	(VP	(S	(VNP_SBJ 형제/NNG + 간/NNB + 이/VCP + 라도/EC)
			(VP 있/VV + 었/EP + 으면/EC))
		(VP	(AP 좀/MAG)
			(VP 낮/VA + 았/EP + 으련만/EC)))
(VP	(VP	(AP 하필/MAG)	
		(VP	(NP_AJT 남/NNG + 들/XSN)
			(VP 같잖/VA + 게/EC)))
	(VP	(DP 단/MM)	
			(VNP 혼자/NNG + 이/VCP + 었/EP + 다/EF + /SF)))

5.6 “우핵 원리(오른쪽 마디 투사)”가 위반되는 경우

· 올해 구문분석에 도입된 구문분석 도구는 두 교점이 상위 교점으로 병합될 때 우측 마디의 구문분석 정보(구문 표지, 기능 표지)가 기본적으로 투사되도록 되어 있다. 실제 분석의 경우에서도 대부분은 이에 따른다. 그러나 그렇지 않은 경우가 종종 있다. 예컨대, 다음과 같은 우측 부가의 경우가 그러한데, 이때에는 부득이 좌핵 투자를 허용할 수밖에 없다.

예) 그가 죽었어, 어제.

⇒ [S [S 그가 죽었어,] [AP 어제.]]

예) 장사꾼들이 다 그렇지 뭐.

⇒ [S [S 장사꾼들이 다 그렇지] [AP 뭐.]]

5.7 띄어쓰기 문제

5.7.1 하이픈

• 이 경우에 구문분석 도구 프로그램이 자동으로 띄어 준다.

예) 북한-미국 막후 협상 중

⇒ 형태소 분석 말뭉치의 어절 배열:

북한-미국 북한/NNP + -/SS + 미국/NNP

막후

협상

중

⇒ 구문 분석 말뭉치의 어절 배열

북한 북한/NNP

- + -/SS

미국 + 미국/NNP

막후

협상

중

· 가운데점도 하이픈의 처리에 준한다.

5.7.2 의존명사

· 구문 분석 도구가 자동으로 띄어줄 수 없으므로, 작업자가 직접 살펴보고 띄어 준다. 이러한 조작이 가해지면 구문 분석 말뭉치는 형태 분석 말뭉치와 달라지게 된다. 그러나 새롭게 독립 어절로 분석된 어절은 그 앞에 “+”를 갖추게 되어 다시 형태 분석 말뭉치로 환원이 가능해진다.

예) 식량이 없는데 그는 쌀을 원조해 주었다.

⇒ 형태소 분석 말뭉치의 어절 배열:

...

없는데 없/VA + 는/ETM + 데/NNB

...

⇒ 구문 분석 말뭉치의 어절 배열

...

없는 없/VA + 는/ETM
 데 + 데*/NNB

...

※본래의 형태분석 말뭉치 내용과 달라진 것에 ‘*’를 붙여준다. 그 위치는 ‘형태*/표지’와 같이 해 준다.

※ 기호를 수정할 때에는 ‘*’를 붙이지 않는다.

5.7.3 원시말뭉치 오류

· 원시말뭉치 구축시에 생긴 형태 분석 오류가 형태 분석 말뭉치에 그대로 유지된 경우가 있다. 이때 구문 분석에서는 올바른 어절 분리를 해 준다. 다만 형태 분석 말뭉치와의 호환 가능성은 분리된 어절 앞에 남아 있는 “+”로 보장된다.

예) 나는 그것만하여도 다행이라고 생각했다.

⇒ 형태소 분석 말뭉치의 어절 배열:

...

그것만하여도 그것/NP + 만/JX + 하/VV + 아도/EC

...

⇒ 구문 분석 말뭉치의 어절 배열

...

그것만 그것/NP + 만/JX

하여도 + 하*/VV + 아도/EC

...

※본래의 형태분석 말뭉치 내용과 달라진 것에 ‘*’를 붙여준다. 그 위치는 ‘형태*/표지’와 같이 해 준다.

IV. 작업 단계별 절차 및 유의 사항

1. 전처리 단계에서의 유의 사항

1.1 문장 분리

- 독립된 절이나 구를 하나의 문장으로 분리한다.

(이 경우 문장분리기와 같은 프로그램의 지원을 받을 수 있다.)

- 보통, 문장들은 마침표나 물음표 등의 종결기호를 동반한다. 그러나 종종 ‘주어-서술어’를 갖춘 온전한 문장임에도 불구하고 종결 기호를 가지지 않는 경우가 있다. 이럴 경우, 일부러 종결기호를 넣어 주어서는 안 되지만, 하나의 독립된 문장으로 정확하게 분리해 주어야 한다.

- 특히 신문 기사의 경우 제목이 명사나 부사 등으로 끝나는 경우가 빈번하다. 이 경우 역시 독립된 문장으로 처리해야 한다. (이러한 경우는 문장분리기와 같은 프로그램이 올바르게 포착해 주지 못한다. 따라서 작업자가 철저히 살펴보는 수밖에 없다.)

1.2 인용절의 처리 문제

- 인용부호(“”) 사이에 두 개 이상의 독립된 문장이 들어 있는 경우 Ⅲ장의 5.2.1절에 의거하여 전처리 작업을 해 주어야 한다.

- 이러한 사례가 하나의 모문에 연속해서 두 번 이상 나타날 경우가 있다.

예) “너는 자장면 먹을래? 나는 자장면 먹을께.”를 “너는 자장면 먹을래? 나도 자장면 먹을께.”로 수정해야 한다.

이 경우 두 가지 경우를 구별하기 위해 첫째 것에는 Q를 사용하고, 그 다음 인용절에는 U를 사용하고 그 다음으로는 W, Y, Z를 순차적으로 사용한다. 자세한 것은 Ⅲ장의 5.2.3절에 제시되어 있다.

2. 구문 분석 작업 단계에서의 유의 사항

2.1 형태 분석 정보의 오류 검색

- 본격적인 구문 단계에 접어들기 전 형태 분석 오류를 바로 잡을 수 있는 단계가 마련되어 있다. 이 과정에서는 특히 구문표지가 관형사(DP)나 부사(AP)로 되어 있는 어

절을 주의 깊게 살펴야 한다. 다음과 같이 구문분석기가 그 구문표지를 잘못 추출하는 경우가 많이 있다.

예) 형태 분석 어절	구문표지	기능표지
가득/MAG + 하/VV + 아서/EC	AP	
몇/MM + 명/NNB + 이/JKS	DP	SBJ

이러한 오류는 구문분석기가 어절 가운데 제일 먼저 나오는 형태소의 품사표지를 참조하기 때문이다. 이는 구문분석기의 품질 개선과도 관련되어 있지만, 작업자가 계속 주의해 살펴보아야만 될 사항이다.

2.2 구문 분석의 일관성 및 정밀성

· 지침을 명확히 숙지하여 구문 분석에 임해야 한다. 만일 동일한 구문에 대해 작업자마다 다르게 분석한다면, 무엇보다도 그로부터 올바른 규칙을 추출할 수 없다는 문제가 발생한다.

3. 후처리 단계에서의 유의 사항

3.1 구문분석 지침과의 괴리 제거

· 구문분석 작업의 진행 과정에서 불가피하게 지침의 일부가 변경되는 경우가 생긴다. 이때 바뀐 지침에 따라 해당 말뭉치의 수정이 필수적이다. 지침의 수정에 따른 말뭉치의 수정은 한꺼번에 몰아서 하는 것보다는 필요한 때마다 수행해 주는 것이 좀더 효율적이다.

3.2 일관성 검사

· 계획된 말뭉치의 구축이 끝나면 말뭉치들이 균질적인가를 알아보는 일관성 검사를 실시해야 한다. 만일 일관되지 않은 유형들이 검색되면 별도의 프로그램을 통해서든 수작업을 통해서든 이를 균질화하는 작업을 수행해야 한다.

이 보고서는 문화관광부의 21세기 세종계획
“국어 기초자료 구축” 연구 용역 결과 보고서이다.

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

21세기 세종계획 국어 기초자료 구축

2004년 12월 12일 인쇄

2004년 12월 12일 발생

발행자 : 문화관광부
서울시 종로구 세종로 82-1
전화 02-3704-9420

인쇄처 : 현대문화사
전화 02-922-3376

ISBN : 89-7820-124-5 93710
