

7학년 국어교과서의 어휘 통계조사 -노미숙, 김태철 교과서를 중심으로- Statistical Survey of Vocabulary in Korean Textbook for 7th-Grade -Focus on the No's and the Kim's Textbooks-

김철수
서남대학교 컴퓨터정보학과

Cheol-su Kim(chskim@seonam.ac.kr)

요약

본 연구는 중학교 7학년 국어교과서 4종에 나타나는 어휘들에 대한 통계(음절 개수, 음절 종류, 음절 출현 빈도, 어절 개수, 어절 종류, 어절 평균 길이, 어절 출현 빈도 등)를 조사하였다. 전체 한글 음절 개수는 286,801개, 한글 음절 종류는 1,350개, 음절 평균 출현빈도는 212.4회이다. 어절 개수는 109,393개, 어절 종류는 29,356개, 어절 평균 출현빈도는 3.7회이다. 평균 음절 길이는 2.7음절이며, 한글 어절 중 가장 긴 어절 길이는 8이다.

■ 중심어 : | 어휘 | 통계조사 | 음절 | 어절 | 출현빈도 |

Abstract

This paper studied the statistics such as the total number of syllables, the kinds of syllables, the frequency of syllables, the total number of eojeols, the kinds of eojeols(word phrases unique in Korean language), average length of eojeols and the frequency of eojeols in four different Korean textbooks for 7th-grade students. The results of the statistical survey are as follows: the number of syllables was 286,801; the kinds of Korean syllables were 1,350; the average frequency of syllables was 212.4; the number of eojeol was 109,393; the kinds of eojeol were 29,356; the average frequency of eojeol was 3.7; the average length of eojeols was 2.7 syllables, the longest one consist of 8 syllables.

■ keyword : | Vocabulary | Statistical-survey | Syllable | Eojeol | Frequency |

1. 서론

우리는 언어를 바르고 정확하게 사용하기 위하여 적어도 9~12년 이상의 긴 세월동안 많은 시간과 노력을 투자하여 교육을 받고 임하고 습득한다. 국어 교과서는 한국인의 언어 교육을 위해 필요한 가장 중요한 자료이자 대표적인 자료로써 체계적인 국어교육을 위한 어휘

선택은 국어교과서 편찬에 중요한 요소가 된다.

“중학교의 교육은 초등학교 교육의 성과를 바탕으로, 학생의 학습과 일상생활에 필요한 기본 능력을 배양하며, 다원적인 가치를 수용하고 존중하는 민주시민의 자질 함양에 중점을 둔다.”고 개정교육과정[1]에서 기술하고 있다.

국어교과서에 수록된 어휘들에 대한 통계 자료들은

접수일자 : 2014년 02월 28일
수정일자 : 2014년 03월 24일

심사완료일 : 2014년 04월 10일
교신저자 : 김철수, e-mail : chskim@seonam.ac.kr

국어 교육의 목적을 달성하기 위하여 국어교과서 편찬 과정에서 교과서에 수록할 내용 선택 및 범위, 어휘 선택 등에 중요한 기초 자료가 된다. 또한 인간과 기계 사이의 인터페이스로써 더욱 중요시되고 각광받고 있는 음성인식, 정보검색, 기계번역, 문자인식과 같은 분야에서 자연어 처리 기법을 적용하여 처리하고자할 때 필수적인 전자사전 구축을 위한 기초 자료가 된다.

국어교육의 기초과정이라 할 수 있는 초등교육과정을 마무리하고, 학습과 일상생활에 필요한 기본 능력을 배양하기 위하여 새롭게 시작하는 중등교육과정의 첫째인 7학년(중학교 1학년)은 또 다른 의미를 가진다.

본 연구는 7학년 국어교과서 국어1, 국어2에 수록된 어휘들에 대한 통계 자료들을 조사하여 7학년 국어교과서 편찬을 위한 기초 자료로 활용될 수 있도록 한다. 뿐만 아니라 국어교과서에 수록된 내용을 자연어 처리 기법을 활용하여 처리하고자 할 때 전자사전 구축과정에서 전자사전에 포함될 어휘들의 범위, 어휘들의 종류, 어휘들의 발생빈도 등과 같은 어휘 관련 정보 및 어휘에 포함된 음절 종류, 음절발생 횟수 등과 같은 정보들을 활용할 수 있도록 한다. 이를 위해 교과서가 국정 체계에서 검인정 체계로 전환됨에 따라 2개의 출판사에서 발행하는 국어1, 국어2 교과서 4종[2-5]을 대상으로 하였다. 4종 교과서에 수록된 내용들에 대한 연어휘수, 개별어휘수, 어휘 발생빈도, 어휘길이, 음절수, 개별음절수, 음절 발생빈도 등에 대하여 조사하였다.

II. 관련연구

1. 교육과정의 변천[1]

국가교육과정 정보센터의 우리나라 교육과정 원문 인벤토리에 따르면, 1954년에 제정1차 교육과정의 초안이라 할 수 있는 “국민학교·중학교·고등학교·사범학교 교육과정 시간배당기준령”이 제정된 이후, 1963년 2월에 2차 교육과정을 공포하였다. 1973년에 3차, 1981년에 4차, 1987년에 5차, 1992년에 6차, 1997년에 7차 교육과정을 공포하였다. 제7차 교육과정 이후 상시 개정체제로 전환하여 2007년에 2007 개정 교육과정, 2009년에 2009 개정교육과정을 공포하였다. 2009개정 교육과

정의 중학교 1학년 국어교과서는 2013년부터 적용된다.

2. 시군별 중학교 현황

연구대상이 될 교과서 선정을 위하여 전북지역의 행정구역별 중학교 분포, 검인정 교과서 개수, 각각의 검인정 국어교과서별 채택 중학교 현황, 국어교과서 채택 학교들의 중학교 1학년 재학생수 등에 대하여 조사하였다. 전북지역의 행정구역은 6시 8군으로 총 14개의 시군으로 구성되어 있다. 각 시군별 소재 중학교 수는 [표 2-1]과 같다. 전북지역 전체 중학교 수는 208개이며, 전주시 소재 중학교 수는 37개로 도내 전체 중학교 수의 17.8%이다.

표 2-1. 시군별 중학교 현황(가나다순)

시군명(A)	공립(B)	사립(C)	계(D=B+C)	비율(%)
고창군	10	4	14	6.7
군산시	15	4	19	9.1
김제시	7	6	13	6.3
남원시	13	1	14	6.7
무주군	6	0	6	2.9
부안군	9	4	13	6.3
순창군	6	1	7	3.4
완주군	11	3	14	6.7
익산시	16	10	26	12.5
임실군	9	0	9	4.3
장수군	7	0	7	3.4
전주시	28	9	37	17.8
정읍시	12	7	19	9.1
진안군	10	0	10	4.8
합계	159	49	208	100.0

3. 시군별 7학년 학생 수

분석할 교과서 선정을 위한 전단계의 일환으로 시군별 7학년 학생 수를 조사하였다. 2012년 4.1일 기준 전북지역의 7학년 학생수는 22,238명이며 전주시가 9,470명으로 전체의 42.6%를 차지한다[표 2-2].

표 2-2. 시군별 7학년 학생수 현황

시군명	학생수	비율(%)	시군명	학생수	비율(%)
전주시	9,470	42.6	부안군	560	2.5
익산시	3,831	17.2	순창군	234	1.1
군산시	3,470	15.6	무주군	211	0.9
정읍시	1,386	6.2	임실군	179	0.8
남원시	1,034	4.6	진안군	177	0.8
완주군	960	4.3	장수군	163	0.7
고창군	563	2.5	합계	22,238	100.0

4. 검인정 중학교 국어 교과서 발행현황

2013학년도부터 7학년 학생들에 적용되는 2009교육 과정에 따른 검인정 국어 교과서는 전체 16종이며, 각각의 교과서 집필에 참여한 저자들은 11~27명으로 전체 저자 수는 277명, 평균 저자 수는 17.3명이다[표 2-3].

표 2-3. 2009 개정교육과정 국어교과서 출판사 및 저자

출판사	저자	출판사	저자
교학사	남미영외17	비상교육	한철우외17
금성출판사	박경신외14	좋은책신사고	민현식외18
대교	장수익외12	좋은책신사고	우한용외18
두산동아	전경원외11	지학사	방민호외20
두산동아	이삼형외15	창비	이도영외10
미래엔	윤여탁외21	천재교과서	김종철외15
비상교과서	이관규외18	천재교육	노미숙외15
비상교육	김태철외26	천재교육	박영목외14

5. 국어교과서 채택 현황

2013학년도 7학년 학생들의 국어 교육을 위해 전주시 전체 중학교에서 채택하고 있는 국어 교과서 현황은 [표 2-4]와 같다. 교과서별 채택 현황은 노미숙[4][5]이 2,661명, 김태철[2][3]이 1,771명으로 노미숙 또는 김태철 교과서를 채택한 학교의 학생수 비율이 전체의 41.0%이고, 채택 학교 수는 37개교 중 18개교로 38.6%로 나타났다. 조동길의 경우는 2013년도가 7학년들에 대한 2009교육과정을 적용하는 첫해인데, 3개 학교에서 2007 교육과정인 기존 교과서를 채택하였다.

표 2-4. 교과서 채택 현황

출판사	저자	채택 학교수	채택 학생수	채택학교 비율(%)	채택학생 비율(%)
천재교육	노미숙	10	3,062	27.0	32.3
비상	김태철	8	1,771	21.6	18.7
천재교육	박영목	5	1,276	13.5	13.5
미래엔	윤여탁	3	918	8.1	9.7
창비	이도영	4	789	10.8	8.3
창비	김상욱	2	717	5.4	7.6
비상	조동길	3	492	8.1	5.2
두산동아	전경원	1	225	2.7	2.4
좋은책신사고	민현식	1	220	2.7	2.3
합계		37	9,470	100.0	100.0

6. 중학교 국어교과서 어휘 관련 연구

국어교과서 어휘 분석과 관련한 많은 연구가 있었으며, 중학교 1학년 국어교과서 관련 어휘 분석 연구로 [6][7]을 들 수 있다. [6]은 제 7차 교육과정에 따른 중학교 국어교과서 1학년 1학기를 대상으로 언어회수 24,818개, 개별어회수 4,812개에 대하여 빈도별, 자모별, 품사별, 조어형태별, 어원별, 의미별로 분석하여 그 목록을 제시하고 있다. [7]은 7차 교육과정의 중학교 국어교과서 어휘를 학년별 조사 분석을 시도한 것으로 중학교 1학년 1학기, 2학년 1학기, 3학년 1학기 국어 읽기 교과서를 대상으로 하였다. 1학년 1학기 3,210개, 2학년 1학기 4,767개, 3학년 1학기 5,623개 총 13,600개의 언어회와 1학년 1학기 1,163개, 2학년 1학기 1,240개, 3학년 1학기 1,674개 총 4,077개의 개별어회에 대하여 빈도별, 품사별, 조어형태별, 의미영역별로 분석하였다. 개별어회 평균빈도는 3.34로 제시하고 있다.

III. 어휘 통계 조사

1. 어휘 통계조사 기준 및 용어

7학년 국어 교과서 4종에 대한 어휘 통계 조사를 위한 어휘 분석 기준 및 용어는 다음과 같다.

- ① 음절분석은 분석 범위내의 한글 음절뿐 만 아니라 한자, 영어, 숫자도 포함하였다. 숫자의 경우 아라비아숫자 1개를 1음절로 하고, 특수문자는 제외한다.
- ② 어휘는 띄어 쓴 문자열을 하나의 어휘로 함을 원칙으로 한다. 성명은 띄어쓰기 되어 있어도 붙여서 하나의 어휘로 한다.
- ③ 한글과 외국어, 한글과 특수문자, 한글과 숫자가 혼합된 어휘는 하나의 어휘로 한다.
- ④ 접사가 붙은 파생어는 별개의 어휘로 한다.
- ⑤ 사투리와 외국어도 독립된 단위로 한다.
- ⑥ 본용언과 분리된 어절의 보조용언은 별개의 단위로 한다. 단 본용언과 보조용언이 합하여 굳어진 것은 하나의 단위로 한다.
- ⑦ 각 교과서에 대한 분석 범위는 교과 내용이 실제

시작되는 쪽에서 종료되는 쪽까지 한다. 목차와 부록은 제외한다.

- ⑧ 용어정의 : “조사대상이 되는 자료에 출현하는 모든 조사 단위의 어휘 수”를 언어휘(N), “중복되어 나타난 것을 헤아리지 않은 개별적인 어휘 수”를 개별어휘(V)로 정의한다[8].

연구 대상이 되는 교과서를 간략히 지칭하기 위하여 K_1[2], K_2[3], N_1[4], N_2[5]로 표기하기로 한다.

2. 언어휘

4종 교과서에 나타나는 언어휘수를 조사하였다. 4종 교과서에 나타나는 언어휘수는 109,393개로, 각 교과서에 나타나는 평균 언어휘수는 27,348개이다. K_2 교과서가 31,617개로 가장 많고, N_2 교과서가 22,535개로 가장 적다. 4종 교과서 전체에 대한 각 장(Chapter)의 평균 언어휘수는 5,758개이다. 가장 많은 언어휘수를 가지는 장은 N_1 교과서의 5장으로 10,035개이고, 가장 적은 언어휘수를 가지는 장은 N_1 교과서의 1장으로 3,491개이다[표 3-1].

표 3-1. 교과서별 출현 언어휘수

구분	K_1	K_2	N_1	N_2	합계	평균
1장	4,016	5,995	3,491	5,459	18,961	4,740
2장	8,381	7,159	4,279	5,117	24,936	6,234
3장	3,539	9,558	5,305	7,385	25,787	6,447
4장	5,400	3,910	4,986	4,574	18,870	4,718
5장	5,809	4,995	10,035		20,839	6,946
합계	27,145	31,617	28,096	22,535	109,393	27,348
평균	5,429	6,323	5,619	5,634	21,879	5,758

109,393개의 언어휘에 포함된 전체 음절수는 291,989개로 평균 음절길이는 2.67이다. 2음절 또는 3음절로 구성된 어휘가 전체의 68.32%이다[표 3-2]. 1음절로 구성된 어휘가 11.77%로 상대적으로 높은 편인데 이는 단일 숫자로 구성된 어휘들이 많은 것에 기인한다. 한글 어휘 중 가장 긴 어휘는 8음절 구성된 어휘로 28개이다. 9~11음절로 구성된 어휘는 영어 어휘 5개이다.

국어교과서이지만 교과서 내용 전체가 국어로만 구성되어 있는 것은 아니다. 어휘는 한글로만 구성된 어

표 3-2. 어절 길이

어절길이	K_1	K_2	N_1	N_2	합	비율(%)
1	3,464	3,881	3,086	2,443	12,874	11.77
2	9,555	11,461	9,795	7,865	38,676	35.36
3	8,634	10,460	9,288	7,675	36,057	32.96
4	4,218	4,535	4,492	3,434	16,679	15.25
5	998	1,210	1,115	889	4,212	3.85
6	218	22	248	197	685	0.63
7	47	42	62	26	177	0.16
8	10	5	7	6	28	0.03
9			3		3	0.00
10	1				1	0.00
11		1			1	0.00
합	27,145	31,617	28,096	22,535	109,393	100.00

휘와 숫자와 한글이 결합된 어휘, 숫자로만 구성된 어휘, 영어로 구성된 어휘, 한자로 구성된 어휘, 기타 어휘로 분류해 볼 수 있다. 한글로만 구성된 어휘들을 다시 첫음절의 초성에 따라 어휘들을 분류할 수 있다. 4종 교과서에 수록된 어휘들의 첫음절 초성에 따라 분류하면 [표 3-3]과 같다.

표 3-3. 어휘 첫음절 초성에 따른 어휘분류

구분	어절수	비율(%)	구분	어절수	비율(%)
ㄱ	14,431	13.2	ㅊ	3,145	2.9
ㄴ	6,249	5.7	ㅋ	546	0.5
ㄷ	10,745	9.8	ㅌ	1,542	1.4
ㄹ	417	0.4	ㅍ	1,933	1.8
ㅁ	7,877	7.2	ㅎ	6,910	6.3
ㅂ	8,994	8.2	중성	206	0.2
ㅅ	12,028	11.0	숫자	2,830	2.6
ㅇ	20,519	18.8	영어	141	0.1
ㅈ	10,702	9.8	한자	178	0.2
			합계	109,393	100.0

[표 3-3]에서 보는 것처럼 “ㅇ”으로 시작하는 어휘가 18.8%로 가장 많고 ㄱ, ㄷ, ㅈ..... 순으로 나타났다. 숫자로 시작되는 어휘는 2.6%로 ㅍ, ㅌ, ㅋ, ㄹ 로 시작하는 어휘보다 많이 나타났다. 국어 교과서이지만 한자나 영어로 시작하는 어절이 0.3%를 차지한다.

3. 개별어휘

4종 교과서에 나타나는 언어휘들에 대한 개별어휘를 조사하였다. 각 교과서에 나타나는 개별어휘수는 [표 3-4]와 같다.

표 3-4. 교과서별 개별어휘수

구분	K_1	K_2	N_1	N_2	합	평균
1장	1,898	2,822	1,376	2,340	8,436	2,109
2장	3,947	3,102	2,011	2,470	11,530	2,883
3장	1,561	4,354	2,576	4,037	12,528	3,132
4장	2,345	3,910	2,418	2,069	10,742	2,686
5장	3,145	2,568	4,999		10,712	3,571
합계	12,896	16,756	13,380	10,916	53,948	14,380
평균	2,579	3,351	2,676	2,729	11,335	2,839
전체	9,546	10,931	10,528	8,676	39,681	9,920

4종 교과서 각 장에 나타나는 개별어휘들의 합은 53,948개이며, 각 교과서에 평균적으로는 13,487개의 개별어휘가 나타난다. 개별어휘수가 가장 많은 장은 N_1 교과서의 5장으로 4,999개이며, 가장 적은 장은 N_1 교과서의 1장으로 1,376개이며, 각 장별 평균 개별어휘수는 2,839개이다.

동일한 개별어휘가 여러 장에 공통으로 나타날 수 있으므로 장별 그룹이 아닌 4종 교과서 전체를 하나로 통합하여 조사한 개별어휘수는 283,56개이다. 가장 많은 개별어휘수를 가지는 교과서는 K_2 교과서로 10,931개이고, 가장 적은 개별어휘수를 가지는 교과서는 N_2 교과서로 8,676개이다.

4종 교과서에 나타나는 109,393개의 연어휘에 대한 개별어휘수가 28,356개이므로 28,356개의 개별어휘는 평균적으로 3.86회 나타난다. 28,356개의 개별어휘 중 가장 많이 발생하는 어휘는 “보자”로 적게는 279회에서 많게는 352회 나타나 평균적으로는 332.8회 나타났다. 이 어휘는 4종 교과서 모두에서 가장 많이 발생하는 어휘이기도 하다. 28,356개의 개별어휘 중 상위 10위 안에 속하는 한글 어휘들의 순위 및 발생 횟수는 [표 3-5]와 같다.

4종 교과서에 나타나는 28,356개의 개별어휘 중 4종 교과서에 공통으로 나타나는 개별어휘는 1,386개(4.9%)이고, 26,972개(95.1%)의 개별어휘는 4종 교과서에 분산되어 나타난다. 그러나 공통으로 나타나는 1,395개(4.9%)의 개별어휘들의 발생빈도를 보면, 연어휘 109,393개 중 50,154개로 45.8%나 된다. 분산되어 나타나는 26,972개(95.1%)의 개별어휘 발생 빈도는 59,239회로 54.2%이다. 이는 4.9%의 일부 어휘들이 집중적으로

로 나타나 전체 교과서의 45.8%를 차지함을 의미한다.

표 3-5. 발생 횟수 상위 10위 안의 한글 어휘

순위	단어	K_1	K_2	N_1	N_2	합계
1	보자	349	352	351	279	1,331
2	수	245	212	203	166	826
3	이	229	203	127	120	679
4	있는	125	140	137	102	504
5	내용을	125	68	181	85	459
6	그	93	171	104	87	455
7	있다	59	95	139	123	416
8	글을	107	45	104	90	346
9	다음	87	98	89	69	343
10	한	71	84	117	64	336

4종 교과서에 분산되어 나타나는 26,972개의 개별어휘 중 17,308개의 개별어휘는 단 1번 나타난다. 이는 한종의 교과서에서만 나타나고 나머지 3종의 교과서에서는 나타나지 않음을 의미한다. 개별어휘 관점에서 보면 28,356개의 전체 개별어휘 중 17,308개의 개별어휘가 한번만 출현하므로 61.0%의 개별어휘가 한 교과서에서만 나타남을 의미한다. 연어휘 관점에서 보면 109,393개의 연어휘 중 17,309개의 개별어휘는 한번만 출현하므로 15.8%의 어휘는 한 교과서에서만 나타난다고 볼 수 있다. 28,356개의 개별어휘들의 발생횟수에 따른 구간별 개별어휘수, 개별어휘 발생 횟수 및 구성 비율은 [표 3-6]과 같다.

표 3-6. 발생순위 구간별 개별어휘 및 연어휘

발생횟수 구간	개별어휘		연어휘	
	개수	비율(%)	개수	비율(%)
100회 이상	85	0.3	18,743	17.2
50 ~ 99	187	0.7	12,824	11.7
10 ~ 49	1,549	5.4	29,349	26.8
5 ~ 9	1,946	6.9	12,575	11.5
4	1,081	3.8	4,324	4.0
3	1,867	6.6	5,604	5.1
2	4,333	15.3	8,666	7.9
1	17,308	61.0	17,308	15.8
합계	28,356	100.0	109,393	100.0

[표 3-6]에서 보는 것처럼 4종 교과서의 연어휘 109,393개에 대한 28,356개의 개별어휘 중 100회 이상 발생하는 85개(0.3%)의 개별어휘가 18,743번 발생하여

전체 교과서의 17.2%를 차지한다. 그렇지만 28,358개의 개별어휘 중 4회 이하 발생하는 24,590개(86.7%)의 개별어휘들은 4종교과서에 35,902회 발생하여 전체 교과서의 32.8% 정도 밖에 차지하지 않는다. 4회 이하 발생하는 개별어휘는 각각의 개별 교과서에서 평균적으로 1~0.25회 출현하는 어휘이다.

4. 음절

언어휘 109,393개를 구성하는 음절수를 조사하였다. 4종 교과서를 구성하는 음절수는 291,989개로 어휘 평균 길이(음절수)는 2.67이며, 각 장의 평균 음절수는 15,368개이다[표 3-7]. 가장 많은 음절수를 가지는 교과서는 K_2 교과서로 83,047개이며, 가장 적은 음절수를 가지는 교과서는 N_2 교과서로 60,791개이다. 4종 교과서를 구성하는 장별 평균 음절수는 K_2 교과서가 16,609개로 가장 많고, K_1 교과서가 14,413개로 가장 적다.

표 3-7. 각 장별 음절수

구분	K_1	K_2	N_1	N_2	합	평균
1장	10,630	16,047	9283	14,720	50,680	12,670
2장	23,031	19,348	11734	13,763	67,876	16,969
3장	8,873	24,668	14119	20,109	67,769	16,942
4장	13,660	9,891	13625	12,199	49,375	12,344
5장	15,869	13,093	27,327		56,289	18,763
합계	72,063	83,047	76,088	60,791	291,989	77,688
평균	14,413	16,609	15,218	15,198	58,398	15,538

N_2 교과서가 K_1 교과서에 비해 전체 음절수가 적음에도 불구하고 평균 음절수는 N_2 교과서가 많은 이유는 N_2 교과서가 4개의 장으로 구성되어 있고 나머지 3종교과서는 5개의 장으로 구성되어 있기 때문이다.

5. 개별음절

109,393개의 언어휘를 구성하는 291,989개의 음절들에 대한 개별 음절을 교과서별 및 장별로 조사하였다. 가장 많은 개별 음절을 가지는 교과서는 N_1 교과서로 1,265개이며, 가장 적은 개별음절을 가지는 교과서는 K_1 교과서로 1,171개이다. 장별로 비교했을 때 가장

많은 개별음절을 가지는 장은 N_1 교과서의 5장으로 959개이며, 가장 적은 개별 음절을 가지는 장은 N_1 교과서의 1장으로 558개이다[표 3-8].

표 3-8. 개별음절수

구분	K_1	K_2	N_1	N_2	합	평균
1장	558	729	632	678	2,597	649.3
2장	822	724	635	678	2,859	714.8
3장	602	893	705	919	3,119	779.8
4장	722	591	702	619	2,634	658.5
5장	766	705	959		2,430	810.0
합계	3,470	3,642	3,633	2,894	13,639	3,409.8
평균	694.0	728.4	726.6	723.5	2872.5	717.8
전체	1,171	1,188	1,265	1,208	4,832	1,208.0

4종교과서 전체에 나타나는 291,989개의 음절들에 대한 1.725개의 개별음절들의 발생회수에 따른 상위 10위 안에 속하는 음절들의 발생회수는 [표 3-9]와 같다.

표 3-9. 발생 순위 10위 이내 음절

순위	음절	K_1	K_2	N_1	N_2	합계	비율
1	이	2,564	2,665	2,812	2,242	10,283	3.5
2	다	1,643	2,327	1,971	1,623	7,564	2.6
3	는	1,708	2,245	2,058	1,509	7,520	2.6
4	을	1,870	2,013	1,942	1,556	7,381	2.5
5	하	1,371	1,511	1,502	1,220	5,604	1.9
6	에	1,310	1,555	1,463	1,253	5,581	1.9
7	의	1,204	1,612	1,123	1,280	5,219	1.8
8	가	1,156	1,526	1,188	1,148	5,018	1.7
9	어	1,152	1,354	1,409	948	4,863	1.7
10	고	1,184	1,336	1,340	948	4,808	1.6
	:	:	:	:	:	:	
1725	
합계		72,063	83,047	76,088	60,791	291,989	100.0

가장 많이 발생하는 음절은 “이”로 10,283회 발생하여 전체 음절의 3.5%를 차지한다. 1위인 “이” 음절을 포함한 10개 음절(이,다,는,을,하,에,의,가,어,고)의 발생회수는 63,841회로 전체 교과서의 21.9%를 차지한다. 11~20위 사이의 음절(지,기,자,해,나,를,보,은,아,한)을 포함하면 20개의 음절이 33.9%를 차지한다.

109,393개의 언어휘는 291,989개의 음절로 구성되고, 291,989개의 음절은 1,725개 개별음절로 구성되므로 1,725개의 개별음절은 평균적으로 169.3회 출현한다[표

3-10].

표 3-10. 개별음절수 및 음절 출현회수

구분	개별 음절수	비율(%)	출현횟수	비율(%)	평균출현횟수
한글	1,350	78.3	286,801	98.22	212.4
한글자소	40	2.3	476	0.16	11.9
한자	246	14.3	405	0.14	1.6
영어대문자	22	1.3	117	0.04	5.3
영어소문자	22	1.3	253	0.09	11.5
숫자	10	0.6	3,251	1.11	325.1
기호	13	0.8	207	0.07	15.9
특수문자	22	1.3	479	0.16	21.8
합계	1,725	100.0	291,989	100.00	169.3

1,725개의 개별 음절 중 한글이 1,350개이며 한자가 246개이다. 1,350개의 한글 개별음절은 286,801회 나타나므로 평균적으로 212.4회 나타난다.

291,989개의 전체 음절 중 한글 음절 개수가 286,801개 이므로 98.22%가 한글로 구성되어 있다. 숫자음절은 3,251회 나타나 1.11%를 차지한다.

6. 영어 어휘 및 음절

영어 관련 어휘들은 주로 시사, 과학 관련 내용을 다루는 부분에서 주로 나타난다. 4종 교과서에 나타나는 영어 어휘수는 140개로, 각 교과서마다 평균 35개의 영어 어휘가 나타난다[표 3-11].

표 3-11. 영어 어휘수 및 음절수

구분	K_1	K_2	N_1	N_2	합	평균
어휘수	18	72	34	16	140	35.00
음절수	79	184	127	34	424	106.00
평균길이	4.39	2.56	3.74	2.13	3.03	3.03

가장 많은 영어 어절을 포함하는 교과서는 K_2 교과서로 72회이고 가장 적은 영어 어휘를 포함하는 교과서는 N_2 교과서로 16회이다. K_2 교과서가 평균보다 큰 차이를 나타내는 요인을 다양한 시나리오 장면을 나타내는 기호로 “S#”이 32회나 나타나기 때문이다.

140개의 어휘를 구성하는 음절수는 424개로 어휘당 평균 길이는 3.03이다. 각 교과에 나타나는 영어 개별어

휘들의 합은 75개로 평균적으로 18.5개의 영어 개별어휘들이 나타난다. 4종교과서 전체를 대상으로 한 영어 개별어휘수는 73개로 하나의 영어 어휘만 3종의 교과서에서 공통으로 나타나고 나머지 72개의 개별어휘는 각 교과서에서 분산되어 한 교과서에서만 나타났다.

7. 한자 어휘 및 음절

우리가 사용하는 어휘는 한자어휘와 밀접한 관계가 있다. [12]에 따르면 국어대사전[10]의 표제어 440,594개 중 303,951개의 표제어에서 한자가 출현하여 68.99%의 표제어에 한자가 포함되어 있다. 한글 전용운동으로 대부분의 영역에서 한자 사용이 자취를 감추다시피 하였다가 한자 혼용을 주장하는 목소리가 커지고 있다. 이를 반영하듯 국어교과서의 어휘학습 단원에서 한자가 나타난다.

4종 교과서에 나타나는 한자 어휘수는 203개로 각 교과서에 평균적으로 50.8개의 한자 어휘가 나타난다. 한자 어휘가 가장 많이 나타나는 교과서는 N_2 교과서로 89개이며 가장 적은 한자 어휘가 나타나는 교과서는 K_2 교과서로 25개이다. 203개의 한자 어휘에 나타나는 한자 음절수는 405개로 교과서마다 평균적 101.3개의 한자가 출현한다. 가장 많은 한자 음절이 나타나는 교과서는 N_2 교과서로 201글자이며 가장 적은 한자 음절이 출현하는 교과서는 K_1 교과서이다. 한자 어휘수 및 한자 음절수는 [표 3-12]와 같다.

표 3-12. 한자 어휘수 및 음절수

구분	K_1	K_2	N_1	N_2	합계	평균
어휘수	33	25	56	89	203	50.8
음절수	41	47	116	201	405	101.3
평균길이	1.2	1.9	2.1	2.3	2.0	2.0

4종 교과서에 나타나는 203개의 한자 어휘들에 대한 각 교과서별 한자 개별어휘수 합은 192개로 평균 48개의 개별어휘가 나타난다. 4종 교과서 전체를 대상으로 한 203개의 한자 어휘들에 대한 개별어휘수가 174개이므로 개별어휘들이 평균적으로 1.2회 나타난다. 각 교과서별 한자 개별어휘수는 [표 3-13]과 같다.

표 3-13. 각 교과서별 한자 개별어휘

구분	K_1	K_2	N_1	N_2	합계	평균	전체
개별어휘수	29	24	56	83	192	48.0	174
출현어휘수	33	25	56	89	203	50.8	203
평균출현회수	1.1	1.0	1.0	1.1	1.1	1.1	1.2

8. 대표 문자 및 어휘

교과서의 내용을 기술하는데 있어 특정 지역, 명칭, 숫자를 사용하지 않고 대표문자(기호)를 사용한 어휘(예: 00명)에 대하여 조사하였다. 대표문자를 사용한 어휘수는 60개로 평균 15개의 대표어휘가 나타난다.

대표어휘가 가장 많이 나타나는 교과서는 K_2로 23개이며, 가장 적은 대표어휘가 나타나는 교과서는 N_1로 8개이다. 각 교과서에 나타나는 개별 대표어휘수 합은 30어절이며 전체 교과서를 대상으로 한 개별대표어절은 27개이다. 대표어절에 사용되는 문자는 전각문자 2글자(○,×)와 숫자 1글자(0)이다. 특정 장소, 건물명 사람명을 대표하여 사용하는 대표문자는 “○” 과 “×”을 사용하고 숫자의 자리수를 대표하여 사용하는 문자는 숫자 “0”을 사용하였다.

표 3-14. 대표 어휘수 및 대표 개별어절수

구분	K_1	K_2	N_1	N_2	합계	평균	전체
출현어절수	17	23	8	12	60	15	60
개별어절수	7	10	4	9	30	7.5	27

IV. 결론

본 논문은 초등교육과정을 마무리하고 중등교육과정으로 진입하는 첫째인 7학년 국어 교과서에 포함된 어휘들에 대한 통계를 조사하였다. 국어교과서에 수록된 어휘들에 대한 보다 객관적인 조사를 위하여 2개의 출판사에서 발행한 7학년 검인정 국어교과서 국어1 국어2 4종을 기준으로 하여 조사하였다. 조사결과 다음과 같은 결과를 얻었다.

4종의 교과서에 수록된 연어휘수는 109,393개로 각 교과서에 수록된 평균 연어휘수는 27,348개이며, 교과서 내의 장별 평균 연어휘수는 5,758개이다. 개별어휘수

는 각각의 교과서를 독립된 기준으로 했을 때의 개별어휘수 합은 39,681개로 평균적으로 9,920개이다. 가장 많은 개별어휘수를 가지는 교과서는 K_2 교과서로 10,931개이며 가장 적은 개별 어휘수를 가지는 교과서는 N_2 교과서로 8,676개이다. 4종 교과서를 하나로 통합했을 때의 개별어휘수는 28,356개이다. 28,356개의 개별어휘 중 4종 교과서에 공통으로 나타나는 개별어휘는 1,386개(4.9%)이며 나머지 26,972개(95.1%)의 개별어휘는 4종 교과서에 분산되어 나타난다. 그러나 1,395개(4.9%)의 개별어휘들의 발생 빈도는 109,393개 중 50,154개로 45.8%나 된다. 28,356개의 개별어휘 중 가장 많이 발생하는 어절은 “보자”로 교과서별로 279회~352회 발생한다. 28,356개의 개별어휘 중 단 1번 나타나는 개별어휘는 17,308개로 전체 개별어휘의 61.03%이다. 개별어휘에서 많이 발생하는 어휘 또는 4종 교과서에서 공통으로 발생하는 어휘들은 일상생활 및 교육과정에서 중요한 어휘로 볼 수 있다.

연어휘수 109,393개에 대한 전체 음절수는 291,989개로 어휘당 평균음절수는 2.67이다. 2음절 또는 3음절로 구성된 어절이 전체 어절의 68.32%이다. 가장 많은 음절을 가진 어절은 순수 한글 어절인 경우 8음절이며, 영단어인 어절은 9~11음절이다.

109,393개의 어절을 구성하는 291,989개의 음절들에 대한 개별음절은 1,725개이며 이 중 한글 1,350개, 한자 246개, 영어알파벳 대소문자 44개, 숫자 10개, 한글 불완전 자소, 기호, 특수문자이다. 1,350개의 한글 개별음절은 초등학교 6학년 국어교과서의 한글 개별음절 1,290개[9] 비하여 60개 많으며, 국어대사전[10]에 포함된 개별음절이 2,463개[11] 임을 고려할 때 대사전에 등록된 표제어에 나타난 개별음절의 54.8%의 개별음절을 이용한 어절들을 사용하고 있다.

가장 많이 발생하는 음절은 “이” 음절로 10,283회 발생하여 전체 음절의 3.5%이다. 1위~10위까지의 상위 10개 음절(이,다,는,을,하,에,의,가,어,고)의 발생횟수는 63,841회로 전체 교과서의 21.9%이고, 11~20위 사이의 음절(지,기,자,해,나,를,보,은,아,한)을 포함하면 20개의 음절이 33.9%나 되어 교과서의 1/3이 20개의 음절로 구성되어 있다.

한자로 구성된 전체 어휘수는 203개로 평균적으로 50.8개의 한자 어휘가 나타난다. 전체 교과서를 대상으로 한 한자 개별어휘수는 192개로 평균 48개의 개별어휘가 나타난다. 한글, 한자 외에 영어음절, 기호음절 또는 특수문자 등이 나타난다. 국어 교과서이지만 한글, 한자, 숫자, 영어, 특수문자 등 다양한 언어 및 글자들이 나타난다.

4종 교과서를 하나로 통합한 개별어휘수가 28,356개이고 개별한글 글자가 1,350개 이므로 16종의 국어1과 국어2에 수록된 개별어휘수는 28,356개 이하가 될 것이고, 개별한글 글자도 1,350글자 이하가 될 것이다. 개별어휘들을 형태소단위로 분리한다면 이보다 적은 어휘수가 된다. 따라서 7학년 국어 교육과정에서 익혀야할 어휘수, 개별한글 글자 같은 통계자료와 중등교육과정에서 익혀야할 기준과 비교하여 교과서 분량 및 내용을 조절할 수 있다. 또한 개별어휘들을 형태소 단위가 아닌 개별어휘를 전자 사전에 삽입한다면 전자사전의 엔트리수는 28,356개가 되고 형태소 단위로 한다면 이보다 적어진다. 전자사전에 저장할 어휘들에 대한 발생빈도 또는 발생 확률 등 필요한 정보들을 엔트리 정보에 추가하여 저장하여 활용할 수 있다. 16종의 국어 교과서를 대상으로 한다면 언어어휘수는 추가된 과목수에 비례하여 증가하지만 개별어휘수는 과목수가 추가될수록 추가되는 개별어휘수는 크게 줄어든다.

본 연구에서는 2개의 출판사에서 발행한 국어1, 국어2 4종 교과서를 선택하여 어휘들의 통계를 조사하였다. 1종의 교과서 분석 작업에도 많은 시간과 노력이 소요되지만, 채택 가능한 교과서가 16가지이고 각각이 1, 2로 구성되어 있어 32종임을 고려한다면 보다 정확하고 객관적인 통계정보를 위해서 분석 대상 확장을 고려해 볼만하다.

- [3] 김태철 외 26, 국어2, 비상교육, 2013(3).
- [4] 노미숙 외 15, 국어1, 전체교육, 2013(3).
- [5] 노미숙 외 15, 국어2, 전체교육, 2013(3).
- [6] 임오심, *중학교 1학년 국어 교과서의 어휘 분석*, 동국대학교 교육대학원 석사학위논문, 2002.
- [7] 박주화, *중학교 국어 교과서의 어휘 분석 연구*, 한양대학교 교육대학원 석사학위논문, 2006.
- [8] 김광해, *국어 어휘론 개설*, 집문당, 1993.
- [9] 김철수, “초등학교 6학년 국어교과서의 어휘 통계조사”, *한국콘텐츠학회논문지*, Vol.12, No.5, 2012.
- [10] 국립국어연구원, *표준국어대사전*, 두산동아출판사, 1999.
- [11] 김철수, 김양범, “대용량 전자사전 구축을 위한 국어 대사전의 통계정보”, *한국콘텐츠학회논문지*, Vol.7, No.6, pp.60-68, 2007.
- [12] 김철수, “국어대사전의 표제어에 나타나는 한자 정보”, *한국콘텐츠학회논문지*, Vol.10, No.4, pp.1-9, 2010.

저 자 소 개

김 철 수(Cheol-Su Kim)

정회원



- 1987년 2월 : 전북대학교전산통계학과(이학사)
- 1989년 2월 : 전북대학교전산통계학과(이학석사)
- 1998년 8월 : 전북대학교전산통계학과(이학박사)
- 1995년 3월 ~ 현재 : 서남대학교 컴퓨터정보학과 교수
<관심분야> : 한국어처리, 정보검색, 지식표현, U-Learning

참 고 문 헌

- [1] <http://www.ncic.re.kr/nation.kri.org4.inventoryList.do>
- [2] 김태철 외 26, 국어1, 비상교육, 2013(3).