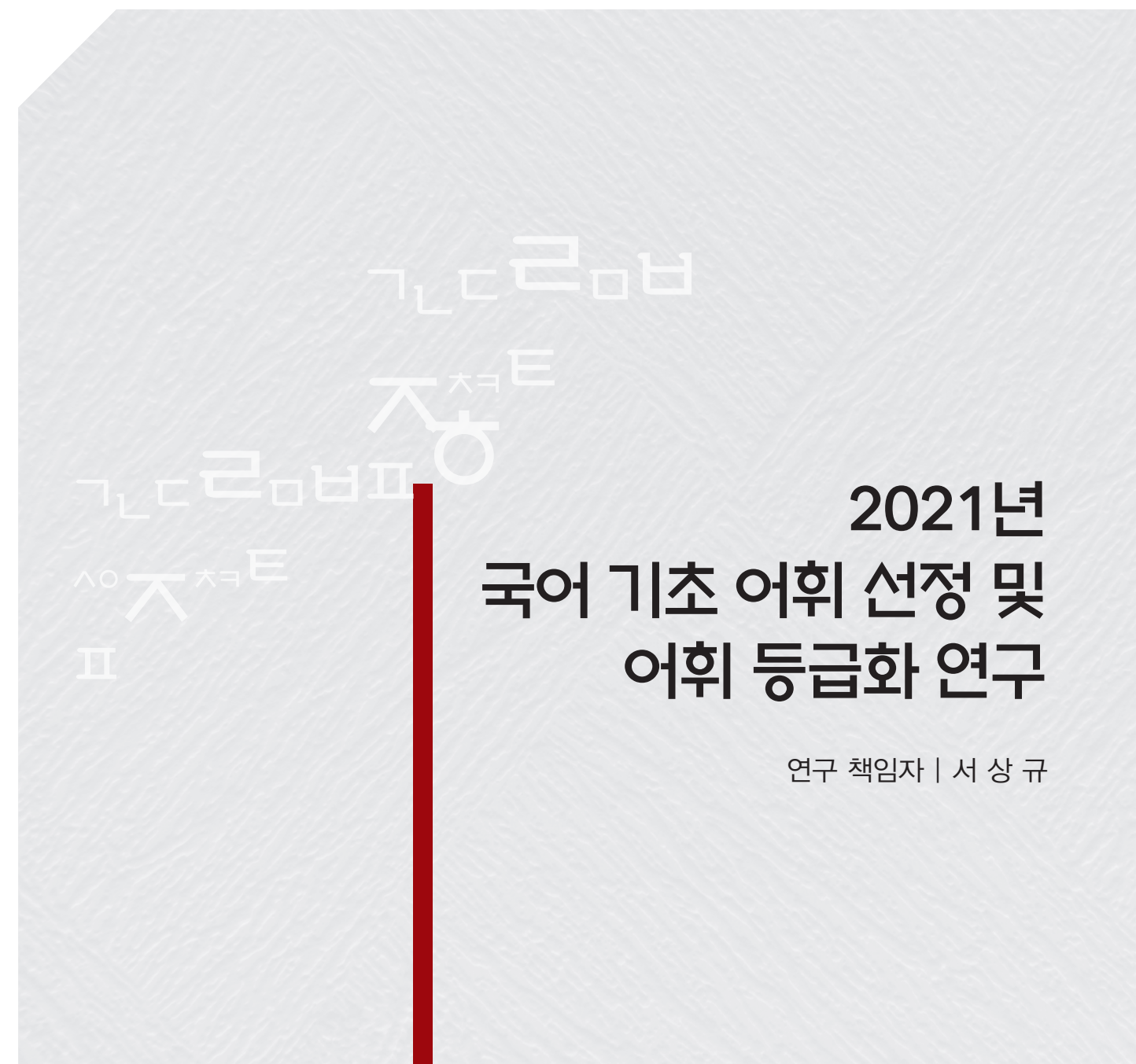




NATIONAL INSTITUTE OF KOREAN LANGUAGE



2021년 국어 기초 어휘 선정 및 어휘 등급화 연구

연구 책임자 | 서 상 규

국립국어원 2021-01-44

발 간 등 록 번 호
11-1371028-000890-01

2021년 국어 기초 어휘 선정 및 어휘 등급화 연구

연구 책임자
서 상 규

제 출 문

국립국어원장 귀하

국립국어원과 체결한 연구 용역 계약에 따라 ‘2021년 국어 기초 어휘 선정 및 어휘 등급화 연구’에 관한 최종 보고서를 작성하여 제출합니다.

■ 사업 기간: 2021. 05. 06 - 2021. 12. 21

2021년 12월 21일

사업 책임자: 서상규(연세대학교)

사업 관리 기관	연세대학교 산학협력단
사업 책임자	서상규(연세대학교)
사업 참여자	김한샘(연세대학교)
	봉미경(연세대학교)
	황은하(배재대학교)
	이병규(서울교육대학교)
	윤천탁(한국교원대학교)
	가은아(공주대학교)
	강범일(연세대학교)
	윤영민(연세대학교)
	유근선(연세대학교)
	박혜진(연세대학교)
	천성호(연세대학교)
	오태환(연세대학교)

<사업 수행자>
연세대학교 산학협력단

사업 책임자

서상규(연세대학교)

사업 참여자

연세대학교 총 8명	봉미경
	김한샘
	강범일
	윤영민
	유근선
	박혜진
	천성호
	오태환
서울교육대학교 총 1명	이병규
공주대학교 총 1명	가은아
한국교원대학교 총 1명	윤천탁
배재대학교 총 1명	황은하

2021년 국어 기초 어휘 선정 및 어휘 등급화 연구

본 사업의 목적은 국민 전체를 대상으로 한국어의 사용 양상을 파악하여 수준별 어휘 목록을 구축하기 위한 기초 단계로서 국어 기초 어휘를 선정하고 1등급 어휘 목록을 제시하는 것이다. 이로써 교과서 등의 교육 자료 개발 및 감수, 국가고사 출제, 국민의 국어 능력 평가를 비롯한 각종 기초 조사 등의 근거로 활용될 수 있는 수준별 어휘 목록 구축에 이바지하고자 하였다. 이에 따른 주요 과업과 연구의 성과는 다음과 같다.

○ 기초 어휘 목록 선정용 자료 수집 및 정비

- 1~3차년도(2017년, 2018년, 2020년) 연구에서 수립된 방법론에 따른 기초 어휘 목록 선정용 자료(말뭉치)의 정비
- 등급화 작업을 위한 한국어 어휘 목록 자료 수집 및 정비
- 기반 자료 보완 연구 및 추가 자료 확보

○ 기초 어휘 범위 및 등급 제시

- 수집 자료를 통한 말뭉치의 어휘량 비교 검토
- 기초 어휘 범위 제시
- 기초 어휘 등급 제시
- 결과물에 대한 정량적, 정성적 평가 방법과 검증 기준 마련

○ 기초 어휘 예비 목록 및 1등급 어휘 목록 제시

- 기초 어휘 예비 목록 제시: 기초 어휘 전체 예비 목록 제시
- 1등급 어휘 목록 제시: 기초 어휘 예비 목록 중 1등급 어휘 목록 제시
- 국민의 어휘 능력 조사 방법론 검토 및 시범 적용
- 예비 목록 선정 과정과 결과에서 나타난 시사점 제시

4차 산업혁명 등 최근 대규모 말뭉치의 구축 및 분석, 정제 등에 대한 수요와 관심이 높아지고 있고, 그 기술이 발전해 가고 있는 현 시점에 국민 전체를 대상으로 한 한국어 사용 양상에 대한 전체적 조망이 필요하다. 이에 따라 ‘2021년 국어 기초 어휘 선정 및 어휘 등급화 연구’는 교육 분야, 공공 분야, 산업 분야 등 다양한 분야에 활용할 수 있는 어휘 목록 및 어휘 등급화를 제안하기 위하여 수행되었다. 본 사업의 결과물은 국내외에서 활용할 수 있는 언어 자료 자체로서, 또는 다양한 언어 자료를 생산, 활용, 평가할 수 있는 보조 자료로서 사용될 수 있는 자료로서 의의를 갖는다.

주요어 : 기초 어휘, 어휘 등급화, 기본 어휘, 말뭉치, 학령 전 어휘, 사용자 어휘 조사, 정량적 분석, 정성적 평정

차 례

제1장 사업 개요

1. 사업 목적	2
2. 사업 범위	4
3. 사업 수행	5
3.1. 기초 어휘 목록 선정 및 등급화 방법론 연구	6
3.2. 말뭉치 분석 방법론 연구	9
3.3. 기초 어휘에 대한 정성적 평정 및 활용	11
4. 사업 결과	12

제2장 선행 연구

1. 말뭉치 구축 및 말뭉치 기반 어휘의 계량적 연구	15
2. 기초 어휘 관련 연구	23
3. 어휘 등급화 관련 연구	28
4. 말뭉치 기반 사전 구축 및 사전 표제어 선정 관련 연구	38
5. 기타 어휘 선정 관련 연구(아동 어휘, 영어 어휘)	44

제3장 사업 수행 내용

1. 사업 추진 일정	54
2. 세부 과업 개요	55
2.1. 기초 어휘 목록 선정 및 등급화 방법론 연구	56
2.2. 말뭉치 분석 방법론 연구	56
2.3. 기초 어휘에 대한 정성적 평정 및 활용	56
3. 기초 어휘 선정 및 어휘 등급화 방법론	58
3.1. 기초 어휘 및 1등급 어휘의 개념과 범위	58
3.2. 기존 어휘 등급 목록 검토 및 정제	61
3.3. 정량적·정성적 등급화 방안 적용	61
3.4. 기초 어휘 예비 목록 및 1등급 어휘 목록 제시	65

차 례

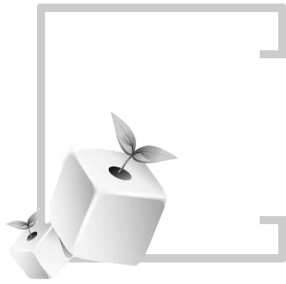
4. 말뭉치 분석 방법론	66
4.1. 말뭉치의 설계 및 구성	66
4.2. 말뭉치 정제 및 기초 어휘 후보 목록 추출	70
4.3. 기초 어휘 선정용 최소 말뭉치 규모 분석	75
5. 기초 어휘에 대한 정성적 평정 및 활용	82
5.1. 정성적 평정을 위한 후보 선정	82
5.2. 어휘 전문가 평정	83
5.3. 어휘 사용자 평정(교수 학습 현장 설문)	84
5.4. 교육 전문가 평정	87
5.5. 평정 결과 반영 및 최종 선정	90
6. 최종 결과물	91
6.1. 기초 어휘 등급화안	91
6.2. 1등급 어휘 목록	92
6.3. 기초 어휘 예비 목록	92

제4장 결론

1. 사업 요약	94
2. 의의 및 기대 효과, 활용 방안	94
3. 향후 연구 및 정책 제안	95

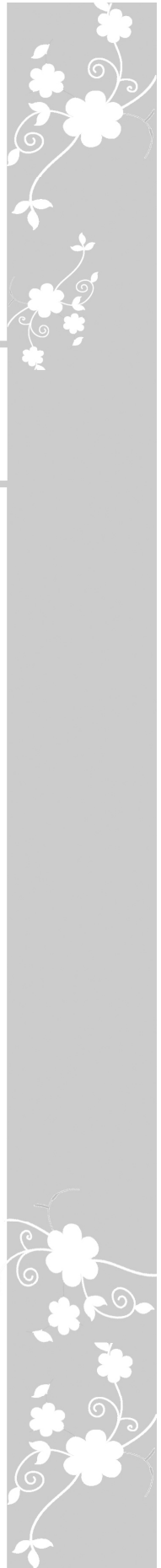
부록

1. 회의록 요약	99
2. 참고문헌 목록	102
3. 선정된 1등급 어휘 목록	106



제 1 장

사업 개요



1. 사업 목적

본 사업의 목적은 국민 전체를 대상으로 한국어의 사용 양상을 파악하여 수준별 어휘 목록을 구축하기 위한 기초 단계로서 국어 기초 어휘를 선정하고 1등급 어휘 목록을 제시하는 것이다. 이로써 교과서 등의 교육 자료 개발 및 감수, 국가 고사 출제, 국민의 국어 능력 평가를 비롯한 각종 기초 조사 등의 근거로 활용될 수 있는 수준별 어휘 목록 구축에 이바지하고자 한다. 이를 위한 세부 목표는 다음과 같다.

- 기초 어휘 목록 선정용 자료 수집 및 정비
- 기초 어휘 선정 방법론 제시
- 기초 어휘 예비 목록 및 1등급 어휘 목록 제시

4차 산업혁명 등 최근 대규모 말뭉치의 구축 및 분석, 정제 등에 대한 수요와 관심이 높아지고 있고, 그 기술이 발전해 가고 있는 현 시점에 국민 전체를 대상으로 한 한국어 사용 양상에 대한 전체적 조망이 필요하다. 이에 따라 ‘2021년 국어 기초 어휘 선정 및 어휘 등급화 연구’는 교육 분야, 공공 분야, 산업 분야 등 다양한 분야에 활용할 수 있는 어휘 목록 및 어휘 등급화를 제안하기 위하여 수행되었다. 본 사업의 결과물은 국내외에서 활용할 수 있는 언어 자료 자체로서, 또는 다양한 언어 자료를 생산, 활용, 평가할 수 있는 보조 자료로서 사용될 수 있는 자료로서 의의를 갖는다. 사업 결과물의 다양한 의의는 다음과 같다.

먼저, 한국어 사용 양상은 곧 대한민국 국민의 언어 사용 양상과 크게 다르지 않다. 국민들이 한국어를 어떻게 사용하고 있는지 알아볼 수 있는 다양한 지표가 있는데, 그 중 하나가 바로 어휘 사용과 관련된 것이다. 특히 기초 어휘는 그 사용량이 매우 높고 기초 어휘를 통해 다른 어휘나 개념을 이해하게 되므로 중요도 역시 아주 높다고 할 수 있다. 국민들의 어휘 사용 양상을 관찰하는 것은 국민들의 언어 사용 양상을 기반으로 한, 대표성을 가지는 말뭉치를 구축하고 그 말뭉치에서의 어휘 사용 양상을 살펴보는 것과 같다고 할 수 있다. 따라서 국민 전체를 대상으로 한 한국어의 사용 양상을 파악하기 위해 균형 말뭉치를 구축하는 것은 필수적이다. 또한 이를 기반으로 하여 언어 사용 양상을 파악할 수 있는 한국어 기초 어휘를 선정하는 것은 국민들의 한국어 사용 양상을 관찰할 수 있게 해 준다.

또한 본 사업에서는 교과서 등의 교육 자료 개발 및 감수, 국가 시험 출제, 국어(또는 한국어) 능력 평가 등에 기초 자료로서 쓰일 만한 수준별 어휘 목록 구축을 제공한다.

즉, 국민 전체의 한국어 사용 양상 파악을 통하여 교육 분야뿐만 아니라 국어 정책이나 공공 언어 전반에 이용할 수 있는 수준별 한국어 기초 어휘 목록을 구축하는 것을 목표로 하였다. 기본적으로는 어휘 교육과 관련하여 학년별·과정별 교육 목표를 설정하고, 학생들의 어휘 수준을 평가하는 지표로서 등급화된 어휘 목록을 활용할 수 있다. 이에 따라 본 사업에서는 수준 및 필요에 맞는 기초 어휘 목록 선정과 그 등급화를 실현하고자 하였다. 뿐만 아니라 기초 어휘 선정은 국민의 일반적인 문해력에 맞추어 사용되는 공공 언어의 사용과도 관련이 있다. 문해력에는 어휘에 대한 이해도 포함되는데, 본 사업에서 선정한 기초 어휘 목록은 공공 언어 사용 및 평가 전반에도 사용할 수 있을 것이다.

아울러 본 사업은 1~3차년도(2017년, 2018년, 2020년)에 한국어 기초 어휘 목록 선정 및 등급화를 목적으로 진행된 관련 연구를 바탕으로 한다. 1~3차년도 연구에서는 기초 어휘 목록 선정 및 등급화를 위한 이론적 배경을 제시하고 기반 자료를 수집·보완하였다. 따라서 본 사업은 1~3차년도 연구 결과물에 바탕을 두고 진행되었다. 또한 선정된 기초 어휘를 통해 1~3차년도 연구 결과물을 검증하는 역할도 할 수 있을 것으로 기대된다.

2. 사업 범위

본 사업은 기초 어휘 목록 선정용 자료 수집 및 정비, 기초 어휘 선정 방법론 연구 및 설계, 기초 어휘 예비 목록 및 1등급 어휘 목록 제시의 단계로 진행되었다. 각각의 세부 내용은 다음과 같다.

먼저, 기초 어휘 목록 선정용 자료를 수집하고 정비하였다. 이를 위하여 기초 어휘의 범위 및 선정 단위를 확정하였으며, 기존 어휘 등급 목록을 통합하여 기존 목록의 동형어를 검토하는 등 어휘 목록을 정제하였다. 또한 기초 어휘 목록 선정 및 등급화 방법론과 관련한 선행 연구들을 바탕으로 정성적·정량적 선정 및 등급화 방법론을 제시하였다.

특히 정량적 어휘 선정 방법론을 적용하기 위하여 기초 어휘 선정용 말뭉치를 설계 및 구성하였으며, 구축된 말뭉치를 정제하여 기초 어휘 후보 목록을 추출하였다. 이 과정에서 기초 어휘 선정용 최소 말뭉치 규모를 분석하는 과정을 수행하였다.

정량적 기준으로 추출된 기초 어휘 후보 목록으로 정성적 평정을 진행하였는데, 평정 주체는 어휘 전문가, 어휘 사용자, 교육 전문가의 세 주체로 설정하여 어휘 조사를 진행하였다. 수행된 어휘 조사 결과를 바탕으로 하여 최종 1등급 어휘 목록을 선정하였으며, 사업 수행 과정에서의 결과를 바탕으로 하여 기초 어휘 등급화안, 기초 어휘 후보 목록 또한 제시하였다.

3. 사업 수행

본 사업은 연구 내용에 따라 어휘 목록 연구팀, 말뭉치 자료 분석팀, 정성적 평정팀으로 분과를 나누어 연구 인력을 구성하고 과업을 수행하였다. 또한 연세대 언어정보연구원의 인력, 장비 자원을 활용하여 전산 지원팀을 따로 꾸려 각 분과의 과업 수행을 지원하도록 하였다. 각 분과의 과업 수행 흐름을 그림으로 보이면 아래와 같다.

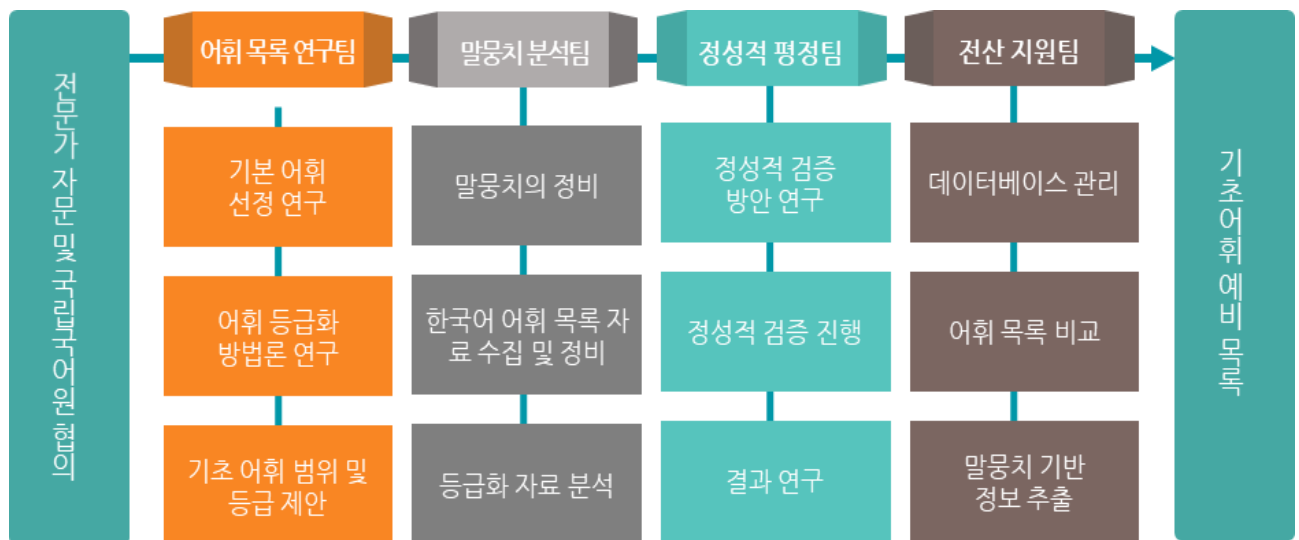


그림 1 분과별 과업 수행 조직도

또한 이와 관련하여 각 분과가 맡은 과업 수행 내역을 보이면 아래와 같다.

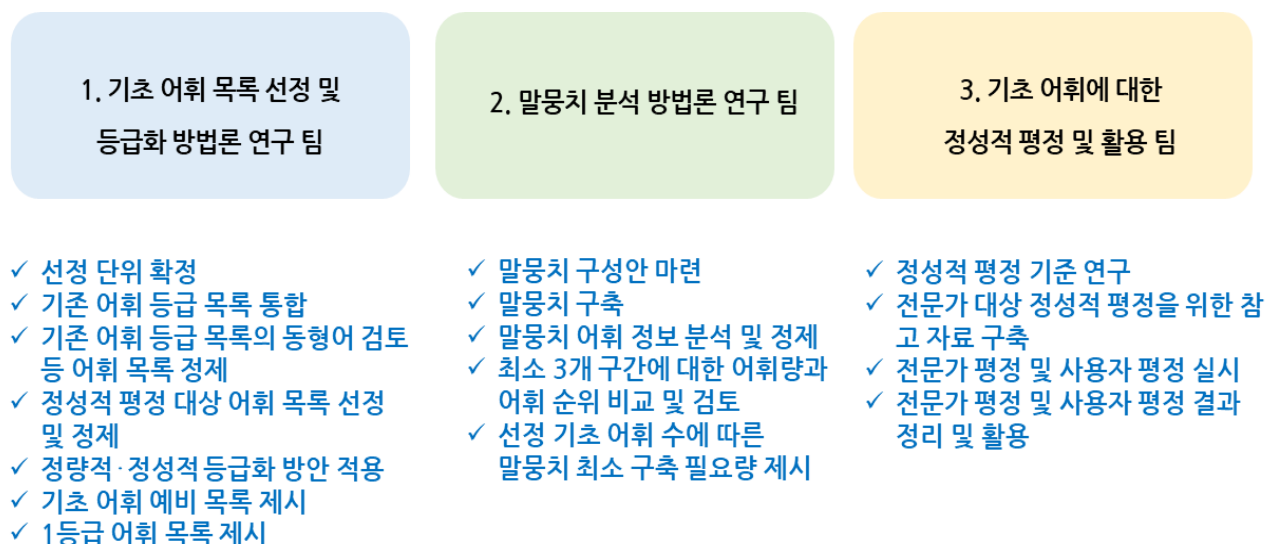


그림 2 분과별 과업 수행 내용

이 장에서는 각 분과의 과업 내용을 절로 나누어 좀 더 살펴보도록 한다.

3.1. 기초 어휘 목록 선정 및 등급화 방법론 연구

1) 기초 어휘의 개념, 범위 및 선정 단위 확정

본 사업에서 정의한 기초 어휘의 개념과 범위는 다음과 같다. 먼저, 기초 어휘는 한국어를 사용하여 일상 언어생활을 하는 데 꼭 필요한 어휘의 목록으로 정의하였으며 범용적 목적의 어휘 목록을 선정하고자 하였다. 그 중 가장 필수적이고 가장 많이 쓰이는 어휘 5,000여 개를 1등급 어휘로 선정하였다.

다음으로, 선정 단위는 단어로 정하였다. 또한 동형어가 구분된(표준국어대사전의 현재 웹사전 번호) 어휘 목록을 만들고자 하였다. 선정 단위가 단어이므로 조사, 어미 등의 문법소, 구 단위는 제외하였다. 고유명사는 인지도가 높은 일부만 포함하였다.

2) 기존 한국어 어휘 목록 자료 검토 및 정제

먼저, 등급화 선정을 위한 참조 자료로서의 기선정 어휘 목록을 수집하여 검토하였다. 검토한 어휘 목록을 보이면 아래와 같다.

목록명 또는 사업명	규모	비고
한국어 교육을 위한 기초 어휘 (서상규 외, 1998)	기본어휘 후보 목록 5,000개	빈도수, 교과서 중복 출현도, 기존 어휘 목록 등과의 중복 출현도 등을 종합적으로 고려
연세 한국어 사전 표제어 목록(1998)	약 52,000개	사전편찬용 말뭉치를 구축하여 말뭉치 사용 빈도 14이상인 단어로 구성.
현대 한국어의 어휘 빈도(서상규, 1998)	-	연세말뭉치 말뭉치 빈도 7 이상의 어휘와 빈도 정보 수록
한국어 교육 기초 어휘 의미 빈도 사전 개발(서상규 외, 2000)	1,087개의 낱말	100만어절 말뭉치에 의미주석하여 의미별 빈도를 계산
국어 교육용 등급별 어휘 1,2등급 (김광해, 2003)	약 5,000개	
외국인을 위한 한국어 학습	약 5,000개	공통 중요어휘(2,028개), 한국어교과서

사전(서상규 외, 2004~2006)		중요어휘(2,025개), 기본어휘목록 및 사전 표제어의 공통 중요어휘(1,977개) 표시
한국어 기본어휘 의미 빈도 사전 (서상규, 2014)	약 7,000개	100만어절 말뭉치에 의미주석하여 의미별 빈도를 계산
한국어 구어 빈도 사전(서상규, 2015)	33,701개	1992년~2005년 사이에 녹음된 실제 구어 자료(연세구어말뭉치, 약 100만 어절)의 어휘 빈도/ 정규화된 빈도수 반영
연세초등국어사전 제3판(연세대 언어정보연구원, 2018)	약 4만 개	개정 교육 과정 초등학교 교과서 어휘 반영 (2,000 항목의 중요 어휘 목록 포함)
연세초등국어사전 제4판(연세대 언어정보연구원, 2020)	약 4만 개	개정 교육 과정 초등학교 교과서 어휘 반영 (2,000 항목의 중요 어휘 목록 포함)
초등학교 교과서 출현 어휘에 대한 예문 선정(교육과정평가원, 2021)	약 34,000개	교과서 출현 정보 및 각종 기초 어휘 목록 분석 정보에 기반하여 3개의 학년군으로 등급화한 초등학생용 어휘

표 1 수집·검토한 어휘 목록 자료

본 사업팀의 위의 어휘 목록들을 교차 검토하여 중복되는 어휘와 중복되지 않는 어휘들을 검토하였다. 또한 기존 어휘 등급 목록의 동형어 구분자가 각자 상이하기 때문에 교차 검토를 위해 본 사업의 동형어 구분과 일치하도록 어휘 목록을 정제하는 작업을 하였다.

3) 정량적 · 정성적 등급화 방안 연구

다음의 어휘 목록들을 통합하여 기초 어휘 후보 목록을 작성하였다.

- 2억 어절 균형 말뭉치 빈도 목록
- 2억 어절 균형 말뭉치 텍스트 분포 및 산포도를 반영한 기초 어휘 점수 순위 목록
- 연세초등사전 등재 여부 및 연세 초등 국어사전 중요 어휘(약 3,000개)
- 서상규(2014) 한국어 기본어휘 목록 (약 3,000개)
- 김광해(2003) 국어 교육용 등급별 어휘 1, 2등급(약 5,000개)
- 교육과정평가원 등급화 어휘 (초등 교과서 학년군별어휘 목록 약 20,000개)

또한, 말뭉치를 통한 기초 어휘의 정량적 선정 방법론을 통해 균형 말뭉치 점수 순위가 5,000위 내외이나 다른 등급 목록에서 높은 등급에 속하는 어휘, 균형말뭉치점수 순위가 5,000위 이상이나 기존 등급화 목록에서 낮은 등급에 속하는 어휘 약 7,600개를 평정 후보 어휘로 선정하였다.

이때, 균형 말뭉치 점수 순위가 모두 5,000위 이내이고 기존 등급화 목록 모두에서 1등급에 준하는 낮은 등급에 속하는 어휘는 1등급 어휘로 우선 선정하고 평정 대상에서 제외하였는데 약 2,400개의 어휘이다.

그리고 앞서 선정한 약 7,600개의 평정 후보를 바탕으로 하여 1) 어휘 전문가 평정, 2-1) 사용자 조사(교수 학습 현장 설문), 2-2) 교육 전문가 평정을 수행하였는데 각각의 평정 과정을 간략히 제시하면 다음과 같다. 먼저, 어휘 전문가 평정 단계에서는 학령전기 자녀가 있는 어휘 전문가들이 약 7,600개 어휘를 검토하여 1등급 1,945개, 평정 필요 어휘 1,580개를 선별하였다. 다음으로 사용자 조사(교수 학습 현장 설문) 단계에서는 초등학교 1학년생, 초등학교 1학년 교사, 유치원 교사들이 평정 필요 어휘 1,580개를 검토하여 1등급 어휘 1,028개를 선별하였다. 그리고 교육 전문가 평정에서는 국어 교육 전문가들이 선정 및 평정 목록을 검토하여 삭제 검토 목록을 작성하였으며, 또한 유아용 교재, 교안 등을 참고하여 어휘장 보충 어휘, 유아용 어휘의 추가 목록을 작성하여 어휘 목록을 보완하였다.

이러한 정량적·정성적 기초 어휘 선정 결과를 바탕으로 어휘 목록 연구팀에서 검토 의견에 따라 일부 어휘를 추가 또는 삭제하여 최종적으로 1등급 어휘 총 5,000개를 선정하였다.

정성적 평정 및 선정 과정을 그림으로 간단히 나타내면 다음과 같다.

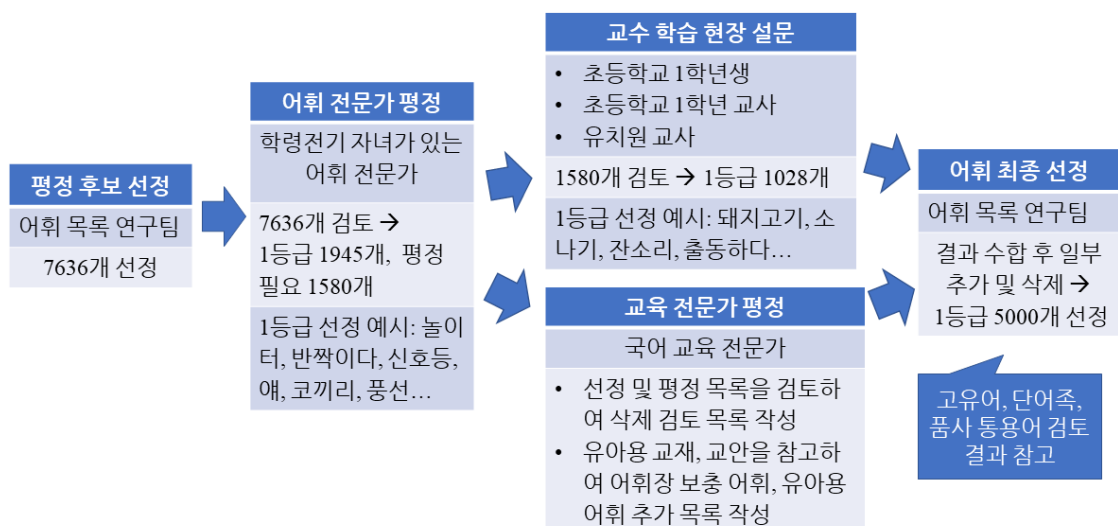


그림 3 어휘의 정성적 평정 및 최종 선정 과정

4) 기초 어휘 등급화안 제시

1~3차년도(2017년, 2018년, 2020년) 연구 결과를 바탕으로 본 사업의 말뭉치 분석 연구 결과, 국내외 교육용 기초 어휘 선정 현황, 전문가 자문 등을 토대로 한 전체 기초 어휘 등급(구간)을 제시하였다.

연령 (만)	학년	어휘 수	어휘 수 (누적)	보유 어휘 목록	등급(제안)	
4-5세	어린이집	2,000	2,000	‘연세초등국어사전’ 중요 어휘 2,000	1	별 ***
5-6세	유치원	3,000	5,000	‘A Frequency Dictionary of Korean’ 문구어 고빈도 어휘 5,000		
				‘한국어 교육을 위한 기초 어휘’ 5,000		
7세	초1	1,000	6,000		2	별 **
8세	초2	1,500	7,500	‘한국어 기본어휘 의미 빈도 사전’ 기본어휘 표제어 목록 7,000		
9세	초3	2,500	10,000		3	
10세	초4	3,000	13,000			
11세	초5	4,500	17,500		4	
12세	초6	5,500	23,000			
13세	중1	5,500	28,000		5	별 *
14세	중2	5,500	34,000	‘초등 어휘에 대한 초기 등급화 데이터’ 34,000		
15세	중3	6,000	40,000	‘연세초등국어사전’ 표제어 목록 40,000		

표 2 어휘 등급화안(제안)

5) 1등급 어휘 목록 및 기초 어휘 후보 목록 제시

어휘 등급화안에 따라 정량적 수치와 정성적 평정 결과를 종합한 1등급 어휘 목록 및 기초 어휘 후보 목록을 제시한다.

3.2. 말뭉치 분석 방법론 연구

1) 말뭉치 구성안 마련 및 구축

말뭉치의 필수 요건이라 할 수 있는 ‘균형성’과 ‘대표성’은 말뭉치의 활용 목적에 따라 다르게 구현될 수 있는데, 본 연구에서는 기초 어휘의 개념에서 출발하는 동시에 활용 가능한 기구축 언어 자원을 점검하여, 기초 어휘 후보 목록의 추출에 합당하고 가능한 말뭉치를 설계하고 구축하는 것을 기본 원칙으로 하였다. 본 사업에서는 2억 어절 규모의 기초 어휘 선정용 말뭉치를 구성하였으며 문어-구어 매체 간 균형, 장르 구분의 세분화 정도 등을 고려하여 말뭉치를 구성하고자 하였다. 특히 발달 단계 자료를 포함하여 어린아이들이 사용하는 어휘를 포착할 수 있도록 하였다. 시기적으로는 맞춤법이 어느 정도 고정된 1945년 이후 자료로 구성하되 시기별로 고르게 많이 사용된 어휘를 기초 어휘로 선정하고자 하였다.

2) 말뭉치 어휘 정보 분석 및 정제

구성한 말뭉치는 형태이미 분석을 수행하여 동형어를 구분할 수 있도록 하였으며 분석의 정확률을 높이기 위해 고빈도 어휘 빈도 목록에 포함된 어휘 중 중의성이 있는 어형에 대해서는 오류 교정 작업을 실시하였다.

3) 기초 어휘 후보 목록 추출

구성한 말뭉치를 바탕으로 하여 어휘의 빈도, 범위, 산포도 등의 정보를 기초 어휘 점수 공식에 대입하여 어휘 점수를 산출하였다. 어휘 점수와 빈도 등을 종합적으로 고려하여 기초 어휘 후보 목록을 추출하였으며, 이 중 빈도 순위와 기초 어휘 점수 순위가 크게 다른 경우에는 정성적 평가의 대상으로 처리하였다.

4) 선정 기초 어휘 수에 따른 말뭉치 최소 구축 필요량 제시

본 사업의 목표 중 하나인 최소 3개 구간에 대한 어휘량 및 어휘 순위의 비교·검토를 수행하였으며 구체적으로는 1억 어절 균형 말뭉치, 2억 어절 균형 말뭉치, 20억 어절 규모 말뭉치를 대상으로 비교를 진행하였다. 말뭉치별 어휘량 및 고빈도 실질어휘의 누적 비율 등을 비교한 결과, 1억 어절 규모의 균형 말뭉치로도 기초 어휘 후보 목록을 선정할 수 있다는 결론을 도출하였다.

3.3. 기초 어휘에 대한 정성적 평정 및 활용

1) 정성적 평정 기준 연구

정성적 평가 기준을 마련하기 위하여 선행연구를 검토하였으며 이를 바탕으로 하여 기초 어휘, 특히 1등급 어휘의 선정에 대해 어휘 전문가 평정, 사용자 조사, 교육 전문가 평정 등을 수행하여 선정하고자 하였다.

2) 전문가 대상 정성적 평정을 위한 참고 자료 구축

정성적 평정의 후보가 될 어휘 목록은 기존 어휘 등급 목록과 균형 말뭉치의 정량적 분석 결과를 종합하여 선정하였다. 후보 목록 선정에는 『연세초등국어사전』 등재 표제어 목록, 서상규(2014), 김광해(2003), 교육과정평가원 등급화 어휘 등을 활용하였으며 어휘 자료를 교차 검토하여 1등급 어휘 후보 목록 약 1만 개를 추려 내었다.

3) 전문가 평정 및 사용자 평정 실시

이상의 평정 후보 목록 약 7,600개에 대하여 전문가 집단과 어휘 사용자 집단을 대상으로 어휘 친숙도 조사를 단계적으로 실시하였다. 전문가 집단은 어휘 전문가와 국어 교육 전문가로 구성하였고, 어휘 사용자 집단은 교수 학습 현장의 초등학교 1학년 교사, 유치원 교사, 초등학교 1학년 학생으로 구성하였다. 정성적 평정의 1단계는 평정할 후보 목록 약 7600개, 정확히는 7636개에 대한 어휘 전문가의 평정이다. 1등급 어휘는 주로 학령전기에 습득하여 지속적으로 사용하는 어휘이므로 이 평정에 참여할 어휘 전문가는 7세 자녀를 두고 있어 1등급 어휘 사용자의 언어생활을 관찰 또는 확인 가능하면서 사전 편찬 경험이 풍부한 연구자로 구성하였다. 어휘 전문가 평정을 통해 다음 단계 평정 후보 목록을 선정한 뒤, 초등학교 1학년 교사 및 학생, 유치원 교사를 대상으로 하여 어휘 사용자의 친숙도를 조사하기 위한 교수 학습 현장 설문을 실시하였다. 다음으로 교육 전문가 평정 단계에서는 앞선 단계에서 확보된 어휘 목록을 검토하여 목록에서 삭제되는 것이 적절하다고 생각하는 어휘를 선정하거나 목록에 추가되는 게 바람직하다고 생각하는 어휘를 발굴하는 작업을 거쳤다.

4. 사업 결과

본 사업의 결과물은 다음과 같다.

1) 기초 어휘 등급화안

기초 어휘 전체를 의무교육인 중학교를 이수하였을 때 학습하거나 습득할 수 있는 어휘로 보고 그에 따른 어휘량은 약 4만여 개로 산정한다. 또한, 교육 과정을 감안하여 학령전기 1등급, 초등 2,3,4등급, 중등 5등급의 5개 등급으로 등급화할 것을 제안한다.

2) 1등급 어휘 목록

약 5,000개의 1등급 어휘 목록을 선정한다. 1등급 어휘 목록은 말뭉치를 통한 정량적 분석 결과 및 기존 어휘 등급 목록, 전문가의 정성적 평정 결과, 사용자 조사 결과 등을 아울러 선정한 결과물이다.

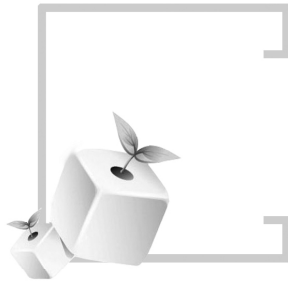
3) 기초 어휘 후보 목록

약 40,000여 개의 기초 어휘 후보 목록을 제시한다. 기초 어휘 후보 목록은 본 사업의 결과로서 마련된 기초 어휘 선정을 위한 말뭉치 구성안과 적정 규모 정보에 기반하여 정량적 기준으로 선정한 결과물이다.

4) 연구의 의의 및 활용 방안

본 사업의 결과물을 통해 기존에 구축된 다양한 국어 자료의 정비를 통해 기초 어휘 목록의 타당성을 제고할 수 있다. 둘째로, 기존 자료 이외의 기초 어휘 목록의 검증 및 활용에 도움이 될 만한 추가 자료를 확보하였다. 다음으로, 수준별 어휘 등급화 목록 작성을 위한 방법론을 수립하였다. 학계나 업계의 수준별 어휘 등급 목록은 주로 여러 유형의 어휘 목록을 통합하여 도출하는 메타계량법에 의존하여 작성해 왔으나 실제 사용 양상을 반영하는 정량적 검증, 교육 현장 전문가의 정성적 검증을 통해 새로운 방법론에 의한 어휘 등급별 목록을 구축하고자 하였다. 또한 본 사업의

결과물은 교과서 개정 및 보완, 어휘 교육 등에 활용되어 국민의 언어생활 향상에 기여할 수 있다. 무엇보다 본 과제의 정교한 연구 방법론을 활용하여 한국어 어휘 사용 양상을 파악하는 연구 및 조사가 지속적으로 이루어진다면 한국어 어휘 사용 양상의 변화 등을 파악하는 데도 큰 도움이 되리라 생각된다.



제 2 장

선행 연구



1. 말뭉치 구축 및 말뭉치 기반 어휘의 계량적 연구

말뭉치를 활용한 어휘의 계량적 연구에는 서상규 외(1998), 서상규(2006, 2017), 김한샘(2003, 2005), 장경희 외(2012) 등이 있다. 먼저, 서상규 외(1998)에서는 말뭉치 분석을 통해 한국어 교육을 위한 교육용 기초 어휘를 선정하였고, 서상규(2006)에서는 학습 사전 편찬과 기본 어휘 선정에 위한 말뭉치 활용 방안을 모색하였으며, 서상규(2017)에서는 한국어 빈도 조사의 성과를 분석하고 언어 연구와 빈도 조사에서 말뭉치를 효율적으로 활용하기 위해 방안을 제시하였다. 다음으로, 김한샘(2003)은 국어교육용 어휘의 단계별 선정 사업의 일환으로, 한국 현대 소설에 대한 어휘 조사가 이루어졌으며, 김한샘(2005)에서는 현대 국어 사용 빈도에 대한 조사가 진행되었다. 또한 장경희 외(2012)는 구어 어휘에 관한 연구로, 초·중·고등학생들이 실제 사용하고 있는 어휘를 형태적, 의미적으로 분석하였다.

○ 서상규·남윤진·진기호(1998), 외국어로서의 한국어 교육을 위한 기초 어휘 선정-기초 어휘 빈도 조사 결과, 문화관광부·한국어 세계화 추진 위원회.

연구 목적 및 필요성	- 대규모의 실제 언어 자료(말뭉치)에 대한 국어학적, 정보학적, 계량적 분석을 통하여 외국어로서의 한국어 교육을 위한 교육용 기초 어휘를 선정하는 것을 목적으로 한다.
연구 방법	1) 기본 어휘 선정 방법 ① 기본 어휘의 개념을 수립한다. ② 말뭉치를 대상으로 한 어휘 빈도 조사를 한다. - 대규모(4,000만 어절 이상)의 연세 말뭉치의 어휘 빈도를 조사한다. - 교육 분야 텍스트의 어휘 빈도를 조사한다: 한국어 교육용 말뭉치를 구성하고 어휘의 빈도를 추출한다. - 위의 두 가지 어휘 빈도 자료의 상관성을 분석하고, 아울러, 기존의 본격적인 어휘 빈도 결과의 대표적인 ‘조선어 빈도수 사전’의 어휘 빈도의 분포와 대조한다. ③ 어휘의 분포와 단계를 분석한다. - 어휘 사용률에 바탕을 둔 어휘 분포를 분석한다. - 어휘 증가율에 따른 구간을 분석한다. - 기본 어휘 후보 목록을 확정한다: 약 5,000어휘, 사용률 80%이상.

	<p>④ 어휘 목록을 검증하고 보완한다.</p> <ul style="list-style-type: none"> - 국어 교육을 위한 기본 어휘 자료와 대조, 검증, 보완한다. - 한국어 교재의 어휘 사용의 분석 결과와 대조한다. - 한국어 사전의 ‘정의’에 나타나는 풀이말 용어의 어휘 빈도 조사 결과와 대조, 검증, 보완한다. - 기본 어휘의 중복도에 의해 중요도를 설정한다. <p>⑤ 한국어 교육의 단계의 목표에 따라, 기본 어휘 구간과 목록을 확정한다.</p> <p>2) 어휘 선정 과정</p> <p>① 본 연구에서는 기본 어휘 선정 작업을 위한 기초 자료로서 약 5,000개의 어휘 목록을 ‘기초 어휘 후보군’으로 추출하였다.</p> <p>② 기초 어휘 후보군 중에서 교과서, 조선어 빈도수 사전, 한국어 교재 등의 검증 자료들 가운데 하나에라도 나온 어휘 목록을 추출하였다. 그 결과, 기초 어휘 후보군에 속하면서, 검증 자료에서 1번 이상 출현하는 어휘가 후보군 중 약 48.9%인 2,337개로 나와 기초 어휘 후보군이 개연성이 있음을 시사하였다.</p> <p>③ 한국어 교육용 말뭉치 약 100만 어절을 대상으로 한 어휘 빈도표와 ‘기초 어휘 후보군’의 어휘 목록의 일치도를 살펴보았다. 앞의 검증 자료와의 합집합에 속한 어휘 2,337개 중 교육용 말뭉치에서 확인되지 않은 것은 단 2개에 불과하다.</p> <p>④ 기초 어휘 후보군 중에서 교과서, 조선어 빈도수 사전, 한국어 교재에 모두 중복되어 나온 어휘 목록을 교집합으로 하여 살펴보았다.</p> <p>⑤ 교육용 말뭉치에서 대상 어휘 500개에 대한 30여만 개의 용례 중 약 10만 개의 용례를 분석 대상으로 하여 의미 빈도 조사를 실시하였다.</p>
연구 결과	<ul style="list-style-type: none"> - 기초 어휘는 지금까지 주관적 가치에 의해 선정되는 경우가 많았으나 이 연구는 객관적인 방법론을 이용하여 기초 어휘를 선정하고자 하였다. 또한 기초 어휘 선정의 실제에 있어서 필요한 여러 가지 실질적인 자료와 방법론을 세밀하게 제시하고 있는데, 이러한 방법론은 실제 어휘 선정에 활용할 수 있다. - 특히 말뭉치의 계량적인 분석을 이용해 기초 어휘를 선정함에 있어서 문제가 될 수 있는 어휘 수 결정과 같은 여러 가지 수치나 기준의 객관성과 신뢰성을 검증하고 있어 활용도가 크다. - 또한 외국어로서의 한국어 교육을 위한 기본 어휘라는 특수 목적 기본 어휘 선정의 방법론을 제시하고 있고 말뭉치 구축 기준에 대해서도 자세하게 언급하고 있어 분야별 기본 어휘 선정에 대한 연구를 진행함에 있어 참고가 될 수 있다.

○ 김한샘(2003), 한국 현대 소설의 어휘 조사 연구, 국립국어원.

연구 목적 및 필요성	<ul style="list-style-type: none"> - 본 사업은 국어교육용 어휘의 단계별 선정 사업의 일환으로, 그 최종 목표는 국어교육 현장에서 이용할 수 있는 신뢰도 있는 국어교육용 기본 어휘 목록을 단계별로 확정하는 것이다. 이는 대량의 문헌 자료를 대상으로 한 어휘 사용 실태 조사 결과와 기존 연구 검토 결과, 전문가 자문을 종합하는 매우 방대한 작업이다. - 어휘 분석이 완료된 소설 문헌 약 100만 어절을 대상으로 하여 국어 사용 빈도 조사 결과를 제시하고자 하였다.
연구 방법	<ul style="list-style-type: none"> - 초판이 1900년 이후 출판된 소설을 연구 대상으로 한다. - 표본 추출 방법으로 확률 통계적 방법과 판단 추출법을 사용하였다. 표본 후보를 형태 분석하여 빈도를 구하고, 1,000만 어절 말뭉치에서 누적 빈도 90%인, 상위 5,000여 개의 목록과 비교하여 그 사용률을 지표로 활용하였다. - 표본 선정 원칙 <ul style="list-style-type: none"> ① 번역 자료는 조사 대상에 포함하지 않는다. ② 외국어, 방언, 전문 용어 등이 많이 섞인 문헌은 배제한다. ③ 한문체이거나 고어 문체인 것은 배제한다. ④ 다양한 표본이 포함되도록 하기 위해서 한 작가의 작품이 3편을 초과하지 않도록 선정한다. ⑤ 입력 오류가 적고 파일의 수집이 용이한 '21세기 세종계획'의 자료를 최대한 활용한다.
연구 결과	<ul style="list-style-type: none"> - 자소, 음절, 어절 등 각종 언어학적 단위에 대하여 각각 가나나순 목록과 빈도순 목록을 제시하였다. - 일반 어휘의 빈도순 목록에서는 누적 빈도를 전체 일반 어휘의 수로 나누어 백분율로 표시한 점유율을 함께 제시하였다.

○ 김한샘(2005), 현대 국어 사용 빈도 조사, 국립국어원.

연구 목적 및 필요성	<ul style="list-style-type: none"> - 국어교육용 어휘의 단계별 선정 사업의 일환이다. - 1900년 이후 출판된 문헌을 대상으로 한 현대 국어 사용 빈도 조사에 해당된다.
연구 방법	<ul style="list-style-type: none"> - 되도록 다양한 문헌을 포함하여 조사 결과가 편중되지 않도록 하기 위하여 표본으로 선정된 문헌의 앞부분부터 약 5,000 어절씩 끊어 300만 어

	절을 목표로 문헌 자료를 구성하였다. 크기가 5,000어절에 못 미치는 표본이 상당수 포함되어 있어 최종적으로 672개의 표본, 300만 어절의 문헌 자료를 수집하여 어휘 조사를 진행하였다.
연구 결과	<ul style="list-style-type: none"> - 300만 어절의 어휘 분석 말뭉치를 분석하여 자모, 음절, 일반 어휘, 조사, 어미, 어절 구성, 구, 어휘 범주, 활용형, 규범 오류형 등의 빈도를 조사하였다. - 목록에서 순위, 빈도, 항목, 풀이, 품사 혹은 범주 등을 제시하였다.

○ 서상규(2006), 한국어 학습 사전 편찬과 기본 어휘의 선정을 위한 기초연구, <朝鮮語研究> 3, 朝鮮語研究會 編, くろしお出版.

연구 목적 및 필요성	<ul style="list-style-type: none"> - 이 연구의 목적은 말뭉치를 분석함으로써 얻을 수 있는 여러 가지 기초적 언어 정보가 한국어 교육을 위한 기본 어휘의 선정을 위한 연구, 한국어 학습 사전의 기본 어휘 표제어나 중요 어휘의 선별 과정에 과연 도움을 줄 수 있는지, 만일 그렇다고 한다면 어떠한 방법으로 이를 활용할 수 있는지 모색하는 것이다.
연구 방법	<ol style="list-style-type: none"> 1) 어휘 빈도 목록상의 중요 어휘 추출 <ul style="list-style-type: none"> - 연세 말뭉치, 한국어 교육 표준 말뭉치, 한국어 교재 말뭉치, 초등학교 전체 교과서 말뭉치, 연세 한국어 사전 뜻풀이 말뭉치 등 5개의 말뭉치 분석 후 어휘 목록을 작성하였다. - 이들의 중복도를 조사하여 중복도 3 이상의 2,051개를 중요어 후보로 선정하고, 고유명사와 복합수사를 제외한 2,028개를 ‘말뭉치 사용 빈도에서의 최종 중요어’로 선정하였다. 2) 한국어 교재 사용 어휘에서 중요 어휘 선정 <ul style="list-style-type: none"> - 한국어 교재 14종에서 중요어 후보를 2,006개를 선정한다. (A) - 교재의 전체 텍스트를 대상으로 중복되는 어휘 목록 작성 후 중복도 6 이상의 어휘 2,407개를 중요어 후보로 선정한다. (B) - 한국어 교재 전체 어휘 빈도를 조사하여 빈도 41 이상의 어휘 2,021개를 중요어 후보로 선정한다. (C) - 위의 세 후보 목록(A, B, C)을 대조하여 중복도 2 이상의 어휘 2,025개를 ‘한국어 교재에서 추출된 중요어’로 선정하였다. 3) 기본 어휘 목록, 사전의 중요 어휘 선정 <ul style="list-style-type: none"> - 기본 어휘 목록 10종 및 학습 사전의 중요어 목록 4종을 통합하여 중복도를 조사하고 중복도 7 이상의 어휘 1,977개를 ‘기본 어휘 목록 및 학습 사전 중요 어휘 목록에서의 중요도 높은 어휘’로 선정하였다.

	<p>4) 학습 사전의 중요 어휘 목록</p> <ul style="list-style-type: none"> - 앞서 작성된 세 종류의 중요 어휘 목록을 대조, 통합하여 얻은 총 2,995개의 단어 목록을 한국어 학습 사전에서의 중요 어휘 목록으로 제안한다.
연구 결과	<ul style="list-style-type: none"> - 말뭉치에 근거한 객관적 빈도 목록, 교재 분석 결과, 기존의 다양한 어휘 목록 및 사전 표제어 목록 등 이용 가능한 모든 어휘 자료를 망라하여 기본 어휘를 선정하였다. 또한 기본 어휘 선정 절차를 자세히 설명함으로써 최종 어휘 목록이 선정된 근거를 명확히 하고 있다. - 이 연구에서는 어휘 활용의 주체와 한국어 교육의 주체들의 문제를 고려해 볼 때 앞서서 이루어진 성과는 물론이고 이제부터 작성될 기본 어휘 목록은 ‘그 자체로서 완성된 닫힌 목록’이 아니라 추후 그를 활용하는 목적이나 여건에 따라서 얼마든지 수정, 보완, 재조정이 가능한 형태의 ‘열린 목록’으로 이해하는 것이 바람직하다고 지적하고 있는데, 이는 앞으로의 어휘 조사 연구에 시사하는 바가 크다. 내국인을 위한 교육용 기본 어휘 선정에 있어서도 확정적인 어휘 목록이 아니라 다양한 목적에 따라 수정, 보완, 재조정하여 이용할 수 있는 어휘 목록으로 구성해야 할 것이다.

○ 서지영(2007), 말뭉치를 활용한 중학교 국어과 학습사전 편찬을 위한 기초 연구, 경기대 교육대학원 석사학위논문.

연구 목적 및 필요성	<ul style="list-style-type: none"> - 이 연구에서는 중학교 국어 교과서 말뭉치를 활용하여 중학교 국어과 학습 사전을 편찬하는 데 필요한 표제어의 선정과 정보 기술 방법을 알아보고, 그 예를 제시하여 학습자들의 국어과 학습에 도움이 되는 국어과 학습 사전 편찬에 도움을 주고자 하였다.
연구 방법	<p>1) 조사 대상</p> <ul style="list-style-type: none"> - 제7차 교육과정의 7, 8, 9학년(중1~중3) 국어 교과서 (총6권) <p>2) 조사 방법</p> <ul style="list-style-type: none"> - 교과서 말뭉치를 구축 - 깜짝새를 이용하여 품사별, 어절별 색인 작성, 어절별 용례 사전 작성 - 학년별, 빈도별, 품사별 어휘 분석 - 표제어 선정 - 국어과 학습 사전을 편찬하기 위한 지침 마련 <p>3) 조사 단위</p> <ul style="list-style-type: none"> - 교과서에 수록된 그대로 입력했다. - 띄어쓰기, 구어체, 비표준어, 준말, 외래어 등도 교과서를 따랐다.

연구
결과

4) 어휘수 : 개별어휘 수 15,431개

1) 학년별 어휘 수

	중1-1	중1-2	중2-1	중2-2	중3-1	중3-2	전체
항목	5698	5651	5920	4874	6818	6037	15431

2) 빈도별 어휘 수

빈도	개별항목 수	누적항목 수	빈도 수	누적빈도 수
12763~1001	56	56(00.4%)	187961	187,961(46.3%)
1000~101	419	475(03.1%)	105047	293008(72.2%)
100~11	2584	3059(19.8%)	77927	370935(91.4%)
10~5	2372	5431(35.2%)	16372	387307(95.4%)

- 고빈도 어휘 56개(0.4%)가 전체의 46.3%를 차지한다. 빈도 5 이상의 어휘 (35.2%)는 전체의 95.4%를 차지한다.

- 빈도 1~4회의 어휘 항목 수는 전체의 64.8%에 해당하므로, 국어 교과서 절반 이상이 빈도수 4회 이하의 어휘들로 구성되어 있음을 알 수 있다. 학습 사전 편찬에 저빈도어 표제어가 반영되어야 한다.

3) 품사별 분석

- 일반명사가 전체의 61.0%를 차지한다.

- 고유명사는 모두 754개가 나타나며 사용 빈도를 고려하여 표제어로 선정해야 한다.

4) 제6차 교육과정의 중학교 국어 교과서 어휘(최진영 2001)와 비교

	중1-1	중1-2	중2-1	중2-2	중3-1	중3-2	전체
6차	3748	3739	4283	4339	4862	4207	12648
7차	5698	5651	5920	4874	6818	6037	15431

- 평균 1600개 증가했다. (신조어가 다수 추가)

	6차		7차	
빈도	어휘 수	비율	어휘 수	비율
100 이상	164	1.28	479	3.10
60-99	0	1.52	286	1.85
50-59	195	1.52	142	0.92
...				
2	2088	16.50	2733	17.71
1	5882	46.50	5013	32.49
합계	12648	100	15431	100

- 평균 상대 빈도 비율은 6차가 8.36회, 7차가 26.28회이다.

- 빈도 1회인 어휘 항목 수가 절반에 가까운 46.5%에서 32.49%로 줄어든 것은 긍정적이다.

○ 장경희·이삼형·이필영·김명희·김태경·김정선·전은진(2012), 초·중·고등학생의 구어 어휘 조사, 지식과 교양.

연구 목적 및 필요성	- 이 연구에서는 한국 초·중·고등학생들이 실제 사용하고 있는 구어 어휘를 형태적, 의미적 측면에서 심층적으로 살펴보는 것을 목적으로 하였다.
연구 방법	<ul style="list-style-type: none"> - 한양대학교 한국교육문제연구소의 ‘연령별 구어 말뭉치’를 분석 대상으로 한다. - 분석 대상 자료의 화자 수는 총 478명(남 230명, 여 248명), 학년별 남녀 각 20명이다. - 녹음은 학교 교실 등 빈 공간에서 조사 대상자 둘 또는 세 명이 짝을 지어 자유롭게 일상적인 대화를 나누도록 하였고, 조사 분량은 각 대화 세트별로 1회에 1시간 동안 연속으로 녹음하는 것을 원칙으로 하였다. - 녹음 조사 후 음성 자료를 문자로 전사하여 기계 가독형 말뭉치로 전환하였다. 그 후 형태 정보 주석, 의미 정보 주석을 하여 분석할 자료 형태로 가공하였다. - 학생들의 어휘 사용 특징이 파악될 수 있도록 단순 사용 빈도 외에도 사용 화자 수, 형태·의미의 하위 유형 등 여러 차원에서 분석을 수행하였다. 특히 고빈도 어휘를 중심으로 학교급별로 그 사용 추이를 분석하여 학교급에 따른 변화 양상을 구체적으로 살펴보았다.
연구 결과	<ul style="list-style-type: none"> - 초·중·고등학생이라는 특정 집단의 언어를 대상으로 한다는 점에서 언어학적 의의가 있다 할 수 있다. - 어휘 사용에 대한 관찰을 통하여 그 사용자의 인지 특성이나 그들이 속한 집단의 문화적 특성 등을 가늠할 수 있다는 점에서 교육적 의의가 있다.

○ 서상규(2017), 한국어 빈도 조사와 말뭉치 주석, <한글> 316, 한글학회, 71-120.

연구 목적 및 필요성	- 이 연구에서는 그간 말뭉치를 바탕으로 이루어진 한국어 빈도 조사의 성과를 분석하여 언어 연구와 빈도 조사에 말뭉치를 효율적으로 활용하기 위해 필요한 요건과 문제점을 구체적으로 밝히고자 하였다. 더불어 빈도 조사의 정밀성과 신뢰도를 좌우하는 요인은 말뭉치의 분석과 주석의 방법과 태도임을 밝혀 주석 말뭉치 가공의 새로운 방향성을 모색하였다.
연구 방법	<p>1) 한국어 빈도 조사의 흐름</p> <ul style="list-style-type: none"> - 한국어 연구에서 계량적 분석 방법에 의한 실제적 성과를 보여준 것은 최현배(1930)이며, 낱말 빈도 조사는 최현배·이승화(1955, 1956)에서 처음 이루어졌다. - 1930년 이래 이제까지의 한국어 빈도 조사 16종을 제시하여 ①빈도 조사의 단위와 목적, ②빈도 조사의 대상과 구성, ③말뭉치 규모와 빈도 조사 결과의 크기,

	<p>④빈도 조사 결과의 구성과 제시 방법을 분석하였다.</p> <p>2) 낱말의 용법과 빈도 조사의 정밀도</p> <ul style="list-style-type: none"> - 낱말 가운데 빈도 조사 결과에 큰 영향을 미치는 품사통용어(동형어 또는 동음이의어), 준말과 구어형, 문법적 변이형태를 최현배·이승화(1956), 문영호 외(1993), 서상규(1998), 조남호(2002), 김흥규·강범모(2004), 김한샘(2005), 강범모·김흥규(2009), 서상규(2015)에서 어떻게 다루었는가를 살폈다. <p>3) 빈도 조사를 위한 말뭉치 주석의 문제</p> <ul style="list-style-type: none"> - 주석 말뭉치를 바탕으로 한 대표적인 빈도 조사인 조남호(2002), 김흥규·강범모(2004), 김한샘(2005), 강범모·김흥규(2009), 서상규(2015)의 주석 말뭉치를 대상으로, 말뭉치 처리 과정의 측면에서 빈도 조사 결과를 좌우하는 요소(수사, 합성어, 고유명사, 생략형의 주석)를 구체적으로 살피고 있다.
연구 결과	<ul style="list-style-type: none"> - 조사 대상의 말뭉치를 하나로 하여 크고 작은 균형 말뭉치를 구성하는 방법에서는 균형 말뭉치 구성의 규모, 표본의 크기와 구성 방법 등에서 충분한 공감대가 이루어지고 있지 않으므로, 되도록 말뭉치가 특정한 사용역을 대표하도록 하는 것이 현실적인 대안이다. - 빈도 조사는 반드시 말뭉치의 주석을 전제로 하는데, 말뭉치의 규모가 커질수록 주석의 정밀성이 저하되어 전체적인 빈도 결과의 신뢰에 문제를 야기할 뿐 아니라, 주석 단계에서 빈도 조사의 목적을 위한 단위 설정과 언어학적 해석 처리가 충분이 이루어지지 않는 문제를 해결해야 하며, 조사 결과의 검증이나 후속 연구를 위해서는 조사 바탕이 된 주석 말뭉치를 공개할 필요가 있다. - 빈도 결과를 제시할 경우 반드시 정규화한 결과를 함께 제시하는 것이 필요하다. - 서상규(2015)에 제시된 바와 같이, 조사 어미의 변이형태나, 모든 어휘의 구어형, 방언형, 축약형 등을 한자리에 묶어서 볼 수 있도록 하는 제시 방법을 발전시켜 나가는 것이 바람직하다.

2. 기초 어휘 관련 연구

기초 어휘 관련 연구로는 임지룡(1991), 한국어능력 평가위원회(1997), 이익환·서상규(2002), 장현진·전희숙(2014) 등이 있다. 임지룡(1991)은 어휘 교육을 위한 기초 어휘 목록을 추출하기 위한 연구이며, 한국어능력 평가위원회(1997)에서는 한국어교육 및 한국어 능력 평가의 모형 수립을 위해 현대 한국어의 기본 어휘 목록을 제시한 바 있다. 또한 이익환(2002)는 기본 어휘 선정 및 사용 실태 조사를 위한 말뭉치를 구축하려는 목적하에 진행되었으며, 장현진·전희숙(2014)는 초등학생의 어휘 지도를 위한 기초 자료 제공에 목적을 둔 연구이다.

- 임지룡(1991), 국어의 기초어휘에 대한 연구, <국어교육연구> 23, 국어교육학회, 87-132.

연구 목적 및 필요성	- 어휘교육을 위한 기초적인 목록을 추출하기 위해 이 연구는 기초어휘론을 택하여 연구를 진행하였다. 이는 의사소통을 위해 사전에 제시된 모든 어휘를 알 필요는 없으며, 일정 정도의 어휘만으로 의사소통이 가능하다는 점에 기반하며, 어휘 빈도수의 조사결과를 보면 어휘수와 사용빈도수의 관계가 반비례하기 때문이다. 또한, 우리말은 배의성이 짙은 개방적 조어력을 지닌 언어이기 때문이다.
연구 방법	<p>1) 어휘조사의 현황 및 기초어휘의 선정원리</p> <p>기초 어휘를 선정하기 위해 절충적 방법을 사용한다. 객관적 방법으로는 어휘 빈도 조사의 고빈도어 참조하면서, 필자의 주관적 판정을 바탕으로 1,500개의 어휘를 분류, 선정한다.</p> <p>2) 고빈도어의 추출</p> <ul style="list-style-type: none"> - 문교부(1956), 우리말 말수 사용의 찾기 조사 - 서정국(1969), 국어기본어휘의 연구-국민학교 국어과 교과서 - 이응백(1972), 국민학교 학습용 기본어휘 - 이연섭, 권경안, 정인지(1980), 한국 아동의 어휘발달 연구(1) - 정우상(1987), 국민학교 교과서 어휘연구 - 국어연구소(1988), 중학교 교과서 어휘(국어) - 임지룡(미간), 제5차 국민학교 국어교과서의 어휘 빈도 조사 : 어휘 빈도조사 대략 상위 빈도어 10% 내 범위, 그 중 제 5차 교육과정에 따른 국민학교 국어교과서 고빈도어 997개는 교육용 기본어휘 <p>3) 의미 분야의 설정</p>

	<ul style="list-style-type: none"> - 필자의 주관적 판정을 가미하여 의미 분야별로 1,500개의 어휘를 분류 및 선정 - 사람에 관한 어휘(인체, 정신, 부류, 기타) - 의식주에 관한 어휘(의생활, 식생활, 주생활, 생필품) - 사회생활에 관한 어휘(사회조직, 제도/관습, 교통/통신, 공공시설, 경제분야) - 교육 및 예체능에 관한 어휘(교육일반, 언어, 문학, 체육/오락, 음악, 미술) - 자연계에 관한 어휘(천체, 지리, 지형, 자연현상, 동물, 식물, 광물) - 감각 및 인식에 관한 어휘(일반부류, 공간, 시간, 수량, 추상) - 동작에 관한 어휘 - 상태에 관한 어휘 - 기타(대명사, 의존명사, 부사, 보조동사/형용사, 관형사)
연구 결과	<ul style="list-style-type: none"> - 기초어휘는 언어생활에서 빈도수가 높고 분포가 넓으며, 이차조어의 근간이 되는 최소한의 필수어를 말한다. 광범위한 어휘 가운데 기초언어를 설정하게 되는 근거로는 언어집단에 필수적이며 공통적으로 사용될 뿐 아니라 이차조어의 근간이 되는 어휘가 제한되어 있기 때문이다. - 이러한 바탕 위에서 이 글에서는 객관적 방법에 의해 선정된 고빈도어와 주관적 방법에 의한 의미 분야를 절충하여 내용어 1,500개의 기초어휘를 선정하였다. 여기서 취한 고빈도 어는 어휘 빈도조사 가운데서 대략 상위 빈도어 10% 이내의 범위에 해당하며, 그 중 제 5차 교육과정에 따른 국민학교 국어 교과서 고빈도어 997개는 교육용 기본어휘인데, 또한 이를 10개의 의미 분야로 분류하였다. - 이 글에서 제시한 1500개의 기초어휘를 중심으로 한 합성과 파생의 이차 어휘를 포함하면 대략 5,000개의 어휘로 확장될 것으로 본다. 이만한 양의 내용어와 조사 및 어미와 같은 기능어 100여개가 붙으면 일상 언어 생활의 80%~90% 정도를 무리 없이 수행해 갈 수 있을 것으로 추정된다. 아울러 이글에서 선정한 기초 어휘의 목록은 어휘연구 및 교육, 사전 및 교과서 편찬에 참고가 될 수 있을 것이다.

○ 한국어능력 평가위원회(1997), 한국어 능력 평가용 기본 어휘표.

연구 목적 및 필요성	<ul style="list-style-type: none"> - 한국어교육 및 한국어 능력 평가의 모형 수립에서의 기초적 자료로서 작성된 현대 한국어의 기본 어휘 목록 제시하였다.
----------------------	--

연구 방법	<ul style="list-style-type: none"> - 연세대학교 한국어 사전 편찬실에서 만든 ‘현대 한국어 총 어휘 빈도표’를 사용하여 기본 어휘표를 작성하였다. 이 연구소에 비치된 약 4,300만 어절의 말뭉치를 형태소 분석하여 얻어낸 자료이다. - 형태소 분석기의 정확도가 대체로 90% 전후로 일컬어지고 있으나, 실제 언어자료를 분석하였을 경우 적지 않은 오류를 포함하는 것으로 확인되어 사후 처리를 통해 오류를 제거하였다. 																																																																															
연구 결과	<ul style="list-style-type: none"> - 총 10,740개 어휘를 제시하고 있으며, 품사에 따라 그 비율을 보면 명사가 약 57%, 동사가 약 20%, 형용사가 약 7%, 부사가 6% 순으로 높은 비율을 차지하고 있다. <table border="1"> <thead> <tr> <th colspan="2"></th><th>어휘 항목 수</th><th>비율</th><th>누계</th></tr> </thead> <tbody> <tr> <td rowspan="4">체언</td><td>명사</td><td>6,134</td><td>57.113%</td><td>6,134</td></tr> <tr> <td>대명사</td><td>64</td><td>0.595%</td><td>6,198</td></tr> <tr> <td>의존명사</td><td>185</td><td>1.723%</td><td>6,383</td></tr> <tr> <td>수사</td><td>49</td><td>0.456%</td><td>6,432</td></tr> <tr> <td rowspan="4">용언</td><td>동사</td><td>2,213</td><td>20.605%</td><td>8,645</td></tr> <tr> <td>형용사</td><td>799</td><td>7.439%</td><td>9,444</td></tr> <tr> <td>지정사(이다)</td><td>1</td><td>0.009%</td><td>9,445</td></tr> <tr> <td>보조용언</td><td>41</td><td>0.382%</td><td>9,486</td></tr> <tr> <td colspan="2">부사</td><td>698</td><td>6.499%</td><td>10,184</td></tr> <tr> <td colspan="2">관형사</td><td>71</td><td>0.661%</td><td>10,255</td></tr> <tr> <td colspan="2">감탄사</td><td>49</td><td>0.456%</td><td>10,304</td></tr> <tr> <td rowspan="4">조사와 어미 등</td><td>조사</td><td>72</td><td>0.670%</td><td>10,376</td></tr> <tr> <td>어미</td><td>317</td><td>2.952%</td><td>10,693</td></tr> <tr> <td>선어말어미</td><td>7</td><td>0.065%</td><td>10,700</td></tr> <tr> <td>접미사</td><td>40</td><td>0.372%</td><td>10,740</td></tr> <tr> <td colspan="2">합 계(총어휘수)</td><td>10740</td><td>99.997%</td><td></td></tr> </tbody> </table>						어휘 항목 수	비율	누계	체언	명사	6,134	57.113%	6,134	대명사	64	0.595%	6,198	의존명사	185	1.723%	6,383	수사	49	0.456%	6,432	용언	동사	2,213	20.605%	8,645	형용사	799	7.439%	9,444	지정사(이다)	1	0.009%	9,445	보조용언	41	0.382%	9,486	부사		698	6.499%	10,184	관형사		71	0.661%	10,255	감탄사		49	0.456%	10,304	조사와 어미 등	조사	72	0.670%	10,376	어미	317	2.952%	10,693	선어말어미	7	0.065%	10,700	접미사	40	0.372%	10,740	합 계(총어휘수)		10740	99.997%	
		어휘 항목 수	비율	누계																																																																												
체언	명사	6,134	57.113%	6,134																																																																												
	대명사	64	0.595%	6,198																																																																												
	의존명사	185	1.723%	6,383																																																																												
	수사	49	0.456%	6,432																																																																												
용언	동사	2,213	20.605%	8,645																																																																												
	형용사	799	7.439%	9,444																																																																												
	지정사(이다)	1	0.009%	9,445																																																																												
	보조용언	41	0.382%	9,486																																																																												
부사		698	6.499%	10,184																																																																												
관형사		71	0.661%	10,255																																																																												
감탄사		49	0.456%	10,304																																																																												
조사와 어미 등	조사	72	0.670%	10,376																																																																												
	어미	317	2.952%	10,693																																																																												
	선어말어미	7	0.065%	10,700																																																																												
	접미사	40	0.372%	10,740																																																																												
합 계(총어휘수)		10740	99.997%																																																																													

○ 이익환·서상규(2002), 기본 어휘 선정 및 사용 실태 조사를 위한 기초 연구, 국립국어원.

연구 목적 및 필요성	<ul style="list-style-type: none"> - 이 연구는 ‘한국어의 기본 어휘를 선정하고 그 사용 실태를 조사’하기 위한 기초 자료로서의 말뭉치를 실제적으로 구축하기 위한 연구를 목적으로 한다. 			
연구 방법	<ul style="list-style-type: none"> - 본 연구의 목적을 충족하는 말뭉치를 구축하기 위해, 확률·통계적인 방법과 논리적인 방법을 병행함으로써 구축 지원 도구가 제공해야 할 일관성과 정확성을 보완하고자 하였다. - 장르 구성 비율에 있어서 기본으로 삼은 세종 균형 말뭉치의 구성비에 대한 문제점을 확인하고 재구성 하였다. 특히 신문 장르에 대해서는 발행년도, 일간신문 			

연구
결과

/전문신문, 주제(하위장르) 지면 할애율, 인터넷 조회율 등을 조사하여 재구성하였다. 또한 각 표본들의 문제점을 기술하여 보완 지침을 명시하였다.

- 구축된 말뭉치는 원시 말뭉치와 주석 말뭉치를 모두 확보하였다.
- 전체 표본 목록은 1243개이며 파일이 확보된 1121개이다.
- 실제 구축된 말뭉치의 구성비는 아래와 같다.

제1분류	제2분류	제3분류	어절수	구축어절 수	파일 수	구축파일 수
구어(10%)	순구어(70%)		420,000	425,352	84	96
	준구어(30%)		180,000	158,041	36	40
문어(90%)	신문(20%)	기타/대화/인터뷰(10%)	120,000	130,403	24	26
		문화/매체/과학(25%)	300,000	271,685	60	54
		사설/칼럼(20%)	240,000	241,226	48	48
		스포츠/연예/취미/생활(15%)	180,000	150,841	36	30
		정치/사회/경제/외신/종합(30%)	360,000	271,637	27	54
		총류		140,582		28
	책,상상(25%)	동화(10%)	150,000	148,645	30	33
		중·장편(50%)	750,000	774,551	150	155
		단편(20%)	300,000	281,411	60	63
		희곡, 대본(20%)	300,000	299,423	60	66
	책,정보(35%)	교육자료(10%)	210,000	206,644	42	52
		사회(15%)	315,000	320,969	63	64
		예술/취미/생활(15%)	315,000	316,135	63	63
		인문(20%)	420,000	411,515	84	82
		자연(10%)	210,000	215,831	42	43
		체험기술(15%)	315,000	316,266	63	63
		총류(5%)	315,000	317,307	63	63
		잡지(10%) 합계	600,000	601,975	120	120
	총 합계		6,000,000	6,000,439	120	1243

- 장현진·전희숙(2014), 초등학생 교육용 기초 어휘 선정 연구: 저학년 중심으로, <언어치료연구> 23(1), 한국언어치료학회, 157-170.

연구 목적 및 필요성	<ul style="list-style-type: none"> - 본 연구는 초등학생을 위한 효과적인 어휘지도를 위한 기초자료를 제공하는 데 목적이 있다. 예비연구에서는 초등학생 관련 선행 연구를 토대로 기초 어휘 533개를 1차 선정하였다. 예비연구에서 선정한 어휘가 어휘 교수에 적절한지 타당도 검증을 위해 현재 초등학생을 지도하는 교사 10명을 대상으로 어휘 중요도를 5점 척도로 평정하도록 하였다.
연구 방법	1) 연구 대상 (총 2305개) <ul style="list-style-type: none"> - ‘수용·표현 어휘력 검사’(김영태 등, 2009)의 검사 어휘 - 한국어 학습용 어휘 결과 보고서(조남호, 2003)의 연구에서 A등급으로 선정된 982개의 어휘

	<ul style="list-style-type: none"> - 초등학교 저학년 어휘 지도 연구(2003)에서 초등학교 저학년 교과서에 출현 빈도 5회 이상인 217개 어휘 - 초등학교 읽기 교과서 어휘 연구(김저옥, 2004)에서 617개 어휘 - 초등학생의 구어에 나타난 어휘 빈도와 분포도 조사(이필영, 김정선, 2008)에서 489개 어휘 <p>2) 예비연구</p> <ul style="list-style-type: none"> - 어휘의 분포적 특성을 살펴보기 위하여 각 어휘의 종류와 개별 단어의 빈도를 조사하였다. 5개의 연구 대상의 전체 어휘 수는 2305개였다. (5종류의 4개 자료에서 공통되는 어휘: 71개, 3개의 자료에서 공통되는 어휘: 231개, 2개 자료에서만 나타난 어휘: 232개, 이 중 2개 이상의 자료에서 공통되는 533개의 어휘를 추출). 533개의 어휘 중 언어치료 전공자 4명이 검토하였다. 인지 및 기능적 수준에 적절하지 않은 어휘를 삭제하고 수정하여 최종 어휘는 530개로 1차 기초어휘를 선정하였다. <p>3) 본 연구</p> <ul style="list-style-type: none"> - 선정된 어휘를 의미 중심으로 검토하고, 검토한 어휘를 현재 초등 어휘를 지도하는 교사들에게 어휘 지도에 적절한 어휘인지 한 번 더 타당도를 검증하였다. - 초등학교 교사들로부터 어휘의 중요도 평정의 평균이 3.0이상인 어휘만을 추출하여 나온 최종 기초학습 어휘 수는 총 511개로 1차 선정한 530개 어휘 중 96.4%에 해당한다. (중요도 평균 3.0미만-총 19개로 삭제) - 511개 3등급 나눔: 1등급: 중요도 평균 4.5이상/367개, 2등급: 중요도 평균 4.0이상/127개, 3등급: 중요도 평균이 3.0이상/17개
연구 결과	<p>1) 품사별 초등학생 기초 어휘 목록</p> <ul style="list-style-type: none"> - 전체 어휘 511개 중 명사 218개(45.6%), 동사 112개(23%), 형용사 58개(11.4%), 부사 58개, 대명사 17개, 관형사 8개, 감탄사 1개, 수사 18개 <p>2) 의미영역별 초등학생을 위한 기초 어휘 목록</p> <ul style="list-style-type: none"> - 동물, 신체, 사람, 사물, 상태, 시간, 위치, 음식, 의류, 장소, 정도, 탈것, 행위, 수, 기타 15개로 나눌 수 있다. - 행위 의미영역(동사가 대부분)과 상태 의미영역(형용사가 대부분)의 비율이 높았다. - 명사의 의미범주 가운데 높은 비율을 차지하는 의미범주는 사물, 음식, 사람, 신체 등 아동의 생활 주변에서 쉽게 접할 수 있는 어휘들로 구성되어 있다.

3. 어휘 등급화 관련 연구

어휘 등급화와 관련된 연구에는 임철성(2002), 김광해(2003), 조남호(2003), 강현화(2014a, 2014b, 2015), 윤혜경(2016), 최소영(2020) 등을 들 수 있다. 먼저 임철성(2002)에서는 초급 한국어 교육용 어휘를 제시하였고, 조남호(2003)에서도 한국어 학습용 기본 어휘 5,965개를 선정하였으며. 김광해(2003)에서는 교육용 어휘 목록을 등급화하여 제시하였다. 또한 강현화(2014a, 2014b, 2015)는 한국어교육용 초급, 중급, 고급 어휘 선정을 위한 일련의 연구이며, 윤혜경(2016)에서는 초급 교육용 구어 어휘 선정을 시도하였다. 최근의 연구인 최소영(2020)에서는 초등학교 교육용 어휘에 대한 선정 및 등급화가 이루어졌다.

- 임철성(2002), 초급 한국어교육용 어휘 선정 연구, <국어교육학연구> 14, 국어교육학회, 355-385.

연구 목적 및 필요성	- 이 연구는 초급 한국어교육용 어휘를 선정하는 것을 목적으로 한다.
연구 방법	- 빈도와 분산을 함께 고려하는 계량의 방법을 사용하여 객관성을 유지하였다. - 가능한 한 직관에 일치하는 결과가 나오도록 하였다. - 어휘의 체계를 고려하였다. - 1,000개 상당의 어휘를 추출하였다.
연구 결과	- 초급 한국어교육용 어휘는 한국어를 외국어로서 배우고자 하는 초급 단계의 학습자들이 익혀야 하는, 교육의 대상이 되는 어휘를 의미한다. - 먼저 연세 말뭉치에서 빈도 5,000 이상인 어휘 500~600개 정도를 추출하였고, 다음으로 국내 초급용 한국어교육용 교재들에 등장하는 어휘에 대한 계량 분석을 실시하여 어휘를 출하였다. - 두 방법을 통해 얻어진 어휘를 바탕으로 연구자의 직관과 연세 빈도를 고려하여 어휘를 조정하여 총 1,038개 어휘를 선정, 제시하였다.

- 김광해(2003), 등급별 국어 교육용 어휘, 박이정.

연구 목적 및 필요성	- 방대한 양의 표제어를 담은 사전에는 실제로 사용되지 않는 유령어(ghost word)들이 많이 포함되어 있다. 어휘는 인지적 발달 과정에 따라 학습되는 양에 확실한 수준 차를 보이므로 선정된 어휘 목록을 중요도에 따라 등급화할
----------------	--

필요가 있다.

- 어휘 평정 과정을 통하여 선정된 등급별 총어휘 목록은 교육 자료, 교과서, 평가 도구 개발 등을 위한 수준을 정하는 데 필요한 객관적인 자료가 될 수 있으며, 학습 보조물 저술, 사전 및 시소러스 편찬 등에도 활용될 수 있다.

연구 방법

1) 절충적인 방법의 필요

- 무한에 가까운 어휘 자료를 모두 처리 대상으로 삼을 수는 없으며, 조사 목적으로 볼 때 중요한 어휘소들의 빈도치가 반드시 높게 나타나지 않을 수도 있으므로 객관적인 계량에도 한계가 있다. 또한 높은 빈도가 곧바로 그 단어의 교육적 중요성을 의미하지는 않을 수 있다.

2) 메타 계량 방법

- 기존에 계량 처리된 자료들을 충분히 구하여 그 분포 상황과 자료의 타당도를 함께 고려하면서 비교함으로써 중요도를 정한다.

- 기존 자료 등장 어휘를 수합하고 기존 자료 누락 어휘를 수습한다.

- 자료의 분포 정보와 자료의 타당도 정보(주관적으로 결정)를 기준으로 중요도를 판정한다. (참고한 기존 자료: 서상규(2000), 연세한국어사전(1998), 신현숙 외(2000), 김흥규 외(2000), 고려대(2000) 등)

3) 등급별 어휘량

- 玉村文郎편(1989)의 연령에 따른 이해 어휘량 조사 결과를 따라서(6세~20세) 1등급에서 7등급까지의 등급을 선정한다.

4) 국어 교육용과 한국어 교육용 등급별 어휘량 일치

연구 결과

어휘량	누계	국어교육용		한국어교육용			비고	
		등급	개념	4구분	6구분	개념	소사전	중사전
1839	1839	1	기초 어휘	초급	1	자국인과 어휘량을 일치시키는 방향으로 조절함		
4228	6067	2	정규 교육 이전	중급	2			
8361	14428	3	정규 교육 개시-사춘기 이전	상급	3			
					4			
19377	33805	4	사춘기 이후	고급	5			
					6			
32946	66751	5	전문화된 지적 성장 단계					
45569	112320	6	저빈도어: 대학 이상, 전문어					
125670	237990	7	누락어: 분야별 전문어, 기존 계량 자료 누락어휘					

- 등급별 어휘수: 총 237,990 개

○ 조남호(2003), 한국어 학습용 어휘 선정 결과 보고서, 국립국어원.

연구 목적 및 필요성	<ul style="list-style-type: none"> - 본 연구는 외국인이 한국어를 배울 때 사용할 한국어 학습용 기본 어휘 5,965개를 선정하였다. 외국인 학습자의 학습 단계를 고려하여 3단계를 구분하였는데 초급 단계(1단계)에 982개, 중급 단계(2단계)에 2,111개, 고급 단계(3단계)에 2,872개를 각각 배정하였다.
연구 방법	<p>1) 자문회의 개최</p> <ul style="list-style-type: none"> - 선정 어휘 수에 관한 논의 *1단계 1000개, 2단계 2000개, 3단계 3000개, 합 6000개 선정 계획) - 조사, 어미를 선정하는데 대한 논의: 조사, 어미는 선정에서 제외하였다. - 여러 품사로 기능하는 단어와 복합어에 대한 논의: 의미 예상 가능한 복합어 선정 및 품사 구분하여 선정 계획 <p>2) 1차 선정</p> <ul style="list-style-type: none"> - 외부 인사(선정 위원 6명) 통한 작업 - 현대 국어 사용 빈도 조사에서 1회 이상 출현 단어 59000여 개-> 빈도 순위 10,452등(출현 빈도 15회 이상) 안에 든 단어만 등급을 판정하는 것으로 결정 -> A~E로 나누어 등급을 매기는 것으로 정하였다. <p>3) 조정 및 최종 검토</p> <ul style="list-style-type: none"> - 기존 발표된 바 있는 한국어 교육 관련 어휘 목록과 비교 : 조현용(2000), 한국어 학습 사전을 만들기 위해 작성한 어휘 목록(서상규, 2002) - 고유명사 목록 전제 재정리 <p>4) 목록 분석</p> <ul style="list-style-type: none"> - 최종 검토 이후 5,965개의 단어 선정을 완료하였다. (명사-동사-부사-형용사 순으로 많음)
연구 결과	<ul style="list-style-type: none"> - 외국인이 한국어를 배울 때 사용할 한국어 학습용 기본 어휘 5,965개를 선정하였다. - 외국인 학습자의 학습 단계를 고려하여 3단계를 구분하였는데 초급 단계(1단계)에 982개, 중급 단계(2단계)에 2,111개, 고급 단계(3단계)에 2,872개를 각각 배정하였다. - 연구원에서는 선정 경위와 목록을 담은 보고서를 발간하였으며 이 보고서를 관련 학자, 기관 등에 제공하였다. (어휘목록 excel파일 포함)

○ 강현화(2014), 한국어교육용 초급 어휘 선정 연구, <문법교육> 21, 한국문법교육학회, 1-26.

연구 목적	- 본 연구는 일반 외국인을 위한 초급 단계의 한국어 교육 어휘를 선정하는 데
-------	---

및 필요성	<p>에 목적이 있다. 그간 한국어 교육계에서는 다양한 어휘 선정의 연구가 진행되어 왔다. 대부분의 연구는 교육용 기본 어휘 혹은 기초 어휘라는 이름으로 연구되었고, 일부는 초급 어휘 혹은 학습용 어휘라는 이름으로 제시되었다. 선행 연구들은 연구의 목적과 방법이 상이하므로 각각의 용어가 표상하는 개념도 동일하지 않고 제시되고 있는 목록의 수도 일치하지 않는 측면이 있다. 선행 연구의 대부분은 특정 교육과정에 초점을 두기보다는 기본 어휘 관점에서의 연구 결과라는 점에서, 이들 연구의 결과를 초급용 어휘에 그대로 적용하기란 쉽지 않다. 이에 본 연구는 선행 연구를 검토하고 차별화된 연구 방법론을 통해, ‘초급’에 적절한 어휘 목록을 선정하였다.</p>
연구 방법	<ol style="list-style-type: none"> 1) 본 연구는 ‘초급 숙달도의 일반 목적의 학습자’를 대상으로 한정하며, 급당 200시간(초급1과 초급2 총 400 시간)의 교육과정을 기반으로 한다. 이는 국내 한국어교육기관의 교육과정의 초급과 TOPIK의 초급, 국제통용 표준교육과정의 초급 기준과 일치한다. 2) 문헌 연구-사례 조사-말뭉치 기반 통계-조사 및 의견 수렴 연구(네 단계 진행) 3) 기존의 어휘 선정 목록과 선행 연구의 이론, 한국어 교재와 한국어 학습 사전 등의 사례 자료를 분석하였다. <ul style="list-style-type: none"> - 연세대학교, 『한국어학당, 연세 한국어』초급 1 2, 연세대학교 출판부 - 서강대학교, 『서강 한국어』초급 1 2, 서강대학교 국제문화원 출판부 - 이화여자대학교, 『이화 한국어』초급 1 2, EPRESS - 고려대학교, 『한국어문화교육센터 재미있는 한국어』초급 1 2, 교보문고 - 경희대, 『한국어』초급 1 2, 경희대학교 출판부 - 공개된 9회-26회 중 변환이 불가능한 2회분을 제외한 16회 분 4) 빈도, 중복도, 친숙도, 활용도 등을 근거로 해서 초급 단계의 한국어 교육 어휘 목록을 선정하였다. 5) 객관적 평정으로 어휘 목록을 확보한 뒤, 어휘 선정의 세부 지침과 전문가 평정을 토대로 최종 어휘를 선정하였다.
연구 결과	<ul style="list-style-type: none"> - 약 1,836개의 최종 어휘 선정하였다. - 국내 한국어교육 기관과 한국어능력시험의 숙달도에 준한 초급 학습자를 위한 어휘 선정하였다. - 객관적인 어휘 선정을 위해서 일반 말뭉치 외에 한국어 교육용 말뭉치(교재 및 토픽)를 구축하였다. - 교수자 집단과 학습자 집단의 친숙도 조사와 전문가 어휘 평정을 통해 객관

	<p>적 방법이 가지는 한계를 극복하고자 했으며, 어휘군별 어휘 선정 지침을 마련하였다.</p> <ul style="list-style-type: none"> - 선정된 초급 어휘가 선행 연구나 한국어 교육 현장에서 제시되는 어휘량과 부합하는지 여부를 검증하였다. - 본고는 비교적 자세하게 선정 절차를 논의하여, 어휘선정 방법의 과정을 밝히고 이를 통해 유사 논의가 활발히 이루어지는 데에 일조하고자 하였다.
--	---

○ 강현화(2014), 한국어교육용 중급 어휘 선정에 대한 연구, <외국어로서의 한국어교육> 40, 연세대학교 언어연구교육원 한국어학당, 1-48.

연구 목적
및 필요성

- 지금까지의 어휘 선정과 관련된 연구는 주로 초급 단계 혹은 변별적 특성이 명확히 드러나는 고급 단계에 집중되어 왔으나 중급 단계의 어휘 선정에 관한 연구는 거의 이루어진 바가 없다. 중급 단계에서의 어휘 사용 범위가 포괄적이고 개념 규정이 어렵기 때문이다. 그러나 교육 과정의 운영 및 평가를 위해서는 체계적인 방법과 절차를 기반으로 한 중급 어휘 선정이 필요하다. 이에 본 연구는 ‘중급’의 숙달도, ‘일반 외국인’이라는 대상의 특성을 고려한, 한국어 교육 현장에서의 활용도가 높은 중급 어휘 선정을 목적으로 진행되었다.

연구
방법

1) 중급 어휘 선정 절차

① 객관적 지표에 의한 어휘 평정

- 선행 연구의 목록 유형별 빈도 및 목록 간 중복도

- 한국어 교육 말뭉치, 일반 말뭉치의 빈도 및 어휘 사용 범위

어휘 군별 평정 지침과 객관적 지표에 따른 연구진 평정

② 절충적 어휘 평정(참고)

- 한국어 교사의 어휘 친숙도

- 어휘 전문가의 평정을 통한 목록 보완

2) 후보 목록의 예비 선정

구분	목록 설정	참조 목록 수	타당성 검증 후 남은 어휘
목록1	‘초급’ 예비 어휘 목록 중 초급 선정 어휘 1,836개를 제외한 목록	2,545	2,149
목록2	중급 한국어 교재에만 있는 목록 (교재 어휘-목록1-초급 1,836)	10,231	272
목록3	중급 토픽에만 있는 어휘 (토픽 어휘-목록1-교재-초급 1,836)	5,330	30

목록4	선행연구에서 선정한 목록에만 있는 어휘 (각 목록-목록1-초급 1836)	4,022	1,842
목록5	국제통용 한국어 표준 모형(2011) 초·중급, 조남호(2003)의 누락 목록	-	895
목록6	1단계 초급 관련어 목록	-	972
목록7	어근, 합성어, 파생어 목록	-	485
목록8	주제별 어휘 목록	-	24
목록9	제거 목록의 재검토	-	138
합계		23,362	6,807

3) 본 목록의 평정

- 앞선 단계에서 선정된 6,807개의 목록을 대상으로 본격적인 평정을 하여 ‘중급’ 단계의 어휘 목록을 선정하는 단계이다.
- 중급 단계의 어휘 특성을 고려한 최적의 목록을 선정하기 위하여 ① 어휘군별 특성에 따른 평정, ② 참조 지표에 의한 평정, ③ 어휘 빈도와 어휘 선정 지침에 의한 평정의 3차례의 평정 절차를 거쳤다(3,884개의 어휘가 선정됨).

4) 어휘 친숙도 조사 및 전문가 평정

- 객관적 지표를 활용하여 선정된 어휘 목록에 대한 현장 적용 가능성을 점검하기 위한 단계로, 선정된 어휘 목록에 대한 한국어 교사의 어휘 친숙도 조사와 전문가 평정을 실시하였다.

연구 결과

- 총 3,855개의 중급 단계 어휘 목록 선정하였다.
- 어휘 신규 목록 외에 초급 어휘 1,836개 중 104개의 다의 항목 145개와 중급 선정 어휘 3,855개 중 1,162개 어휘의 다의 항목 620개가 포함된 것을 고려하면, 약 765개의 다의 항목이 추가되는 것이다.
- 본 연구에서는 한국어 교육 현장에서 활용 가능한 어휘 교육 자료의 마련이라는 실용적인 목적하에 실행되었다.
- 어휘 선정 연구의 방법론 면에서 대안을 제시하고자 함. 본 연구에서는 말뭉치 언어학에 기반을 둔 계량적 방법과 선행 연구의 메타 분석, 어휘 교육의 성과에 따른 중급의 교육 과정에 부합하는 어휘 고려, 한국어 교육 및 어휘 전문가의 어휘 평정 등 다양한 연구 방법을 적절하게 활용하여 객관성과 신뢰성을 갖추고자 하였다.

○ 강현화(2015), 한국어교육용 고급 어휘 선정에 대한 연구, <외국어로서의 한국어교육> 42, 연세대학교 언어연구교육원 한국어학당, 1-28.

연구 목적	- 그간 고급 어휘 선정에 관한 연구는 이론적으로나 실제적으로 거의 이루어진
-------	--

및 필요성	<p>바가 없으며, 고급 교재 개발도 체계화된 고급 어휘 실러버스에 기초하고 있다고 보기 어렵다. 이 연구의 목적은 ‘고급’의 숙달도, ‘일반 외국인’이라는 대상의 특성을 고려한 고급 어휘를 선정하는 것이다.</p>
연구 방법	<p>1) 고급 어휘 선정 절차</p> <ul style="list-style-type: none"> ① 참조 목록 정비를 통한 예비 어휘 목록 선정(8,867개) <ul style="list-style-type: none"> - 고급 단계의 어휘 목록 선정을 위한 참조 목록 설정(29,000개) - 참조 목록 정비를 통한 예비 어휘 목록 선정(29,000개→8,867개) ② 목록의 타당성 검사(8,867→4,788개) <ul style="list-style-type: none"> - [1차]객관적 지표를 바탕으로 한 목록의 타당성 검사(8,867→4,469개) - [2차]고급 교육과정의 주제별 어휘 목록 검토(4,469→4,788개) ③ 어휘의 사용 범위 및 빈도 분석 <ul style="list-style-type: none"> - 한국어 교육 말뭉치(교재, TOPIK)에서의 어휘 사용 중복도 및 빈도 분석 - 강범모·김흥규(2009), 조남호(2002) ④ 조급 어휘의 특성과 객관적 지표에 의한 어휘 평정(4,788→4,986개) <ul style="list-style-type: none"> - [1차]어휘군별, 품사별 편정(4,788→4,778개) - [2차]객관적 지표에 의한 평정(4,778→4,699개) - [3차]어휘 빈도와 어휘 선정 지침에 의한 목록 보완(4,699→5,117개) - [4차]누락 목록 보완(5,117→4,986개) ⑤ 안면 타당성 검사(4,950개) <ul style="list-style-type: none"> - 어휘 전문가의 자문 및 평정 - 평정 결과에 의한 목록 보완(4,986→4,950개) ⑥ 일반인 외국인을 대상으로 한 고급 단계의 최종 어휘 목록 확정 <ul style="list-style-type: none"> - 어휘 선정 지침에 따른 최종 점검 및 보완 <p>2) 평정 지침: 단계별 평정 4단계에서 어휘군별로 평정의 근거가 된 지침을 제시하였다.</p>
연구 결과	<ul style="list-style-type: none"> - 총 4,950개의 고급 단계 어휘 목록 선정하였다. - 본 연구는 한국어능력시험 수준의 숙달도에 기대어 고급 어휘를 선정하는 것을 목표로 진행되었다. - 초급에 비해 중급이나 고급에 이르게 되면 한국어 교재 간 어휘 중복도는 매우 낮고 교재의 주제도 상이하지만, 본 연구는 참조 자료 간에 공통적으로 나타나는 고급용 어휘들을 선정함으로써, 향후 교재 개발이나 평가에 활용할 수 있는 어휘 실러버스를 마련하고자 하였다.

- 윤혜경(2016), 한국어 교육용 구어 어휘 선정 연구-초급 수준의 일반명사를 중심으로, <인문과학연구> 50, 강원대학교 인문과학연구소, 65-88.

연구 목적 및 필요성	<ul style="list-style-type: none"> - 구어말뭉치 자료는 2011년 11월에 세종계획에서 구어균형말뭉치를 구축하여 배포하였다. 이는 현대 한국어의 총체적 현실을 반영하는 대규모의 국가균형 말뭉치로 2001년~2005년 사이에 녹음된 실제의 구어 자료이다. 이러한 구어말뭉치의 구축에 기대어 구어 연구가 본격적으로 시작되었으며, 국제한국어교육학회에서는 2013년에 한국어교육에서의 구어 문법을 주제로 학술대회를 개최하기도 하였다. 그러나 구어에 대한 연구는 대부분 문법과 관련된 연구이며, 구어 어휘에 대한 연구는 아직도 미미한 실정이다. 구어 어휘에 대한 연구는 구어 어휘 자체가 그 속성이 불규칙하고, 사용자에 따라 변수가 많으며, 그 규모가 방대하기도 하지만 이미 구축된 대규모의 말뭉치를 기반자료로 필요하기 때문에 개인적 차원으로 연구하기에는 한계가 있다. - 본 연구는 이러한 문제점을 해결하기 위한 일환으로, 한국어교육 현장에서 ‘초급’ 단계의 ‘일반목적의 학습자’를 대상으로 일상생활에서 자연스럽게 원활하게 의사소통할 수 있는 능력을 키워주기 위한 교육용 구어 어휘를 선정하는 데에 그 목적이 있다.
연구 방법	<ol style="list-style-type: none"> 1) 구어와 구어성에 대한 개념 정리 2) 초급 일반명사 어휘목록 선정: 강현화(2012)의 '일반 외국인을 위한 초급 어휘목록' 자료를 기초 자료로 선정하여 품사별로 분류하였다. (초급 1,836개) 3) 세종계획의 구어형태분석말뭉치(2011)에서 분석 가능하도록 재정리하였다. <ul style="list-style-type: none"> - 첫째, 세종계획(2011)과 마찬가지로 단어의 다의적 의미와 동음이의어를 구분하지 않고 하나의 단어로 처리하였다. 둘째, 강현화(2012)에서는 간혹 두 개의 품사를 하나의 단어로 처리하고 있는데, 본 연구에서는 품사가 다른 단어는 각 품사에 따라 구분하여 다르게 처리하였다. 4) 이러한 과정을 통해서 강현화(2012)의 초급 어휘목록에서 ‘명사’에 해당하는 1,835개의 어휘를 재정리하여 ‘일반명사’에 해당하는 1,107개의 어휘를 선정하였다. 5) 초급의 일반명사에 해당하는 1,107개의 어휘를 한마루2.0(2011)의 검색기를 활용하여 본 연구의 구어말뭉치와 문어말뭉치에서 빈도와 사용률을 분석하였다.

	<ul style="list-style-type: none"> - ‘얘기’는 가장 구어성이 강한 단어이며 점수는 0.002096점이고, ‘세계’는 가장 문어성이 강한 단어로 점수는 0.000692점이다. - ‘직장’이나 ‘값’과 같이 구어말뭉치나 문어말뭉치에서 빈도가 높은 단어지만 구어성 점수나 문어성점수가 낮은 어휘는 구어와 문어에서 두루 쓰이는 고빈도 공통어휘라고 할 수 있다. 이들 어휘는 초급단계의 기초어휘이면서 기본어휘에 가까울수록 많을 것으로 예상된다. - 한국어 모어화자는 말을 할 때에는 ‘엄마, 얘기’를 많이 사용하고, 글을 쓸 때에는 ‘어머니, 이야기’를 많이 사용: 이것은 의미는 같으나 구어성 혹은 문어성 성격이 강한 단어들이 각각 따로 존재한다는 것을 말해준다. - ‘말, 도로’와 같이 구어말뭉치에서 고빈도로 나타나는 단어가 문어말뭉치에서 더 사용률이 높아서 문어성 어휘로 분류: 문어성 어휘라고 해서 ‘글’에서만 사용하는 것이 아니고, 구어성 어휘라고 해서 ‘말’에서만 사용하는 것이 아니라는 것을 보여준다. - ‘교수’와 같은 단어가 구어말뭉치와 문어말뭉치에서 고빈도로 분석: 이는 본 연구의 말뭉치를 구축하는 데 기초자료로 활용한 세종계획말뭉치(2011) 자료가 주로 20대 대학생으로 연령층이나 직업군이 심하게 편중되어 있기 때문에 나타난 결과이다. <p>6) 세종계획구어형태분석말뭉치(2011)의 자료를 기반으로 본 연구의 성격에 맞는 구어말뭉치를 새로이 구축하였다.</p> <p>: 이 구어말뭉치는 세종계획 구어형태분석말뭉치 자료(2011)의 약 60% 정도의 규모로, 주로 20대 대학생의 ‘대화과 독백’을 녹음하여 전사한 자료이다. 이 말뭉치에서 검색되는 일반명사 어휘는 총 10,652개이며 사용률의 합은 0.127131이다, 그 중 누적사용률 80% 에 해당하는 고빈도 어휘는 1,581개이며 빈도합은 126,880개이다.</p>
연구 결과	<ul style="list-style-type: none"> - 본 연구에서 신뢰성과 타당성, 적용가능성이 높은 기존의 교육용 어휘목록 자료를 선정하고, 초급단계의 ‘명사’ 1,835개 어휘를 추출하였다. 또, 본 연구의 구어말뭉치에서 분석할 수 있도록 재정리하여 1,107개의 ‘초급 일반명사’ 어휘를 선정하였다. 본 연구는 위의 두 자료-구어말뭉치와 일반명사 1,107개의 어휘를 기초 자료로, 빈도와 누적사용률을 살펴 ‘구어성 어휘’와 ‘고빈도 어휘’를 분석하고, 고빈도 어휘 647개를 한국어 교육용 어휘로 선정하여 그 목록을 제시하였다. - 앞으로 구어말뭉치에서 검색된 고빈도 일반명사 1,581개의 어휘 중에서 본

	<p>연구에서 선정한 647개 어휘 제외한 934개의 어휘에 대하여 엄선하는 작업이 이루어진다면, 일반명사에 대한 교육용 구어 어휘가 완성될 것이다. 또한 이와 같은 작업들이 총제적으로 이루어진다면 비로소 교육용 구어 어휘에 대한 목록이 완성될 것이다. 이 목록은 한국어 교육현장에 교육과정 설계, 말하기·듣기 교재 편찬, 교육현장의 교수 등에 적용할 수 있는 객관적이고 유용한 자료가 될 것이다.</p>
--	--

4. 말뭉치 기반 사전 구축 및 사전 표제어 선정 관련 연구

말뭉치 기반 사전 구축과 표제어 선정과 관련된 연구에는 서상규(2000), 서지영(2007), 강현화 외(2010), 소강춘 외(2010)이 있다. 서상규(2000)에서는 한국어 기초 어휘의 의미 빈도를 제시하고 있으며, 서지영(2007)에서는 중학교 국어 교과서 말뭉치를 활용하여 표제어 선정 및 기술 방식을 검토하였다. 또한 강현화 외(2010)에서는 한국어 학습자 사전의 표제어 선정을 위해 5만 어휘 목록을 선정하는 과정과 방법론을 정리하였으며, 소강춘 외(2010)에서는 『개방형 한국어 지식 대사전』의 신어 표제어와 관련하여, 선정 원칙과 과정 및 문제점에 대해 논의하였다.

○ 서상규(2000), 한국어교육 기초 어휘 의미 빈도 사전 개발, 국립국어원.

연구 목적 및 필요성	- 교재 개발, 학습 사전 편찬, 교과 과정의 개발 등 한국어교육 분야에서 응용될 수 있는 기반 정보인 한국어 기초 어휘 의미빈도 제시를 목적으 로 한다.																															
연구 방법	1) 표준 한국어 교육용 말뭉치 구성																															
	2) 대상 어휘의 선정																															
	<table><tr><th>품사</th><th>대상 어휘수</th><th>품사</th><th>대상 어휘수</th></tr><tr><td>부사</td><td>123</td><td>수사</td><td>22</td></tr><tr><td>관형사</td><td>23</td><td>의존명사</td><td>75</td></tr><tr><td>감탄사</td><td>20</td><td>형용사</td><td>69</td></tr><tr><td>명사</td><td>481</td><td>동사</td><td>233</td></tr><tr><td>대명사</td><td>21</td><td>보조용언</td><td>20</td></tr><tr><td>합계</td><td></td><td></td><td>1087</td></tr></table>				품사	대상 어휘수	품사	대상 어휘수	부사	123	수사	22	관형사	23	의존명사	75	감탄사	20	형용사	69	명사	481	동사	233	대명사	21	보조용언	20	합계			1087
	품사	대상 어휘수	품사	대상 어휘수																												
	부사	123	수사	22																												
	관형사	23	의존명사	75																												
	감탄사	20	형용사	69																												
	명사	481	동사	233																												
	대명사	21	보조용언	20																												
	합계			1087																												
3) 의미 분류 기준 설정																																
4) 프로그램의 개발																																
5) 품사 주석 말뭉치 구축																																
6) 의미 주석 말뭉치 구축																																
7) 의미 주석 결과 분석																																
8) “기초 어휘 의미 빈도 사전”의 모형 제시																																
연구 결과	- 서상규 외(1998)에서 구성된 ‘한국어교육용 말뭉치’를 더욱 보완하여 1999년에는 총 100만 어절의 표준적 한국어교육용 말뭉치 구성을 완료 하고, 이 말뭉치를 대상으로 2년간 의미 주석 작업을 수행하였다.																															

○ 서지영(2007), 말뭉치를 활용한 중학교 국어과 학습사전 편찬을 위한 기초 연구, 경기대 교육대학원 석사학위논문.

연구 목적 및 필요성	- 이 연구에서는 중학교 국어 교과서 말뭉치를 활용하여 중학교 국어과 학습 사전 을 편찬하는 데 필요한 표제어의 선정과 정보 기술 방법을 알아보고, 그 예를 제 시하여 학습자들의 국어과 학습에 도움이 되는 국어과 학습 사전 편찬에 도움을 주고자 하였다.																																									
연구 방법	1) 조사 대상 - 제7차 교육과정의 7, 8, 9학년(중1~중3) 국어 교과서 (총6권) 2) 조사 방법 - 교과서 말뭉치를 구축 - 깜짝새를 이용하여 품사별, 어절별 색인 작성, 어절별 용례 사전 작성 - 학년별, 빈도별, 품사별 어휘 분석 - 표제어 선정 - 국어과 학습 사전을 편찬하기 위한 지침 마련 3) 조사 단위 - 교과서에 수록된 그대로 입력했다. - 띄어쓰기, 구어체, 비표준어, 준말, 외래어 등도 교과서를 따랐다. 4) 어휘수 : 개별어휘 수 15,431개																																									
연구 결과	1) 학년별 어휘 수 <table><tr><td></td><td>중1-1</td><td>중1-2</td><td>중2-1</td><td>중2-2</td><td>중3-1</td><td>중3-2</td><td>전체</td></tr><tr><td>항목</td><td>5698</td><td>5651</td><td>5920</td><td>4874</td><td>6818</td><td>6037</td><td>15431</td></tr></table> 2) 빈도별 어휘 수 <table><tr><td>빈도</td><td>개별항목 수</td><td>누적항목 수</td><td>빈도 수</td><td>누적빈도 수</td></tr><tr><td>12763~1001</td><td>56</td><td>56(00.4%)</td><td>187961</td><td>187,961(46.3%)</td></tr><tr><td>1000~101</td><td>419</td><td>475(03.1%)</td><td>105047</td><td>293008(72.2%)</td></tr><tr><td>100~11</td><td>2584</td><td>3059(19.8%)</td><td>77927</td><td>370935(91.4%)</td></tr><tr><td>10~5</td><td>2372</td><td>5431(35.2%)</td><td>16372</td><td>387307(95.4%)</td></tr></table> - 고빈도 어휘 56개(0.4%)가 전체의 46.3%를 차지한다. 빈도 5 이상의 어휘 (35.2%)는 전체의 95.4%를 차지한다. - 빈도 1~4회의 어휘 항목 수는 전체의 64.8%에 해당하므로, 국어 교과서 절반 이상이 빈도수 4회 이하의 어휘들로 구성되어 있음을 알 수 있다. 학습 사전 편 찬에 저빈도어 표제어가 반영되어야 한다. 3) 품사별 분석 - 일반명사가 전체의 61.0%를 차지한다.		중1-1	중1-2	중2-1	중2-2	중3-1	중3-2	전체	항목	5698	5651	5920	4874	6818	6037	15431	빈도	개별항목 수	누적항목 수	빈도 수	누적빈도 수	12763~1001	56	56(00.4%)	187961	187,961(46.3%)	1000~101	419	475(03.1%)	105047	293008(72.2%)	100~11	2584	3059(19.8%)	77927	370935(91.4%)	10~5	2372	5431(35.2%)	16372	387307(95.4%)
	중1-1	중1-2	중2-1	중2-2	중3-1	중3-2	전체																																			
항목	5698	5651	5920	4874	6818	6037	15431																																			
빈도	개별항목 수	누적항목 수	빈도 수	누적빈도 수																																						
12763~1001	56	56(00.4%)	187961	187,961(46.3%)																																						
1000~101	419	475(03.1%)	105047	293008(72.2%)																																						
100~11	2584	3059(19.8%)	77927	370935(91.4%)																																						
10~5	2372	5431(35.2%)	16372	387307(95.4%)																																						

- 고유명사는 모두 754개가 나타나며 사용 빈도를 고려하여 표제어로 선정해야 한다.

4) 제6차 교육과정의 중학교 국어 교과서 어휘(최진영 2001)와 비교

	중1-1	중1-2	중2-1	중2-2	중3-1	중3-2	전체
6차	3748	3739	4283	4339	4862	4207	12648
7차	5698	5651	5920	4874	6818	6037	15431

- 평균 1600개 증가했다. (신조어가 다수 추가)

빈도	6차		7차	
	어휘 수	비율	어휘 수	비율
100 이상	164	1.28	479	3.10
60-99	0	1.52	286	1.85
50-59	195	1.52	142	0.92
...				
2	2088	16.50	2733	17.71
1	5882	46.50	5013	32.49
합계	12648	100	15431	100

- 평균 상대 빈도 비율은 6차가 8.36회, 7차가 26.28회이다.

- 빈도 1회인 어휘 항목 수가 절반에 가까운 46.5%에서 32.49%로 줄어든 것은 긍정적이다.

○ 강현화·신자영·원미진(2010), 한국어 학습자 사전 표제어 선정을 위한 자료 구축 및 선정 방법에 관한 연구, <한국사전학> 16, 한국사전학회, 7-29.

연구 목적
및 필요성

- 본 연구의 목적은 한국어 학습자 사전의 표제어 선정을 위한 5만 어휘 목록 선정 과정의 방법론을 정리하고, 학습 사전 기술을 위한 적절한 어휘 선정 방법을 모색하는 데에 있다.

연구
방법

1) 학습자 사전 기초 자료 구축을 위한 기본 어휘 목록

	자료유형	연구자(연도)	어휘 수
기구축 자료	학습용 어휘목록	조남호 학습용 기초어휘(2003)	6,000
		조남호 보고서(2002)	59,000
		서상규(2004)	7,440
	사전	신현숙(2000)	16,926
		김하수 외(2007)	2,000
		언세사전	50,000
		초등국어사전	29,651
		한국어그림사전	3,985
		국립국어원 문법사전	1,385
신규구축 자료	말뭉치	한국어교재 목록	8,808
		TOPIK 시험 문제	9,874
		계	195,069

2) 어휘군 보충

유형	어휘 수	비고
접사 목록	5,595	교재 추출 접사 및 파생어
개념어	1,078	(신체어 669+색채어 200+수량단위 209) 1078
본빛말/준말 큰말/작은말	588	표준 목록 중 기본 어휘
문법용어	163	문법사전 표제어, 사전 목록
사고도구어	925	신명선 목록 정제
가표제어	9,442	표준의 가표제어 목록 중 조남호(2003)에 원표제어가 있는 것만 추출
계	17,791	

3) 어휘 선정 절차 및 방법

- ① 1단계 판정절차: 두 명의 연구자가 품사별로 나누어 5만 어휘에 필요한 항목은 확정으로 A, 판단이 애매한 경우는 보류로 B, 불필요하다고 생각하는 항목에는 C라는 표지를 붙여 1차적으로 판정을 한다. 이때 엑셀을 이용하여 품사별, 접사별 등 다양한 방법으로 분류하여 연구자의 직관을 재차 확인하는 절차를 거친다. 이러한 1차 판정의 결과로 대략 만개 미만의 C 항목이 삭제되었다.
- ② 2단계 판정절차: 1단계 절차를 통해 C 항목이 삭제된 7만여 어휘를 다시 9명의 연구원이 나누어 A, B, C 판정을 내리되, B항목을 중점적으로 검토하여 삭제 혹은 선정 여부를 결정하였다.
- ③ 3단계 판정절차: 판정된 C 항목 중에 필요한 항목이 없는지 다시 검토하고, B 판정을 받은 어휘 목록을 모아 모든 연구자가 다시 판정을 내리는 2단계 작업을 통하여 최종 판정을 내리게 된다.
- ④ 의미적으로 관련된 쌍을 이루는 어휘나 파생관계에 있는 어휘 등을 최종적으로 검점, 보완하는 방법을 사용하여 사전항목의 최종 목록을 결정하게 된다.

연구 결과

- 본 연구는 기존의 학습자 사전의 표제항의 어휘 목록이 몇 천 개에서 만개 정도로 한정되어 있는 것에 비해 한국어 학습자 사전의 규모가 5만개의 어휘 규모로 확장되어야 함을 밝히고 현재까지 이루어진 학습자를 위한 어휘 목록의 작업을 총합하여 어휘 목록을 확정지었다는 데 의의가 있다.
- 어떻게 5만 어휘를 선정하는 것이 타당한 방법인가에 대한 방법론적인 접근을 시도하면서 현재까지 이루어진 어휘 선정 방법의 장단점을 고려하

	여 고급 학습자의 목적에 맞는 5만의 어휘를 선정하였다는 점에 있어서 일정 부분 의의가 있다.
--	--

- 소강춘·이래호·주경미(2010), 개방형 한국어 지식 대사전 신어 분과의 표제어 선정과 그 실제, <한국사전학> 20, 한국사전학회, 52-85.

연구 목적 및 필요성	- 본 연구는 『개방형 한국어 지식 대사전』의 하위 분과인 신어 분과의 표제어 선정 원칙과 선정 과정, 표제어 선정 과정에 나타난 문제점을 소개하고 있다.															
연구 방법	<div>1) 신어 표제어 선정 원칙</div> <div>① 현실 언어에서 활발하게 쓰이는 말을 선정한다.</div> <div>② 국어로 정착한 말을 선정한다.</div> <div>③ 현실 언어에서 자주 나타나는 비표준어도 표제어로 선정한다.</div> <div>④ 순화어의 경우 그 쓰임이 활발하면 신는다.</div> <div>⑤ 『표준국어대사전』 등재어 중 어휘 체계의 빈자리를 보이는 어휘를 찾아 적극 선정한다.</div> <div>2) 신어 표제어 선정 과정</div> <div>① 기반 자료</div> <table><tr><th>자료 유형</th><th>자료 목록</th><th>항목/어절 수</th></tr><tr><td rowspan="4">어휘 목록</td><td>1994~2010 신어 자료집</td><td>27,877 항목</td></tr><tr><td>현대 국어 빈도 조사</td><td>10,000 항목</td></tr><tr><td>네이버 오픈 국어사전(신어/유행어/사투리)</td><td>39,233 항목</td></tr><tr><td>2008~2010년 신어 후보어</td><td>약 7만 항목</td></tr><tr><td>말뭉치</td><td>2007~2010년 네이버 제공 언론 기사</td><td>약 12억 어절</td></tr></table> <div>② 말뭉치의 수집과 전처리 과정</div> <div>자료수집: 시조어 조사기 이용 → 자료의 구조화: 태그 제거, 세종기초말뭉치의 기본 형식으로 수정 → 자료 가공: 어미·조사 제거 → 표제어 대조: 원시 표제어 목록과 대조 → 신어 1차 후보 추출</div> <div>③ 표제어 선정</div> <div>3) 표제어 선정의 난점과 문제점</div> <div>- 합성어와 구의 구별</div> <div>- 일반어와 전문어의 구별</div> <div>- 표제어 형태 분석의 문제</div> <div>- 접미사와 의존명사의 구별</div>	자료 유형	자료 목록	항목/어절 수	어휘 목록	1994~2010 신어 자료집	27,877 항목	현대 국어 빈도 조사	10,000 항목	네이버 오픈 국어사전(신어/유행어/사투리)	39,233 항목	2008~2010년 신어 후보어	약 7만 항목	말뭉치	2007~2010년 네이버 제공 언론 기사	약 12억 어절
자료 유형	자료 목록	항목/어절 수														
어휘 목록	1994~2010 신어 자료집	27,877 항목														
	현대 국어 빈도 조사	10,000 항목														
	네이버 오픈 국어사전(신어/유행어/사투리)	39,233 항목														
	2008~2010년 신어 후보어	약 7만 항목														
말뭉치	2007~2010년 네이버 제공 언론 기사	약 12억 어절														
연구 결과	- 신어의 표제어를 선정하는 원칙과 과정은 이미 편찬된 사전들과는 구별되는 나름의 원칙과 과정을 갖고 있다고 볼 수 있다. 그러나 표제어 선정															

	<p>과정에서의 문제점은 『개방형 한국어 지식 대사전』 신어 분과에서 선정한 표제어에서도 여전히 존재한다.</p> <ul style="list-style-type: none"> - 지금까지 집필한 표제어들은 전체 4만 항목 중에서 명사가 52%를 차지하고, 전문어가 23%, 일반 구가 20%이고 약 5%가 나머지 품사다. 그 수가 폐쇄적인 품사(수사, 관형사 따위)는 신어(미등재어)가 거의 없지만, 개방적인 품사는 계속하여 보충해야 할 필요가 있다. - 표제어 분석에서 나타나는 문제점, 특히 분석의 비일관성 및 오분석들은 좀 더 세심한 검토가 필요하다. - 집필된 표제어의 뜻풀이도 여전히 보완이 필요하다.
--	---

5. 기타 어휘 선정 관련 연구(아동 어휘, 영어 어휘)

- 김경철·채미영·홍정선(2003), 유아용 언어 수준 평가 준거 추출을 위한 기초 연구 - 이야기 꾸미기 활동 평가를 중심으로-, 열린유아교육연구 8(1), 203-223.

연구 목적 및 필요성	- 본 연구의 목적은 유아의 언어 수준을 평가하기 위한 준거를 추출함으로써 교사의 평가 활동을 도울 수 있는 수행평가 준거 초안을 구성하는 데 있다.
연구 방법	(1) 연구 대상: 경상도, 충청도, 경기도에 거주하고 있는 만 4세(38명), 5세(42명) 유치원아 80명이 꾸민 이야기 227편 (2) 연구 절차: 1) 이야기 꾸미기 활동, 2) 교사 훈련하기, 3) 이야기 꾸미기 활동 실시하기 (3) 자료 분석: 1) 분류 기준 설정, 2) 분류 기준에 따른 1차 분류, 3) 1차 분류 자료 재검토, 4) 분석자간 신뢰도, 5) 최종 평가 준거 초안 설정
연구 결과	(1) 어휘 수준 평가 준거 -1수준에서 유아는 명사와 동사 같은 기본적인 품사를 사용한다. 부사, 관형사, 형용사 등은 거의 나타나지 않는다. -2수준에서 유아는 명사와 동사뿐만 아니라 부사와 관형사, 형용사 등을 종종 사용한다. 유아는 때때로 표현적인 어휘를 사용한다. -3수준에서 유아는 부사, 관형사, 형용사 등의 표현적인 어휘를 다양하게 사용한다. (2) 문장 수준 평가 준거 -1수준에서 유아는 주어와 서술어가 하나씩 있는 단순한 문장인 홑문장을 사용한다. -2수준에서 유아는 홑문장과 두 개의 간단한 문장이 연결어미로 이어진 문장을 사용한다. -3수준에서 유아는 홑문장과 겹문장을 다양하게 사용한다. 특히 연결 어미로 이어진 문장뿐만 아니라 한 문장 안에 관형절과 명사절 등이 있는 ‘안은 문장’을 사용하기도 한다.

○ 장현진·전희숙·신명선·김효정(2013), 영유아의 기초 어휘 선정 연구, 언어치료연구 22(3), 169-187.

연구 목적 및 필요성	- 본 연구에서는 기존에 유아 어휘 목록으로 제시된 한국판 맥아더-베이츠 의 사소통발달평가(배소영, 광금주, 2011), 언어/읽기 발달 촉진을 위한 우리말 카드(배소영 등, 2013), 한국어 학습용 어휘 연구(조남호, 2003)의 어휘목록 에서 객관적 방법에 의한 고빈도어와 주관적 방법에 의하여 의미영역별로 분류를 하여 검토과정을 거쳐 기초 어휘를 선정하였다.																				
연구 방법	(1) 연구 대상: 이 연구의 연구 대상은 1차 선정된 기초어휘목록의 타당성을 검증 하기 위하여 영유아를 대상으로 보육 또는 교육을 실시하고 있는 교사 9명을 대 상으로 설문조사를 실시하였다. (2) 연구 도구: 이 연구의 기초어휘 목록은 예비연구를 토대로 작성하였다. 설문지 는 604개의 어휘를 14개의 의미영역으로 나누어 작성한 뒤, 언어치료전공 박사 4명에게 내용과 타당성을 물어 문항을 수정 검토하여 최종 설문지를 작성하였다. (3) 자료 처리: 1차 선정된 기초 어휘들의 영유아 언어지도에 있어서 중요도를 5점 리커트 척도로 평정하였다. 평정한 값의 평균을 산출하여, 평균값을 이용하여 평 점이 3.0이상인 어휘만을 추출하여 중요도별, 품사별, 의미유형별로 분석하여 제 시하였다.																				
연구 결과	(1) 영유아를 위한 기초어휘목록 -영유아 교사들로부터 어휘의 중요도 평정의 평균이 3.0이상인 어휘만을 추출하 여 나온 최종 기초학습 어휘 수는 총 588개로 1차 선정한 604개 어휘 중 97.4% 에 해당한다. 중요도 평균이 3.0미만으로 ‘영유아를 위한 기초어휘목록’에서 삭제 된 어휘는 총 16개이며, 2.6%에 해당한다. <table><tr><th>등급</th><th>중요도 평균</th><th>어휘수</th><th>비율(%)</th></tr><tr><td>I</td><td>4.0이상</td><td>230</td><td>39.1</td></tr><tr><td>II</td><td>3.5이상 4.0미만</td><td>239</td><td>40.6</td></tr><tr><td>III</td><td>3.0이상 3.5미만</td><td>119</td><td>20.2</td></tr><tr><td>전체</td><td>3.0이상</td><td>588</td><td>100.0</td></tr></table> <p><중요도에 따른 등급별 영유아 기초어휘목록></p> (2) 품사별 영유아 기초어휘목록 -전체 어휘 588개 중 명사 346개, 동사 150개, 형용사 56개, 부사 17개, 대명사 11개, 관형사 3개, 감탄사 2개, 수사 1개로 나타났다. 명사의 비율은 58.8%로 전체 어휘의 절반 이상을 차지하고 있었으며, 다음으로 동사가 25.9%, 형용사가	등급	중요도 평균	어휘수	비율(%)	I	4.0이상	230	39.1	II	3.5이상 4.0미만	239	40.6	III	3.0이상 3.5미만	119	20.2	전체	3.0이상	588	100.0
등급	중요도 평균	어휘수	비율(%)																		
I	4.0이상	230	39.1																		
II	3.5이상 4.0미만	239	40.6																		
III	3.0이상 3.5미만	119	20.2																		
전체	3.0이상	588	100.0																		

	<p>9.5%로 명사, 동사, 형용사 세 품사의 비율이 94.2%로 기초어휘의 대부분을 차지하는 것으로 나타났다.</p> <p>(3) 의미 영역별 영유아를 위한 기초어휘목록</p> <p>-15개 의미범주에 따라 588개의 기초어휘를 분류하였을 때, 동사가 대부분을 차지하는 행위 의미영역과 형용사가 대부분을 차지하는 상태 의미영역의 비율이 높았다. 명사의 의미분류로는 일상사물, 음식, 사람, 신체, 장소, 동물, 의류, 자연, 시간, 위치, 탈것, 기타의 순으로 나타났다. 명사의 의미범주 가운데 높은 비율을 차지하는 의미범주는 사물, 음식, 사람, 신체 등으로 아동의 생활주변에서 쉽게 접할 수 있는 어휘들로 이는 학령기 아동이나 성인의 기초어휘와는 차이가 있음을 알 수 있다.</p>
--	---

- 윤경선·이유미(2014), 유아 한글 교육용 어휘 목록 선정을 위한 연구, 어문론집 59, 65-84.

연구 목적 및 필요성

- 이 연구의 목적은 유아 한글 교육용 어휘 목록 개발을 위해 어휘 선정 기준을 설정하고 현존하는 어휘 목록들로 적절성을 확인하며 추후 동화책에 사용된 어휘 분석시 활용될 기초를 마련하는 데에 있다.

연구 방법

<유아 한글 교육용 어휘 선정 기준>

(1) 어휘의 친숙성: 어휘의 친숙성이란 유아가 주변 사물의 친숙한 대상의 이름으로부터 한글을 학습하게 하는 것을 말한다.

(2) 어휘의 습득성: 어휘의 습득성이란 유아가 어휘를 습득하는 인지적인 방식을 의미한다. 기본적으로 유아는 비교적 소리 단위가 큰 것을 먼저 인식하고 점차로 단위가 작은 것을 인식하고, 이는 ‘단어→음절→음소’ 순으로 유아의 인식 능력이 발달해 나감을 뜻한다. 그러므로 언어 교육도 유아의 이러한 자연적인 발달 단계에 맞춰야 적절할 것이다.

(3) 어휘의 복잡성: 어휘의 복잡성이란 어휘가 가진 복잡도를 의미한다.

단계	조합 방식	예
1	단모음 또는 기본자음+단모음	우, 가
2	경음 또는 격음+단모음	뽀, 꼬
3	받침 있는 단모음	악, 강
4	경음 또는 격음+단모음+받침	꼭, 땅
5	이중모음	와, 의, 과
6	이중모음+받침	광, 광
7	경음 또는 격음+이중모음	파, 따, 뛰
8	경음 또는 격음+이중모음+받침	팅

<한글 철자에 따른 8단계 분류>

(4) 어휘의 불변화성: 명사가 아닌 동사의 경우 다양하게 활용을 하는데 이것

	은 명사처럼 구상적 대상이 아닌 관계로 한글 학습시 유아가 어려움을 갖게 된다. 하지만 한글의 문자적 인식 교육이라는 측면에서 결국 체득한 명사를 통해서 그것을 활용해 동사 어휘도 읽을 수 있게 된다.
연구 결과	<ul style="list-style-type: none"> - 영아의 어휘를 제외하고 어휘의 복잡도를 중심으로 하여 총 417개의 어휘를 산출하였다. 영유아 교육용 어휘와 다문화 교육용 어휘를 바탕으로 하여 우선 형태 기준인 복잡도에 따라 분류하였고, 그 다음 교육용 어휘의 기준에 따라 분류하여 1단계 1-3단계, 2단계 1-3단계, 3단계로 총 7단계로 나누었다. - 한글 교육을 위한 어휘는 의미적으로 친숙하고 형태적으로 습득이 수월해야 한다. 의미적 친숙도가 높고 형태적 복잡도가 낮은 어휘가 사용된 동화책을 먼저 접하고 친숙도가 낮고 형태적으로 어려운 어휘가 사용된 동화책을 나중에 접할 때 보다 효과적인 교육이 가능하다.

○ 윤미선·차재은·오재혁(2014), 사용 화자 수를 중심으로 본 학령전기 아동의 고빈도 용언, 언어치료연구 23(4), 73-88.

연구 목적 및 필요성	<ul style="list-style-type: none"> - 본 연구의 목적은 학령전기 아동의 자발화에 나타난 고빈도 용언 목록을 사용 화자 수의 관점에서 분석하고 제시하여 언어장애 아동의 언어 중재와 언어 교육의 기본 자료로 제공하는 것이다.
연구 방법	<p>(1) 자료 선정</p> <ul style="list-style-type: none"> - 본 연구에서는 아동 언어에 나타나는 용언의 특징을 사용화자 수별로 살펴보기 위해 나사렛 말뭉치의 아동 구어 말뭉치를 이용하였다. - 대상 아동은 서울, 경기, 충청 지역에 거주하며 청각과 언어 발달에 문제가 없는 아동으로, 아동의 부모에게 녹음기를 장착할 수 있도록 제작된 앞치마와 녹음기(Click Voice S300)를 지급한 뒤 일주일 동안 가정에서 일상적인 생활 및 놀이 시간에 자연스럽게 발화한 것을 10시간 이상 녹음하였으며 자료의 수집과 전사는 과제 수행 기간인 2011년-2013년 사이에 지속적으로 이루어졌다. - 녹음된 자료는 나사렛 말뭉치의 전사 규칙에 따라 언어치료학과 학부생전사자들이 1차 전사를 한 후, 대학원생 전사자들이 2차 전사를 통해 전사 자료를 수정하고 검증하였다. <p>(2) 자료 정련 및 분석</p> <ul style="list-style-type: none"> - 이 연구에서는 2차 전사가 완료된 나사렛 말뭉치 자료 중에서, 아동과 성인이 1:1로 상호작용하는 상황의 연속된 300여 발화를 무작위로 추출하여 3인의 국어학 박사가 3차 전사를 하였다. 3차 전사를 한 결과 연구 대상

	<p>구어 말뭉치의 규모는, 아동 1인당 평균 343.1 발화, 총 65,787어절이었다.</p> <ul style="list-style-type: none"> - 이 연구에서는 자동 태깅과 수동 정련의 2차 과정을 통해 자료를 분석하였다. 자동 태깅은 KSTARS(Kwak & Chang, 2014)에 내장된 형태소 분석기를 이용하였고, 이 결과를 웹의 표준국어대사전을 기준으로 수동 정련하여 최종 분석 자료를 얻었다. - 이렇게 정련된 자료를 대상으로 사용 아동의 비율에 따른 유형 수와 사용 빈도를 분석하고 구체적 용언 목록을 제시하였다. 사용 아동 비율의 구간은 각 연령대에서 해당 용언을 사용한 아동의 비율에 따라 1-24%(4명 이하), 25-49%(5-9명), 50-74%(10-14명) 75-100%(15-20명)인 네 개 구간으로 나누었고, 필요한 경우 상위 75-100% 구간을 75-89%, 90-100%의 두 구간으로 하위 구분하였다.
연구 결과	<ul style="list-style-type: none"> - 2-5세의 각 연령대에서 50% 이상의 아동이 사용하는 동사와 형용사의 유형 수는 구간 별로 완만하게 증가하였으나, 50% 미만의 아동이 사용하는 용언의 유형 수는 급격히 증가하였다. 또한 사용 화자 비율 구간 모두에서 2-3세 사이에 유형 수의 증가가 두드러졌다. - 2-5세의 모든 연령대에서 사용 화자 수 90-100% 구간에 속하는 고빈도 동사는 ‘하다, 되다, 가다, 주다, 보다’, 고빈도 형용사는 ‘있다, 없다, 이렇다’, 보조 동사는 ‘보다, 주다’이다. 그러나 보조 형용사는 이 구간에 속하는 어휘가 없었으며, ‘싶다’가 3세-5세에 걸쳐 75~89%의 아동이 사용하여 사용 화자 수 비율이 가장 높았다. - 사용 화자 수 기준 고빈도 용언 목록과 출현 빈도 기준 고빈도 용언 목록은 대체로 유사하였으나, 차이가 있는 어휘들이 있었다. - 본 연구에서 제시한 고빈도 용언 목록은 의사소통 장애 아동의 중재에서 목표 어휘 선정, 보완대체 의사소통의 어휘 선정에 유용한 기초 자료로 사용할 수 있을 것이다.

○ 이문복·신동광(2015), 2015 영어과 교육과정 기본 어휘 목록 개발, 영어교과교육 14(4), 115-134.

연구 목적 및 필요성	- 2015 영어과 교육과정 개정에서는 기존 어휘 목록의 제한점을 보완하여 완성도를 높이는 것을 개정의 목적 중 하나로 설정하여 보다 체계적이고 과학적인 방법을 사용하여 기본 어휘 목록을 개선하고자 하였다
연구	(1) 2015 영어과 교육과정 기본 어휘 개선 방향 및 개발 절차

방법

1) 어휘 추출을 위한 코퍼스 구성

코퍼스 종류	문어/구어	코퍼스 규모
Freiburg-Brown Corpus	미국 문어	100만 단어
Freiburg-LOB Corpus	영국 문어	100만 단어
Australian Corpus of English	호주 문어	100만 단어
British National Corpus Written Sampler	영국 문어	100만 단어
Wellington Written Corpus	뉴질랜드 문어	100만 단어
Corpus of Contemporary American English Spoken Sampler	미국 구어	100만 단어
British National Corpus Spoken Sampler	영국 구어	100만 단어
Wellington Spoken Corpus	뉴질랜드 구어	100만 단어

〈2015 영어과 교육과정 기본 어휘 목록 개발에 사용된 코퍼스〉

2) 2015 영어과 교육과정 기본 어휘 선정 기준

-친숙도(Familiarity), 사용범위(Range), 빈도수(Frequency)

3) 어휘 선정 기준 적용 절차

-1차 선정 단계: 1차 선정에는 원어민의 어휘 사용 양상을 최대한 반영하기 위해 빈도수>사용범위>알파벳 순서로 선정 기준을 적용하였다. 이를 통해 원어민과 의사소통에서 가장 유용하게 사용될 수 있는 3,500 단어를 선정하였다. 이후 3,500 단어에 포함된 외래어 200개를 제외하여 예외 어휘로 설정하였다. 또한 기본형을 기준으로 하여 굴절이나 파생 변이형으로 구분되어 제시된 기존의 2009 영어과 교육과정 기본 어휘를 단어군(word family)로 묶어 통일하였다. 이러한 어휘 단위 조정 절차를 거쳐 1차 선정 단계에서는 기본형 기준으로 3,088개를 선정하였다.

-2차 선정 단계: 2차 선정의 목적은 빈도수는 상대적으로 낮지만 초등영어 교육에서 유용한 어휘나 한국인 학습자에 친숙하여 난이도 등급의 정보에 반영하기 위함이다. 이를 위해 초, 중, 고 영어교사를 대상으로 친숙도 조사와 초등 권장어휘 추천의 결과를 적용하였다. 이를 모두 반영한 적용 방식은 우선순위로 어휘의 친숙성>초등 권장 추천>사용범위>빈도수>알파벳 순서로 적용하여 최종 어휘를 선정하였고 선정 기준의 적용 순서를 다 각도로 분석했을 때 위의 순서가 타당성과 효율성 면에서 가장 뛰어났다.

-3차 선정 단계: 2차 선정 단계를 거쳐 800개의 초등 권장 어휘, 1,800개의 중등 일반선택 권장 어휘(누적 2,600개), 400개의 진로·전문교과 I 권장 어휘로 총 3,000개의 어휘를 선정한 후 초중등 교사의 최종 검토를 통하여 800개 중 38개의 초등어휘를 ‘eraser’, ‘homework’과 같은 교실영

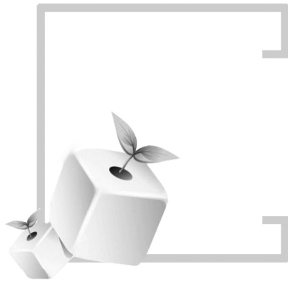
	어와 ‘basketball’이나 ‘elephant’와 같은 초등 중심 주제 어휘로 교체하였다.
연구 결과	<ul style="list-style-type: none"> - 2015 영어과 교육과정 기본 어휘는 진로선택·전문교과 I 를 제외하고 초등 및 중등 일반 선택 교과에서 교과서의 기본 어휘 반영률을 기존 80%에서 90%로 보다 엄격하게 적용하여 학습자의 학습 부담을 줄이고자 노력한 점이 가장 중요한 개정의 초점이었다. - 2015 영어과 교육과정 기본 어휘 목록 개발 과정을 기존 교육과정의 기본 어휘 목록 개발 과정과 비교하여 볼 때 가장 두드러진 차이점은 어휘 목록 개발에 있어 대규모 코퍼스를 활용하고 객관적인 어휘 선정 기준을 적용하여 독자적으로 영어과 교육과정 기본 어휘 목록을 개발하였다는 점이다. 또한 어휘 친숙도라는 어휘 선정 기준을 반영하여 한국인 학습자의 학습 특성을 어휘 목록 개발에 고려하였다는 점에서도 시사하는 바가 크다. - 그러나 제한점으로는 2015 영어과 교육과정의 핵심 취지 중 하나가 영어 학습 부담 경감이라는 점에서 어휘에 대한 통제가 강화되고 이를 보완하기 위해 외래어 규정과 같은 예외 규정을 확대 유지했다는 점을 들 수 있다. 또한 친숙도 측정을 위한 학교급별 다양한 수준의 대규모 학생 표집이 쉽지 않아 교사들만을 대상으로 친숙도를 측정한 점도 본 연구의 한계라고 할 수 있다.

○ 신동광(2018), 영어 기본 어휘 선정 기준의 고찰, 학습자중심교과교육연구 18(2), 381-401.

연구 목적 및 필요성	<ul style="list-style-type: none"> - 이 연구의 목적은 2015 개정 영어과 교육과정 기본 어휘 개발에 적용된 방식은 물론 2015 기본 어휘 개발에도 고려되지 못한 방식을 포함한 다양한 영어 어휘 선정의 기준과 적용 방식을 고찰하여 보다 타당한 어휘 목록 개발을 위한 시사점을 도출하는 데 있다.
연구 방법	<p>(1) 2015 개정 영어과 교육과정 기본 어휘 선정 기준</p> <ul style="list-style-type: none"> -2015 개정 영어과 교육과정 기본 어휘는 국내에서는 처음으로 코퍼스 분석에 기반하여 영어과 교육과정의 기본 어휘를 개발한 사례로서 그 의미가 있다. -기본 어휘 선정 기준으로는 빈도수, 사용범위, 친숙도, 초등권장 어휘 추천이 사용되었다. -2015 개정 영어과 교육과정 기본 어휘 개발에 적용된 독특한 특징 중 하

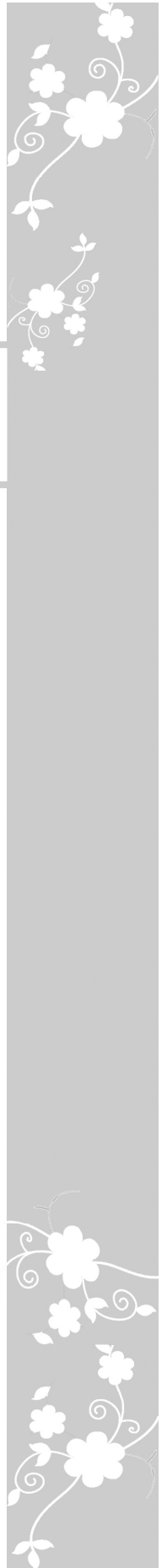
	<p>나는 친숙도와 초등권장 어휘 추천이라는 주관적인 기준을 반영했다는 것이다.</p> <p>(2) 어휘 선정 기준 적용 방식</p> <p>-본 연구에서는 2015 개정 영어과 교육과정에서 적용된 방식 외에도 아직 적용된바 없는 추가적인 어휘 선정 기준의 적용 방식을 포함하여 총 8개의 방식을 제안하고 비교·분석하였다.</p>		
	순위 확정 방식	기준 적용 절차	비고
	유형1	친숙도>초등권장 어휘추천>사용범위>빈도수 순차 적용 순위	
	유형2	t점수 변환 후 가중치를 곱한 값들의 합을 기준으로 한 순위	친숙도×0.4+초등권장 어휘추천×0.3+사용범위×0.2+ 빈도수×0.1
	유형3	t점수 변환 후 가중치를 곱하지 않은 값들의 합을 기준으로 한 순위	
	유형4	원점수에 가중치를 곱한 값들의 합을 기준으로 한 순위	친숙도×0.4+초등권장 어휘추천×0.3+빈도수×0.2+사 용범위×0.1
	유형5	원점수에 가중치를 곱하지 않은 값들의 합을 기준으로 한 순위	친숙도+초등권장 어휘 추천+사용범위+빈도수
	유형6	빈도수>사용범위>초등권장 어휘 추천>친숙도 순차 적용 순위	
	유형7	t점수 변환 후 가중치를 곱한 값들의 합을 기준으로 한 순위	빈도수×0.4+사용범위×0.3+초 등권장 어휘 추천×0.2+친숙도×0.1
	유형8	원점수에 가중치를 곱한 값들의 합을 기준으로 한 순위	빈도수×0.4+사용범위×0.3+초 등권장 어휘 추천×0.2+친숙도×0.1
연구 결과	<p>- 본 연구에서는 ‘빈도수,’ ‘사용범위,’ ‘친숙도,’ ‘초등권장 어휘 추천’ 4개의 어휘 선정 기준을 바탕으로 각 수치의 순차 적용과 합산 적용 방식, 원점수 적용과 t점수 적용 방식, 그리고 각 기준별 가중치 적용 여부 등을 포함한 8개방식의 어휘 선정방식을 비교하였다.</p> <p>- 그 결과, 각 방식을 적용하여 선별된 전체 3,000개의 어휘가 커버하는 텍스트 포괄 범위는 약 89%로 거의 동일하였다.</p> <p>- 그러나 교육과정 어휘 등급을 기준으로 비교한 결과 800개의 초등권장 어휘 수준에서는 8개의 방식 간에 상당한 차이를 보였다. 그 중에서도 유형 8, 즉 어휘 선정 기준의 각 원점수에 가중치를 달리 적용하여 수치를 합산하는 방식이 기초 어휘인 초등권장 어휘 선정에서는 가장 효율적인</p>		

	것으로 나타났다. 따라서 추후 새로운 어휘 목록 개발에는 유형 8과 같은 합산 방식과 가중치 적용 방식을 보다 다각도로 검토하여 적용해 볼 필요 가 있다.
--	--



제 3 장

사업 수행 내용



1. 사업 추진 일정

사업 추진 과정은 다음과 같다.

작업 기간 작업 내용	5월	6월	7월	8월	9월	10월	11월	12월
기존 자료 분석 및 기초 연구	○	○	○					
말뭉치 구성, 분석 및 정제		○	○	○	○			
어휘 추출 및 검토			○	○	○			
어휘 등급화 방법론 수립		○	○	○	○	○	○	
예비 어휘 목록 검토 및 보완					○	○	○	○
정성적 검증 방법론 수립 및 수행					○	○	○	
최종 어휘 목록 작성 및 검토						○	○	○
최종 결과물 산출 및 최종 보고								○

표 8 사업 추진 일정

2. 회의 일정¹⁾

	6월	7월	8월	9월	10월	11월	12월
중간보고/최종보고							
공동연구진 회의							
자문 회의							

표 9 회의 수행 일정

1) 구체적인 회의 내용은 부록을 참고.

2. 세부 과업 개요

본 사업에서는 사업의 목적을 달성하기 위하여 크게 세 분과로 연구 인력을 구성하여 과업을 수행하였다. 세 분과는 각각 기초 어휘 목록 선정 및 등급화 방법론 연구, 말뭉치 분석 방법론 연구, 기초 어휘에 대한 정성적 평정 및 활용이다. 각 분과의 과업과 과업 수행 흐름을 간단히 보이면 아래와 같다.

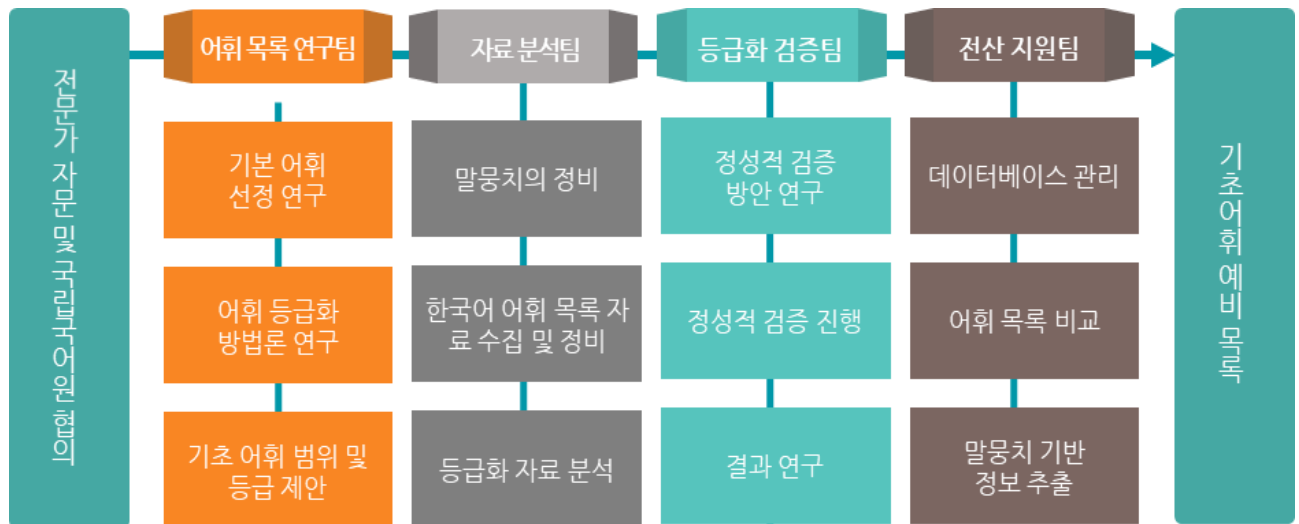


그림 4 분과별 과업 수행 조직도

아래는 각 분과의 과업 수행 내역은 간단히 보인 것이다.

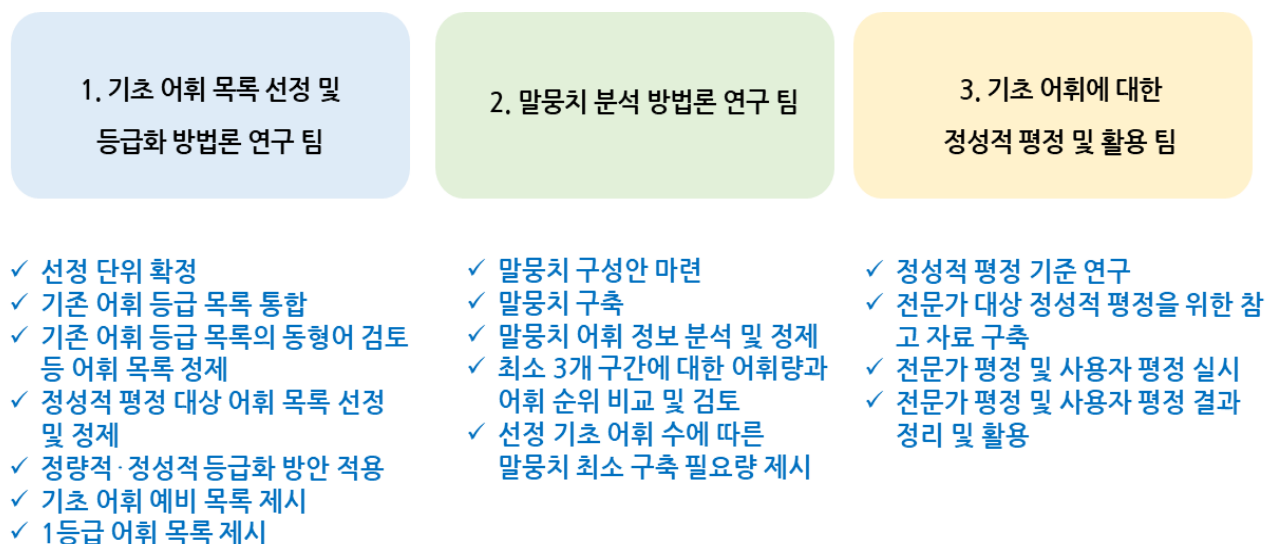


그림 5 분과별 과업 수행 내용

이 장에서는 이 사업의 과업 수행 내용에 대해서 이 세 분과로 나누어 자세히 서술하도록 한다.

2.1. 기초 어휘 목록 선정 및 등급화 방법론 연구

기초 어휘 목록 선정 및 등급화 방법론 연구 분과에서 수행한 과업은 다음과 같다. 먼저, 선정 단위를 확정된 뒤 기존의 어휘 등급 목록을 통합한다. 이때 기존 어휘 등급 목록은 여러 주체가 다양한 기준과 분류를 통해 목록을 선정하고 등급화한 것이기 때문에 어휘 목록 간 동형어 대조를 함께 수행하여야 한다. 그 후 정량적·정성적 등급화 방안을 적용하여 기초 어휘 예비 목록을 제시한 뒤, 이중 1등급에 해당하는 어휘 목록을 제시하도록 한다. 이를 간단히 나타내면 다음과 같다.

- 기초 어휘의 범위 및 선정 단위 확정
- 기존 어휘 등급 목록 통합
- 기존 어휘 등급 목록의 동형어 검토 등 어휘 목록 정제
- 정량적·정성적 등급화 방안 적용
- 기초 어휘 예비 목록 제시
- 1등급 어휘 목록 제시

2.2. 말뭉치 분석 방법론 연구

말뭉치 분석 방법론 연구 분과에서는 본 사업에서 활용할 말뭉치 구성안을 마련하고, 그에 따라 말뭉치를 구축하였다. 그리고 말뭉치 어휘 정보 분석 및 정제를 진행한 뒤, 말뭉치 구간을 나누어 어휘량과 어휘 순위를 비교·검토하였다. 마지막으로 검토 결과에 따라 선정 기초 어휘 수에 따른 말뭉치 최소 구축 필요량을 제시하였다. 이를 간단히 나타내면 다음과 같다.

- 말뭉치 구성안 마련
- 말뭉치 구축
- 말뭉치 어휘 정보 분석 및 정제
- 최소 3개 구간에 대한 어휘량과 어휘 순위 비교 및 검토
- 선정 기초 어휘 수에 따른 말뭉치 최소 구축 필요량 제시

2.3. 기초 어휘에 대한 정성적 평정 및 활용

기초 어휘에 대한 정성적 평정 및 활용 분과에서는 어휘 등급화를 위한 정성적 평정 기준을 연구하고, 전문가 대상 정성적 평정을 위한 참고 자료를 구축하였다. 이에 따라 전문가 평정 및 사용자 평정을 실시한 뒤 그 결과를 정리하여 어휘 등급화에 활용하였다. 이를 정리하면 다음과 같다.

- 정성적 평정 기준 연구
- 전문가 대상 정성적 평정을 위한 참고 자료 구축
- 전문가 평정 및 사용자 평정 실시
- 전문가 평정 및 사용자 평정 결과 정리 및 활용

3. 기초 어휘 선정 및 어휘 등급화 방법론

3.1. 기초 어휘 및 1등급 어휘의 개념과 범위

1) 기초 어휘의 개념 및 범위

기초 어휘를 선정한다고 할 때 선행되어야 할 일은 기초 어휘의 개념을 정의하는 일이다. 어떤 어휘를 기초 어휘로 보느냐에 따라 기초 어휘 선정의 방법론이 달라지고 결과물 역시 크게 달라지게 될 것이기 때문이다. 기존 연구들에서 기초 어휘는 크게 두 가지로 의미로 사용이 된다. 하나는 비교언어학이나 역사언어학의 도구가 되는 어휘의 목록으로 고대 국어에서부터 현재까지 쓰일 뿐더러 언어가 달라져도 크게 목록이 달라지지 않는, 다른 어휘의 조어 핵이 되는 단일어, 고유어로서의 어휘를 말한다. 이들 어휘는 의사소통뿐만 아니라 문화, 자연 환경의 기초가 되는 개념이나 어휘의 목록을 말한다(Swadesh, 1951). 다른 하나는 공시적 관점에서 일상 생활에 꼭 필요한 어휘 중 가장 빈도가 높고 가장 기초적인 2,000~3,000개의 최상위 빈도 어휘를 가리키기도 한다(김광해, 2003).

기초 어휘와 혼용되는 용어 중에는 ‘기본 어휘’나 ‘필수 어휘’ 등과 같은 용어도 있는데 특히 ‘기초 어휘’와 ‘기본 어휘’가 구분 없이 사용되는 경우도 흔하다. ‘기본 어휘’와 ‘기초 어휘’가 구분 없이 사용되는 경우에는 주로 일상 언어생활을 영위하는 데 꼭 필요하다고 생각되는 기본적인 어휘를 뜻하는데 좁게 정의하는 경우에는 어떤 목적과 특정 기준에 따라 선정된 주요 어휘들의 목록으로 보기도 한다. 주로 어휘 빈도 자료에 근거하여 목적 지향적으로 선정된 어휘들을 기본 어휘라고 하는 경우가 많다.

본 과제에서는 교육 및 언어 정책 등에 폭넓게 활용될 범용 목적의 어휘 목록을 선정한다는 전제하에 기초 어휘를 일상 언어생활을 영위하는 데 꼭 필요한 어휘라고 정의하고 연령과 사용역에 크게 구애받지 않고 사용되는 그야말로 한국어 생활에 기초가 되는 어휘라고 본다. 기초 어휘 등급화에 대한 기존 용역(1~3차년도 사업)의 연구 결과에 따르면 어휘 등급화 대상 목록을 약 5만여 개로 제안하고 있다. <연세 말뭉치>를 토대로 빈도 14 이상의 어휘만을 수록한 <연세한국어사전>(1998)의 표제어 수는 약 5만여 개이고 김광해(2003)에서 선정한 국어교육용 어휘 규모가 약 6~7만여 개인 것을 근거로 하였다. 이를 표로 정리한 것은 다음과 같다.

연령 (만)	학년	어휘 수	어휘 수 (누적)	보유 어휘 목록	등급		
4-5세	어린이집	2,000	2,000	‘연세초등국어사전’ 중요 어휘 2,000	1	1	별 ***
5-6세	유치원	3,000	5,000	‘A Frequency Dictionary of Korean’ 문구어 고빈도 어휘 5,000	1	2	
				‘한국어 교육을 위한 기초 어휘’ 5,000			
7세	초1	1,000	6,000		2	3	별 **
8세	초2	1,500	7,500	‘한국어 기본어휘 의미 빈도 사전’ 기본어휘 표제어 목록 7,000	2	4	
9세	초3	2,500	10,000		3	5	
10세	초4	3,000	13,000		3	6	
11세	초5	4,500	17,500		4	7	
12세	초6	5,500	23,000		4	8	
13세	중1	5,500	28,000		5	9	별 *
14세	중2	5,500	34,000	‘초등 어휘에 대한 초기 등급화 데이터’ 34,000	5	10	
15세	중3	6,000	40,000	‘연세초등국어사전’ 표제어 목록 40,000	6	11	
16세	고1	4,000	44,000		6	12	별 없 음
17세	고2	4,000	48,000		7	13	
18세	고3	2,000	50,000	‘연세한국어사전’ 말뭉치 기반 중사전 표제어 목록 52,000	7	14	

표 10 기존 등급화안(1~3차년도 과제)

본 사업에서는 선행 과제에서 제안한 등급화안을 수용하되, 규모를 조정하여 등급화할 기초 어휘의 규모를 약 40,000여 개로 제안하고자 한다. 본 사업에서 선정할 기초 어휘가 기능 어휘와 고유명사를 제외한다는 점을 고려할 때 선행 과제의 기초 어휘 규모와 크게 다르지는 않다. 오기와라 히토시(2016)에서 조사한 바에 따르면 일본 대학생 4학년의 사용 어휘는 약 3만 어, 이해 어휘는 약 4만 5천이라고 하며 한국의 의무 교육에 해당하는 중학교 때까지 학습하거나 습득하는 어휘의 양은 약 4만여 개라고 한다. <연세초등국어사전>의 표제어 수 역시 4만여 개 정도라는 점을 감안하여 기초 어휘의 규모를 4만여 개 정도로 본 것이다.

2) 1등급 어휘의 개념

등급별로 기초 어휘를 평정할 때에는 어휘의 난도에 따라, 관련된 세계 지식의 범위에 따라 등급을 나눌 때에 대응되는 연령이나 학령을 설정하는 경우가 많다. 이 과제에서의 기초 어휘는 다양한 분야에서의 활용을 전제로 한 범용의 한국어 기초 어휘를 의미하지만 일반적인 방식에 따라 선행 연구에서 표4와 같이 제시한 연령뿐만 아니라 교육을 고려한 학령을 등급에 대응하여 선정하였다. 올해 결과로 도출하는 1등급의 어휘는 연령으로는 만으로 4~6세, 학령으로는 초등학교에 입학하기 전인 어린이집에서 유치원에 다니는 아동의 수준을 고려하였다.

이 시기는 교육학 분야에서 학령전기 혹은 전학령기로 일컬어지는 단계이다. 교육심리학자인 비고츠키는 이 시기의 아동이 ‘나’에 대한 인식을 하게 되고 경험의 주체가 된다고 하였다. 학령전기 이전의 아동이 자신이 원하는 것만 배우고 학령기에 접어들면 교사의 프로그램으로 학습하게 되는 데에 반해 과도기로서의 학령전기에는 자생적이면서도 반응적인 교수학습이 이루어지는 시기이다. 교사가 교육하는 것 중에서 아동이 원하는 것을 배울 수 있다는 것이다. 어휘 교육이라는 개념이 가능해지는 시기이다. 최초로 세계관의 윤곽이 생기면서 세계, 자연, 사회, 어린이 자신에 대한 일반적인 표상이 가능하게 된다. 수와 양에 대한 개념이 잡히는 시기이기도 하다.

한국 아동의 학령전기 어휘에 대한 연구는 국어교육, 언어치료 분야에서 이루어졌다. 김경철·채미영·홍정선(2003)에서는 아동 어휘의 수준을 평가하기 위한 기준에 대해 논의하였고, 윤경선·이유미(2014)에서는 한글 교육을 위한 어휘 목록 선정과 관련해 기초 연구를 수행하였다. 장현진·전희숙·신명선·김효정(2013)에서는 학령전기 아동의 사용 명사 어휘에서 높은 비율을 차지하는 의미범주는 사물, 음식, 사람, 신체 등으로 생활에서 쉽게 접할 수 있는 어휘임을 확인하였고, 윤미선·차재은·오재혁(2014)에서는 고빈도 용언의 목록을 추출하여 언어 중재에 목표 어휘 선정, 보완대체 의사소통의 어휘 선정에 활용하고자 하였다. 이 과제에서는 선행 연구에서 제시한 어휘 목록들을 정성적으로 검토하여 정량적 어휘 추출 결과를 보완하여 학령전기에 해당하는 1등급 어휘 목록을 완성하였다.

3) 기초 어휘 선정 단위 확정

본 사업 팀에서는 기초 어휘로 선정하는 언어 단위를 단어로 선정하였다. 따라서 문법소(조사, 어미), 접사, 구 이상의 단위는 제외한다. 고유명사는 인지도가 높은 일부만 포함한다. 또한 동형어를 구분하여 선정하였는데 동형어 구분을 위해서는 국립국어원의 <표준국어대사전>의 동형어 번호를 따랐다. 구와 단어의 구분 역시 <표준국어대사전>

을 참고하였다. 또한 ‘어근’, ‘어근+되다’, ‘어근+하다’는 각각 빈도와 용법이 다르기 때문에 별개의 단어로 인정하였다.

3.2. 기존 어휘 등급 목록 검토 및 정제

본 사업 팀에서는 초등국어사전의 표제어, 초등 교과서 학년군별 어휘 목록, 서상규(2014)의 기본 어휘 목록, 김광해 어휘 등급 목록 등 다양한 어휘 목록을 통합하여(약 45,000개) 검토하고 정제하였다.

본 과제에서 비교 검토한 어휘 목록 자료는 다음과 같다.

- <연세초등국어사전> 등재 어휘 및 사전 수록 중요 어휘(약 3,000개)
- 서상규(2014), 한국어 기본 어휘 목록(약 3,000개)
- 김광해(2003), 국어교육용 등급별 어휘 1, 2등급(약 5,000개)
- 교육과정평가원 등급화 어휘(초등 교육과정 교과서 학년군별 어휘 약 20,000개)

이때 각각의 목록은 동형어의 구분, 선정 단위, 선정 기준 및 범위가 각기 다르기 때문에 이에 대한 검토 및 정제가 필수적이다. 이에 따라 본 사업 팀에서는 동형어 구분을 현재 <표준국어대사전>의 웹사전 동형어 번호에 맞게 정제하고 고유명, 북한어, 방언 등의 삭제 후보 어휘를 선별하는 등 기구축 어휘 목록에 대한 검토와 정제를 수행하였다.

3.3. 정량적·정성적 등급화 방안 적용

본 사업 팀에서는 기초 어휘를 선정할 때 정량적 방법과 정성적 방법을 모두 활용하는 절충적 방법을 선택하였다. 이때 정량적 방법은 대상 말뭉치에서의 빈도나 분포 등의 수치적 자료와 기존 등급화 어휘 목록을 기준으로 하였고, 정성적 방법은 전문가 및 사용자 평정 결과를 기준으로 삼고자 하였다.

1) 정량적 등급화 관련 선행 연구

논문	선정/검증	요약
임지룡(1991)	선정	<p>어휘 빈도 조사의 고빈도어 참조</p> <ul style="list-style-type: none"> - 문교부(1956), 우리말 말수 사용의 찾기 조사 - 서정국(1969), 국어기본어휘의 연구-국민학교 국어과 교과서 - 이응백(1972), 국민학교 학습용 기본어휘 - 이연섭, 권경안, 정인지(1980), 한국 아동의 어휘발달 연구(1) - 정우상(1987), 국민학교 교과서 어휘연구 - 국어연구소(1988), 중학교 교과서 어휘(국어) - 임지룡(미간), 제5차 국민학교 국어교과서의 어휘 빈도 조사
임철성(2002)	선정	<p>연세 말뭉치에서 빈도 5,000 이상인 어휘 500~600개 정도를 추출, 다음으로 국내 초·중·고용 한국어교육용 교재들에 등장하는 어휘에 대한 계량 분석을 실시하여 어휘를 추출</p>
김광해(2003)	선정	<p>메타 계량 방법</p> <ul style="list-style-type: none"> - 기존에 계량 처리된 자료들을 충분히 구하여 그 분포 상황과 자료의 타당도를 함께 고려하면서 비교함으로써 중요도를 측정 - 기존 자료 등장 어휘를 수합하고 기존 자료 누락 어휘를 수습
조남호(2003)	선정	<p>1) 자문회의 개최</p> <ul style="list-style-type: none"> - 선정 어휘 수에 관한 논의 *1단계 1000개, 2단계 2000개, 3단계 3000개, 합 6000개 선정 계획) - 조사, 어미를 선정하는데 대한 논의: 조사, 어미는 선정에서 제외하였다. - 여러 품사로 기능하는 단어와 복합어에 대한 논의: 의미 예상 가능한 복합어 선정 및 품사 구분하여 선정 계획 <p>2) 1차 선정</p> <ul style="list-style-type: none"> - 외부 인사(선정 위원 6명) 통한 작업 - 현대 국어 사용 빈도 조사에서 1회 이상 출현 단어 59000여 개 -> 빈도 순위 10,452등(출현 빈도 15회 이상) 안에 든 단어만 등급을 판정하는 것으로 결정-> A~E로 나누어 등급을 매기는 것으로 정하였다. <p>3) 조정 및 최종 검토</p> <ul style="list-style-type: none"> - 기존 발표된 바 있는 한국어 교육 관련 어휘 목록과 비교 : 조현웅(2000), 한국어 학습 사전 만들기 위해 작성한 어휘 목록(서상규, 2002) - 고유명사 목록 전제 재정리 <p>4) 목록 분석</p> <ul style="list-style-type: none"> - 최종 검토 이후 5,965개의 단어 선정을 완료하였다. (명사-동사-부사-형용사 순으로 많음)
강현화(2014a)	선정	<p>기존의 어휘 선정 목록과 선행 연구의 이론, 한국어 교재와 한국어 학습 사전 등의 사례 자료를 분석</p> <ul style="list-style-type: none"> - 연세대학교, 『한국어학당, 연세 한국어』 초급 1 2, 연세대학교 출판부 - 서강대학교, 『서강 한국어』 초급 1 2, 서강대학교 국제문화원 출판부 - 이화여자대학교, 『이화 한국어』 초급 1 2, EPRESS

		<ul style="list-style-type: none"> - 고려대학교, 『한국어문화교육센터 재미있는 한국어』 초급 1 2, 교보문고 - 경희대, 『한국어』 초급 1 2, 경희대학교 출판부 - 공개된 9회-26회 중 변환이 불가능한 2회분을 제외한 16회 분
강현화(2014b)	선정	<p>객관적 지표에 의한 어휘 평정</p> <ul style="list-style-type: none"> - 선행 연구의 목록 유형별 빈도 및 목록 간 중복도 - 한국어 교육 말뭉치, 일반 말뭉치의 빈도 및 어휘 사용 범위 어휘 군별 평정 지침과 객관적 지표에 따른 연구진 평정
서상규(2014)	검증	서상규(2006, 2013, 2014)의 <한국어 기본어휘>와 연세대학교 언어 연구교육원 한국어학당의 <1급단어목록>을 대조함으로써, 기본어휘 목록의 타당성을 양쪽의 면에서 모두 살펴봄
장현진, 전희숙(2014)	선정	<p>선행 연구를 토대로 기초 어휘 1차 선정</p> <ul style="list-style-type: none"> - ‘수용·표현 어휘력 검사’(김영태 등, 2009)의 검사 어휘 - 한국어 학습용 어휘 결과 보고서(조남호, 2003)의 연구에서 A등급으로 선정된 982개의 어휘 - 초등학교 저학년 어휘 지도 연구(2003)에서 초등학교 저학년 교과서에 출현 빈도 5회 이상인 217개 어휘 - 초등학교 읽기 교과서 어휘 연구(김저옥, 2004)에서 617개 어휘 - 초등학교의 구어에 나타난 어휘 빈도와 분포도 조사(이필영, 김정선, 2008)에서 489개 어휘
강현화(2015)	선정	<p>고급 어휘 선정 절차</p> <ol style="list-style-type: none"> ① 참조 목록 정비를 통한 예비 어휘 목록 선정 <ul style="list-style-type: none"> - 고급 단계의 어휘 목록 선정을 위한 참조 목록 설정 - 참조 목록 정비를 통한 예비 어휘 목록 선정 ② 목록의 타당성 검사 <ul style="list-style-type: none"> - [1차]객관적 지표를 바탕으로 한 목록의 타당성 검사 - [2차]고급 교육과정의 주제별 어휘 목록 검토 ③ 어휘의 사용 범위 및 빈도 분석 <ul style="list-style-type: none"> - 한국어 교육 말뭉치(교재, TOPIK)에서의 어휘 사용 중복도 및 빈도 분석 - 강범모·김흥규(2009), 조남호(2002) ④ 조급 어휘의 특성과 객관적 지표에 의한 어휘 평정 <ul style="list-style-type: none"> - [1차]어휘군별, 품사별 편정 - [2차]객관적 지표에 의한 평정 - [3차]어휘 빈도와 어휘 선정 지침에 의한 목록 보완 - [4차]누착 목록 보완
고길수, 이성현(2019)	검증	<p>실용도와 그것의 점유율을 활용하는 귀납적 방법을 제안함. 이때 실용도는 기본어휘 중 교재에서 실제로 사용되는 것의 비율로서, 교재에서 사용되는 기본 어휘의 수를 기본어휘 전체의 수로 나눈 값. 점유율은 교재에서 기본 어휘가 차지하는 비율로서 그 값은 기본어휘의 출현수를 교재의 전체 출현수로 나누어 계산함.</p>

표 11 정량적 등급화 관련 선행 연구

2) 정성적 등급화 관련 선행 연구

논문	선정/검증	요약
임지룡(1991)	선정	<p>의미 분야의 설정</p> <ul style="list-style-type: none"> - 필자의 주관적 판정을 가미하여 의미 분야별로 1,500개의 어휘를 분류 및 선정 - 사람에 관한 어휘(인체, 정신, 부류, 기타) - 의식주에 관한 어휘(의생활, 식생활, 주생활, 생필품) - 사회생활에 관한 어휘(사회조직, 제도/관습, 교통/통신, 공공시설, 경제분야) - 교육 및 예체능에 관한 어휘(교육일반, 언어, 문학, 체육/오락, 음악, 미술) - 자연계에 관한 어휘(천체, 지리, 지형, 자연현상, 동물, 식물, 광물) - 감각 및 인식에 관한 어휘(일반부류, 공간, 시간, 수량, 추상) - 동작에 관한 어휘 - 상태에 관한 어휘 - 기타(대명사, 의존명사, 부사, 보조동사/형용사, 관형사)
임철성(2002)	선정	연구자의 직관과 연세 빈도를 고려하여 어휘를 조정하여 총 1,038개 어휘를 선정, 제시
김광해(2003)	선정	자료의 분포 정보와 자료의 타당도 정보(주관적으로 결정)를 기준으로 중요도를 판정. (참고한 기존 자료: 서상규(2000), 연세한국어사전(1998), 신현숙 외(2000), 김흥규 외(2000), 고려대(2000) 등)
강현화(2014a)	선정	<p>1) 빈도, 중복도, 친숙도, 활용도 등을 근거로 해서 초급 단계의 한국어 교육 어휘 목록을 선정하였다.</p> <p>2) 객관적 평정으로 어휘 목록을 확보한 뒤, 어휘 선정의 세부 지침과 전문가 평정을 토대로 최종 어휘를 선정하였다.</p>
강현화(2014b)	선정	<p>객관적 지표에 의한 어휘 평정</p> <ul style="list-style-type: none"> - 선행 연구의 목록 유형별 빈도 및 목록 간 중복도 - 한국어 교육 말뭉치, 일반 말뭉치의 빈도 및 어휘 사용 범위 - 어휘 군별 평정 지침과 객관적 지표에 따른 연구진 평정 - 교수자 집단과 학습자 집단의 친숙도 조사와 전문가 어휘 평정을 통해 객관적 방법이 가지는 한계를 극복하고자 했으며, 어휘군별 어휘 선정 지침을 마련하였다.
장현진, 전희숙(2014)	선정	<ul style="list-style-type: none"> - 선정된 어휘를 의미 중심으로 검토하고, 검토한 어휘를 현재 초등 어휘를 지도하는 교사들에게 어휘 지도에 적절한 어휘인지 한번 더 타당도를 검증하였다. - 초등학교 교사들로부터 어휘의 중요도 평정의 평균이 3.0이상인 어휘만을 추출하여 나온 최종 기초학습 어휘 수는 총 511개로 1차 선정한 530개 어휘 중 96.4%에 해당한다. (중요도 평균 3.0미만-총 19개로 삭제)
강현화(2015)	선정	<p>안면 타당성 검사</p> <ul style="list-style-type: none"> - 어휘 전문가의 자문 및 평정 - 평정 결과에 의한 목록 보완

표 12 정성적 등급화 관련 선행 연구

3) 본 과제에서 수행한 정량적·정성적 등급화 절차²⁾

① 정량적 결과 및 기존 어휘 목록을 통합한 기초 어휘 후보 목록 작성

: 균형 말뭉치 점수 순위가 모두 5,000위 이내이고 기존 등급화 목록 모두에서 1등급에 준하는 낮은 등급에 속하는 어휘는 우선 선정한다.

② 정성적 평정

- 어휘 전문가 평정, 사용자 조사(교수 학습 현장 설문, 초등학생 및 초등교사), 교육 전문가 평정을 수행하여 선정한다.

③ 어휘 최종 선정

- 상기의 결과를 바탕으로 어휘 목록 연구팀에서 평정 및 전문가 검토 의견에 따라 일부 어휘를 추가 또는 삭제하여 1등급 어휘 총 5,000개를 선정한다.

3.4. 기초 어휘 예비 목록 및 1등급 어휘 목록 제시

기초 어휘 예비 목록 및 1등급 어휘 목록은 본 사업의 최종 결과물로, 본 사업의 수행 과정에서 연구한 기초 어휘 선정 및 어휘 등급화 방법론을 바탕으로 하여 선정하였다. 각 목록의 실제 자료는 6장에 제시한다.

2) 본 절에서는 정량적, 정성적 등급화 절차에 대해 간략히 소개한다. 정량적·정성적 등급화 방법에 대한 구체적인 내용은 각각 4장과 5장에 제시한다.

4. 말뭉치 분석 방법론

말뭉치 분석 방법론 연구 분과에서 수행한 과업은 말뭉치의 설계 및 구성(4.1), 말뭉치의 정제 및 기초 어휘 후보 목록의 추출(4.2)과 기초어휘 선정에 필요한 말뭉치의 적정 규모 연구(4.3)로 나누어 기술한다.

4.1. 말뭉치의 설계 및 구성

1) 말뭉치 설계를 위한 고려 사항

이전의 기초 어휘 선정용 말뭉치(이삼형, 2017/2018/2020)는 사용, 접근 및 수집이 가능하고 쉬운 자료를 모아서 말뭉치를 구축하고 이로부터 얻어진 결과를 바탕으로 말뭉치의 장르를 조정하는 상향식(bottom-up)으로 접근하되, 말뭉치의 규모의 확장에 많은 공을 들였던 것으로 보인다. 1차년도인 2017년 사업에서는 약 28억 어절의 문어 중심의 말뭉치를 구축하고, 2018년의 2차년도 사업에서는 위의 말뭉치에 자기소개서, 영화 시놉시스, 여행기, 가사, 강의, 뉴스, 법령, 판례, 홈쇼핑 등의 구어와 준구어 중심의 자료 약 16억 6천만 어절을 추가 구축함으로써 약 45억 어절의 말뭉치를 확보한 것으로 나타난다. 말뭉치의 규모 확장은 계속되어, 3차년도인 2020년에는 국립국어원이 2019년에 구축한 문어, 구어, 메신저 대화, 웹 말뭉치를 추가하여 약 63억 어절 규모의 말뭉치를 확보하였다. 상향식 접근이다 보니 말뭉치의 장르는 233종까지 늘었으며 장르 간 균형성은 고려되지 않았다.

말뭉치의 규모 확장이 과연 필요한 것일까 하는 의문에서 출발하여, 본 사업에서는 먼저 1억 어절에서 9억 어절까지 1억 어절씩 규모를 늘려 가면서 어휘량과 어휘 빈도 순위의 상관성을 관찰하였다. 그 결과, 어절별 말뭉치의 상위 10,000개의 어휘 순위 간 상관성이 98.4%에서 99.9%까지 매우 높은 것을 확인할 수 있었다.

	1억	2억	3억	4억	5억	6억	7억	8억	9억
1억	1	0.983521	0.98493	0.98595	0.986333	0.987258	0.987272	0.987852	0.987975
2억	0.983521	1	0.991116	0.992192	0.993424	0.99329	0.993846	0.994157	0.994199
3억	0.98493	0.991116	1	0.994785	0.995417	0.995875	0.996021	0.996424	0.996592
4억	0.98595	0.992192	0.994785	1	0.996654	0.996909	0.997548	0.997599	0.997911
5억	0.986333	0.993424	0.995417	0.996654	1	0.997634	0.997978	0.998349	0.998489
6억	0.987258	0.99329	0.995875	0.996909	0.997634	1	0.998382	0.99875	0.998932
7억	0.987272	0.993846	0.996021	0.997548	0.997978	0.998382	1	0.999093	0.999339
8억	0.987852	0.994157	0.996424	0.997599	0.998349	0.99875	0.999093	1	0.999641
9억	0.987975	0.994199	0.996592	0.997911	0.998489	0.998932	0.999339	0.999641	1

표 13 상위 10,000개 어휘 순위에 대한 샘플 말뭉치별 일치도

이는 어휘 관찰을 위한 연구에서 말뭉치의 규모가 1억 어절을 넘는 경우, 말뭉치의 규모가 더는 어휘량이나 빈도 순위에 영향을 미치는 요인으로 작용하지 않음을 뜻한다.

따라서, 올해 사업에서는 규모의 추가 확장보다는 말뭉치의 ‘균형성’과 ‘대표성’을 확보하는 데 치중하였다. 말뭉치의 필수 요건이라 할 수 있는 ‘균형성’과 ‘대표성’은 말뭉치의 활용 목적에 따라 다르게 구현될 수 있는바, 본 연구에서는 기초 어휘의 개념에서 출발하여, 기초 어휘 후보 목록의 추출에 합당한 말뭉치를 설계하고 구축하는 것을 기본 원칙으로 하였다.

기초 어휘는 일상 구어에서 많이 사용하며, 어린아이들도 아는 어휘인 동시에 연중이 오랫동안 사용해 오고 다양한 장르에서 두루 많이 사용하는 어휘라고 할 수 있다. 이러한 기초 어휘를 추출하기 위해서는 말뭉치의 구성에서 다음과 같은 몇 가지를 고려할 필요가 있다.

첫째, 구어와 문어의 비율을 비슷하게 구성하여, ‘일상 구어에서 많이 사용하는 어휘’를 추출할 수 있도록 한다.

둘째, 발달 단계 아동의 언어 이해 및 생성 자료를 최대한 많이 포함시킴으로써, ‘어린아이들도 아는 어휘’를 추출 가능케 한다.

셋째, 시기적으로는 현대 한국어를 한글맞춤법이 어느 정도 고정된 이후인 1945년에서 2010년대에 이르는 언어 자료를 고루 포함시킴으로써, ‘오랫동안 사용해 온’ 어휘의 추출을 돕는다.

넷째, 장르간 균형성을 확보하여 ‘두루 많이 사용하는 어휘’를 추출할 수 있도록 한다. 또한, 장르의 구성은 이전 연도 사업에서 223개로 지나치게 세분화한 문제를 보완한다. 한편, ‘두루 많이 사용하는 어휘’는 장르에 한정하지 않고, 다양한 텍스트에서 두루 쓰이는 어휘로 해석할 수도 있으므로, 말뭉치에 최대한 많은 텍스트를 포함시키기 위해 책처럼 단위 텍스트의 어절수가 큰 매체나 장르에 한해 일정한 규모로 샘플링을

한다.

다섯째, ‘두루 많이 사용하는 어휘’는 장르에 한정하지 않고, 다양한 텍스트에서 두루 쓰이는 어휘로 해석할 수도 있으므로, 말뭉치에 최대한 많은 텍스트를 포함시키기 위해 단행본처럼 단위 텍스트의 어절수가 큰 매체나 장르에 한해 일정한 규모로 샘플링을 한다.

2) 가용 언어 자원의 정비 및 말뭉치의 구성

올해 사업을 위한 말뭉치는 위의 말뭉치 설계 원칙에서 출발하되, 이미 구축된 말뭉치에서 텍스트를 선별해서 재구성하는 방식으로 구성한다. 이를 위해 가용 말뭉치의 매체, 장르, 시기, 발달 단계 자료 여부 등 정보를 모두 정리하는 작업을 우선하였다.

가용 자원은 국립국어원의 <모두의 말뭉치>를 기본으로 사용하되, 부족한 발달 단계 아동의 언어 자료와 (준)구어 자료를 추가하였다. 발달 단계 자료는 연세대학교 언어정보연구원에서 보유한 초등 교과서, 초등 작문 등과 한국정보화진흥원의 AI 학습용 데이터 중에 자유 대화(유아, 소아)의 음성 전사 데이터를 정제하여³⁾ 포함시켰다. 가용 말뭉치 자료의 정리 결과는 다음의 표와 같다.

3) 한국정보화진흥원의 자유대화(유아, 소아) 음성 전사 데이터는 스크립트를 작성하여 낭독한 자료가 일부 포함되었기에 완전히 일치하는 발화가 대량 포함되어 있다. 따라서 문장 단위로 중복되는 텍스트를 제거하는 작업을 수행하였다.

매체	장르	연령대	장르	확보 말뭉치	비율
문어	신문		신문(1920~2010)	1,085,846,133	51.58%
	잡지		잡지	19,802,912	0.94%
	상상		소설 (및 수필-모두의 말뭉치)	231,019,939	10.97%
	비상상		수필	7,736,262	0.37%
			학술 및 교양(연세 '산문')	523,850,043	24.88%
			수기 및 전기	775,023	0.04%
			교육자료(1900, 1950~1990년대)	1,624,921	0.08%
		발달 단계	교과서(1940~2015)	2,955,113	0.14%
	상상		초등 작문(생성)	643,641	0.00%
			아동도서	1,497,655	0.07%
준구어			준구어 - 자유 발화(유아, 소아 7~10세)	192,107	0.01%
			준구어 - 성인	9,875,056	0.47%
웹, SNS			메신저	4,894,866	0.23%
			웹	72,470,779	3.44%
구어	대화		공적 대화	104,413,707	4.96%
			사적 대화	8,711,096	0.41%
	독백		공적 독백	2,303,339	0.11%
			사적 독백	0	0.00%
				2,105,100,998	100.00%

표 14 기초 어휘 선정용 균형 말뭉치(2억 어절) 구성

표 14에서 보이는 것처럼, 현대 한국어 말뭉치의 규모는 이미 수십억 어절 규모를 자랑하나 공개되어 사용 가능한 구어 말뭉치는 규모가 1억 어절을 넘지 않는다. 발달 단계 자료 또한 그 규모가 약 640만 어절에 불과하다.

결론적으로, 구어와 문어의 비율을 5:5로 구성⁴⁾할 수 있는 최대 규모의 말뭉치는 2억 어절이라는 계산이 나온다. 1억 어절이라는 수치가 절대 규모로는 적지 않으나 시기적으로 2010년대 말의 자료에 편중되어 있는 문제가 있기 때문에, 이를 보완하기 위하여 1948년 이해 모든 회의록이 공개된 국회 회의록⁵⁾을 가공하여 포함시킴으로써 구어 자료의 시기적인 균형을 맞추기로 한다. 그리하여 확보한 발달 단계 자료를 모두 포함시키는 외에, 2억 어절 규모 안에서 최대한 장르와 시기의 균형을 맞추는 방향으로 말뭉치의 설계와 재구성 방안이 결정되었다.

4) 참고로, Nation & Web(2011)의 영어 기초 핵심 어휘 추출에 사용한 말뭉치의 문어와 구어 비율은 4:6이고, 2015 영어과 교육과정 기본 어휘 목록 선정에 사용한 말뭉치의 문어와 구어 비율은 7:3이다.

5) 다음은 url 주소이다. <http://likms.assembly.go.kr/record>

	매체	장르	세부 장르 유형(18개)	1940	1950	1960	1970	1980	1990	2000	2010	2020	합계	세부 장르 비율	매체 장르 비율
문어	신문	비상상	신문	1885424	4569739	4847805	4788143	4761054	4753607	4975573	4309063	0	34890408	17.4%	17.4%
		잡지	잡지	0	397749	806043	1722037	1204238	1733262	3038556	0	0	8901885	4.5%	4.5%
		책	상상	279995	639996	2349988	2349977	2349967	2399967	4144946	484996	0	14999832	7.5%	7.5%
		비상상	수기및전기	0	0	0	0	6686	1491074	0	0	0	1497760	0.7%	11.9%
			수필	113359	159321	1184146	948298	691028	4633301	0	0	0	7729453	3.9%	
			학술및교양	244997	954987	2229975	2229990	2229978	2334963	2964975	1389992	0	14579857	7.3%	
		비상상 (발달단계)	교과서	0	362315	147039	416993	74305	1886118	0	0	0	2886770	1.4%	3.2%
			교육자료	0	0	0	38182	185869	1233198	0	0	0	1457249	0.7%	
			아동도서	0	0	19823	126195	499811	851797	0	0	0	1497626	0.7%	
			초등작문	0	0	0	0	0	0	0	643641	0	643641	0.3%	
준구어		발달단계	자유 발화(유아,소아)	0	0	0	0	0	0	0	192101	0	192101	0.1%	5.0%
			성인	0	0	0	0	0	0	3949935	5924803	0	9874738	4.9%	
메신저			카톡 등 SNS 대화 메시지	0	0	0	0	0	0	0	1000692	0	1000692	0.5%	0.5%
구어	대화	공적	국회회의록	200322	202742	200159	199933	200830	201170	197244	1983014	199991	3585405	1.8%	49.9%
			회의, 토론, 대답 등	0	0	0	0	0	0	0	85121313	0	85121313	42.6%	
	독백	사적	사적대화	0	0	0	0	0	0	0	8710817	0	8710817	4.4%	
			사적독백	0	0	0	0	0	0	124000	0	0	127409	0.1%	
			공적독백	0	0	0	0	0	0	13479	2289835	0	2303314	1.2%	
			연대별 규모	2724097	7286849	11784978	12819748	12203766	21518457	19412260	112050267	199991	200000270	100.0%	
			연대별 비율	1.4%	3.6%	5.9%	6.4%	6.1%	10.8%	9.7%	56.0%	0.1%	100.0%		100.0%
				23.40%								76.60%			

표 15 기초 어휘 선정용 균형 말뭉치(2억 어절) 구성

표 15에서 보이는 것처럼, 2억 어절의 기초 어휘 선정용 말뭉치의 특징은 다음의 몇 가지로 요약할 수 있다.

첫째, 매체간 균형, 즉 문어와 구어를 5:5로 포함시킴으로써, 일상 구어에서 자주 쓰이는 어휘를 관찰할 수 있도록 하였다.

둘째, 장르를 18개로 구분하여, 이전 연구에서의 지나친 장르 세분화와 불균형의 문제를 해소하였다.

셋째, 발달 단계 자료(초등 교과서, 일부 준구어 자료 포함)를 수집하여 2.6% 포함함으로써, 어린아이들도 아는 어휘를 포착할 수 있도록 하였다.

넷째, 최대한 많은 텍스트 포함을 위한 표본 추출(‘상상’과 ‘비상상-학술/교양’의 경우 5,000어절씩)을 통해서, 다양한 텍스트에서 두루 많이 쓰이는 말을 찾아낼 수 있도록 하였다.

다섯째, 시기적인 구성에서는 어문 규범이 비교적 일관되게 사용되기 시작한 1945년 이후 자료로 구성하되, 연대별로 자료가 충분한 신문, 잡지, 소설, 학술 및 교양, 교과서, 국회회의록 등은 최대한 고른 분포로 구성함으로써, 오랜 시기 사용된 어휘를 관찰할 수 있도록 하였다. 결과적으로 1990년 이전의 자료가 23.4%, 1990년대 이후의 자료가 76.6% 포함되었다.

4.2. 말뭉치 정제 및 기초 어휘 후보 목록 추출

1) 형태의미 분석 및 정제

4.1.에서 구성한 말뭉치는 어휘 정보 추출을 위해 울산대에서 개발한 u-tagger 최신 버전을 사용하여 형태의미 분석을 수행한다. 분석의 정확률을 높이기 위해, 기초 어휘 후보 목록을 추출하는 데 필요한 상위 10,000위까지의 어휘 빈도 목록에 대해서는 오류 교정 작업을 실시하였다.

구체적으로는 형태의미 분석 말뭉치에서 1) 장르별로 용례 100개씩 추출하고 2) 일괄 처리가 가능한 오분석 어형을 찾아서, 3) 출현 조건 등을 정리하고, 4) 수정 방안을 제시하며, 5) 이를 토대로 해당되는 오류를 일괄 수정하였다.

다음은 격조사 '의'가 명사로 오분석된 경우인데, 용례 색인을 살펴본 결과 주로 띄어쓰기 오류가 있는 문맥에서 앞뒤에 체언이 오는 출현 조건을 기술한 예이다.

장르	파일명	출현 조건	오분석 어형	수정안	예시
	의_05NNG	*/NNG(NP, NNP)+의_05/NNG+*/NNG(NP, NNP)	의_05NNG	의/JKG	청소년/NNG+의_05/NNG+독서_03/NNG
	의_05NNG	*/NNG(NP, NNP)+의_05/NNG+*/NNG(NP, NNP)	의_05NNG	의/JKG	사람/NNG+의_05/NNG+거룩하/VA+고/EC

2) 기초 어휘 후보 목록 추출 및 분석

말뭉치를 통한 기초 어휘 후보 목록의 정량적 추출은 빈도, 범위, 산포도 외에 기초 어휘 점수 산출 공식을 활용한다.

먼저, 빈도(Frequency)는 하나의 어형(type)이 얼마나 많이 쓰이는지에 대한 측정값으로서, 절대빈도, 상대빈도, 누적빈도, 장르별 빈도를 추가 산출하여 정성 판단의 참고 자료로 활용할 수 있도록 하였다.

다음으로, 범위(range)는 하나의 어형(type)이 얼마나 골고루 쓰이는지에 대한 측정값이며 장르 범위를 산출하는 것이 일반적인데, 이 사업에서는 텍스트 범위를 추가 산출하여 장르 구분이 18개로 줄면서 그 변별력이 낮아진 문제를 해결하고자 하였다.

끝으로, 산포도(Dispersion)는 해당 단어가 한 말뭉치 안에서 일정한 빈도를 유지하는 정도에 대한 측정값인데, 범위의 한계를 보완하는 성격을 지니며, 장르 산포도에 더해 텍스트 산포도를 추가 산출하여 참고 자료로 삼았다.

다음은 2억 어절의 말뭉치에서 추출한 고빈도 어휘의 특성을 보이기 위해 1~30위 어휘의 빈도 순위를 무작위 샘플링을 통한 대규모 말뭉치의 빈도 순위와 비교한 것이다.

빈도 순위	어휘	품사	말뭉치 1(2억 어절)의 어휘 정보				무작위 샘플링을 통한 말뭉치 빈도 순위					
			빈도	장르 범위	텍스트 범위	텍스트 범위 순위	1억	3억	5억	7억	9억	20C
1이다	VCP		8403800	18	301517	1	1	1	1	1	1	1
2하다_01	VV		4801713	18	216124	4	3	3	3	3	3	3
3것_01	NNB		4544727	18	226788	3	2	2	2	2	2	2
4있다_01	VA		2879845	18	184670	6	16	16	16	16	16	23
5되다_01	VV		2349448	18	163859	7	6	6	6	6	6	5
6그_01	MM		2218588	17	97979	23	9	9	9	9	9	8
7있다_01	VX		2048239	18	201686	5	4	4	4	4	4	4
8이_05	MM		1470108	17	125520	12	13	12	12	12	12	12
9수_02	NNB		1358272	18	127866	11	7	7	7	7	7	9
10보다_01	VV		1213906	18	96972	25	20	21	21	21	22	24
11없다_01	VA		1163002	18	121548	16	10	10	10	10	10	10
12그렇다	VA		1128181	18	55959	76	40	40	40	40	40	73
13않다	VX		1119693	18	129415	10	8	8	8	8	8	7
14하다_01	VX		1089732	18	143208	8	11	11	11	11	11	13
15아니다	VCN		1068240	18	88822	29	15	15	15	15	15	16
16갈다	VA		982904	18	108380	18	22	20	20	20	20	22
17거_01	NNB		903605	18	36693	193	1251	1270	1270	1261	1263	2644
18사람	NNG		842198	18	72959	44	14	14	14	14	14	14
19가다_01	VV		840180	18	71372	46	76	78	78	78	78	62
20때_01	NNG		782451	18	93081	27	23	23	23	24	24	20
21보다_01	VX		772217	18	65514	58	18	18	18	18	18	21
22또	MAG		758069	18	85826	30	39	39	39	39	39	36
23주다_01	VX		747579	18	79601	35	17	17	17	17	17	26
24좀_02	MAG		730903	18	37819	185	63	62	62	62	62	159
25이제_01	MAG		726705	18	38756	176	64	61	61	61	61	184
26이렇다	VA		684778	18	43146	135	863	902	902	897	893	1050
27그런_01	MM		681695	17	44279	131	49	51	51	50	50	72
28이런_01	MM		673607	17	48721	109	45	44	44	45	45	90
29대하다_02	VV		661691	18	130891	9	25	25	25	25	25	27
30안_02	MAG		654642	18	49488	102	35	35	35	35	35	45

표 15 2억 어절 균형 말뭉치와 타 말뭉치의 어휘 정보 비교

표에서 보이는 것처럼, 2억 어절 균형 말뭉치의 고빈도 어휘의 특성을 무작위 샘플링 말뭉치의 고빈도 어휘 목록과 비교해 보면, 주로 구어성이 높은 어휘(빨간색 상자로 표시)의 순위가 크게 상향 조정되었다. 예로, 구어에서 간투사로도 사용되는 부사와 명사 품사 통용어인 ‘좀, 이제, 또, 인제, 지금’ 등이 있으며, ‘이-/그-’로 시작하는 지시사로 불리는 ‘이, 이렇다, 그래서, 그런, 이런’ 등이 있다. 또, 기본 용언(파란색 상자로 표시)인 ‘이렇다, 어떻다. 보다, 가다, 주다’ 또한 빈도 순위가 다소 상향 조정된 것을 알 수 있다.

이러한 결과는 기초 어휘 선정 목적에 부합하는 순위 조정으로 보이며, 대체로 구어 말뭉치의 비율을 확대한 데서 비롯된 것으로 풀이된다.

3) 기초 어휘 점수 산출

빈도, 범위, 산포도 등 어휘 정보를 하나의 기초 어휘 점수로 공식화하려는 시도가

적지 않다⁶⁾. 다만, 이전 보고서에서는 기초 어휘 점수를 빈도 1-30위에 대해서만 보이고 있는데, 어휘 점수 공식의 적용에 따라서 어휘의 목록과 순위가 크게 달라지지 않는 것으로 나타났다. 기초 어휘 점수 목록을 1~30위에 한정해 보이고 있는 탓에, 기초 어휘 점수 산출 공식의 타당성 검토가 어려운 문제가 있으므로, 여러 공식 중에 빈도의 가중치를 상대적으로 크게 부여한 2018년 기초 어휘 점수 공식(이하 ‘점수(2018)’로 줄여서 씀.)을 이용하기로 한다⁷⁾.

결과적으로, 빈도 순위나 점수(2018)로 보아 모두 1등급 어휘에 속하는 어휘 목록을 보이면 다음과 같다. 기초 어휘 특성상 어미, 조사, 접사는 제외한 목록이다.

6) 1~3차년도에 제안된 기초 어휘 점수 공식의 유형과 특성을 정리하면 다음과 같다.

연도	어휘점수 유형	공식	특징
2018	점수2018	$0.265 \times \text{빈도} + 0.64 \times \text{범위} + 0.095 \times \text{산포도}$	한국어 모어 화자의 직관 테스트 결과와 비교
2020	점수2020-1	$0.2 \times \text{빈도} + 0.75 \times \text{범위} + 0.05 \times \text{산포도}$	목록당 3회씩 경사하강법을 이용해, 장르 범위의 가중치를 2018점수보다 더 높게 설정함.
	점수2020-2	$0.67 \times \text{로그빈도} + 0.28 \times \text{범위} + 0.05 \times \text{산포도}$	(수치 조정) 소수의 단어는 빈도가 지나치게 높은 반면, 대부분의 단어는 빈도가 매우 낮고 차이가 없기 때문에, 빈도값에 대해 로그 변형을 시행함. 로그 변형을 하면 높은 빈도에 해당하는 값이 낮아지는 반면 낮은 빈도값의 차이가 보다 유의미하게 차별화될 수 있으며, 결과적으로, 빈도의 가중치가 매우 높아지고 범위의 가중치가 상당히 낮아짐.
	점수2020-3	$0.2 \times \text{빈도Z} + 0.75 \times \text{범위Z} + 0.05 \times \text{산포도Z}$	(가중치 다변화) 끝에 Z가 붙은 것은 해당 변수를 표준화(평균을 빼고 표준편차로 나눔)한 값을 입력한 결과를 의미
	점수2020-4	$0.67 \times \text{로그빈도Z} + 0.28 \times \text{범위Z} + 0.05 \times \text{산포도Z}$	(가중평균에 따른 정렬) 빈도의 편향성을 억제하기 위해 빈도에 로그 변형을 한 후 가중치를 새롭게 적용

표 16 1~3차년도 기초 어휘 점수 산출 공식

7) 범위의 적용 단위인 장르 구분을 3차년도에서는 233개로 구분하고, 올해 사업에서는 18개로 구분하고 있으므로, 그 격차로 인해 공식을 그대로 적용하는 것이 과연 타당한 것인지에 대한 의문도 남아 있으며 추후 연구 과제로 남긴다.

빈도 순위	어휘_의미	품사	빈도	비율	범위 (장르)	범위 (텍스트)	산포도 (장르)	점수 (2018)- 장르	점수 (2018)- 순위	순위변동 (빈도순위- 2018점수순위)	공적대화	국회의록	메신저	사적대화	사적독백	소설	신문	잡지	학술및교양
1	이다	VCP	8431310	2.15%	18	305163	99.56696	29.93204	1	0	3368453	125150	48837	277257	4572	652317	1528191	406688	683602
11	하다_01	VV	4813824	1.23%	18	218634	99.70452	17.20373	14	-3	2420942	110682	26448	326813	4524	290285	563448	136972	260527
14	것_01	NNB	4557864	1.16%	18	229581	99.76222	17.24248	13	1	1854610	99097	15957	202235	2797	330561	838203	201173	368574
20	있다_01	VA	2888423	0.74%	18	187087	99.83231	11.49819	20	0	1347629	66138	9427	140727	1932	148984	455288	103712	239388
25	되다_01	VV	2355072	0.60%	18	165777	99.84819	9.535445	26	-1	1240713	49109	7090	113874	1715	113253	298531	70957	146416
26	그_01	MMD	2223275	0.57%	17	98838	99.84257	8.673612	29	-3	1292550	24533	4195	182692	3296	142129	136849	72679	127006
28	네_03	IC	2190416	0.56%	18	25036	99.7364	6.643643	39	-11	2116389	131	2095	6010	113	6345	929	1469	987
29	있다_01	VX	2057325	0.52%	18	204500	99.89552	9.193223	27	2	712202	51957	3164	40350	710	196396	544228	119465	161535
38	이_05	MMD	1473255	0.38%	17	127064	99.88904	6.386249	41	-3	905299	31172	1214	24901	564	62911	167329	43368	87709
40	예_06	IC	1376268	0.35%	18	22415	99.83441	5.092511	61	-21	1329625	8429	184	3000	185	3588	835	954	833
41	수_02	NNB	1363212	0.35%	18	129806	99.92819	6.873874	37	4	560504	26134	2967	40851	579	91645	241666	68985	143942
44	보다_01	VV	1216790	0.31%	18	98090	99.9249	6.185188	43	1	617073	22820	9246	105003	1181	76940	115198	38733	62869
46	없다_01	VA	1166840	0.30%	18	123092	99.94941	6.452454	40	6	375795	23633	6186	48178	695	127563	208991	65358	97385
47	그렇다	VA	1129516	0.29%	18	56342	99.90439	5.411315	55	-8	780260	15010	5653	76226	899	55694	36367	22340	23021
48	않다	VX	1123570	0.29%	18	131118	99.94413	6.164436	44	4	428253	21887	2638	34002	364	118691	205940	62854	91274
49	하다_01	VX	1094515	0.28%	18	145031	99.95484	6.250422	42	7	291576	24754	6971	37706	450	93988	259967	64667	119888
51	아니다	VCN	1070715	0.27%	18	89908	99.93615	5.782868	48	3	522302	18060	6828	49757	662	74567	107337	46139	67837
56	여_02	IC	985427	0.25%	18	25833	99.89411	4.595654	73	-17	852086	16	926	84777	1116	1519	1724	351	522
57	갈다	VA	985278	0.25%	18	109458	99.94083	5.103055	53	4	480312	12114	7279	99426	893	61803	120512	32526	53520
58	우리_03	NP	961000	0.24%	18	77447	99.9484	5.550646	51	7	423545	29655	3620	59237	494	46498	120584	34724	63088
59	나_03	NP	943279	0.24%	18	55177	99.97052	5.912913	47	12	150658	3168	21582	112907	856	150869	50278	57420	48729
60	거_01	NNB	905340	0.23%	18	36916	99.92842	4.926364	67	-7	569750	1513	12490	140467	1496	23882	10111	10330	4286
63	아_02	IC	863486	0.22%	18	30899	99.90831	4.353985	79	-16	739660	315	8599	57409	601	7270	5851	2100	1600
65	사람	NNG	844891	0.22%	18	73824	99.96198	5.325995	57	8	306310	9301	3378	54341	1134	92443	86789	43953	76956
66	가다_01	VV	842509	0.21%	18	72078	99.95991	5.275397	58	8	314140	6688	15378	125907	1339	77185	60808	27419	28479

표 17 빈도 순위와 점수(2018)가 모두 1등급에 속하는 어휘 예시

빈도는 매우 높으나 점수(2018)에서는 그 순위가 크게 낮아져서 1등급 어휘에서 멀어지는 어형들도 있는데, 이를 표로 보이면 다음과 같다.

빈도 순위	어휘	품사	의미	빈도	비율	범위 (장르)	범위 (텍스트)	산포도(장르)	점수(2018)	점수 (2021)- 순위	순위변동 (빈도순위- 점수순위)	공적대화	국회의록	메신저	사적대화	소설	신문	잡지	학술및교양
262	그니까	MAG		167,073	0.043%	10	14,916	99.9809909	1.32024953	32,626	-32,364	147,171	0	1,145	16,896	33	6	5	7
531	예예	IC	01	73,295	0.019%	12	10,325	99.99053991	1.487696125	25,642	-25,111	73,111	0	4	33	27	1	11	1
685	ㅋ	NNG		53,817	0.014%	12	825	99.99306956	1.452691755	25,643	-24,958	0	3	53,568	0	19	74	34	28
686	회담	NNG		53,764	0.014%	14	17,644	99.99536459	1.882355273	13,247	-12,561	19,794	799	0	32	184	31,375	618	427
707	위원장	NNG		51,603	0.013%	14	16,404	99.99580631	1.88154288	13,248	-12,541	28,049	2,624	0	25	381	18,719	618	529
781	아당	NNG		46,405	0.012%	14	12,249	99.99618206	1.866493818	13,249	-12,468	25,406	1,730	0	270	148	17,221	603	456
832	민주당	NNP		42,706	0.011%	14	11,609	99.99634554	1.853803344	13,250	-12,418	24,553	892	1	242	47	15,929	339	453
838	문재인	NNP		42,390	0.011%	9	6,126	99.9949668	0.860841166	49,671	-48,833	39,060	748	7	755	1	1,578	2	115
952	당국	NNG	02	36,796	0.009%	14	23,600	99.99626641	1.827038155	13,254	-12,302	4,240	1,068	0	4	328	29,027	670	742
963	박근혜	NNP		36,133	0.009%	9	6,032	99.99571161	0.848303693	49,672	-48,709	33,263	485	3	246	5	1,873	6	185
978	어	IC		35,829	0.009%	8	11,364	99.99570216	0.653759877	60,858	-59,880	33,312	0	0	1,876	1	1	0	1
1,029	정당	NNG	07	33,680	0.009%	14	9,421	99.99733903	1.834174312	13,253	-12,224	18,363	701	2	165	148	11,108	1,068	1,502
1,054	총리	NNG	01	32,569	0.008%	14	8,524	99.99727208	1.836825748	13,251	-12,197	13,037	4,749	0	60	46	12,618	605	305
1,164	쫌	MAG	01	29,385	0.008%	13	5,255	99.99742685	1.624429848	19,355	-18,191	18,013	0	155	9,878	37	22	3	1
1,184	김정은	NNP		28,814	0.007%	9	3,578	99.99651951	0.832430717	49,673	-48,489	26,965	53	12	169	5	1,565	0	7
1,194	논란	NNG		28,530	0.007%	14	10,241	99.99709192	1.807453959	13,256	-12,062	22,451	365	15	149	50	4,309	336	409
1,230	중공	NNP	02	27,706	0.007%	13	10,207	99.9965945	1.601383532	19,358	-18,128	53	297	0	1	44	26,389	144	622
1,234	티브	NNG		27,570	0.007%	11	9,197	99.99691183	1.220724191	32,627	-31,393	23,967	0	218	2,325	7	14	4	2
1,240	그니까	MAJ		27,424	0.007%	8	7,786	99.99683385	0.639419934	60,859	-59,619	24,541	0	35	2,240	2	0	0	0
1,260	구신	MMN		26,717	0.007%	14	11,201	99.99706723	1.799235272	13,262	-12,002	22,779	0	7	689	95	2,375	25	61
1,337	의혹	NNG		24,961	0.006%	14	7,618	99.99761272	1.80213233	13,259	-11,922	18,308	792	0	16	347	4,561	304	237
1,344	정치적	MMA		24,824	0.006%	13	10,630	99.99860308	1.62736632	19,354	-18,010	8,275	685	0	92	354	7,101	2,790	4,257
1,357	트럼프	NNP		24,555	0.006%	11	3,130	99.99720366	1.213404168	32,628	-31,271	21,655	70	7	216	31	2,428	12	28
1,364	여당	NNG	01	24,438	0.006%	13	8,412	99.99801041	1.614415978	19,356	-17,992	13,220	1,148	0	174	113	9,016	254	275
1,404	외교	NNG	01	23,541	0.006%	14	10,531	99.99822704	1.807777709	13,255	-11,851	9,545	1,148	0	96	105	10,769	482	634

표 18 빈도 순위는 1등급에 속하나, 점수(2018) 순위는 크게 낮아지는 어휘 예시

‘그니까, 예예, 쫌’ 등처럼 구어성이 짙은, 즉 특정 매체, 장르에만 그 사용이 치우치는 어형이 있는가 하면, ‘박근혜, 문재인, 김정은, 트럼프’ 등처럼 2010년대 이후에 특히 그 사용 빈도가 늘어난 정치인 등의 고유명사, ‘중공, 의혹, 여당, 외교, 야당, 총리’ 등 정치 분야 어휘가 대부분이다. 이들은 직관상으로도 1등급 기초어휘로 보기에 어려운 어형들이므로 점수(2018)의 타당성에 힘을 실어준다.

한편, 빈도는 낮은 편이나 점수(2018)의 순위는 상향 조정되어 1등급 어휘 여부를 판

정해야 하는 어형도 있는데, 이를 표로 보이면 다음과 같다.

빈도 순위	어형	품사	의미	빈도	비율	범위 (장르)	범위 (텍스트)	산포도(장르)	점수(2018)	점수 (2018) 순위	순위변동폭 (빈도순위-점수순위)	공적대화	국회 회 의록	메신저	사적대화	소설	신문	잡지	학술및 교양
5001	못	NNG	01	4,946	0.001%	17	3,759	99.99977541	2.337029313	1946	3055	729	25	159	33	653	1,687	306	374
5002	발음	NNG	01	4,946	0.001%	17	2,580	99.99968685	2.335340591	2011	2991	2,457	4	22	72	242	701	146	345
5003	민주	NNP		4,946	0.001%	16	2,636	99.99964053	2.141393019	4631	372	1,421	51	0	12	68	2,549	320	142
5005	간판	NNG	02	4,944	0.001%	16	3,151	99.99974335	2.143344942	4475	530	1,199	47	10	92	630	1,721	337	297
5008	캠프	NNG		4,937	0.001%	17	2,234	99.99962883	2.334195914	2075	2933	2,533	32	6	117	66	1,642	115	129
5010	생생하다	VA		4,936	0.001%	16	3,615	99.99977537	2.1439213	4426	584	1,471	32	7	61	480	1,082	381	600
5012	하여간	MAG		4,933	0.001%	16	2,802	99.99978446	2.144081804	4413	599	1,518	114	21	286	684	522	351	267
5016	출입	NNG		4,926	0.001%	16	3,535	99.99971301	2.142689477	4532	484	1,420	60	8	28	470	1,899	183	278
5017	예선	NNG	02	4,925	0.001%	16	2,967	99.99949164	2.138464297	4950	67	735	11	2	15	14	3,924	69	28
5020	휘두르다	VV	01	4,922	0.001%	16	3,323	99.99976634	2.143689398	4451	569	1,303	19	4	26	1,210	989	392	311
5026	차림	NNG		4,916	0.001%	16	3,202	99.99975896	2.143522982	4467	559	542	1	4	9	1,354	1,425	470	310
5030	성공적	NNG		4,908	0.001%	16	3,664	99.99972826	2.142903367	4507	523	1,361	176	2	19	125	1,733	298	726
5031	내버려다	VV		4,906	0.001%	16	3,437	99.99980583	2.144373872	4388	643	812	53	10	93	1,284	507	464	355
5032	새다	VV	01	4,905	0.001%	16	3,459	99.99978119	2.143899765	4433	599	1,279	26	26	100	1,214	769	364	222
5035	장작	NNG		4,904	0.001%	16	2,451	99.99974442	2.143194437	4484	551	840	22	4	52	259	1,712	441	1,056
5036	기대다	VV	01	4,903	0.001%	17	3,181	99.99977101	2.336761563	1956	3080	825	15	14	69	1,664	483	565	420
5037	멀리	NNG	01	4,903	0.001%	16	3,628	99.99981531	2.144541868	4374	663	1,196	14	49	286	980	631	323	363
5043	확신하다	VV		4,897	0.001%	16	3,961	99.99970699	2.142450842	4555	488	821	241	3	8	339	2,234	226	508
5046	명	NNG	02	4,892	0.001%	16	3,592	99.99956071	2.139640213	4803	243	694	15	10	14	162	3,416	80	209
5047	원장	NNG		4,889	0.001%	16	1,389	99.99957683	2.139934772	4782	265	3,336	1	16	280	122	373	194	140
5048	전설	NNG	04	4,887	0.001%	16	2,637	99.99979287	2.144045591	4416	632	1,215	6	18	95	467	993	241	899
5049	고도	NNG	08	4,882	0.001%	16	3,372	99.99973282	2.142879164	4510	539	960	113	1	20	132	1,861	336	884
5050	금지하다	VV		4,881	0.001%	17	3,619	99.99965343	2.334425659	2065	2985	1,305	170	4	12	65	2,489	181	419
5051	잔치	NNG	01	4,879	0.001%	16	2,927	99.99978902	2.143938019	4424	627	882	8	13	63	643	1,492	273	419
5058	굶다	VA		4,873	0.001%	16	3,082	99.99978246	2.143787288	4440	618	1,438	7	15	65	1,082	585	384	323

표 19 빈도 순위는 1등급 밖이나, 점수(2018) 순위는 크게 높아지는 어휘 예시

이처럼 빈도 순위와 기초어휘 점수 순위가 크게 다른 경우는 전문가에 의한 정성 평가 대상으로 분류된다.

4.3. 기초 어휘 선정용 최소 말뭉치 규모 분석

본 사업의 내용 중 하나가 최소 3개 구간에 대한 어휘량 및 어휘 순위를 비교, 검토하여 기초 어휘 선정을 위한 말뭉치의 적정 규모를 제시하는 것이다.

1) 비교 대상 말뭉치 구성: 1억 어절 균형 말뭉치와 20억 어절 말뭉치

2억 어절의 균형 말뭉치와 비교하기 위해 다른 말뭉치 2개가 더 필요한데, 비교 대상 말뭉치의 구성은 양적인 차이와 질적인 차이를 두는 두 가지 방향으로 구성할 수 있다. 전자는 양적으로 규모를 확장하는 방안이 되겠고, 후자는 균형성과 대표성을 고려한 말뭉치로 구성하는 방안이다. 전자에 관해서는, 이미 무작위로 말뭉치를 구성하는 경우, 1억 어절 이상의 말뭉치에서 9억 어절까지는 빈도 목록의 구성이 크게 다르지 않다는 점을 확인한 바가 있다. 따라서 본 연구는 양적으로 연구진에서 확보한 최대 규모인 20억 어절 말뭉치를 하나의 비교 대상으로 삼는 한편, 두 번째 방안, 즉 균형성과 대표성을 고려한 말뭉치를 또 다른 비교 대상으로 삼기로 한다. 균형성과 대표성을 고려한 말뭉치는 가용 한국어 말뭉치를 활용하는 경우에 2억 어절이 구성 가능한 최대 규모이므

로, 2억 어절보다 작은 1억 어절로 구성하되, 비교 가능성(comparability)을 높이기 위해 매체·장르 구성은 대체로 2억 어절의 그것을 따르기로 한다.

이상에서 기술한 비교 대상 세 말뭉치의 특징을 표로 정리하면 다음과 같다.

	1억 말뭉치	2억 말뭉치	모두의 말뭉치
규모(어절)	1억	2억	20억 4451만어
샘플(개)	약 15만	약 40만	약 630만 개
매체 - 문어:구어	5:5	5:5	94:6
시기 - 균형 고려	○	○	X(주로 2010년대)
장르 - 구분	18	18	6

표 20 비교 대상 말뭉치의 특징 비교

아래에 비교 대상 말뭉치의 구성을 구체적으로 살펴보도록 한다. 먼저, 1억 어절 말뭉치의 구성은 다음과 같다.

사용역	매체	장르	장르	1940	1950	1960	1970	1980	1990	2000	2010	2020	합계	세부 장르 비율	매체.장르 비율	
문어	신문	비상상	신문	1885424	2055651	2055618	2055821	2055727	2055809	2055722	2055622	0	16275394	16.3%	16.3%	
		잡지	비상상	잡지	0	209892	418576	873128	620926	1017888	1555517	0	0	4695927	4.7%	4.7%
		책	상상	소설	149999	329998	1189992	1189985	1189981	1209989	2084968	254996	0	7599908	7.6%	7.6%
			비상상	수필	68245	102380	608570	484794	399139	2362485	0	0	0	4025613	4.0%	
				학술 및 교양(연세 '산문')	134997	489993	1129985	1124996	1129987	1179982	1494991	704996	0	7389927	7.4%	
	(발달단계)		수기 및 전기	0	0	0	0	6686	799947	0	0	0	806633	0.8%		
		비상상	교과서	0	194881	96489	226130	52989	960885	0	0	0	1531374	1.5%		
			교육자료	0	0	0	38182	111649	635966	0	0	0	785797	0.8%		
			조중 작문(생성)	0	0	0	0	0	0	0	643641	0	643641	0.6%		
			아동도서	0	0	19823	74985	261904	450656	0	0	0	807368	0.8%		
준구어		준구어 - 자유 발화(유아, 소아 7~10세)	0	0	0	0	0	0	0	0	96053	0	96053	0.1%	3.9%	
		준구어 - 성인	0	0	0	0	0	0	1987576	2975367	0	4962943	5.0%	5.0%		
웹, SNS		메신저	메신저	0	0	0	0	0	0	0	510360	0	510360	0.5%	0.5%	
	구어	대화	공적	국회의회의록	102780	137232	194121	124440	141178	201170	197244	458089	199991	1756245	1.8%	
			기타 회의, 토론, 대담 등	0	0	0	0	0	0	0	42573396	0	42573396	42.6%		
		사적	사적 대화	0	0	0	0	0	0	0	4367799	0	4367799	4.4%		
	독백	공적	공적 독백	0	0	0	0	0	0	13479	1158223	0	1171702	1.2%		
		사적	사적 독백											0.0%	49.9%	
		연대별 규모				2341445	3520027	5713174	6192461	5970166	10874777	9389497	55798542	199991	100000080	100.0%
연대별 비율				2.3%	3.5%	5.7%	6.2%	6.0%	10.9%	9.4%	55.8%	0.2%				
								23.7%				76.3%				

표 21 1억 어절 균형 말뭉치의 구성

2억 어절 말뭉치와의 비교해, 1억 어절 균형 말뭉치는 다음과 같은 특징을 지닌다.

첫째, 비교 가능성을 높이기 위해 2억 어절 말뭉치의 구성(매체, 장르, 시기별 균형)을 최대한 유지하였다.

둘째, 발달 단계 자료는 교과서만 일부 샘플링하고, 교육 자료, 작문 등은 전부 포함 시킴으로써 2.6%(2억 어절)에서 3.9%로 상향 조정되었다.

셋째, 최대한 많은 텍스트 포함을 위한 표본 추출('상상'과 '비상상-학술/교양'의 경우 5,000어절씩)을 진행하였으며, 전체 샘플 수 약 15만 개가 포함되었다.

다음은, '모두의 말뭉치'의 구성이다.

매체	장르	규모	비율	매체.장르비율
문어	신문	1156321881	56.6%	
	도서 등	682509647	33.4%	89.9%
웹, 메신저	웹	72424758	3.5%	
	메신저	4912743	0.2%	3.8%
구어	일상대화 외	125303198	6.1%	
	일상대화	3042229	0.1%	6.3%
	합계	2044514456	100.0%	100.0%

표 22 ‘모두의 말뭉치’의 구성

2억 어절 말뭉치와 비교하면, ‘모두의 말뭉치’의 특성은 다음처럼 요약된다.

첫째, 20억 어절이 넘는 대규모 말뭉치이며, 공개된 최대 규모 한국어 말뭉치이다.

둘째, 발달 단계 자료가 포함되지 않았다.

셋째, 샘플링을 하지 않고 텍스트 전체를 포함시키는 방식으로 구성되었다.

넷째, 문어와 구어의 비율이 94:6이다.

2) 말뭉치별 어휘량 및 고빈도 실질어휘의 누적 비율 비교

1억 어절과 2억 어절의 균형 말뭉치, ‘모두의 말뭉치’ 등 세 말뭉치에 사용된 어형은 그 규모에서 꽤 큰 차이를 보인다. 다시 말해, 말뭉치의 규모에 비례해 출현 어형의 규모가 커지는 것을 알 수 있다. 그러나 본 연구가 지향하는 말뭉치는 한국어의 모든 어휘를 탐색하고 연구하는 것이 아닌 기초 어휘 선정에 있으므로 비교 대상 어휘를 실질어휘 중에 고빈도 어휘로 좁혀서 비교하면 그 결과는 매우 달라진다.

다음은 실질어휘 규모별 세 말뭉치에서의 누적 비율을 표로 정리해 보인 것이다. 실질어휘는 u-tagger 결과에서 조사, 어미, 접사를 제외하고, 비율 산정에서는 분모를 어절이 아닌 실질형태소 수로 계산하였다.

실질어휘 type(개)	누적 비율		
	1억 말뭉치	2억 말뭉치	모두의 말뭉치
5,000	81.39%	80.59%	75.97%
10,000	87.47%	86.60%	83.84%
20,000	92.26%	91.31%	89.95%
30,000	94.50%	93.51%	92.72%
40,000	95.81%	94.79%	94.32%
50,000	96.68%	95.64%	95.36%
60,000	97.30%	96.24%	96.10%
70,000	97.76%	96.69%	96.63%
80,000	98.12%	97.04%	97.06%
90,000	98.41%	97.32%	97.39%
100,000	98.64%	97.54%	97.66%

표 23 실질어휘 어형의 규모 구간별 누적 비율

위의 표에서 실질어휘 5,000개의 어휘, 즉 1등급 후보 어휘의 누적 비율은 1억, 2억 말뭉치에서 각각 81.39%, 80.59%로 비슷하고, 모두의 말뭉치에서는 75.97%로 나타난다. 그러나 실질어휘의 규모를 상위 빈도 50,000개⁸⁾, 즉 기초 어휘 후보로 그 규모를 확대하면, 세 말뭉치에서의 누적 비율은 모두 95%를 넘는 것을 알 수 있다.

3) 말뭉치별 고빈도 실질어휘 목록의 순위 비교

다음은 모두의 말뭉치에서 추출한 1~100위까지의 실질어휘의 1억 어절, 2억 어절 말뭉치에서의 순위를 표로 정리해 보인 것이다.

8) 본 연구에서 기초 어휘의 규모는 약 4만 개로 계획하고 있으나, 이를 질적으로 선별하기 위한 후보 목록은 양적으로 추출하며 보다 규모가 큰 5만 개로 한다.

어형	품사	의미	20억 순위	2억 순위	1억 순위	어형	품사	의미	20억 순위	2억 순위	1억 순위	어형	품사	의미	20억 순위	2억 순위	1억 순위
이다	VCP		1	1	1	한	MMN	01	36	39	38	안	MAG	02	71	37	32
것	NNB	01	2	3	3	중	NNB	04	37	87	88	거	NNB	01	72	22	23
하다	VV	01	3	2	2	보다	VX	01	38	27	28	미국	NNP	03	73	109	112
있다	VA	01	4	4	4	가다	VV	01	39	25	24	이번	NNG	01	74	135	138
있다	VX	01	5	8	8	통하다	VV		40	169	177	이런	MMD	01	75	35	40
수	NNB	02	6	11	11	파르다	VV	01	41	134	142	다른	MMD		76	126	133
되다	VV	01	7	5	5	않다	VA		42	64	64	가장	MAG	01	77	164	165
등	NNB	05	8	71	74	더	MAG	01	43	59	60	그러나	MAJ		78	115	120
하다	VX	01	9	16	17	개	NNB	10	44	127	131	생각하다	VV		79	77	83
년	NNB	02	10	42	43	말	NNG	01	45	47	50	그리고	MAJ		80	62	61
않다	VX		11	15	16	일	NNG	01	46	46	51	기업	NNB	01	81	486	500
없다	VA	01	12	13	14	종다	VA	01	47	52	47	정도	NNG	11	82	95	93
일	NNB	07	13	41	42	크다	VA	01	48	89	91	이상	NNG	05	83	197	203
그	MMD	01	14	6	6	또	MAG		49	28	30	데	NNB	01	84	104	106
대하다	VV	02	15	36	39	그렇다	VA		50	14	15	그것	NP		85	58	57
이	MMD	05	16	9	10	억	NR	04	51	259	279	곳	NNG	01	86	211	219
위하다	VV	01	17	63	65	만들다	VV		52	92	95	잘	MAG	02	87	78	75
그	NP	01	18	48	48	지나다	VV		53	128	134	대통령	NNG		88	97	107
보다	VV	01	19	12	12	오다	VV	01	54	45	44	후	NNG	08	89	187	197
아니다	VCN		20	17	18	씨	NNB	07	55	54	56	먹다	VV	02	90	81	69
사람	NNG		21	24	29	경우	NNG	03	56	90	92	전	NNG	08	91	106	111
나	NP	03	22	21	20	알다	VV		57	51	49	변	NNB	04	92	76	70
이	NP	05	23	72	76	나오다	VV		58	56	55	시다	VX		93	100	97
원	NNB	01	24	91	96	문제	NNG	06	59	61	62	오다	VX	01	94	146	156
말하다	VV		25	69	73	못하다	VX		60	111	116	보이다	VV	01	95	143	155
받다	VV	01	26	57	58	지역	NNG	03	61	273	289	뒤	NNG	01	96	183	199
때	NNG	01	27	26	25	한국	NNP	05	62	103	114	앞	NNG		97	119	118
컬	NNB	02	28	120	137	자신	NNG	01	63	182	204	그런	MMD	01	98	34	37
같다	VA		29	19	19	네	IC	03	64	7	7	시장	NNG	04	99	487	498
주다	VX	01	30	29	26	밝히다	VV		65	294	347	시	NNB	10	100	130	140
명	NNB	03	31	74	80	할게	MAG		66	185	186						
만	NR	06	32	166	184	및	MAG		67	249	272						
때문	NNB		33	50	53	경부	NNG	08	68	108	108						
우리	NP	03	34	20	21	두	MMN	01	69	83	86						
지다	VX	04	35	66	67	사업	NNG	04	70	483	487						

표 24 말뭉치별 고빈도 실절어휘 목록의 순위 비교

표 24에서 초록색으로 표시한 부분은 1억 어절과 2억 어절의 순위인데 큰 차이가 없다. 다시 말해, 균형성과 대표성의 기준이 일치한 경우, 규모가 1억, 2억 어절로 두 배의 차이가 나도 고빈도 어휘의 순위는 크게 다르지 않음을 확인할 수 있다. 이에 반해, ‘모두의 말뭉치’에서만 유독 높은 빈도를 보이는 어형(빨간색 표시)들이 발견되는데 수사 ‘만, 억’, 동사 ‘통하다, 밝히다’, 명사 ‘지역, 사업’ 등이 있다.

비교 대상 빈도 목록의 규모를 기초 어휘 후보 목록에 포함시킬 수 있는 1~50,000위까지 확대하여 비교한 결과는 다음과 같다.

먼저, ‘모두의 말뭉치’ 빈도 목록에는 있으나 2억 어절의 빈도 목록에는 없는 어형은 274개였으며 품사별로 살펴보면 다음과 같다.

일반명사는 178개로, 65.0%를 차지하였으며, 주로 ‘먹스타그램, 셀스타그램, 립스타그램, 견스타그램, 냥스타그램, 세젤귀, 선팔맛팔’ 등의 인터넷 신어가 주를 이루었다. 또, ‘다이어트식단, 원화값, 결합상품, 스포츠유틸리티차량, 지속가능경영, 녹색경영’과 같은 합성어, ‘성료, 순이자, 방폐물’과 같은 한자어도 일부 포함되었다. 가장 많은 비율을 차지하는 품사는 고유명사로, 총 90개였으며 약 33.9%를 차지하였다. ‘하이투자증권, 한화케미칼, 하나금융투자, 스포츠동아, 씽큐’ 등이 그 예이다. 이밖에, 외래어 파생어로 ‘포스팅하다, 레이어드하다’가 있었다.

요약하면, 모두의 말뭉치 1~50,000 고빈도 어휘 목록에는 있으나, 2억 어절에는 나타

나지 않은 어형에는 기초 어휘로 판단되는 어휘는 발견되지 않았다.

이번에는, 서로 다른 규모의 균형 말뭉치의 고빈도 목록을 비교해 보기로 한다. 2억 어절 말뭉치에서 1~50,000위 목록에는 있으나, 1억 어절 말뭉치에는 없는 어형은 92개였으며, 이 중에 고유명사가 88개로, 95.7%를 차지하였다. ‘꼭정이, 두식, 똥예, 강필, 금수원...’ 등이 그 예이다. 그밖에 4개 공백 어형은 일반명사인데, ‘왕고모님, 암흑물질, 상업02, 입양특례법’ 등이 포착되었는데, 이들 또한 기초 어휘로 판단되는 어휘는 아닌 것으로 확인되었다.

결론적으로, 기초 어휘 후보 목록을 약 50,000개로 상정할 경우, 20억 어절에서 추출한 어휘 목록은 2억 어절 빈도 목록보다 274개의 어휘가 많으나, 이 중에 기초 어휘로 판단되는 어휘는 발견되지 않았다. 이로부터 알 수 있는 사실은, 말뭉치의 규모가 일정 규모 이상인 경우, 기초 어휘 평정을 위한 말뭉치로서의 적합성은 더이상 말뭉치의 규모와 무관하다는 것이다.

한편, 2억 어절에서 추출한 어휘 목록은 1억 어절의 빈도 목록보다 92개의 어휘가 많으나, 이 중에도 기초 어휘로 판단되는 어휘는 발견되지 않았다. 한 발 나아가, 이 두 말뭉치의 1~50,000위의 순위 상관은 86%로, 상관 관계가 매우 높은 것도 확인할 수 있었다. 이는 다름 아니라, 말뭉치의 균형성과 대표성의 구현의 기준을 비슷하게 설정할 경우, 2억 어절과 1억 어절 말뭉치에서 기초 어휘 후보 목록을 추출한 결과는 크게 다르지 않음을 시사한다.

요약해 보면, 1억, 2억 어절의 균형 말뭉치와 20억 어절의 균형을 고려하지 않은 대규모 말뭉치 셋 중에, 기초 어휘 후보 목록의 추출에는 1억 어절 균형 말뭉치로도 충분히 원하는 기초 어휘 후보 목록을 포착할 수 있다는 결론을 얻게 된다.

4) 기초 어휘 선정용 말뭉치 연구에 대한 제언

기초 어휘 선정을 위한 향후 사업에서는 말뭉치 규모의 추가 확장보다는, 말뭉치의 균형성을 강화하고 형태이미 분석의 정확성을 향상하는 것이 우선될 필요가 있다.

말뭉치의 균형성 강화는 시기적인 균형성 강화 및 발달 단계 자료의 규모 확대 두 가지로 구체화된다. 1, 2억 어절의 말뭉치의 구성표에, 시기별 구성을 보면 여전히 2000년대 이후의 자료에 편중되어 있는데, 이 문제를 해결하기 위해 2000년대 이전의 구어 자료를 확보할 수 있는 방안이 고려되어야 한다. 또한, 발달 단계 아동의 자료, 특히 생성 자료를 최대한 확보하여 그 비율을 늘리는 것이 매우 필요하다. 올해 사업에서는 이전의 말뭉치에는 없던 발달 단계 자료를 최대한 포함시켰으나, 그 비율이 여전히 5%

에 못 미치는 한계가 있다. 이 중에 일부 생성 자료는 한국정보화진흥원에서 AI 학습용으로 구축한 데이터로서, 스크립트를 작성해 낭독함으로써 실제성이 낮은 문제가 있는 외에도, 구어 전사의 원칙이 여타의 국어원 구어 말뭉치의 그것과는 다른 문제를 안고 있다. 따라서 앞으로 발달 단계 자료를 최대한 확보하여 말뭉치에 반영하는 것이 기초 어휘 선정이라는 본 사업의 목표에 보다 접근할 수 있는 방법이 될 것으로 기대된다.

5. 기초 어휘에 대한 정성적 평정 및 활용

5.1. 정성적 평정을 위한 후보 선정

정성적 평정의 후보가 될 어휘 목록은 기존 어휘 등급 목록과 균형 말뭉치의 정량적 분석 결과를 종합하여 선정하였다. 후보 목록 선정에 활용한 어휘 자료는 다음과 같다.

번호	어휘 자료 명칭	자료 유형
1	『연세초등국어사전』 등재 표제어 목록(약 35,000개) 및 중요 어휘(약 3,000개)	기존 어휘 등급 목록
2	서상규(2014), 한국어 기본 어휘 목록(약 3,000개)	
3	김광해(2003), 국어 교육용 등급별 어휘 1, 2등급(약 5,000개)	
4	교육과정평가원 등급화 어휘 (등급화 제외 기초 어휘 약 500개, 초등 교육과정 교과서 학년군별 어휘 목록 약 20,000개)	
5	2억 어절 균형 말뭉치 어휘 빈도	정량적 분석 자료
6	2억 어절 균형 말뭉치의 텍스트 분포 및 산포도를 반영한 기초 어휘 점수	

표 25 기초 어휘 선정 기반 자료

<표 2>의 어휘 자료를 교차 검토하여 1등급 어휘 후보 목록 약 1만 개를 추려 내었다. 1만 개 후보 어휘는 교육과정 평가원의 등급화 제외 기초 어휘, 어휘 등급 목록들에서 1등급 수준에 속하는 어휘(자료 1의 중요 어휘 3000개, 자료 2의 중요도 1과 2에 속하는 어휘 2000개, 자료 3의 1등급 어휘), 균형 말뭉치 점수 순위가 5000위 안에 속하는 어휘, 점수 순위가 5000위~10000위 안에 속하면서 2등급 혹은 1-2학년군에 속하는 어휘를 모두 통합하여 선정하였다.

추려 낸 1만 개 후보 목록을 정성적 평정을 실시할 어휘와 정성적 평정 없이 1등급으로 선정할 어휘로 구분하였다. 1만 개 후보 목록 중 정성적 평정 없이 1등급으로 우선 선정한 기준은 다음과 같다.

첫째, 교육과정평가원의 등급화 제외 기초 어휘 약 500개는 1등급으로 우선 선정하고 정성적 평정 대상에서 제외하였다. (이 목록은 초등학교 교육용 어휘의 등급화 목록 중 학령전기 습득 어휘로 판정하여 초등 교육과정 등급화 데이터에서 제외한 어휘 목록이다. 이 어휘들은 다른 등급화 목록과 말뭉치 고빈도 어휘 목록에서도 중복 출현한 어휘들이다.)

둘째, 어휘 점수가 5,000위 이내에 속하면서 기존 목록 1등급 수준의 어휘 목록에 중복 출현한 어휘도 1등급으로 우선 선정하였다. 기존 1등급 수준의 어휘와의 중복 출현만을 기준으로 삼고, 기존 2등급 어휘와의 중복, 1-2학년 어휘와의 중복은 우선 선정 기준에서 제외하였다.

위 기준에 따라 우선 선정한 1등급 어휘 수는 약 2,400개이다. 이 어휘들은 정성적 평정에서 제외하고 여기에 속하지 않은 약 7,600개를 정성적 평정 대상으로 삼았다.

지금까지 설명한 과정을 도식화해 보면 다음과 같다.

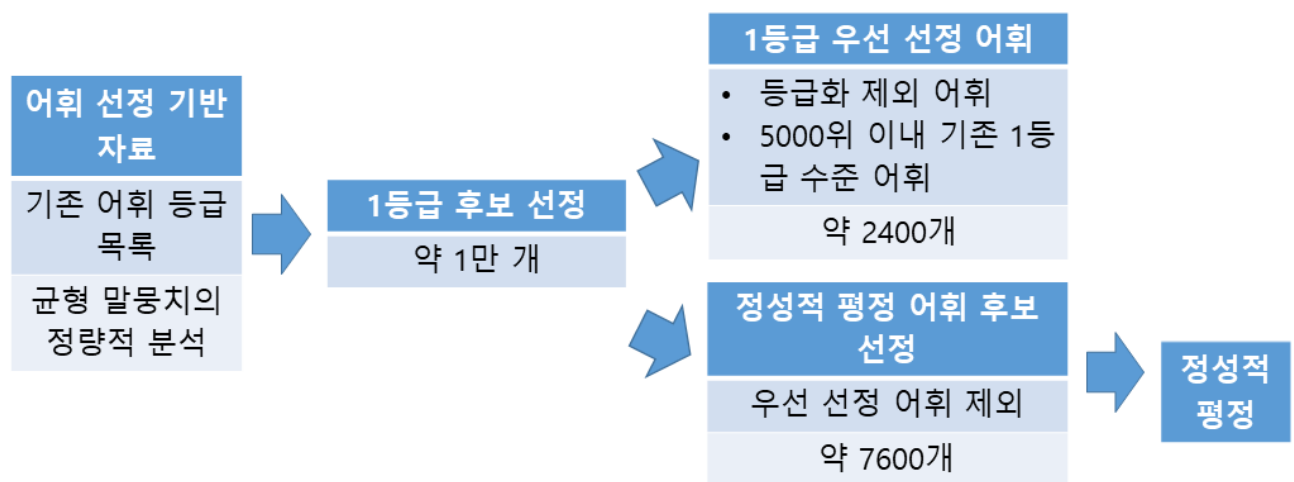


그림 8 정성적 평정 어휘 후보 목록 선정 과정

이상의 평정 후보 목록 약 7,600개에 대하여 전문가 집단과 어휘 사용자 집단을 대상으로 어휘 친숙도 조사를 단계적으로 실시하였다. 전문가 집단은 어휘 전문가와 국어 교육 전문가로 구성하였고, 어휘 사용자 집단은 교수 학습 현장의 초등학교 1학년 교사, 유치원 교사, 초등학교 1학년 학생으로 구성하였다. 학령전기에 해당되는 어휘 사용자인 7세 아동을 대상으로 한 조사도 고려해 볼 수 있겠으나, 아동을 대상으로 다량의 어휘에 대한 친숙도를 단기간에 조사하는 것은 현실적으로 쉬운 일이 아니므로 학령전기가 갓 지난 초등학교 1학년 학생과 1학년 교사, 7세 아동을 지도하는 현직 유치원 교사를 조사 대상에 포함하였다.

5.2. 어휘 전문가 평정

정성적 평정의 1단계는 평정할 후보 목록 약 7600개, 정확히는 7636개에 대한 어휘 전문가의 평정이다. 1등급 어휘는 주로 학령전기에 습득하여 지속적으로 사용하는 어휘이므로 이 평정에 참여할 어휘 전문가는 7세 자녀를 두고 있어 1등급 어휘 사용자의

언어생활을 관찰 또는 확인 가능하면서 사전 편찬 경험이 풍부한 연구자로 구성하였다. 정성적 평정에 이 단계를 포함한 것은 어휘 7636개는 교수 학습 현장 설문 조사를 실시하기에 매우 방대한 규모이기 때문이다. 평정 이전 단계에서 평정 후보 목록을 정량적으로 선정할 때 그 규모를 줄이는 방법도 있으나, 정량적으로 범위를 줄이는 과정에서 제외되어 1등급 어휘로 선정되지 못하는 어휘를 최소화하기 위하여 후보 목록의 범위를 넓게 정하였다.

어휘 전문가의 평정 단계에서는 후보 목록 7,636개를 검토한 결과, 1,945개를 1등급으로 선정하고 1,839개를 추가 평정 후보로 선정하였으며, 나머지 목록은 추가 평정 없이 1등급 어휘에서 제외하였다. 동일 어휘를 2명이 검토하여 2명 모두 학령전기 어휘에 동의하면 1등급으로 선정하고, 2명 모두 반대하면 1등급에서 제외하였다. 2명의 의견이 불일치하거나 추가 평정이 필요하다고 판정하면 추가 평정 후보로 선정하였다. 이 단계에서 1등급으로 선정된 어휘의 예로는 ‘놀이터, 반짝이다, 신호등, 코끼리, 풍선’ 등이 있다.

5.3. 어휘 사용자 평정(교수 학습 현장 설문)

본 사업 팀에서는 어휘 전문가의 평정 결과로 다음 단계 평정 후보 목록으로 선정된 1,839개의 어휘에 대하여 어휘 사용자의 친숙도를 조사하기 위한 교수 학습 현장 설문을 실시하였다. 설문 참여 집단은 초등학교 1학년 교사, 초등학교 1학년 학생, 유치원 교사로 구성하였다.

1) 초등학교 교사 대상 조사

교수 학습 현장 설문 조사 첫 단계는 초등학교 1학년의 담임을 맡고 있는 현직 교사 15명을 대상으로 실시하였다. 참여 교사의 연령은 29세에서 60세까지 고르게 분포하였다. 또한 국어교육 석사 이상으로서 다수의 현장 조사 연구 경험을 지니고 있어 어휘 조사의 취지를 이해하고 연구팀과 소통하면서 조사 방식을 충분히 숙지한 교사 3명이 자문위원으로 참여하였다. 현직 교사인 자문위원 3명이 각각 참여 교사 5명씩을 담당하여 조사 지침을 직접 설명하고 지침 관련 질의에 응답해 가며 조사를 진행하였다.

참여 교사에게는 구글 스프레드 시트 형식의 어휘 목록을 배포하여 해당 어휘가 학령전기 어휘이면 ‘Y’, 그렇지 않으면 ‘N’으로 온라인상에서 입력하게 하였다. 배포 목록에는 조사 어휘의 품사, 원어, 동형어 번호, 동형어 구분 길잡이말을 함께 제공하였다. 온

라인 어휘 시트 링크는 교사마다 개별적으로 배포하여 다른 교사의 응답을 볼 수 없도록 하여 간섭을 최소화하고자 하였으며, 동일 어휘는 소속 학교가 다른 교사가 평정하도록 배정하였다.

교사 대상 조사는 1차와 2차, 두 번에 걸쳐서 진행하였다. 1차 조사에서는 어휘 전문가가 추린 추가 평정 후보 1,839개에 대하여 실시하였는데, 한 어휘에 대하여 3명이 응답하게 하고 전원 Y로 응답한 어휘는 1등급 어휘로 선정하였다.

2차 조사는 1차 평정 결과 중 응답이 불일치한 목록 일부에 대하여 실시하였다. 2차 조사에서는 한 어휘에 대하여 5명이 응답하게 하고, 이 조사에서 전원 Y로 응답한 어휘는 1등급으로 선정하였다.

교사 대상 1차, 2차 조사를 통하여 1등급으로 선정된 어휘는 후보 1,839개 중 918개이다. 1차, 2차 조사에서 모두 응답자 반 이상이 'N'으로 응답한 어휘는 1등급 어휘에서 제외하였고 그 밖의 어휘 662개는 다음 단계의 평정 후보로 재선정하였다. 교사 대상 조사 단계에서 1등급으로 선정된 어휘의 예로는 '낙서, 목욕하다, 보라색, 장미꽃, 책가방' 등이 있다.

다음 그림은 교사 대상 조사 시트의 일부이다.

A	C	D	E	F	G	H	I	J	K
평정번호	어형	표준어	초등용어	초등원어	초등길잡이말	교사그룹	차수	평정(Y/N/P)	평정자메모
1528	미녀	0 명사	美女			1 1회차	y		
1529	키	4 명사			얼쇠	1 1회차	y		
1530	기대감	0 명사	期待感			1 1회차	y		
1531	딜다	0 동사				1 1회차	y		
1532	결승전	0 명사	決勝戰			1 1회차	y		
1533	컬러	1 명사	영 color			1 1회차	y		

그림 9 초등학교 교사 대상 조사 구글 스프레드 시트

2) 초등학교 1학년 학생 대상 조사

다음 단계인 초등학교 1학년 학생 대상 조사는 앞서 진행된 초등학교 교사 대상 1차 조사 결과의 응답 불일치 목록 중 일부인 480개에 대하여 실시하였다. 480개 조사 목록은 응답 불일치 목록 중에서 3명 중 2명이 Y로 응답한 어휘로 구성하였으며, 이미 3명 중 2명의 의견이 일치하였으므로 교사 조사를 추가로 진행하지 않고 다른 집단인 1학년 학생들이 실제로 알고 있는지 조사하기로 한 것이다. 조사에 참여한 초등학생은 교사 대상 조사에 참여한 교사 중 6명의 학급에 각각 속해 있으며 학생 수는 모두 134명이다. 교사 대상 조사와 마찬가지로 자문위원인 현직 교사 3명이 학생 조사를 실시하

는 담임교사들을 2명씩 담당하여 조사 지침을 설명하고 소통하며 조사를 진행하였다.

학생 대상 조사도 1차와 2차, 두 번에 걸쳐 진행하였다. 1차 조사에서는 '들어 본 적 없음'과 '의미를 앎' 중 하나에 응답하게 하여 전체 응답자 중 의미를 아는 학생의 비율을 계산하고 이를 기준으로 2차 조사 여부를 결정하였다. 1차 조사 결과에서 의미를 아는 학생의 비율이 40~70%인 어휘는 2차 조사를 진행하고, 70% 이상인 어휘는 1등급 어휘로 선정하는 것을 기본 지침으로 하였다. 또한 의미를 아는 학생 비율이 70% 이상이라도 조사 결과가 미심쩍거나 동형어의 의미와 혼동하여 잘못 응답했을 가능성이 보이는 어휘도 2차 조사 어휘에 포함하였으며, 이는 사업 팀과 자문위원 교사, 담임교사의 논의를 거쳐 결정하였다.

2차 조사는 어휘의 의미를 적거나 해당 어휘를 넣어 짧은 문장을 작성하는 과제를 수행하는 방식으로 진행하였다. 문항 응답지는 담임교사가 10점 만점을 기준으로 하여 채점하였으며, 어휘의 의미를 잘 알고 있다고 판단하면 10점, 어느 정도 알고 있다고 판단하면 5점, 모른다고 판단하면 0점으로 처리하고 동일 어휘에 대한 응답자의 평균 점수가 8점 이상인 어휘를 1등급으로 선정하였다. 1등급 어휘 최종 선정 단계에서 전문가의 재평정을 통하여 목록 보정 및 기준 점수 조정이 필요하다.

학생 대상 조사의 경우, 480개 후보 중 1차 조사에서 175개, 2차 조사에서 31개가 1등급 어휘로 선정되었으며 선정 어휘의 예로는 '위치, 현관문, 난방, 가스, 수고하다, 원장' 등이 있다.

다음 그림은 각각 초등학교 대상 1차, 2차 조사 문항의 일부이다.

		이 낱말을 책이나 주변에서 본 적 있나요?		이 낱말의 뜻이 무엇인지 알고 있나요?	
1	일반인	본 적 있어요 (<input checked="" type="radio"/>)	없어요 (<input type="radio"/>) ✓	뜻을 알아요 (<input checked="" type="radio"/>)	잘 모르겠어요 (<input type="radio"/>)
2	품	본 적 있어요 (<input type="radio"/>)	없어요 (<input checked="" type="radio"/>)	뜻을 알아요 (<input checked="" type="radio"/>)	잘 모르겠어요 (<input type="radio"/>) ✓
3	애인	본 적 있어요 (<input checked="" type="radio"/>)	없어요 (<input checked="" type="radio"/>) ✓	뜻을 알아요 (<input type="radio"/>)	잘 모르겠어요 (<input checked="" type="radio"/>)
4	설득하다	본 적 있어요 (<input type="radio"/>)	없어요 (<input type="radio"/>)	뜻을 알아요 (<input checked="" type="radio"/>)	잘 모르겠어요 (<input type="radio"/>)
5	일생	본 적 있어요 (<input type="radio"/>)	없어요 (<input checked="" type="radio"/>) ✓	뜻을 알아요 (<input type="radio"/>)	잘 모르겠어요 (<input checked="" type="radio"/>) ✓
6	환상	본 적 있어요 (<input type="radio"/>)	없어요 (<input checked="" type="radio"/>) ✓	뜻을 알아요 (<input type="radio"/>)	잘 모르겠어요 (<input checked="" type="radio"/>)
7	전원	본 적 있어요 (<input type="radio"/>)	없어요 (<input checked="" type="radio"/>) ✓	뜻을 알아요 (<input checked="" type="radio"/>)	잘 모르겠어요 (<input type="radio"/>)
8	세포	본 적 있어요 (<input checked="" type="radio"/>)	없어요 (<input type="radio"/>) ✓	뜻을 알아요 (<input checked="" type="radio"/>)	잘 모르겠어요 (<input type="radio"/>)
9	광경	본 적 있어요 (<input type="radio"/>)	없어요 (<input checked="" type="radio"/>)	뜻을 알아요 (<input type="radio"/>)	잘 모르겠어요 (<input checked="" type="radio"/>) ✓

〈그림 10〉 초등학교 대상 1차 조사 문항 예시

12	유치되다	남말의 뜻 남말을 넣어 문장 쓰기	○ 온도가 유지되다.
13	대비	남말의 뜻 남말을 넣어 문장 쓰기	○ 불나는데 대비해요
14	당분간	남말의 뜻 남말을 넣어 문장 쓰기	○ 당분간 집에 있어요
15	충격적	남말의 뜻 남말을 넣어 문장 쓰기	○ 충격적이게 했어요
16	출입	남말의 뜻 남말을 넣어 문장 쓰기	○ 문이 닫아해요.
17	집중되다	남말의 뜻 남말을 넣어 문장 쓰기	○ 공부가 집중되다.

12	유치되다	남말의 뜻 남말을 넣어 문장 쓰기	×
13	대비	남말의 뜻 남말을 넣어 문장 쓰기	○ 대풍이 올 대면 저 대비해요
14	당분간	남말의 뜻 남말을 넣어 문장 쓰기	○ 엄마가 마트가다 올게 당분간 집에 있어
15	충격적	남말의 뜻 남말을 넣어 문장 쓰기	×
16	출입	남말의 뜻 남말을 넣어 문장 쓰기	×
17	집중되다	남말의 뜻 남말을 넣어 문장 쓰기	○ 나는 공부할 때 집중되다

〈그림 11〉 초등학교 대상 2차 조사 문항 예시

3) 유치원 교사 대상 조사

유치원 교사 대상 조사는 유치원 현직 교사 8명을 대상으로 실시하였다. 조사 어휘는 앞서 진행된 초등학교 교사 대상 1차, 2차 조사 결과의 응답 불일치 목록 중 182개 어휘이다. 교사 2차 조사나 학생 조사를 실시하지 않은 목록, 교사 2차 조사 결과 중 선정되지 않았으나 비교적 Y 응답이 많았던 어휘 목록이 그 대상이 된다.

유치원 교사 조사 방식은 초등학교 교사와 동일하게 온라인 어휘 시트를 제공하여 진행하되, 7세 아동이 알고 있는 어휘에 대해서 Y로 응답하도록 하였다. 동일 어휘에 대해서 4명이 응답하였는데 4명 전원이 Y로 응답한 어휘를 검토한 결과 ‘작곡가’, ‘척추’ 등의 어려운 어휘가 다수 포함되어 있었다. 또한 ‘오십’에 대한 Y 응답은 3명인데 ‘육십’에 대한 Y 응답은 4명이어서 교사 응답의 전원 일치율 기준으로 선정하기 어렵다고 판단하여 어휘 전문가의 재평정을 추가로 진행하였다. 재평정은 교사 3명 이상이 Y로 응답한 어휘 372개에 대하여 실시했으며, 이 어휘들에 대하여 어휘 전문가 2명이 모두 Y로 응답한 어휘 69개만을 1등급 어휘로 선정하기로 하였다. 선정된 어휘의 예로는 ‘맞히다, 몽땅, 오십, 육십, 행사’ 등이 있다.

유치원 교사 대상 조사에서는 어휘 조사 지침을 이해하여 현장에서 어휘 조사를 안내하며 협조해 줄 교사를 섭외하는 것이 쉽지 않아서 본 연구자들이 직접 온라인 설명을 통하여 조사를 진행하였는데, 이 점이 이번 조사에서의 아쉬운 점이다. 따라서 어휘 전문가의 재평정을 통하여 이를 보완하고자 하였다.

5.4. 교육 전문가 평정

다음으로 교육 전문가 평정 단계에서는 앞선 단계에서 확보된 어휘 목록을 검토하여 목록에서 삭제되는 것이 적절하다고 생각하는 어휘를 선정하거나 목록에 추가되는 게 바람직하다고 생각하는 어휘를 발굴하는 작업을 거쳤다.

어떤 기준으로 삭제 대상을 선별하고 목록에 추가할 단어를 선정하였는지를 예와 함께 제시하면 다음과 같다.

1) 삭제 대상 어휘 선정 기준

먼저 삭제 대상 어휘를 선정할 때, 피아제의 인지 발달 이론을 근거로 하여 전조작기(2-7세) 단계의 인지 특성을 반영하고자 노력하였다. 이를 위해 아동의 발달적 특성에 비추어 봤을 때, 매우 추상적이며 일상생활에서의 사용 빈도가 비교적 낮은 것으로 판단되는 어휘를 삭제했다. 또한 학령 전 아동의 일상생활이나 학습 상황과 밀접하게 관련되어 있는 개념 어휘에 해당되지 않는다고 판단되는 어휘들을 삭제 후보로 선정하였다.

※<참고 자료> 전조작기(2-7세) 단계의 인지 특성

- 지각이 아동의 사고를 지배하는 시기
- 급속한 언어발달을 겪으며, 상징적 사고가 발달
- 정신적 조작을 수행할 수 있는 능력을 갖추. 즉, 개, 고양이, 곰 등 다른 종류의 동물을 분류할 수 있는 아동은 정신적 조작을 수행한다고 볼 수 있음.
- 개념 획득 능력이 빠르게 성장하고 다양한 개념을 형성함.
- 구체적 사물에 대한 개념 획득은 용이하나, 추상적 개념의 습득은 한정되어 있음. 즉, 공정성, 민주주의, 에너지 등과 같은 추상적인 개념은 획득하지 못함.

세부적으로 특정 기준을 적용했을 때 삭제하는 게 적절하다고 판단한 어휘를 몇 가지 예로 들면 다음과 같다.

○ 3-4, 5-6학년 의사소통 한국어, 학습 도구 한국어에 포함되어 있는 것 중에 학령전기 어휘에 포함된다고 보기 어려운 것은 삭제 후보로 선정

(예) 범위, 미터, 참여, 참여하다, 측정하다, 특징, 판단, 판단하다 등

○ 자모는 학령전기 학습 대상이 아니며 이들을 말로 익히는 것이기 때문에 삭제 후

보로 선정

(예) 아, 야, 어, 여

2) 목록 추가 어휘 발굴 기준

먼저 목록에 추가할 어휘를 선정할 때, 학령 전 아동의 일상생활이나 학습 상황과 밀접하게 관련되어 있는 개념 어휘라고 판단되는 것 중에, 목록에 빠진 것을 발굴하여 목록에 추가하고자 했다. 어린이집이나 유치원에서의 학습 상황이나 가정에서 부모와의 학습 상황에서 자주 쓰이는 단어, 일상생활에서 접하는 빈도가 높은 단어 등도 목록에 추가하고자 했다.

세부적으로 특정 기준을 적용했을 때 추가하는 게 적절하다고 판단한 어휘를 몇 가지 예로 들면 다음과 같다.

○ 유아용 교재, 1-2학년, 3-4학년의 의사소통 한국어, 학습 도구 한국어 중 빈도가 높은 것 중 누락된 것을 추가, 유아용 교재 및 초등학교 선생님들 10여 명이 한국 초등학교 저중고 학년의 학교생활, 일상생활, 교과학습에 요구되는 기본적인 어휘를 선정하여 집필한 교재의 목록과 교차 검토하여 특히 저학년 어휘 중 누락된 경우 추가

(예) 색종이, 스케치북, 색연필, 크레파스 등

○ 많이 사용되는 나라 이름을 추가

(예) 미국, 일본, 중국

3) 참고 자료

어휘 검토에 참고한 자료는 다음과 같다.

- 공주대학교 부설 유치원 교육 계획안(2년 분량)
- 유아 어휘 및 초등 입문기 관련 책자
 - 똑똑한 유아 어휘(웅진 주니어 편), 1,2,3단계로 구분, 각 단계별 3권씩 제작(1단계: 생활 어휘, 동요·동시, 학습어휘 / 2,3단계: 생활 어휘, 이야기 속 어휘, 학습 어휘)
 - 초등교과서어휘능력(아울북 편), 예비단계(예비 초등~초등 1학년), 총 4권으로 제작, 취학 전 알아야 할 어휘로 1,500개 제시
 - 어휘가 독해다(EBS 편), 입문(1·2학년) 권에 교과서 어휘 126개 수록

- 김경령 외(2019). 신나는 한국어(유아용 활동지), 하우출판사
- 시중에서 사용되는 어휘력 검사지
- Receptive and Expressive Vocabulary Test(REVT)의 수용 어휘력 검사지, 표현 어휘력 검사지 확보(예: 표현 어휘력 검사에서 5~6세 관련 어휘로 ‘시소, 놀이터, 병아리, 찢술, 주다, 경찰, 무겁다/무거운, 네모 등 30여 개 단어 제시)
- 표준 한국어 교재의 목표 어휘
- 초등학생을 위한 표준 한국어 저학년 의사소통 1,2,3,4 (익힘책 제외)
- 초등학생을 위한 표준 한국어 고학년 의사소통 1,2,3,4 (익힘책 제외)
- 초등학생을 위한 표준 한국어 학습도구 1~2학년용
- 초등학생을 위한 표준 한국어 학습도구 3~4학년용
- 초등학생을 위한 표준 한국어 학습도구 5~6학년용

5.5. 평정 결과 반영 및 최종 선정

우선 선정 어휘 목록과 평정을 통해 선정된 목록을 전문가 평정 결과를 참조하여 재검토하고 목록을 보정하여 1등급 어휘 5,000개를 최종 선정하였다.

목록 보정 단계에는 다음의 과정이 포함되어 있다.

첫째, 주요 어휘장 목록을 추려서 누락 어휘가 있는지 확인하고 보완하였다. 예를 들어, 숫자와 요일이 모두 포함되어 있는가 하는 등의 확인을 말한다.

둘째, 말뭉치는 공적 텍스트의 비중이 높으므로 고유어 빈도가 상대적으로 낮게 나타날 수 있다는 점을 감안하여 고유어 목록을 따로 검토하여 고유어 기초 어휘가 누락되지 않도록 확인하였다.

셋째, 빈도와 어휘 점수, 평정 결과의 경계에 있는 어휘들에 대하여 어휘 전문가 및 교육 전문가의 정성적 평정을 다시 실시하여 최종 선정하였다.

6. 최종 결과물

본 사업은 앞서 제시한 기초 어휘 목록 선정 및 등급화 방법론 연구 결과와 말뭉치 분석 방법론 연구 결과, 기초 어휘에 대한 정성적 평정 결과를 종합적으로 고려하여 최종 결과물을 산출하였다.

6.1. 기초 어휘 등급화안

연령 (만)	학년	어휘 수	어휘 수 (누적)	보유 어휘 목록	등급(제안)	
4-5세	어린이집	2,000	2,000	‘연세초등국어사전’ 중요 어휘 2,000	1	별 ***
5-6세	유치원	3,000	5,000	‘A Frequency Dictionary of Korean’ 문구어 고빈도 어휘 5,000		
				‘한국어 교육을 위한 기초 어휘’ 5,000		
7세	초1	1,000	6,000		2	별 **
8세	초2	1,500	7,500	‘한국어 기본어휘 의미 빈도 사전’ 기본어휘 표제어 목록 7,000		
9세	초3	2,500	10,000		3	
10세	초4	3,000	13,000			
11세	초5	4,500	17,500		4	
12세	초6	5,500	23,000			
13세	중1	5,500	28,000		5	별 *
14세	중2	5,500	34,000	‘초등 어휘에 대한 초기 등급화 데이터’ 34,000		
15세	중3	6,000	40,000	‘연세초등국어사전’ 표제어 목록 40,000		

표 26 어휘 등급화안(제안)

본 사업에서는 등급화할 기초 어휘의 규모를 약 40,000여 개로 제안하고자 한다. 본 사업에서 선정할 기초 어휘가 기능 어휘와 고유명사를 제외한다는 점을 고려할 때 선행 과제의 기초 어휘 규모와 크게 다르지는 않다. 오기와라 히토시(2016)에서 조사한 바에 따르면 일본 대학생 4학년의 사용 어휘는 약 3만 어, 이해 어휘는 약 4만 5천이라고

하며 한국의 의무 교육에 해당하는 중학교 때까지 학습하거나 습득하는 어휘의 양은 약 4만여 개라고 한다. <연세초등국어사전>의 표제어 수 역시 4만여 개라는 점을 감안하여 기초 어휘의 규모를 4만여 개 정도로 본 것이다. 초등학교부터 중학교 수준 어휘까지 수록한 <보리국어사전>의 표제어도 4만여 개로 비슷한 규모이다.

기초 어휘의 규모를 5만여 개로 늘리면 고등학교 수준 어휘를 포함하게 되는데, 고등학교 어휘 단계는 기본적인 언어생활을 위한 기초 어휘를 습득하는 단계를 넘어 전문 어휘, 교양 어휘를 학습하는 단계에 해당되므로 이 어휘들은 국어 기초 어휘로 선정하여 등급화하는 것보다는 교과별, 전문 영역별로 선정하는 것이 적절하다.⁹⁾

6.2. 1등급 어휘 목록

최종 선정한 1등급 어휘 5,000개의 목록은 보고서의 부록으로 붙이고 결과물 파일로 제출한다. 부록에는 어휘와 표준국어대사전 동형어 번호를 함께 제시하고, 제출 결과물에는 품사 정보도 함께 보인다.

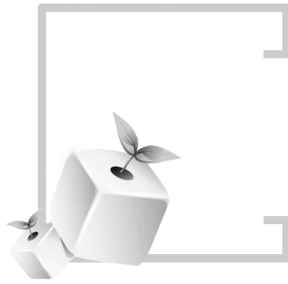
1등급 선정 어휘 5,000개의 2억 어절 말뭉치에서의 누적 비율은 68.69%이다. 앞서 표 23에서 살펴본 실질어휘 상위 빈도 5000개의 누적 비율인 80.59%보다 낮은 것은 정성 평정에 의하여 빈도가 낮은 어휘가 선정되면서 어휘 목록이 조정되었기 때문이다.

6.3. 기초 어휘 예비 목록

기초 어휘 예비 목록으로는 1등급 선정 어휘 5,000개를 포함하여 39,398개를 선정하였다. 제출한 예비 목록 어휘의 2억 어절 말뭉치에서의 누적 비율은 92.40%로, 표 23에서 살펴본 상위 빈도 40,000개의 누적 비율인 94.79%와 비교적 유사하다. 정량적 기준으로 뽑은 40,000개의 누적 비율보다 조금 낮은 것은 1등급 어휘에 정성적 평정 결과가 반영되었기 때문이다. 한편, 기초 어휘 목록 전체를 정량적 기준으로 상위 50,000개로 늘릴 경우 2억 어절에서의 누적 비율은 95.64%임을 보아 상위 40,001~50,000위 사이의 어휘가 차지하는 텍스트 점유율은 크게 높지 않음을 알 수 있다.

예비 목록은 정성적 평정을 완료한 목록이 아니기 때문에 향후 단계적 어휘 등급화를 통하여 어휘 구성은 달라질 수 있다.

9) 김창원(2012)에서는 “고등학교 어휘교육은 초·중학교의 어휘교육과 성격 및 수준을 달리 해야 한다. 생활 어휘·기능 어휘 중심의 기초·기본 교육 단계를 지나 개념 중심의 교양 어휘·전문 어휘를 습득하고, 사전적·문맥적 의미 이해를 넘어선 함축적·문화적 의미의 활용 단계로 나아가며, 한자어와 외국어를 포괄하는 확산적 어휘교육이 이루어지는 단계가 고등학교 교육이다.”라고 설명하고 있다.



제 4 장

결론



1. 사업 요약

본 사업은 기초 어휘 목록 선정 및 어휘 등급화를 목적으로 하여 수행되었다. 본 사업의 결과물은 크게 기초 어휘 등급화안, 선정한 기초 어휘 후보 목록, 그중 1등급 어휘 확정 목록 이렇게 세 가지이다. 이 결과를 위해 기존 어휘 등급 수집 및 정제, 말뭉치 구성 및 구축, 분석 등 자료 정비와 어휘 사용자와 전문가로 구성된 평정이 함께 진행되었다.

2. 의의 및 기대 효과, 활용 방안

본 과제 수행의 의의 및 기대 효과는 다음과 같다.

본 과제는 현대 한국어 사용 양상을 공시적인 관점에서 포착하고자 한 연구로 본 과제의 연구 결과는 언어 사실에 대한 객관적 기술 자체에 큰 의의가 있다. 본 과제의 연구 결과는 순수하게 학문적 목적에서 활용될 수 있을 뿐만 아니라 실용적 목적에서도 교육이나 언어 정책 수립 등 전반에 걸쳐 다양하게 활용될 수 있다는 점에서 파급력이 크다. 특히 본 과제는 수행 과정에서 기초 어휘 목록 선정용 자료의 수집과 정비, 기초 어휘 범위 및 등급 제시, 1등급 어휘 목록 제시 등에 대해 객관적이고 구체적인 절차와 방법을 마련하였으며, 이에 따라 정량적·정성적 근거를 확보하고자 하였다. 이 과정에서 특정 연구 목적의 말뭉치 규모에 대한 이론적·실증적 기준을 제시하였다. 본 사업의 대상이 되는 기초 어휘 1등급인 5,000여 개의 경우뿐만 아니라 기초 어휘 전체의 목록을 선정하기 위해서도 약 1억 어절 규모의 말뭉치로 충분하다는 것은 시사하는 바가 크다. 말뭉치의 규모보다 구성과 균형성이 더 타당성 있는 결과를 얻을 수 있다는 점 역시 후속 연구에 기여할 수 있는 바가 크다고 할 수 있다. 정성적 평정 역시 여러 번의 교차 검증을 통해 방법론을 정교화시켰으며 그 과정을 구체적으로 제시하고 있다. 앞으로의 기초 어휘 연구를 위한 방법론 정립에 큰 도움이 되리라 생각된다. 또한 ‘기초 어휘’를 연속성과 확장성이라는 특성을 통해 개념화하고 이를 가시화하기 위한 팀별 수행 결과가 과업의 목표를 이루는 데 기여하였다고 할 수 있다.

구체적인 의의 및 기대 효과는 다음과 같다.

기존에 구축된 다양한 국어 자료의 정비를 통해 기초 어휘 목록의 타당성을 제고할

수 있다. 발주 기관인 국립국어원에서 구축한 모두의 말뭉치, 국어교육용 어휘 선정에 위한 기초 연구와 어휘 분석 말뭉치 등의 기존 자료를 기초 어휘 목록 선정에 활용하였다. 둘째로, 기존 자료 이외의 기초 어휘 목록의 검증 및 활용에 도움이 될 만한 추가 자료를 확보하였다. 사업 수행 기관에서 보유하고 있거나 연구 진행 중인 어휘 자료를 추가로 확보하여 과제 수행 결과로 선정한 기초 어휘 목록의 검증에 활용함으로써 동 분야 연구 성과의 통합을 이루어낼 수 있도록 하였다.

다음으로, 수준별 어휘 등급화 목록 작성을 위한 방법론을 수립하였다. 학계나 업계의 수준별 어휘 등급 목록은 주로 여러 유형의 어휘 목록을 통합하여 도출하는 메타계량법에 의존하여 작성해 왔으나 실제 사용 양상을 반영하는 정량적 검증, 교육 현장 전문가의 정성적 검증을 통해 새로운 방법론에 의한 어휘 등급별 목록을 구축하고자 하였다.

또한 본 사업은 교과서 등의 교육 자료 개발 및 감수, 국가 고사 출제, 국민의 국어 능력 평가를 비롯한 각종 기초 조사 등의 객관적 근거를 제시하고자 하였다. 특히 수준별 어휘 등급 목록의 활용처가 주로 국어 교육 현장임을 고려할 때 교육 단계별 교수자 집단의 검증을 거친 어휘 목록은 교육 분야에 객관적 어휘 지표로 활용 가능할 것으로 기대된다.

또한 본 사업의 결과물은 교과서 개정 및 보완, 어휘 교육 등에 활용되어 국민의 언어생활 향상에 기여할 수 있다. 등급별로 평정된 수준별 어휘는 교과서 기술용 메타언어의 난도 조정, 도서의 이독성 판정 등에 사용되어 정규교육을 포함한 국어교육 콘텐츠의 개발 및 선택에 기여함으로써 국민 언어생활의 향상에 실질적으로 기여할 수 있을 것으로 보인다.

무엇보다 본 과제의 정교한 연구 방법론을 활용하여 한국어 어휘 사용 양상을 파악하는 연구 및 조사가 지속적으로 이루어진다면 한국어 어휘 사용 양상의 변화 등을 파악하는 데도 큰 도움이 되리라 생각된다.

3. 향후 연구 및 정책 제언

본 과제 수행 결과를 바탕으로 향후에 진행하여야 할 연구 과제 및 정책에 대해 제언하면 다음과 같다.

1) 자료 관련 연구 및 정책 제언

먼저, 사용하는 말뭉치의 균형성을 강화할 필요가 있다. 본 과제의 연구 결과에 따르

면 기초 어휘 규모를 40,000개로 상정하고, 그 후보 목록을 약 50,000만 개로 추출할 경우, 1억, 2억, 20억 규모의 말뭉치에서 모두 95%가 넘는 순위 일치도를 보였다. 구간별 어휘량 비교를 통해 살펴보았을 때 기초 어휘 선정에 위해서는 말뭉치의 규모보다 균형성 강화가 더 필수적으로 나타났다. 말뭉치의 균형성 강화는 발달 단계 자료의 규모 확대 및 시기적 균형성 두 가지로 나누어 볼 수 있다. 현재 참조 말뭉치에서 발달 단계 자료가 차지하는 비중은 약 5%로 만족할 만한 규모는 아니다. 따라서 아동 대상 자료나 아동 발화 자료 등 발달 단계 자료를 보완하여야 할 필요가 있다. 또한 말뭉치의 구성이 20세기에 치우쳐 있는 경향이 있어 2010~2020년대의 최신 말뭉치 자료를 보완할 필요가 있다. 구어 자료의 확충 역시 필요한데 기초 어휘 선정에 있어 말뭉치 자료가 차지하는 비중은 매우 크지만 어휘 선정 과제에서 말뭉치 구축까지 함께 하기는 현실적으로 불가능하기 때문에 이러한 자료 구축은 별개의 사업 과제로 분리될 필요가 있다고 본다.

2) 용어 관련 제안

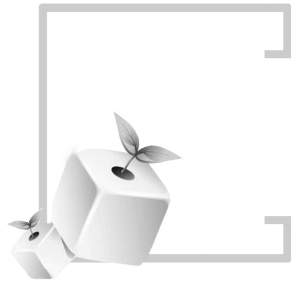
본 과제의 과제명이 ‘기초 어휘 선정 및 어휘 등급화 연구’로 설정되어 있고 기존 1~3년차 과제에서 기초 어휘로 약 4-5만여 개의 어휘를 등급화 대상으로 삼고 있는 것으로 미루어 볼 때 국립국어원에서 제시하는 ‘기초 어휘’란 일상 언어생활에서 자주, 널리 쓰이는 필수적인 어휘로 생각된다. 그러나 학계에서 ‘기초 어휘’는 언어 생활에서 가장 기초가 되는 2,000~3,000여 개의 최상위 빈도 어휘를 가리키거나 통시적 관점을 아울러서 고대 국어에서부터 현재까지 쓰이는, 조어의 핵이 되는 단일어, 고유어로서의 어휘 목록을 가리키는 경우가 많아 용어의 사용에 혼란이 있을 수 있다. 학계에서 통상적으로 일상 언어생활을 영위하는 데 꼭 필요하다고 생각되는 기본적인 어휘를 ‘기본 어휘’라고 하는 점을 고려하였을 때, 그리고 기초(基礎)와 기본(基本)의 의미를 고려해보았을 때 본 과제는 ‘기본 어휘’를 선정하는 것으로 보아야 한다. 따라서 본 과제에서 선정하여야 할 어휘들의 목록은 ‘기초 어휘’가 아닌 ‘기본 어휘’라고 지칭하여야 함을 제안하고자 한다.

3) 사업 진행 관련 제안

본 과제의 목적이 한국어 기초 어휘의 사용 양상을 파악하고 이에 따라 이들 기초 어휘를 등급화하는 것이라고 볼 때 본 과제는 연속성이 있는 과제라고 할 수 있다. 본 과

제에서는 1등급 어휘를 선정하였으므로 2, 3, 4등급, 또는 그 이상의 등급의 기초 어휘 등급 선정 사업이 계속 이루어져야 할 것이다. 기초 어휘 선정에 대한 기본 원칙은 1등급부터 그 이상의 모든 기초 어휘 등급에 동일하게 적용되어야 하므로 본 과제에서 제시한 방법론을 비슷하게 적용하되, 본 사업 결과를 검토·보완하여 향후 기초 어휘 선정 사업에서 연계하여 수행할 수 있도록 할 필요가 있다. 즉, 기초 어휘 선정 사업은 1등급부터 모든 단계에 걸쳐 유기적으로 수행·보완되어야 하므로 과제 수행 방법과 원칙에 일관성을 기하되, 각 단계별로 적절한 연구 규모, 연구 기간 등을 조정하여 수행될 필요가 있다. 특히 등급이 올라갈수록 어휘량이 많아지고 난이도도 높아지기 때문에 정성적 평정 등에 인력과 시간이 더 소요될 수 있다는 점을 고려하여 적절한 사업 계획을 수립할 필요가 있다.

이 외에도 정성적 평정에 있어 개인 간의 격차를 줄이는 방안에 대해 좀 더 정밀한 연구가 필요하다. 개인의 어휘 평정은 자의성이 개입될 수밖에 없는데 이러한 자의성을 최대한 배제하고 객관적인 평정 결과를 얻기 위해서는 평정 과정이나 관련 지침이 섬세하게 고안될 필요가 있다. 현재 어휘 전문가, 어휘 사용자, 국어 교육 전문가, 사업 팀의 순으로 평정 작업을 진행하였는데, 학습자 평정과 교사 집단의 평정 순서를 바꾸거나 각 평정 주체별로 평정한 결과를 해석할 때 어떤 집단의 평정 결과에 가중치를 둘 것인지 등에 대해 고민이 필요하다고 할 수 있다. 이에 대해서는 후속 연구가 필요하다고 하겠다.



별첨 자료

부록

1. 회의록 요약

1) 중간보고

- 일시 : 6월 30일 수요일 오전 10시~11시
- 참석자 : 서상규, 김한샘, 이병규, 가은아, 윤천택, 황은하, 봉미경, 김선헌, 남미정(국어원 담당자), 신호철, 도원영, 신동광, 조현나(국어원), 이운영(국어원)
- 과제 수행 현황 보고 및 향후 과제 추진 일정 보고.
- 사전 미등재어의 선정 문제, 전문가 평정 및 사용자 평정 결과의 반영 방식에 대해서 논의.
- 자문위원(도원영, 신호철, 신동광) 자문 의견 수렴.

2) 최종보고

- 일시 : 12월 16일 목요일 오후 2시~4시
- 참석자 : 서상규, 김한샘, 봉미경, 김선헌, 황은하, 윤천택, 도원영(자문위원), 최정도(자문위원), 신호철(자문위원), 남미정(국어원 담당자), 이운영(국어원 사전팀)
- 과제 수행 결과 보고 및 논의 사항.
- 정량적·정성적 평정 과정 및 결과물 예시, 기초 어휘 등급화 목록 활용 방안, 1등급 이상의 어휘 등급에 대한 제안, 향후 사업 계획 등을 제시.
- 국어원 자문위원(도원영, 신호철, 최정도)의 자문 의견을 수렴.

3) 공동연구진 회의

사업 수행 중 수시로 실무진 회의와 공동연구진 회의를 진행하였다. 실무진 회의에서는 사업 수행의 각 단계에 맞게 사업 수행이 이루어질 수 있도록 세부적인 사항을 논의하였다. 공동연구진 회의에서는 사업 수행 전반에 대한 방향 설정 및 수행 진척도 평가 등 각 실무팀들이 사업 목표에 맞게 세부 과제를 진행할 수 있도록 지속적으로 사업 수행 내용을 공유하였다. 실무진 회의는 각 분과별로 수시로 진행하였으며 실무 자문도 겸하였다. 공동연구진 회의는 한 달에 한 번 주기로 진행하였으며 중간보고회 및 최종보고회를 앞두었을 때에는 추가적으로 실시하였다. 실제 내용은 다음과 같다.

▶ 1차 공동연구진 회의

- 일시 : 6월 30일 수요일 오전 10시~11시

- 참석자 : 서상규, 김한샘, 이병규, 가은아, 윤천탁, 김선혜
- 사업 개요 설명 및 사업 수행 일정 공유. 진행 상황 보고. 각 팀 과업 내용 공유.
- ▶ 2차 공동연구진 회의
- 일시 : 7월 22일 목요일 오후 1시~3시
- 참석자 : 서상규, 이병규, 윤천탁, 가은아, 황은하, 봉미경, 김한샘, 김선혜
- 과제 진행 상황 확인 및 논의사항 공유
- 어휘 분석팀: '서술성 명사+하다'의 처리 문제, 구-합성어 구분 문제, 고유명사 포함 문제, 사전 미등재어 처리 문제 등.
- 말뭉치 분석팀: 말뭉치 구성에서 구어 비율 확대, 발달 단계 자료 확장, 시기별/장르별 균형성 문제 등.
- ▶ 3차 공동연구진 회의
- 일시 : 8월 26일 목요일 오전 10시~12시
- 참석자 : 서상규, 김한샘, 이병규, 가은아, 윤천탁, 황은하, 봉미경, 김선혜, 남미정(국어원 담당자)
- 어휘 분석팀: 초등교사 및 초등생 대상 어휘 평정 기획, 분석 대상 및 단위 확정, 어휘 목록 정제 작업 진행 등.
- 말뭉치 분석팀: 말뭉치 구성안 수정 및 샘플링, 어휘 정보 추출 및 기초어휘 점수 산출, 변수별 가중치 산출 방안 논의 등.
- 평정팀: 전문가 평정 참고 자료 작업, KSL 어휘 목록 활용 방안, 학습도구 어휘 목록/동요 가사 아카이브 등 자료 확보 현황 공유.
- ▶ 4차 공동연구진 회의
- 일시 : 11월 17일 수요일 오전 10시~12시
- 참석자 : 서상규, 김한샘, 이병규, 윤천탁, 황은하, 봉미경, 김선혜, 남미정(국어원 담당자), 박혜진
- 과제 진행 상황 확인 및 논의사항 공유
- 어휘 분석팀: 어휘 계열관계 고려 및 실생활 어휘 보완 필요.
- 말뭉치 분석팀: 초등 작문 자료를 통한 발달 단계 어휘 목록 보완 필요, 말뭉치 비교 구간별 상대 빈도 적용 등.
- 평정팀: 초등학교 학생들 수준 이해/표현 어휘 고려하여 필수 1등급 어휘 작성 예정, 초등 작문 데이터를 통해 현재 목록에 미포함된 어휘 검토.
- ▶ 6차 공동연구진 회의
- 일시 : 12월 1일 수요일 오후 3시~4시

- 참석자 : 서상규, 김한샘, 봉미경, 김선헌, 황은하, 강범일, 윤영민, 천성호, 김세현
- 사용자 어휘 조사 결과 반영 및 교육 전문가 평정 방법론 논의
- 사용자 어휘 조사 결과 10점 만점 기준으로 어느 정도의 점수를 ‘대다수가 알고 있다’고 간주할 것인가.
- 아동의 어휘 교육, 초등 교육의 관점에서 꼭 포함되어야 할 어휘 유형이 있는가.
- 어휘의 정량적 평가 결과와 정성적 평가 결과 사이의 차이가 있는 경우 어디에 어느 정도에 가중치를 둘 것인가.

3) 자문 회의

사업 수행 내용을 평가하고, 보완점 등을 제시할 수 있는 자문위원을 섭외하여 자문 회의를 진행하였다. 특히 사업 초기에 사업의 목적에 맞는 사업 수행 방향을 설정하기 위하여 자문 위원 5인(강현화, 김향희, 남기심, 노명완, 유현경)과 회의를 진행하였다. 간략한 자문 내용은 다음과 같다.

▶ 1차 자문회의

- 일시 : 6월 30일 수요일 오전 11시~12시
- 자문위원 : 강현화, 김향희, 남기심, 노명완, 유현경
- 연구 목적이 선명해질 필요가 있음.
- 연구 목적에 따라 말뭉치 구성을 달리할 필요가 있음. 문어-구어 비율, 교육용 말뭉치 비율 등을 고려해야 함.
- 정량적 평가와 정성적 평가의 가중치를 어떻게 설정할 것인지 논의 필요.
- 활용 범위를 언어병리학, 언어치료, 언어발달 장애 등까지 넓힐 수 있는지 검토 필요.

2. 참고 문헌 목록

1) 논문 및 단행본

- 강현화·신자영·원미진(2010), 한국어 학습자 사전 표제어 선정에 관한 연구, <한국사전학> 16, 한국사전학회, 7-29.
- 강현화(2014a), 한국어교육용 초급 어휘 선정 연구, <문법교육> 21, 한국문법교육학회, 1-26.
- 강현화(2014b), 한국어교육용 중급 어휘 선정에 대한 연구, <외국어로서의 한국어교육> 40, 연세대학교 언어연구교육원 한국어학당, 1-48.
- 강현화(2015), 한국어교육용 고급 어휘 선정에 대한 연구, <외국어로서의 한국어교육> 42, 연세대학교 언어연구교육원 한국어학당, 1-28.
- 고길수, 이성현(2019), 프랑스어 기본어휘의 적절성 검증에 관한 연구 - 프랑스어 교육 및 평가를 위한 프랑스어 기본어휘 DB 구축을 위한 기초 연구, <언어학> 85, 사단법인 한국언어학회, 29-56.
- 고석주·남윤진·서상규(1998), 한국어교육을 위한 기초어휘 의미빈도사전의 개발, <언어정보개발연구> 창간호, 연세대 언어정보연구원.
- 국립국어연구원(2000), 교과서 어휘의 조사단위 연구: 초등학교 국어 교과서를 중심으로: 국립국어연구원.
- 김경철·채미영·홍정선(2003), 유아용 언어 수준 평가 준거 추출을 위한 기초 연구 -이야기 꾸미기 활동 평가를 중심으로-, 열린유아교육연구 8(1), 한국열린유아교육학회, 203-223.
- 김광해(2003), 등급별 국어 교육용 어휘, 박이정.
- 김선혜(2013), <연세 현대한국어사전>의 전문용어 처리의 현황과 과제, <한국사전학회 학술대회 발표논문집> 2013(8), 한국사전학회, 83-96.
- 김창원(2012), 고등학교 어휘교육의 위상과 어휘교육론의 과제, <國語教育學研究> 44, 229-259.
- 김한샘(2003), 한국 현대 소설의 어휘 조사 연구, 국립국어원.
- 김한샘(2005), 현대 국어 사용 빈도 조사, 국립국어원.
- 김한샘, 서상규(2020), “말뭉치의 구축과 활용: 최초의 한국어 말뭉치 ‘연세 말뭉치 I’의 구성과 실제”, 『한국어 말뭉치 구축과 활용의 실제』, 한국문화사.
- 봉미경(2015), 초등학생용 국어사전 개정의 방향과 실제: 『연세초등국어사전』을 중심으로, 한국사전학 25, 한국사전학회.
- 서상규(1998), 한국어 교육을 위한 기초 어휘 선정 보고서, 국립국어원.
- 서상규 편(2003), 「한국어 교육과 학습 사전」, 한국문화사.
- 서상규(2000), 한국어교육 기초 어휘 의미 빈도 사전 개발, 국립국어원.

- 서상규(2001), “말뭉치의 주석과 한국어 기본 어휘 의미 빈도 사전”, [계량언어학 1집], 도서출판 박이정, 57-104
- 서상규(2002), 기본 어휘와 말뭉치 분석, 21세기 한국어교육학의 현황과 과제, 한국문화사.
- 서상규(2003), 한국어 교육용 기본 어휘와 학습사전, 조선어연구 200회기념 제술대표자료집.
- 서상규(2005), “Informatization and Use of Korean Language Data”, The Review of Korean Studies Vol8 No4., 한국학중앙연구원. 2005.12, 71-103.
- 서상규(2006), 한국어 학습 사전 편찬과 기본 어휘의 선정을 위한 기초연구, <朝鮮語研究> 3, 朝鮮語研究会 編, くろしお出版.
- 서상규(2006), “의미 빈도 사전과 어휘 연구”, 『朝鮮學報』 198집, 텐리(일본):조선학회, 33-61.
- 徐尙揆(2006). “韓國語 學習 辭典 編纂과 基本 語彙의 選定을 위한 基礎研究”, 『朝鮮語研究』3, 朝鮮語研究会 編, 동경:くろしお出版
- 서상규(2009), <교육용 기본 어휘 선정을 위한 기초 연구>, 국립국어원.
- 서상규(2017), 한국어 빈도 조사와 말뭉치 주석, <한글> 316, 한글학회, 71-120.
- 서상규·강현화·유현경·김홍범(2001), 「외국인을 위한 한국어 학습 사전 개발(1차년도) 최종 보고서」, 한국어세계화재단·문화관광부
- 서상규·김진웅(2016), “The Development of a Dictionary of the Usage Frequency of Basic Korean Vocabulary and Its Application to Korean Lexicology”, 『언어사실과 관점』40집, 연세대학교 언어정보연구원, 1-30
- 서상규·김진웅·김한샘(2001), “Yonsei Sense Frequency Dictionary based on sense-tagged corpus”, [사전편찬학 연구] 제11집 2호, 19-38
- 서상규·남윤진·진기호(1998), 외국어로서의 한국어 교육을 위한 기초 어휘 선정-기초 어휘 빈도 조사 결과, 문화관광부·한국어 세계화 추진 위원회.
- 서지영(2007), 말뭉치를 활용한 중학교 국어과 학습사전 편찬을 위한 기초 연구, 경기대 교육대학원 석사학위논문.
- 소강춘·이래호·주경미(2010), 개방형 한국어 지식 대사전 신어 분과의 표제어 선정과 그 실제, <한국사전학> 20, 한국사전학회, 52-85.
- 신동광(2018), 영어 기본 어휘 선정 기준의 고찰, <학습자중심교과교육연구> 18(2), 학습자중심교과교육학회, 381-401.
- 안의정(2005), 기존 국어 사전에서의 구어 어휘의 기술, 한국사전학 5, 한국사전학회, 139-159.
- 안의정(2016), '연세 현대한국어사전'의 뜻풀이문의 계량적 분석 연구, 한국사전학,(28), 한국사전학회), 112-131.
- 윤경선·이유미(2014), 유아 한글 교육용 어휘 목록 선정을 위한 연구, <어문론집> 59,

- 중앙어문학회, 65-84.
- 윤미선·차재은·오재혁(2014), 사용 화자 수를 중심으로 본 학령전기 아동의 고빈도 용언, <언어치료연구> 23(4), 한국언어치료학회, 73-88.
- 윤혜경(2016), 한국어 교육용 구어 어휘 선정 연구-초급 수준의 일반명사를 중심으로, <인문과학연구> 50, 강원대학교 인문과학연구소, 65-88.
- 이경남, 박혜림, 이경화(2018), “한글해득을 위한 기초 어휘 선정 연구”, 『청람어문교육』(청람어문교육학회) 65권, 213-235.
- 이삼형, 김정선(2020), “기초 어휘 선정 및 등급화를 위한 동형어 처리 방안 연구”, 『한어문교육』(한국언어문학교육학회) 41권, 103-105.
- 이승재(2020), “인공지능 시대와 우리말 말뭉치”, 『574돌 한글날 기념 전국 국어학 학술대회 발표자료집』(한글학회), 55-60.
- 이문복·신동광(2015), 2015 영어과 교육과정 기본 어휘 목록 개발, <영어교과교육> 14(4), 한국영어교과교육학회, 115-134.
- 이익환·서상규(2002), 기본 어휘 선정 및 사용 실태 조사를 위한 기초 연구, 국립국어원.
- 임지룡(1991), 국어의 기초어휘에 대한 연구, <국어교육연구> 23, 국어교육학회, 87-132.
- 임철성(2002), 초급 한국어교육용 어휘 선정 연구, <국어교육학연구> 14, 국어교육학회, 355-385.
- 장경희·이삼형·이필영·김명희·김태경·김정선·전은진(2012), 초·중·고등학생의 구어 어휘 조사, 지식과 교양.
- 장현진·전희숙·신명선·김효정(2013), 영유아의 기초 어휘 선정 연구, <언어치료연구> 22(3), 한국언어치료학회, 169-187.
- 장현진·전희숙(2014), 초등학생 교육용 기초 어휘 선정 연구: 저학년 중심으로, <언어치료연구> 23(1), 한국언어치료학회, 157-170.
- 장현진(2020), “학령기 언어장애 아동 중재를 위한 기초 어휘 선정 연구”, 『언어치료연구』(한국언어치료학회) 29권 3호, , 69-79.
- 조남호(2003), 한국어 학습용 어휘 선정 결과 보고서, 국립국어원.
- 최정도(2011), “말뭉치를 이용한 사전 편찬에서의 몇 문제에 대하여”, 『언어사실과 관점』(연세대학교 언어정보연구원) 27권, 237-276.
- 최형용(2019), “기초 어휘 선정을 위한 형태소 분석기의 오류 유형에 대하여”, 『언어와 정보사회』(서강대학교 언어정보연구소) 37권, 219-257.
- 한국어능력 평가위원회(1997), 한국어 능력 평가용 기본 어휘표.
- 황은하(2021), “대조분석을 위한 말뭉치의 타당성 연구: 한중 대조분석을 중심으로”, 『이중언어학』(이중언어학회) 82호, 259-286.
- Sun-Hee Lee, Seok Bae Jang, Sang Kyu Seo(2016), *A Frequency Dictionary of*

Korean, Routledge, UK

Swadesh(1951), Basic Vocabulary of Glottochronology. Denver.

2) 사전

국립국어원, 『표준국어대사전(웹사전)』, <https://stdict.korean.go.kr>..

서상규(2014), 『한국어 기본어휘 의미빈도 사전』, 한국문화사.

서상규(2015), 『한국어 구어 빈도 사전 1 (잡기순)』, 한국문화사.

서상규·백봉자·강현화·김홍범·남길임·유현경·정희정·한송화(2006), 『(외국인을 위한) 한국어 학습 사전』, 신원프라임.

연세대학교 언어정보개발연구원(1998), 『연세 한국어 사전』, 두산동아.

연세대학교 언어정보연구원(2018), 『(동아) 연세 초등국어사전(제3판)』, 동아출판.

연세대학교 언어정보연구원(2020), 『(동아) 연세 초등국어사전(제4판)』, 동아출판.

토박이 사전 편찬실(2020), 『보리 국어사전(제3판)』, 보리.

3. 선정된 1등급 어휘 목록

가 ¹	가수 ¹¹	간판 ²	갑작스럽다	거뜰히
가게	가스	간편하다	값	거리 ¹
가격 ³	가슴	간호	갓 ⁵	거리 ²
가구 ⁴	가슴속	간호사	강 ¹	거리 ⁷
가까이	가시 ¹	간호하다	강가 ¹	거미 ²
가깝다	가시다	간하다	강도 ⁶	거미줄
가꾸다	가없다	갈다 ¹	강력 ²	거북 ²
가끔	가운데	갈다 ²	강력하다	거북이
가난 ¹	가위 ¹	갈라지다	강렬하다 ²	거세다
가난하다	가을 ¹	갈매기	강물	거슬리다
가늘다	가장 ¹	갈비 ¹	강산 ¹	거실 ²
가능	가정 ⁶	갈색	강아지	거울 ¹
가능하다	가져가다	갈수록	강제 ¹	거위 ¹
가다 ¹	가져다주다	갈아입다	강하다 ¹	거의 ¹
가두다	가져오다	갈아입히다	강하다 ²	거인 ¹
가득 ¹	가죽	갈아타다	갯가지	거절 ¹
가득하다 ¹	가죽	감 ¹	갯다 ¹	거지 ¹
가라앉다	가지 ¹	감기 ³	같다	거짓
가라앉히다	가지 ⁴	감기다 ¹	같이	거짓말
가래 ⁶	가지다	감나무	같이하다	거짓말쟁이
가렵다	가짜	감다 ¹	값다	거짓말하다
가로 ¹	가축 ²	감다 ²	개 ³	거칠다
가로등	가치 ⁴	감다 ³	개 ⁷	거품
가로막다	각 ¹	감독	개구리	걱정
가루 ¹	각 ²	감동 ²	개나리 ¹	걱정되다
가르다	각각 ¹	감동적	개다 ¹	걱정스럽다
가르치다 ¹	각자 ²	감동하다 ¹	개다 ³	걱정하다
가르침	간 ¹	감사 ⁸	개미 ³	건강 ³
가리다 ²	간 ⁸	감사하다 ⁵	개수 ³	건강하다 ²
가리다 ³	간격 ²	감상하다 ⁴	개운하다 ¹	건너 ¹
가리키다 ¹	간단하다 ²	감시 ²	개울	건너가다
가만	간단히	감싸다	개인 ²	건너다
가만있다	간섭	감옥 ²	개학	건너오다
가만히	간섭하다	감자 ¹	가우똥하다	건너편
가면 ²	간식 ²	감정 ⁶	개	건네다
가름	간장 ¹	감쪽같이	거 ¹	건네주다
가발 ²	간절하다	감추다	거기 ¹	건드리다
가방 ¹	간지럽다	갑갑하다	거꾸로	건물 ³
가볍다	간직하다	갑자기	거대하다	건전지

건조하다 ²	결정 ¹	고래 ¹	곧장	과정 ³
건지다	결정하다 ¹	고루 ¹	곧 ⁸	과학
걸기	걸코	고르다 ¹	곧고루	과학자
걷다 ²	결혼	고르다 ³	곧다	관계 ⁵
걷다 ³	결혼식	고리 ¹	곧목 ¹	관광객
걷다 ⁴	결혼식장	고마움	곧목길	관람
걸다 ²	결혼하다	고마워하다	곧짜기	관심 ¹
걸레	걸치다	고맙다	곧프	관찰 ¹
걸리다 ¹	경고 ⁴	고모 ¹	곰 ³	관찰하다
걸어가다	경기 ¹¹	고모부	곰곰이	관하다 ²
걸어오다 ¹	경기장	고무 ¹	곰팡이	광고 ²
걸음	경례	고무줄	곰다 ²	괜찮다
걸음마	경비 ⁴	고민	곳	괜히
걸치다	경우 ²	고민하다	공 ¹	괴로움
검다 ²	경쟁	고백	공 ⁹	괴로워하다
검사 ³	경주 ⁸	고백하다	공 ¹²	괴롭다
검사하다	경찰 ⁴	고생	공간 ⁵	괴롭히다
검은색	경찰관	고생하다	공감하다	괴물
검정 ¹	경찰서	고소하다 ¹	공격 ²	괴상하다 ²
검정색 ^{미등재}	경찰차	고속버스	공격하다	굉장하다
겹 ²	경치 ²	고슴도치	공기 ⁵	굉장히
겹나다	결	고약하다	공놀이	교과서
겹내다	계곡 ¹	고양이	공동 ²	교도소
것	계단 ⁴	고장 ¹	공룡	교류 ¹
겉 ¹	계란	고장 ⁹	공부 ¹	교사 ⁹
겉모습	계란프라이 ^{미등재}	고정 ⁵	공부방	교실
게 ¹	계산 ¹	고정되다	공부하다	교육하다
게으르다	계산하다	고정하다 ⁵	공사 ²	교체하다
게으름	계속 ⁴	고집 ²	공사장	교통사고
게임	계속되다 ²	고집하다	공원 ³	교환하다 ¹
겨드랑이	계속하다 ³	고추 ¹	공장 ²	교회 ²
겨루다	계시다	고추장	공주 ¹	구 ¹
겨우	계절 ¹	고춧가루	공중 ³	구경 ¹
겨울	계획 ¹	고치다	공짜	구경하다
겨울철	계획하다 ¹	고통스럽다	공책 ¹	구급차
겪다	고 ³	고프다 ¹	공통점	구덩이
견디다	고개 ¹	고함 ¹	공포 ⁸	구두 ¹
결과 ²	고개 ²	고향 ²	공항 ²	구렁이
결국	고구마	곡 ²	공휴일	구르다 ¹
결석 ¹	고기 ¹	곧 ¹	곶감	구르다 ²
결석하다	고등어	곧바로	과거 ²	구름
결심 ¹	고등학교	곧잘	파일 ¹	구멍
결심하다 ¹	고등학생		과자 ²	구박 ³

구별 ²	굽다	그러다	글썩	기린 ²
구별하다	굽다 ¹	그러면	글썩요	기본
구부리다	굽다 ²	그러므로	글씨	기본 ¹
구분 ⁶	굽히다	그러하다	글자	기빠하다
구분하다 ³	궁 ⁴	그런 ¹	굽다	기쁘다
구석 ¹	궁궐	그런데	굽히다	기쁨
구석구석	궁금증	그럼 ¹	금 ²	기사 ²
구슬 ¹	궁금하다	그럼 ²	금 ³	기술 ¹
구십	궁전 ⁵	그렇게	금 ⁴	기술자
구역 ⁴	권 ¹	그렇다	금메달	기억 ²
구역질	권총 ¹	그렇지만	금방 ¹	기억나다
구월	권투 ¹	그루 ¹	금붕어	기억되다
구이 ¹	귀 ¹	그릇 ¹	금세 ¹	기억하다 ²
구출하다 ¹	귀걸이	그리 ¹	금연 ¹	기온
구하다 ¹	귀신 ¹	그리 ²	금요일	기운 ¹
구하다 ³	귀엽다	그리고	금지 ⁴	기울다
국 ¹	귀중하다 ¹	그리다 ²	금 ⁴	기울어지다
국군 ²	귀찮다	그리움	급식	기울이다
국기 ⁷	귀하다	그리워하다	급하다	기저귀
국물	긋속말	그리하여	급히	기적 ³
국민	규칙 ²	그림 ¹	긋다 ¹	기절하다 ²
국밥	규칙적	그림자	기계 ⁶	기지 ⁸
국산	굴	그림책	기관 ¹¹	기지개
국수 ¹	그 ¹	그림다	기념 ²	기차 ¹
국어 ¹	그거	그만 ²	기념일	기차역
국화 ¹	그것	그만두다	기념품	기침 ¹
국화 ⁵	그곳	그만큼	기념하다	기타 ²
군 ⁴	그날	그만하다 ¹	기다 ¹	기특하다
군 ⁵	그냥	그물 ¹	기다랗다	기회 ³
군대 ²	그네 ²	그분	기다리다	긴장
군데	그녀	그사이	기대 ²	긴장되다
군인	그놈	그저	기대감	긴장하다
굳다	그늘	그저께	기대다 ¹	길 ¹
굳어지다	그다음	그전	기대되다	길가
굳히다	그대로	그제	기대하다	길거리
굴 ¹	그동안	그제야	기도 ³	길다 ¹
굴 ³	그때	그중	기도하다 ¹	길다 ²
굴다 ¹	그때그때	그쪽	기도하다 ²	길들이다
굴뚝	그래 ¹	그치다	기둥	길목 ¹
굴러가다	그래도	극장	기러기	길바닥
굴러다니다	그래서	근데	기록 ²	길이 ¹
굴리다	그러나	근처	기르다	길쭉하다
굽다	그러니까	글	기름 ¹	김 ¹

김 ³	깨어지다	깡 ²	끝 ¹	낙서 ³
김 ⁴	깨우다	꽤	끝나다	낙지 ¹
김밥	깨지다	꽤 ¹	끝내	낙타 ²
김장	깨트리다	꾸다 ¹	끝내다	낙다
김치 ¹	깻잎	꾸리다	끝마치다	낙시
김치찌개	꺼내다	꾸미다	끝없이	난로 ¹
김칫국	꺼지다 ¹	꾸역꾸역	끼 ¹	난방 ²
깃발 ²	꺼지다 ²	꾸준하다	끼다 ¹	난쟁이
깃털	꺾다	꾸준히	끼다 ²	날 ¹
깊다	꺾이다	꾸짖다	끼다 ³	날 ²
깊숙이	꺾	꼭	끼어들다	날개 ¹
깊이 ¹	꺾데기	꼭꼭 ¹	끼었다	날다 ¹
깊이 ²	꺾질	꿀	끼우다	날뛰다
까궁	껴안다	꿀꺽	끼이다 ¹	날리다 ¹
까다 ¹	꼬다	꿀다	끼치다 ¹	날리다 ²
까다롭다	꼬리	꿈	나 ³	날씨 ¹
까닭	꼬마	꿈꾸다	나가다	날씬하다
까마귀	꼬박꼬박 ²	꿈속	나날이	날아가다
까맣다	꼬이다 ²	꿈틀거리다	나누다	날아다니다
까먹다	꼬이다 ³	뽕 ¹	나뉘다	날아오다
까불다 ¹	꼬집다	웨다	나다 ¹	날짜 ¹
까지다 ¹	꼬투리	웨매다	나들이 ¹	날카롭다
까치 ¹	꼭 ²	뛰다 ³	나라 ¹	낱다
깎다	꼭 ³	끄다 ¹	나란히	낱 ¹
깎이다 ¹	꼭꼭 ¹	끄덕이다	나르다 ¹	낱 ²
갈끔하다	꼭대기	끄집어내다	나머지	낱 ⁴
갈다	꿀 ¹	끈	나무 ¹	낱기다
갈리다 ¹	꿀찌	끈기	나무라다	낱녀
감감하다	꼼꼼하다	끈끈하다	나물	낱다
감박	꼼꼼히	끊기다	나뭇가지	낱동생
감빡	꼼작	끊다	나뭇잎	낱매
감빡하다	꼼작없이	끊어지다	나비 ³	낱자
감작 ²	콧하다	끝다	나빠지다	낱자아이
깡통	공공 ¹	끌려가다	나쁘다 ¹	낱쪽
깨	꽃다	끌리다	나서다	낱편 ¹
깨끗이	꽃하다	끌어내다	나아가다	낱학생
깨끗하다	꽃 ¹	끌어당기다	나아지다	낱다 ¹
깨다 ¹	꽃게	끌어안다	나오다	낱다 ²
깨다 ²	꽃다발	끌어올리다	나이 ¹	낱비하다
깨닫다	꽃밭	끊다	나중 ¹	낱
깨뜨리다	꽃잎	끊이다	나타나다	낱다
깨물다	꽃집	끔찍하다	나타내다	낱잠
깨어나다	꽉	굉굉	나홀	낱추다

낫설다	넉히다	눔 ¹	눈앞	다시 ¹
낯말 ²	넘겨주다	농구 ⁵	눈치	다시마
낱다 ¹	넘기다	농구공	눌리다 ¹	다양하다
내 ⁴	넘다	농구장	눅다 ¹	다음 ¹
내 ⁷	넘어가다	농담 ¹	눅히다	다음날
내기 ¹	넘어뜨리다	농부 ¹	뉴스	다음번
내년	넘어오다	농사 ¹	느끼다 ²	다이아몬드
내놓다	넘어지다	농사짓다	느낌	다이어트
내다 ²	넘치다	농장 ⁵	느리다	다정하다
내다보다	넋다	농촌	늑대	다지다 ²
내던지다	네 ¹	높다	늘	다짐
내려가다	네 ²	높아지다	늘다	다짐하다
내려놓다	네 ³	높이 ¹	늘리다	다치다 ¹
내려다보다	네모	높이 ²	늘어나다	다투다
내려앉다	넥타이	높이다	늘어놓다	다툼
내려오다	넷	농다	늘어서다	다행
내리다 ¹	넷째	놓아두다	늘어지다	다행히
내리막길	녀석	놓아주다	늘이다 ¹	뒹다
내밀다	년 ²	놓이다	늪다	뒹이다 ¹
내버리다	노란색	놓치다	늪은이	단 ⁷
내보내다	노랑 ¹	놉두다	늪다	단계 ²
내뿜다	노랑다	누구	늪잠	단단하다
내쉬다	노래 ¹	누구누구	님 ¹	단단히
내용 ²	노래방	누나 ¹	다 ³	단둘
내일	노래하다	누님	다가가다	단맛
내주다	노려보다	누다	다가서다	단숨에
내쫓다	노력 ¹	누렁다	다가오다	단정하다 ¹
냄비	노력하다 ¹	누르다 ¹	다녀가다	단지 ⁷
냄새	노인 ¹	눈 ¹	다녀오다	단체 ²
냉면	녹다 ¹	눈 ⁴	다니다	단추 ¹
냉장고	녹색	눈곱	다다르다	단풍 ¹
너 ¹	녹음 ³	눈동자	다듬다	달다 ²
너그럽다	녹음하다	눈물 ¹	다람쥐	단히다
너무	녹이다	눈병	다르다 ¹	달 ⁵
너무나	논 ¹	눈보라	다른	달걀
너무너무	놀다 ¹	눈부시다	다리 ¹	달다 ³
너무하다	놀라다	눈빛 ¹	다리 ²	달다 ⁵
너희	놀랍다	눈사람	다리다 ¹	달다 ⁷
넉넉하다	놀래다	눈싸움 ¹	다리미	달라붙다
넉넉히	놀리다 ¹	눈싸움 ²	다만 ¹	달라지다
널다 ¹	놀리다 ³	눈썰매	다물다	달래다 ¹
널리다 ¹	놀이 ¹	눈썹	다섯	달려가다
넉다	놀이터	눈알	다스리다	달려들다

달려오다	대낮	덕 ⁴	도와드리다 ^{미등재}	동굴
달력	대다 ¹	덕분	도와주다	동그라미
달리기	대단하다	던지다	도움	동그랗다
달리다 ¹	대단히	털	도자기	동네
달리다 ⁴	대답	털다	도장 ¹⁶	동동 ²
달빛	대답하다	털털 ¹	도저히	동물
달아나다	대들다	털컹 ¹	도전 ²	동물원
달콤하다	대로 ¹	털하다	도전하다 ¹	동산 ¹
닭	대머리 ¹	덤벼들다	도착 ¹	동상 ⁷
닭고기	대문 ³	덤비다	도착하다 ¹	동생 ¹
닭다	대변 ¹	덥다	도토리	동시 ²
닭다	대부분	덥석	독 ³	동안 ¹
담 ¹	대신 ³	덧붙이다	독감 ¹	동요 ²
담그다	대신하다	덩어리	독서 ³	동작 ³
담다	대왕	덩치	독서하다 ²	동전 ⁵
담배	대장 ⁴	덮개	독수리	동쪽
담요	대접하다	덮다	독하다	동치미
담임	대청소	덮이다	돈 ¹	동화 ⁶
담장 ¹	대체 ¹	덮치다	돈가스	동화책
담 ²	대체로	데 ¹	돈다 ¹	돼지
담답하다	대추 ¹	데다 ¹	돈보기	돼지고기
담변하다	대충 ¹	데려가다	돌아나다	되게
담장	대통령	데려오다	돌 ¹	되다 ¹
담하다	대포 ²	데리다	돌 ²	되돌다
당근	대표	데우다	돌고래 ³	되돌리다
당기다 ¹	대피하다	도 ⁴	돌다	되돌아가다
당당히	대하다 ²	도구 ⁸	돌돌 ¹	되돌아오다
당분간	대학교	도깨비	돌려보내다	되묻다 ³
당시 ²	대학생	도끼 ¹	돌려주다	되찾다
당신 ²	대화 ⁶	도대체	돌리다 ⁴	되풀이하다 ¹
당연하다 ¹	대화하다 ²	도둑	돌맹이	된장
당연히 ¹	대회 ²	도둑맞다	돌보다	된장국
당장 ²	택 ¹	도둑질	돌아가다	된장찌개
당하다 ¹	더 ¹	도둑질하다	돌아다니다	두 ¹
당황스럽다	더듬다	도로 ¹	돌아보다	두근거리다
당황하다	더러워지다	도로 ⁷	돌아서다	두껍다
닿다 ¹	더럽다	도마 ¹	돌아오다	두다
대 ¹⁶	더럽히다	도망 ²	돌이키다	두드리다
대강 ²	더욱	도망가다	둑다	두들기다
대개 ³	더욱더	도망치다	돛자리	두려움
대결 ²	더위	도서관	둥 ⁹	두려워하다
대결하다 ²	더하기	도시 ³	둥 ¹⁴	두렵다
대나무	더하다	도시락	둥갑 ¹	두르다

두부 ¹	들다 ⁴	따라잡다	떨어지다	뜯어보다
두세	들려오다	따로	떼 ¹	뜯 ¹
두툼하다	들려주다	따로따로 ¹	떼 ²	뜯
둔하다	들르다	따르다 ¹	떼다 ¹	띄우다 ¹
둘 ¹	들리다 ³	따르다 ²	또	띄우다 ²
둘러보다	들리다 ⁵	따스하다	또는	띠 ¹
둘러싸다	들어가다 ¹	따지다	또다시	띠다 ¹
둘러싸이다	들어서다	딱 ²	또렷하다	라디오
둘레	들어오다	딱 ³	또박또박 ¹	라면 ¹
둘째	들어주다	딱딱 ¹	또한	라이터
둥글다	들여놓다	딱딱하다	똑 ³	로봇
둥지	들여다보다	딱지 ²	똑같다	로켓 ²
뒤 ¹	들여보내다	판 ³	똑같이	리 ⁶
뒤덮다	들이다	판판하다	똑똑 ¹	리모컨
뒤덮이다	들이대다 ²	딸 ¹	똑똑하다	리본
뒤바뀌다	들키다	딸기	똑바로	마구 ¹
뒤적거리다	들판 ¹	땀 ¹	똥	마냥 ²
뒤지다 ²	듬뿍	땅 ¹	똥깅	마녀
뒤집다	دت ¹	땅바닥	똥벽똥벽	마늘
뒤집어쓰다	دت이 ¹	땅속	똥똥 ¹	마당
뒤집히다	دت하다	땅콩	똥다	마렵다
뒤쪽 ¹	등 ¹	때 ¹	똥리다	마루 ³
뒤처지다	등 ³	때 ²	똥어지다	마르다 ¹
뒤통수	등 ⁴	때때로	똥똥하다	마리 ¹
뒤편	등 ⁷	때리다 ¹	똥놀다 ²	마무리
뒹면	등교	때문	똥다 ¹	마무리하다
뒹모습	등교하다	땡 ²	똥다 ²	마법
뒹문	등불	떠나가다	뛰어가다	마술 ²
뒹부분	등산	떠나다	뛰어나다	마스크
뒹산	등산하다	떠내려가다	뛰어나오다	마시다
뒹자리	등원 ¹	떠다니다	뛰어내리다	마을 ¹
뒹줄	등원하다	떠돌다	뛰어넘다	마음
뒹굴다	등장 ¹	떠돌아다니다	뛰어놀다 ²	마음껏
드디어	등장하다 ¹	떠들다 ¹	뛰어다니다	마음대로
드라마	디디다	떠들썩하다 ²	뛰어들다	마음먹다
드러내다	따갑다	떠오르다	뛰어오다	마음속
드러눕다	따끔하다	떠올리다	뜨거워지다	마음씨
드레스	따다 ¹	떡 ¹	뜨겁다	마이크
드리다 ¹	따듯하다	떡볶이	뜨다 ¹	마주 ¹
든든하다	따뜻하다	떨구다	뜨다 ⁴	마주치다
들다 ¹	따라가다	떨다 ¹	뜨다 ⁵	마중
들 ¹	따라다니다	떨리다 ¹	뜯기다 ¹	마지막
들다 ¹	따라오다	떨어뜨리다	뜯다	마차

마찬가지	말투	맨몸	메모하다	목욕탕
마치 ³	말하다	맨발	메시지	목욕하다
마치다 ²	맑다	맨손	메아리	목적지
마침 ²	맘	맵다	메우다 ²	몰다
마침내	맘대로	맹세하다	머칠	몰래
마흔	맛 ¹	맏다	면 ⁵	몰려들다
막 ¹	맛나다	머리 ¹	멸치	몰려오다
막 ²	맛보다	머리띠	명 ³	몸 ¹
막내	맛없다	머리카락	명령 ¹	몸무게
막다	맛있다	머릿속	명령하다	몸부림
막대기	망가뜨리다	머물다	몇	몸살
막히다	망가지다	머뭇거리다	몇몇	몸집
만 ⁵	망설이다	먹구름	모금 ¹	몸짓
만나다	망원경	먹다 ¹	모기 ¹	몸통
만두 ¹	망치 ¹	먹다 ²	모니터	몹시
만들다	망치다	먹이	모두 ¹	못 ¹
만만하다 ¹	망하다	먹이다	모든	못 ⁴
만세 ⁴	맞다 ¹	먹히다	모래	못되다
만약	맞다 ²	먼저	모래성	못살다
만일 ¹	맞다 ³	먼지	모래	못생기다
만족 ¹	맞대다	멀다 ²	모르다	못하다
만족하다	맞은편	멀리	모습 ¹	몽둥이
만지다	맞이하다	멀리하다	모시다	몽땅 ¹
만큼	맞추다	멀미	모양 ²	무 ²
만하다	맞히다 ¹	멀쩡하다	모여들다	무겁다
만화 ¹⁰	맞히다 ³	멈추다	모으다	무게
만화책	말기다	멋 ¹	모이다 ¹	무관심하다
많다	말다 ¹	멋대로	모임	무궁화
많아지다	말다 ²	멋있다	모자 ⁷	무기 ³
많이	매 ¹	멋지다	모자라다	무너뜨리다
말 ¹	매끄럽다	멍 ¹	모조리 ¹	무너지다
말 ⁴	매니큐어	멍들다	모통이	무늬
말다 ¹	매다 ¹	멍멍이	모험	무당벌레
말다 ²	매달다	멍청이	목 ¹	무대 ⁵
말다 ³	매달리다	멍청하다	목 ⁶	무더위
말랑말랑하다	매미 ²	멍하니	목걸이 ¹	무덤
말리다 ²	매우 ¹	멍하다	목구멍	무덤다
말리다 ³	매일	메뉴	목도리	무력무력
말소리	매일매일	메다 ²	목마르다	무렵
말썽	매주 ¹	메달	목소리	무료 ¹
말씀	매트 ¹	메뚜기 ²	목숨	무릎
말씀하다	맨 ¹	메롱	목요일	무리 ¹
말없이	맨날	메모	목욕	무리하다

무사하다 ³	물기	미역국	바람개비 ¹	반복하다 ¹
무사히 ²	물놀이 ²	미용사	바로 ²	반성 ¹
무서워하다	물다 ²	미용실	바르다 ¹	반성하다 ¹
무섭다	물들다	미움	바르다 ³	반장 ⁸
무술 ⁴	물들이다	미워하다	바보	반주 ⁴
무슨	물러가다	미지근하다	바쁘다	반죽
무시 ³	물러나다	미치다 ¹	바빠	반지 ²
무시무시하다	물려서다	민족	바싹 ¹	반짍 ¹
무시하다	물려받다	민다	바위 ¹	반짝거리다
무엇	물려주다	밀가루	바이러스	반짝반짝 ¹
무용 ³	물론 ¹	밀다 ¹	바이올린	반짝이다
무작정	물리다 ²	밀려나다	바지 ¹	반쪽
무조건	물리치다	밀려들다	바짝	반찬
무지 ³	물방울	밀려오다	바치다 ¹	반칙
무지개	물속	밀리다 ¹	바퀴 ¹	반하다 ¹
무척 ¹	물안경	밀리다 ²	바퀴벌레	받다 ¹
무치다	물어뜯다	밀어내다	박 ⁵	받치다 ²
묵 ¹	물어보다	밀어붙이다	박다 ¹	받침
묵다	물음 ¹	밀어주다	박물관	발 ¹
묵음	물음표	밀치다	박사 ¹	발가락
묵이다	물총	밟다	박살 ¹	발걸음
문 ⁴	물컵	밀	박수 ²	발견 ¹
문득 ¹	몽치다	밀바닥	박스	발견하다 ¹
문밖	뒹	바가지 ¹	박자	발끝
문어 ¹	뿔	바구니	박쥐	발달되다
문제 ⁶	미국 ³	바깥	박히다	발뒤꿈치
문지르다	미꾸라지	바깥쪽	밖	발등
문짝	미끄러지다	바꾸다	반 ⁵	발레
문턱 ²	미끄럼틀	바뀌다	반 ⁸	발명 ¹
문다 ¹	미끄럽다	바나나	반가워하다	발명가
문다 ²	미끼	바느질	반갑다	발명하다 ¹
문다 ³	미래 ²	바느질하다	반기다	발목 ¹
문어나다	미로 ¹	바늘	반년	발바닥
문히다 ¹	미루다	바다	반대 ²	발버둥
문히다 ²	미리 ¹	바닥	반대쪽	발사 ¹
물 ¹	미리미리	바닥나다	반대편	발사하다
물가 ¹	미소 ⁵	바닷가	반대하다 ¹	발음 ¹
물감 ²	미술	바닷물	만드시	발자국
물개	미술관	바둑	반듯하다	발전되다
물거품	미아 ¹	바라다	반말	발짝 ¹
물건	미안 ¹	바라보다	반바지	발톱
물결	미안하다	바람 ¹	반반 ⁴	발판 ¹
물고기	미역 ²	바람 ²	반복 ¹	발표 ¹

발표하다	방향 ¹	번쩍 ¹	별 ¹	복장 ⁴
발표회	발 ¹	번쩍번쩍 ¹	별것	복다
밝다	배 ¹	번호 ²	별로 ¹	본 ¹
밝혀내다	배 ²	벌 ²	별명 ¹	볼 ¹
밝히다	배 ³	벌 ³	별일	볼 ⁴
밟다	배 ⁸	벌 ⁶	병 ⁴	볼일
밟히다	배고프다	벌금	병 ⁵	볼펜
밤 ¹	배고픔	벌다 ²	병들다	봄
밤 ²	배구 ⁵	벌떡	병실	봄날
밤길	배꼽	벌레	병아리	봄바람
밤낮	배낭 ¹	벌리다 ¹	병원 ²	봉지 ⁶
밤늦다	배달 ²	벌벌 ¹	보건소	봉투
밤새 ²	배달되다	벌써	보관 ¹	봐주다
밤새다	배달원	벌어지다 ¹	보관하다	웁다
밤잠	배달하다	벌어지다 ²	보기 ¹	부끄러움
밤중	배부르다	벌이다	보내다	부끄러워하다
밤하늘	배불리	벌칙	보내오다	부끄럽다
밥 ¹	배우 ¹	범인 ²	보너스	부담스럽다
밥그릇	배우다	범 ¹	보다 ¹	부드럽다
밥맛	배추 ¹	벗겨지다	보다 ²	부들부들 ¹
밥상	배탈 ¹	벗기다 ²	보라 ¹	부딪치다
맛줄	배터리 ¹	벗다	보라색	부딪히다
방 ⁷	백 ⁴	벗어나다	보람	부러워하다
방귀 ¹	백성	베개	보름달	부러지다
방금 ¹	백인 ¹	베다 ²	보리 ¹	부럽다
방망이 ¹	백화점	베란다	보물 ²	부르다 ¹
방문 ²	밴드 ²	베이다	보살피다	부리 ¹
방문 ³	맵	벨 ¹	보석 ³	부리다 ¹
방문하다	맷속	벨트	보이다 ¹	부리다 ²
방바닥	맹맹	벼	보이다 ²	부모 ¹
방법	맬다	벼락	보일러	부모님
방석 ²	머럭	벽 ⁵	보자기 ²	부부 ²
방송 ¹	머룻	벽돌	보태다	부분 ¹
방송국	머리다	벽지 ²	보통	부상 ⁵
방송하다 ¹	머쓱 ²	변경하다	보트	부서지다
방어 ²	머스 ²	변기	보호 ¹	부수다
방울 ¹	머튼	변덕	보호하다	부시다 ²
방울 ²	머티다	변명 ¹	복 ¹²	부엌이
방패 ²	번 ⁴	변명하다 ¹	복도 ⁴	부엌
방학	번갈다	변신 ¹	복수 ³	부인 ¹
방학하다	번개	변신하다 ¹	복숭아	부인 ⁴
방해 ¹	번지다 ¹	변하다	복습하다	부자 ⁶
방해하다	번째	변화하다	복잡하다	부잣집

부족 ¹	불빛	비슷하다 ²	빨개지다	사과 ⁸
부족하다	불쌍하다	비싸다	빨다 ¹	사과나무
부지런하다	불쑥	비웃	빨다 ²	사과하다 ²
부지런히	불안하다	비우다	빨대	사귀다
부채 ¹	불어나다	비웃다	빨래	사나이
부치다 ²	불어넣다	비좁다	빨래하다	사납다
부침개	불어오다	비추다	빨리	사냥
부탁	불쾌하다	비치다	빨리다 ¹	사냥꾼
부탁하다	불타다	비키다	빨리빨리	사냥하다
부풀다	불태우다	비타민	빨아들이다	사다
부풀리다	불편 ¹	비틀거리다	빵 ¹	사다리
부품	불편하다 ¹	비하다	빵집	사라지다
부하 ³	불평 ¹	비행 ³	빼내다	사람
북 ³	불행	비행기	빼놓다	사랑 ¹
북 ⁶	불행하다	빈자리	빼다 ¹	사랑스럽다
북극 ¹	붉다	빈틈없이	빼앗기다	사랑하다
북쪽	붉은색	빌다 ¹	빼앗다	사료 ⁴
분 ¹	붉비다	빌딩	뺏기다	사르르
분 ⁶	붓 ¹	빌려주다	뺏다	사막 ³
분리하다 ³	붓다 ²	빌리다	뺨	사무실
분명 ¹	붕대	빗 ¹	뺨하다 ¹	사물 ¹⁰
분명하다 ¹	붕어 ¹	빗다 ²	뺨하다 ²	사슴 ¹
분명히	불다	빗물	뺨다	사실 ⁴
분수 ¹⁰	불들다	빗방울	뺨 ²	사십
분위기	불이다	빗소리	뺨	사용 ⁴
분필 ³	불잡다	빗자루	뽀뽀	사용법
분하다 ³	불잡히다	빗줄기	뽀내다	사용하다 ³
분홍 ¹	브레이크 ¹	밍밍	뽀다	사월 ²
분홍색	블록 ²	빋다	뽀아내다	사이 ¹
볼 ¹	비 ¹	빋	뽀히다	사이다 ²
불가능	비교 ¹	빋나다	뽀죽하다	사이좋다
불가능하다	비교하다	빋내다	뽀듯하다	사이즈 ¹
불고기	비누	빠뜨리다	뿌리	사인 ¹⁴
불꽃	비눗방울	빠르다	뿌리다	사자 ¹¹
불꽃놀이	비닐	빠져나가다	뿌영다	사장 ¹⁵
불나다	비다 ¹	빠져나오다	뿔 ¹	사전 ²²
불다	비둘기	빠지다 ¹	뿔다	사진 ⁶
불량 ¹	비밀	빠지다 ²	뺨뚫어지다	사진기 ²
불러내다	비바람	빠짐없이	사 ¹⁰	사촌
불러오다	비비다	뺨히	사각형	사탕 ²
불리다 ⁴	비빔밥	빨간색	사건 ¹	사회 ⁷
불리다 ⁷	비상 ¹	빨강	사고 ¹²	사흘
볼만	비슷비슷하다	빨짱다	사과 ⁵	산 ¹

산골 ²	상냥하다	생기다	선택하다	셋째
산골짜기	상대 ⁴	생김새	선풍기	소 ³
산길 ²	상대편	생명	설 ¹	소개 ¹
산더미	상상 ⁷	생선	설거지	소개되다
산불	상상력 ¹	생신 ²	설거지하다	소개하다 ¹
산사태	상상하다 ³	생일 ²	설날	소고기
산산조각	상어 ¹	생일날	설레다	소금 ¹
산소 ¹	상자 ⁸	생활	설마 ¹	소나기
산소 ³	상처 ²	생활하다	설명	소나무
산속	상추 ¹	샤워	설명서	소녀 ²
산책	상쾌하다	샤워하다	설명하다	소년 ¹
산책하다	상태 ¹	삼푸	실사 ¹	소독 ⁴
살 ¹	상품 ³	서늘하다	설탕	소독약
살 ⁴	상하다 ²	서다 ¹	섬 ³	소독하다 ²
살금살금	상황 ²	서두르다	섬섬하다	소란스럽다
살다 ¹	새 ³	서랍	성 ⁵	소름
살리다	새 ⁶	서러움	성격 ²	소리 ¹
살살 ²	새것	서럽다	성공 ¹	소리치다
살아가다	새기다 ¹	서로 ¹	성공하다	소망 ²
살아나다	새까맣다	서로서로	성냥 ¹	소매 ¹
살아남다	새끼 ²	서른	성숙하다	소문 ²
살아오다	새끼손가락	서비스	세 ¹	소문나다
살짝	새다 ¹	서서히	세계 ²	소방관
살찌다	새로	서운하다	세계적	소방서
살피보다	새롭다	서점 ³	세균	소방차
살피다 ¹	새벽 ¹	서쪽	세다 ²	소변 ¹
삶	새싹	서투르다	세다 ³	소스 ¹
삶다	새우 ²	서툰다	세로 ¹	소시지
삼 ⁶	새해	석유 ¹	세면대	소식 ⁴
삼각 ¹	색 ³	섞다	세모 ¹	소용 ⁵
삼각형	색깔	섞이다	세배	소용돌이
삼겹살	색다르다	선글라스	세뱃돈	소원 ⁴
삼다 ²	색연필	선녀 ¹	세상 ¹	소음 ⁶
삼십	색종이	선물 ³	세수 ⁴	소중하다
삼월	색칠	선물하다	세수하다 ²	소중히
삼촌	색칠하다	선반 ¹	세우다	소파 ⁶
삼키다	샌드위치	선생 ¹	세제 ⁴	소풍
삽 ¹	샘 ²	선생님	세차다	소화 ⁶
상 ³	생각 ¹	선선하다	세탁	소화되다 ¹
상 ²³	생각나다	선수 ⁵	세탁기	소화하다 ¹
상가 ⁷	생각되다	선인장 ²	세탁소	속 ¹
상관없다	생각하다	선장 ⁶	센터 ²	속다
상관없이	생겨나다	선택	셋	속도 ¹

속마음	수 ²³	술가락	슬픔	식초
속삭이다	수건	술갈	습관	식탁
속상하다 ²	수고 ¹	술 ¹	승객	식판 ¹
속웃	수고하다 ¹	술잔	승리	식품 ¹
속이다	수다 ¹	숨 ¹	승리하다	식혜
속하다 ²	수도 ⁵	숨기다	승용차	식히다
손 ¹	수도꼭지	숨다	시 ²	신 ¹
손가락	수돗물	숨바꼭질	시 ⁶	신 ²
손길	수레	숨쉬기	시간 ⁴	신 ⁹
손녀	수리 ¹¹	숨차다	시간표	신고 ¹
손님	수리하다 ²	숫자	시계 ²	신고하다 ¹
손대다	수많다	승승 ¹	시골	신기하다 ¹
손들다	수박 ¹	숫불	시금치	신나다
손등 ¹	수백 ⁴	숲	시끄럽다	신다
손목	수백만	쉬다 ³	시내 ¹	신량 ²
손바닥	수비 ⁴	쉬다 ⁴	시내 ³	신문 ¹⁰
손발	수상하다 ⁵	원	시냇물	신문지
손수건	수수께끼	쉽다	시다	신발
손자 ¹	수술 ³	씻	시들다	신발장
손잡다	수십	슈퍼	시범	신부 ⁴
손잡이	수업 ³	슈퍼마켓	시시하다	신부 ¹⁰
손짓	수업하다 ³	숫	시원하다	신비 ²
손톱 ¹	수없이	스님	시월 ¹	신비롭다
술술	수염 ³	스마트폰	시작 ¹	신비하다
술직하다	수영 ²	스무	시작되다 ¹	신선하다 ³
술직히	수영복	스물	시작하다 ¹	신하 ¹
숨	수영장	스스로	시장 ⁴	신호 ¹
숨사탕	수영하다 ¹	스웨터	시집가다	신호등
숨씨	수요일	스위치	시체 ²	신다
숫구치다	수저 ¹	스치다	시커멓다	실 ¹
숫다	수첩 ¹	스케이트	시키다 ¹	실내
숫아나다	수컷	스케치북 ¹	시퍼렇다	실내화 ²
숫아오르다	수확 ²	스키	시합 ¹	실력 ¹
송아지	숙모 ¹	스테이크	시합하다	실망 ²
송이 ¹	숙이다	스티커	시험 ³	실망하다 ²
송충이	숙제 ³	스포츠	식 ²	실수 ¹
송편	숙제하다	스피커	식구 ¹	실수하다
술	순간 ³	슬그머니	식다	실제 ²
쇠 ¹	순간적	슬금슬금	식당	실제로
쇠고기	순서	슬리퍼	식물 ²	실컷
쇼 ¹	순식간	슬쩍	식빵	실패 ²
쇼팽	순진하다 ¹	슬퍼하다	식사 ³	실패하다 ¹
수 ²	순하다 ²	슬프다	식사하다 ²	실험

실험하다	쏟다	아나	아직	알 ¹
싫다	쏟아지다	아노	아침	알다
싫어하다	쭈시다 ¹	아니 ¹	아침밥	알레르기
심각하다 ²	쭈시다 ²	아니 ²	아파트	알록달록하다
심다 ¹	쭈 ³	아니다	아파하다	알리다
심부름	쭈스렵다	아니야	아프다	알림장 ²
심부름하다	쭈쭈	아니요	아픔	알맞다
심심하다 ¹	쓰기	아니하다	아하	알맹이
심장 ²	쓰다 ¹	아들	아홉	알아내다
심판 ¹	쓰다 ²	아래	아흔	알아듣다
심하다	쓰다 ³	아래쪽	악 ²	알아보다
십	쓰다 ⁶	아래층	악기 ⁵	알아주다
십이월	쓰다듬다	아름답다	악당	알아채다
십일월	쓰라리다	아마 ¹	악마 ²	알약
십자가 ¹	쓰러뜨리다	아마도	악몽	암컷
싱겁다	쓰러지다	아무 ¹	악수 ⁶	앗
싱싱하다	쓰레기	아무거	악어 ²	앞
싫다	쓰레기통	아무것	안 ¹	앞길 ¹
싸다 ¹	쓰리다	아무래도	안 ²	앞니
싸다 ²	쓰이다 ¹	아무런	안개	앞두다
싸다 ⁵	쓸다 ²	아무렇다	안경 ³	앞뒤
싸우다	쓸데없다	아무리	안과 ²	앞면
싸움	쓸데없이	아무튼	안기다 ¹	앞바다
싸이다 ¹	쓸모	아물다	안기다 ²	앞부분
씩 ¹	쓸쓸하다	아버님	안내 ¹	앞서다
씩 ²	씩다 ²	아버지	안내하다	앞자리
씩씩 ¹	씩우다 ¹	아빠	안녕	앞장 ¹
씩트다	씨 ¹	아쉬움	안녕하다	앞장서다
쌀	씨 ⁴	아쉬워하다	안녕히	앞줄
쌀밥	씨름	아쉽다	안다	앞집
쌀쌀하다	씨앗	아슬아슬하다	안되다 ²	앞쪽
쌍둥이	씩씩하다 ²	아아 ¹	안방	애 ²
쌍다	썩다	아야 ¹	안심 ⁴	애교 ²
쌍이다	썩기다 ²	아우 ¹	안심하다	애국가
쌍쌍 ¹	썩다	아유 ¹	안쓰럽다	애기 ¹
써내다 ¹	아 ²	아이 ¹	안전 ³	애쓰다
씩 ¹	아가 ¹	아이 ²	안전벨트	액자 ¹
씩다	아기 ¹	아이고	안전하다 ¹	앨범
썰다 ¹	아까	아이스크림	안쪽	앵무새
썰매	아깝다	아저씨	안타깝다	야 ⁴
쏘다	아끼다	아주 ¹	앓다	야구 ²
쏘이다 ²	아낌없이	아주머니	앓히다	야구공
쭈 ²	아내 ¹	아줌마	앓다	야구장

야단 ¹	어떡하다	어찌하다	영영	연결 ¹
야외	어떤	어항 ¹	영키다	연결하다 ¹
야채	어떻다	어허	영터리	연극
약 ⁷	어려움	억 ¹	엇그제	연기 ⁸
약간	어렴풋이	억울하다	엎다	연두색
약국 ²	어렵다	억지 ¹	엎드리다	연락 ²
약사 ⁴	어른 ¹	억지로	엎어지다	연락하다 ²
약속	어리다 ³	언니	엎지르다	연못
약속하다	어리둥절하다	언덕	에너지	연습 ²
약품	어리석다	언어 ¹	에어컨	연습하다 ³
약하다 ¹	어린아이	언제 ¹	에이 ¹	연약하다 ¹
얇다	어린애	언제나	에잇	연주 ⁶
얇전하다	어린이 ¹	언젠가	엘리베이터	연주하다 ³
얇전히	어린이날	엎다	여 ⁵	연주회
양 ³	어린이집	엎다 ¹	여기 ¹	연필
양 ⁵	어마어마하다	언어맞다	여기다	연하다 ¹
양 ¹⁸	어머	언어먹다	여기저기	열 ³
양념	어머니	얼굴	여덟	열 ⁵
양말 ¹	어머니 ¹	얼다	여동생	열다 ²
양면	어머님	얼룩	여든	열리다 ¹
양보 ³	어미 ¹	얼른 ²	여러	열리다 ²
양보하다	어버이날	열리다 ³	여러분	열매 ¹
양복	어부 ³	얼마	여럿	열쇠
양손 ¹	어색하다 ²	얼마나	여름	열심
양쪽	어서 ¹	얼어붙다	여름철	열심히
양치기 ²	어우리지다	얼음 ¹	여보 ¹	열차 ²
양치질	어울리다	엄마	여보세요	엄색하다
양치질하다	어이 ⁵	엄지 ¹	여섯	엄소 ¹
양파	어이구	엄지손가락	여왕	엇 ¹
얄다	어저께	엄청	여우 ¹	엇든다
애 ²	어제 ¹	엄청나다	여유롭다	엇보다
애기	어젯밤	엎다	여자 ²	영 ³
애기하다	어지럽다	엎히다 ¹	여자아이	영 ¹³
어 ²	어지럽히다	없다	여쭙다	영감 ¹
어긋나다	어질다	없애다	여학생	영감님
어기다 ¹	어쨌든	없어지다	여행 ²	영상 ¹
어깨 ¹	어찌다 ¹	없이	여행하다 ¹	영수증
어느	어찌다 ²	엇갈리다	역 ¹¹	영양가
어느새	어찌다가	영금영금	역사 ³	영어 ²
어둠	어찌면	영덩이	역시 ¹	영웅
어둡다	어쩐지	영똥하다	역할	영원하다
어디 ¹	어찌	영망	워다	영원히
어떠하다	어찌나	영망진창	연 ¹⁰	영화 ¹

영화관 ¹	오른발	와이셔츠	요리하다 ²	우주 ²
옆	오른손	완벽하다	요번	우주선 ¹
옆구리	오른쪽	완성 ¹	요사이	우주인
옆방	오른팔	완성되다	요새 ¹	우체국
옆자리	오리 ³	완성하다 ¹	요새 ²	우체통
옆집	오리다 ¹	완전 ¹	요술	우표
예 ⁶	오빠	완전히	요일	운 ⁶
예방 ¹	오십	왕 ⁴	요정 ²	운동 ²
예쁘다	오염	왕비	요즈음	운동장
예상되다	오염되다	왕자 ¹	요즘	운동하다
예순	오월 ¹	왜 ²	욕 ²	운동화
예약	오이 ¹	왜냐하면	욕실	운동회
예약하다	오전 ²	웬지	욕심	운전 ²
예의 ⁷	오줌	외갓집	욕조	운전기사
예전 ¹	오징어	외계인	욕하다	운전사
예절	오토바이	외국 ²	용 ⁸	운전수
예정되다	오해 ²	외국인	용감하다	운전하다
옛 ¹	오해하다	외로움	용궁	운행하다 ²
옛날	오후 ²	외롭다	용기 ²	울다 ¹
오 ²	오히려	외부 ²	용돈	울리다 ²
오 ⁴	옥상 ³	외삼촌	용서 ¹	울부짖다
오가다	옥수수	외숙모	용서하다	울상
오늘	온 ¹	외식 ²	우기다	울음
오늘날	온갖	외우다	우뚝	울음소리
오다	온도	외출	우렁차다	울타리
오뎅	온몸	외치다 ¹	우르르	울통불통하다
오들오들	온통	외할머니	우리 ¹	움직이다
오락실	올라가다	외할아버지	우리 ³	움직임
오래 ¹	올라서다	원발	우리나라	웃기다
오래 ²	올라오다	원손	우리말	웃다
오래가다	올라타다	원쪽	우물 ¹	웃음
오래간만	올려놓다	원팔	우산 ¹	웃음소리
오래도록	올려다보다	요 ³	우선 ²	웅크리다
오래되다	올리다	요거 ¹	우선 ⁷	원 ¹
오래오래 ¹	올림픽	요것	우수하다 ²	원 ²
오래전	올바르다	요구르트	우습다	원래 ¹
오랜	올챙이	요기 ¹	우승 ⁵	원숭이
오랜만	옮기다	요란하다 ¹	우승하다	원인 ²
오랫동안	옮다 ¹	요런 ¹	우연 ²	원장 ⁷
오렌지	옷	요리 ¹	우연히	원피스
오르내리다	옷차림	요리 ⁵	우울하다	원하다 ²
오르다	와 ²	요리사	우유 ¹	월 ²
오르막길	와르르	요리조리 ²	우정 ¹	월 ³

월요일	음식점	이러하다	이어지다	일본 ²
웬	음악 ¹	이런 ¹	이용하다 ¹	일부 ²
웬일	음악회	이런 ²	이웃	일부러
위 ¹	응 ¹	이런저런	이월 ¹	일부분
위 ⁵	응급실	이렇게	이유 ⁴	일어나다
위대하다 ¹	응원	이렇다	이전 ³	일어서다
위로 ²	응원단	이롭다	이제 ¹	일요일
위로하다	응원하다	이루다	이제껏	일월 ¹
위쪽	의견 ¹	이루어지다	이제야	일으키다
위층	의논	이뤄지다	이쪽 ²	일일 ¹
위치 ¹	의논하다	이르다 ¹	이튿날	일일이 ²
위하다	의미 ²	이르다 ²	이틀 ¹	일주일
위험	의사 ¹²	이르다 ³	이해 ⁵	일찍
위험하다	의상 ¹	이름	이해하다 ²	일터
유령 ²	의심 ³	이리 ³	이후 ²	일하다
유리 ⁷	의자 ³	이리로	익다 ¹	일흔
유리창 ¹	의젓하다	이리저리 ²	익다 ²	읽기
유리하다 ¹	의하다 ¹	이리하다 ¹	익히다 ²	읽다
유명 ¹	이 ³	이마 ¹	인간 ¹	읽히다 ¹
유명하다 ¹	이 ⁴	이만 ²	인기 ¹	읽다
유아 ¹	이 ⁵	이만큼	인사 ²	읽어버리다
유연하다 ⁴	이 ⁷	이모 ²	인사말	임금 ¹
유월 ¹	이거 ¹	이모부	인사하다	임금님
유치원	이것	이미 ¹	인삼	임신 ²
유치원생	이것저것	이번 ¹	인생 ¹	임신하다
유행 ²	이곳	이별	인쇄	입
유행하다 ¹	이곳저곳	이분 ¹	인어 ¹	입가
육 ²	이기다 ¹	이불 ¹	인정 ³	입구 ²
육십	이끌다	이비인후과	인정 ⁷	입김
육지 ²	이날	이빨	인제 ¹	입다
율동	이내 ³	이쁘다	인터넷	입맛
웃놀이	이놈	이사 ¹³	인형 ¹	입술
으뜸	이다 ³	이사하다 ¹	일 ¹	입원 ¹
은 ³	이다음	이삿짐	일 ³	입원하다 ¹
은행 ²	이달 ¹	이상 ⁴	일 ⁴	입장 ²
은행나무	이대로	이상 ⁹	일 ⁵	입장권
은혜	이동 ³	이상하다	일거리	입장료
음 ¹	이동하다	이슬	일곱	입장하다 ²
음 ³	이따	이십	일기 ¹²	입학
음료	이따가	이야기	일기장	입학식
음료수	이때	이야기하다	일꾼	입학하다
음식	이러다	이어 ¹	일단 ¹	입히다
음식물	이러이러하다	이어서	일등	잇다 ¹

있다 ¹	자연스럽다	잠자다	재빨리	적어지다
잉크	자유롭다	잠자리 ¹	재우다 ³	적응하다 ²
잇다	자장가	잡다 ¹	재주 ¹	적히다
잊어버리다	자장면	잡수다	재채기	전 ⁶
잎 ¹	자전거	잡아가다	재활용	전 ⁷
앞사귀	자주 ¹	잡아당기다	저 ³	전기 ¹⁴
자 ¹	작가 ¹	잡아먹다	저 ⁴	전날
자 ⁴	작년	잡채	저거	전달 ³
자 ¹¹	작다 ¹	잡혀가다	저것	전달하다 ²
자 ¹⁴	작동 ¹	잡히다 ²	저곳	전등 ⁷
자갈 ¹	작아지다	장 ¹⁷	저금	전망대
자국 ¹	작은딸	장갑 ¹	저금통	전봇대
자그마하다	작은아들	장구 ¹	저금하다	전부 ⁵
자기 ²	작은아버지	장군 ⁴	저기 ¹	전시 ⁴
자꾸 ¹	작전 ¹	장난	저기 ²	전시되다
자꾸만	잔 ²	장난감	저녁	전시장
자꾸자꾸	잔디	장난치다	저녁때	전시회
자다 ¹	잔디밭	장난하다	저녁밥	전염병
자동 ¹	잔뜩	장님	저놈	전쟁
자동차	잔소리	장마 ¹	저러다	전쟁터
자두	잔인하다	장면 ⁴	저런 ¹	전쟁하다
자라나다	잔잔하다 ¹	장미 ⁵	저렇게	전철 ⁴
자라다 ¹	잔치 ¹	장미꽃	저렇다	전체 ¹
자랑 ¹	잘 ²	장바구니	저리 ¹	전통 ⁵
자랑스럽다	잘나다	장사하다 ¹	저리 ²	전투기 ³
자랑하다 ¹	잘다	장소 ⁵	저리다	전하다
자료 ³	잘되다	장수 ¹	저물다	전혀 ¹
자루 ¹	잘리다 ¹	장식 ⁵	저번	전화 ⁷
자르다 ¹	잘못	장식품	저분	전화기
자리 ¹	잘못되다	장식하다 ³	저수지	전화번호
자물쇠	잘못하다	장애물	저장 ³	전화하다 ²
자빠지다	잘살다	장애인	저절로	절 ¹
자살하다 ¹	잘생기다	장점 ²	저쪽	절 ²
자상하다 ¹	잘하다	장화 ²	저축 ³	절대 ⁴
자석 ³	잠 ¹	재 ¹	저희 ¹	절대로
자세 ²	잠그다 ¹	재난	적 ³	절반
자세하다 ¹	잠기다 ¹	재다 ²	적 ¹¹	절벽
자세히	잠기다 ²	재료 ¹	적다 ¹	절약
자신 ¹	잠깐	재미 ¹	적다 ²	절약하다
자신 ²	잠들다	재미나다	적당하다 ²	절하다 ¹
자신감	잠바	재미없다	적당히	젊다
자신만만하다	잠시	재미있다	적시다	젊은이
자연 ¹	잠옷	재빠르다	적어도	젊음

점 ¹⁰	정확히 ¹	조용히	주다 ¹	줄기 ¹
점검	젓	조이다	주로 ¹	줄넘기
점검하다	젓다 ¹	조절 ²	주름 ¹	줄다
점수 ⁶	젓소	조절하다 ²	주말 ²	줄어들다
점심	젓히다 ¹	조종 ⁶	주머니	줄이다
점심때	제 ¹	조종사	주먹	줄줄
점심밥	제 ²⁴	조종하다	주먹밥	줄줄이
점심시간	제거 ³	조직	주무르다	줄짓다
점원 ¹	제거하다	조카	주무시다	줍다
점점 ¹	제대로	존경	주문 ⁴	중간 ¹
접다	제때	존경하다	주문하다 ¹	중국 ¹
접시	제멋대로	존댓말	주방 ⁵	중국집
젓가락	제목 ²	존중	주변 ²	중단되다
젓다	제발 ¹	존중하다	주부 ³	중독
정 ¹⁷	제비 ²	줄다 ¹	주사 ¹³	중심 ¹
정거장	제일 ⁴	줄리다 ¹	주사위	중얼거리다
정글	제자리	졸업	주소 ¹	중요 ²
정답	제정신	졸업생	주스	중요성
정도 ¹¹	제품 ²	졸업식	주어지다	중요하다 ²
정돈 ²	조 ²	졸업하다	주위 ²	중지 ²
정들다	조 ¹⁴	졸음 ¹	주유소	중학교
정류장	조각 ¹	졸졸	주의 ⁷	중학생
정리 ⁹	조각 ⁵	죸 ²	주의하다 ¹	쥐 ²
정리되다	조개 ¹	죸처럼	주인 ¹	쥐다 ¹
정리하다	조그마하다	죸다	주인공	즈음
정말 ¹	조그맣다	죸히다 ¹	주일 ³	즐거움
정말로	조금 ¹	죸 ¹³	주저앉다	즐거워하다
정면 ¹	조금씩	죸교	주전자	즐겁다
정문 ³	조끼 ¹	죸류별	주차 ⁴	즐기다 ¹
정반대	조르다 ²	죸소리	주차장	증거
정상 ²	조립 ²	죸아리	주택	지 ²
정성 ¹¹	조립하다	죸이 ¹	주황 ¹	지각 ⁵
정성스럽다	조명 ⁶	죸이접기	주황색	지갑 ³
정신 ¹²	조사 ²⁹	죸합	죽 ⁶	지계 ¹
정신없이	조사하다 ¹²	죸다	죽다 ¹	지겹다
정원 ⁶	조상 ⁷	죸아하다	죽음 ¹	지구 ³
정육점	조수 ¹	좌석	죽이다 ¹	지구 ⁴
정지 ⁵	조심 ²	좌우 ¹	준비	지금 ²
정지되다 ¹	조심스레	죄 ²	준비되다	지금껏
정지하다 ¹	조심조심	죄송하다	준비물	지나가다
정직하다 ¹	조심하다 ²	죄인	준비하다	지나다
정하다 ³	조심히	주 ²⁴	줄 ¹	지나오다
정확하다 ¹	조용하다 ¹	주고받다	줄 ⁴	지나치다

지난날	직업	짚	찢다	참깨
지난달	직접	짜다 ¹	찢어지다	참다
지난번	진달래	짜다 ²	찢다	참새 ¹
지난주	진료	짜다 ³	차 ⁶	참석하다
지난해	진실 ²	짜장면	차갑다	참외 ¹
지내다	진실하다	짜증	차곡차곡	참을성
지나다	진심 ¹	짹 ¹	차근차근 ²	참치 ¹
지다 ²	진정하다 ¹	짹 ²	차다 ¹	찾길
지다 ³	진정하다 ⁵	짹다	차다 ²	창 ⁷
지다 ⁴	진주 ⁴	짹아지다	차다 ³	창가 ¹
지다 ⁵	진짜	짹 ²	차다 ⁴	창고 ¹
지도 ³	진짜로	짹가리	차도 ¹	창문
지독하다	진찰 ²	짹개다	차라리	창밖
지렁이	진하다 ¹	짹금	차레 ²	창피
지루하다	진흙	짹다 ²	차레차레	창피하다
지르다 ²	질기다	짹 ²	차리다	찾다
지르다 ³	질리다 ¹	짹 ³	차별	찾아가다
지름길	질문	짹 ⁵	차별하다	찾아내다
지방 ⁴	질문하다	짹 ⁶	차분하다 ¹	찾아다니다
지방 ⁷	질서 ³	짹지	차분히	찾아보다
지붕	질질 ¹	짹겨나다	차비 ²	찾아오다
지식 ²	질투	짹기다	차이	채 ⁸
지어내다	짹어지다	짹다	차이다	채 ⁹
지옥	짐 ¹	짹아가다	차이점	채널
지우개	짐승	짹아내다	차지 ¹	채소
지우다 ¹	짐작	짹아다니다	차지하다 ¹	채우다 ³
지은이	짐 ¹	짹아오다	착각 ³	책 ¹
지저분하다	짐계 ¹	짹	착각하다	책가방
지진 ²	집다	짹	착륙	책방 ¹
지치다 ¹	집안 ¹	짹 ²	착용하다	책상 ¹
지켜보다	집안일	짹개 ¹	착착 ³	책임감
지키다	집어넣다	짹그리지다	착하다	책임지다
지팡이	집주인	짹끼기	찬물 ¹	책장 ²
지퍼	집중 ²	짹다 ¹	찬바람	챙기다
지폐	집중하다 ²	짹르다	찬성 ¹	처리
지푸라기	집집	짹푸리다	찬성하다	처리하다
지하	짚 ¹	짹다 ¹	참 ¹	처음
지하도	짚다 ¹	짹다 ²	참 ³	처지다
지하실	짚뱀다	짹히다 ²	참가 ¹	척 ¹
지하철	징 ²	짹 ²	참견	척척 ²
지하철역	징그럽다	짹리다	참고 ¹	척하다
지혜 ²	짚다	짹그리다	참고하다	천 ³
지혜롭다	짚다 ²	짹기다 ¹	참기름	천국 ¹

천둥	총알	측	캔	키 ⁴
천사 ⁴	촬영	층 ³	캠핑	키우다
천장 ²	촬영하다	치과	커다랗다	타다 ¹
천재 ²	최고 ²	치다 ¹	커지다	타다 ²
천천히	최근	치다 ²	커튼	타다 ³
철 ¹	최대한	치다 ⁵	커피	타다 ⁴
철 ⁵	추가 ²	치료	کمکم하다	타오르다
철도	추가하다	치료비	컴퓨터	타이어 ²
철봉	추다 ²	치료하다	컵	탁 ¹
첫	추락 ²	치마 ¹	케이크	탁구
첫날	추락하다 ²	치아 ²	켜다 ¹	탁자 ¹
첫눈 ¹	추석 ¹	치약	컬레 ²	탁탁 ¹
첫째	추수 ²	치우다 ¹	코 ¹	탈 ¹
청바지	추억	치즈	코끼리	탈락 ²
청소 ⁵	추위 ¹	치킨	코스모스 ¹	탈출 ²
청소기 ²	축 ³	칙칙폭폭	코앞	탈출하다 ²
청소하다 ³	축구 ⁴	친구 ²	코트 ¹	탐정 ¹
체온	축구공	친근하다	코피 ¹	탐 ¹
체육	축구장	친절	콧콧	탓
체육관	축복	친절하다	콜라	탕 ²
체조 ²	축제 ¹	친척	콧물	탕탕 ²
체하다 ¹	축축하다	친하다	콩 ¹	태권도
체험	축하	칠 ¹	콩나물	태극기
체험하다	축하하다	칠 ²	광	태도 ³
쳐다보다	출근	칠십	쿨쿨 ²	태양 ¹
쳐들다	출근하다	칠월	콩 ²	태어나다
쳐들어가다	출동	칠판	퀴즈	태우다 ¹
쳐들어오다	출동하다	칠하다	크기	태우다 ²
초 ¹	출발	침 ¹	크나크다	태풍
초 ⁷	출발점	침대 ²	크다	택배
초가집	출발하다	침실	크레파스	택시
초대 ⁵	출석	칫솔	크리스마스	탱크
초대하다 ²	출입문	칭찬	크림	터뜨리다
초등학교	춤 ¹	칭찬하다	큰길	터지다
초등학생	춤추다	카드	큰돈	터트리다
초록 ²	춤다	카레	큰딸	턱 ¹
초록색	충격 ²	카메라	큰소리 ¹	턱 ⁵
초원 ⁴	충분하다 ¹	카페	큰아들	털
초콜릿	충분히	칸 ¹	큰아버지	털다
촉촉하다 ¹	취미 ⁴	칼 ¹	큰일 ¹	털어놓다
촛불	취소 ¹	칼날	큰절 ¹	털밭
총 ³	취소하다 ¹	깜깜하다	큼직하다	텅 ¹
총 ⁶	취하다 ³	캐다	키 ¹	텅텅 ¹

테두리	튼튼하다	팔월	포기 ²	풍경 ¹
테이블	튼튼히	팔짱	포기하다 ¹	풍선 ²
테이프	틀	팔찌	포도 ⁶	풍성하다
텐트	틀다	팔	포장 ¹	프로그램
텔레비전	틀리다 ¹	패다 ³	포장지	프린트
토 ⁴	틀림없다	팬 ¹	포장하다 ¹	플라스틱
토끼	틀림없이	팬티	포크 ¹	피 ²
토마토	툼	팬이 ¹	포함하다 ²	피곤
토막 ¹	퇴우다 ²	퍼내다	폭발 ²	피곤하다
토요일	티 ²	퍼뜨리다	폭발하다 ¹	피나다
토하다	티브이	퍼지다	폭발하다 ²	피다 ¹
톡 ²	티셔츠	퍽 ¹	폭죽	피리 ¹
통 ⁹	티켓	퍽 ²	폭탄	피부 ²
통 ¹¹	팀	펼쩍	폭풍 ¹	피부병
통과	파 ¹	펼 펼 ¹	폰 ^{미등제}	피아노 ¹
통과하다	파괴	평 ¹	표 ²	피어나다
통로	파괴되다	페이지	표 ⁴	피우다
통일 ²	파괴하다	페인트 ²	표 ⁵	피자
통장 ¹	파내다	펜 ¹	표시 ¹	피하다
통조림	파다 ¹	퍼다	표시 ²	피해 ¹
통째	파도	퍼지다	표시하다 ¹	핀
통통하다 ¹	파란색	편 ⁴	표시하다 ²	필수 ²
통하다	파랑 ¹	편 ⁹	표정 ²	필요
통화하다 ²	파랗다	편리하다	표지 ¹	필요하다
퇴근	파리 ¹	편식	표지판	필통
퇴근하다	파묻히다	편안하다 ¹	표현	핏줄
투명 ²	파스 ³	편안히	표현하다	핑 ¹
투명하다 ²	파이다	편지 ²	푸다	핑계
툭 ¹	파이팅	편하다	푸르다	핑크색 ^{미등제}
툭툭 ¹	파티	편히	푸른색	허 ³
튀기다 ¹	파헤치다	펼쳐지다	푸짐하다	하나
튀기다 ²	팍	펼치다	폭 ¹	하나님
튀김 ¹	판 ¹	평등하다	폭신하다	하나하나
튀다	판 ⁶	평범하다	폭폭 ¹	하느님
튀어나오다	판매	평상시	폴 ¹	하늘 ¹
튀기다	판매하다	평생	폴 ²	하늘나라
튜브	팔 ¹	평소	폴다	하늘색
트럭	팔 ²	평일	폴려나다	하다 ¹
트집	팔다	평화 ¹	폴리다	하다 ²
특별	팔다리	평화롭다	폴발	하도 ¹
특별하다	팔딱팔딱	폐 ²	풀어지다	하루 ¹
특별히	팔리다	포근하다	풀이	하루빨리
특이하다	팔십	포기 ¹	품다 ¹	하루하루

하룻밤	한숨 ²	핸드백	현관문	화장실
하마터면	한약	핸드폰	현재 ²	화장지 ¹
하수구	한여름	핸들	협동	화장품
하얏다	한우 ³	햄버거	헛바닥	화재 ¹
하여튼	한입	햇별	형 ¹	화해 ²
하인 ¹	한자 ²	햇빛	형님	화해하다 ¹
하지만	한자리	햇살	형식 ¹	확 ²
하트 ⁴	한쪽	행동	형제 ¹	확실하다
하품 ¹	한참	행동하다	호 ¹	확실히
하필 ²	한판	행복 ²	호기심	확인 ²
하하 ¹	한편	행복하다	호들갑	확인되다
학교	할머니	행사 ¹	호랑이	확인하다
학년	할아버지	행운 ²	호박 ¹	환기 ²
학생	할인 ¹	행주 ¹	호수 ⁷	환영 ²
학습	할퀴다	향 ³	호스	환영하다
학습지	훤다	향기 ¹	호텔	환자 ³
학습하다	함께	향하다	흑 ¹	환하다 ¹
학원 ²	함부로	허 ²	흑시 ¹	활 ¹
한 ¹	합격	허겁지겁	흑시나	활기
한 ⁵	합격하다	허락	흑은	활기차다
한가운데	합창 ¹	허락하다	훈 ²	활동 ²
한가하다 ²	합창단	허리 ¹	훈나다	활동하다 ¹
한걸음	합치다	허리띠	훈내다	활발하다
한겨울	합하다	허물다 ²	훈자 ¹	활발히
한곳	향복하다	허물어지다	훈жат말	활짝
한구석	항상	허벅지	홀로	활활
한국말	항아리	허수아비	홍수 ¹	황당하다
한국어	해 ¹	허영다	회 ⁵	황소 ¹
한국인	해 ⁵	허전하다 ¹	화 ⁶	회 ⁴
한글 ¹	해결	허허 ²	화가 ³	회 ⁹
한꺼번에	해결책	헌	화나다	회복
한눈 ¹	해결하다	헐다 ¹	화내다	회복되다
한동안	해골	험하다	화려하다	회복하다
한두	해내다	형겼	화면 ⁵	회사 ⁴
한때	해마다	해매다	화목하다	회사원
한마디	해바라기 ²	헤어지다	화분 ¹	회색 ¹
한바탕	해변	헤엄치다	화산 ¹	회전 ⁴
한밤중	해설 ³	헤치다	화살	횃수
한번	해수욕장	헬리콥터	화살표	횡단보도
한복	해외여행	헛갈리다	화상 ³	효도
한복판	해적 ²	행구다	화석 ¹	효도하다
한순간	해치다	허 ¹	화요일	효자 ¹
한숨 ¹	해치우다	현관 ¹	화장 ²	후 ¹

후 ²	휘날리다	휴지통	흘러내리다	힘
후 ⁸	휘다	흥 ¹	흘리다	힘겹다
후다닥	휘두르다 ¹	흥내	흙 ¹	힘껏
후딱	휘말리다	흥터	흙탕물	힘내다
후련하다	휘젓다	흐르다 ¹	흥 ³	힘들다
후추	획	흐리다 ²	흥 ⁴	힘세다
후회 ¹	획획	흐트러지다	흥겹다	힘쓰다
후회하다	획체어	흑인	흥미롭다	힘없다
후후 ¹	휩싸이다	흔들거리다	흥분	힘없이
훈련	휴가 ¹	흔들다	흥분하다	힘주다
훈련하다	휴게소	흔들리다	흘어지다	힘차다
훌륭하다	휴대폰	흔적	희다	
훌쩍	휴식 ²	흔하다	희망	
흙치다 ²	휴식하다 ²	흔히	흰머리	
흰하다	휴일	흘러가다	흰색	
훨씬	휴지 ²	흘러나오다	힌트	

사업 책임자	서상규 (연세대학교 국어국문학과 교수)
사업 참여자	김한샘 (연세대학교 언어정보연구원 교수)
	봉미경 (연세대학교 언어정보연구원 전문연구원)
	황은하 (배재대학교 한국어교육과 교수)
	이병규 (서울교육대학교 한국어교육과 교수)
	가은아 (공주대학교 한국어교육과 교수)
	윤천택 (한국교원대학교 한국어교육과 교수)
	강범일 (연세대학교 언어정보연구원 전문연구원)
	윤영민 (연세대학교 언어정보연구원 전문연구원)
	유근선 (연세대학교 언어정보연구원 전문연구원)
	박혜진 (연세대학교 국어학 박사과정)
	천성호 (연세대학교 국어학 박사과정)
	오태환 (연세대학교 국어학 박사과정)
담당 연구원	남미정 (국립국어원)

발행인: 국립국어원장

발행처: 국립국어원

서울시 강서구 금남화로 154

전화 02-2669-9775, 전송 02-2669-9727

인쇄일: 2021년 12월 21일

발행일: 2021년 12월 21일

인 쇄: 연세대학교 POD 센터

※ 이 책은 국립국어원의 용역비로 수행한 ‘2021년 국어 기초 어휘 선정 및 어휘 등급화 연구’ 사업의 결과물을 발간한 것입니다.