

# 초등학교 6학년 국어교과서의 어휘 통계조사

## Statistical Survey of Vocabulary in Korean Textbook for Elementary School 6th-Grade

김종영, 김철수  
서남대학교 컴퓨터정보통신학과

Jong-Young Kim(namkim11@hanmail.net), Cheol-su Kim(21chskim@hanmail.net)

### 요약

본 연구는 초등학교 6학년 국어교과서 4종(6-1 읽기, 6-1 말하기·듣기·쓰기, 6-2 읽기, 6-2 말하기·듣기·쓰기)에 나타나는 어휘들에 대한 통계(전체 음절수, 음절종류, 음절 출현빈도, 어절 개수, 어절 종류, 어절 평균길이, 어절 출현빈도, 품사 등)를 조사하였다. 한글 음절수는 194,683개, 음절종류는 1,290개, 평균 음절 출현빈도는 150.9회이다. 어절 개수는 70,185개, 어절 종류는 22,647개, 어절 평균 출현빈도는 3.1회이다. 평균 음절 길이는 2.8음절이며, 가장 긴 어절은 10음절이다. 품사는 읽기 교과는 명사가 말하기·듣기·쓰기교과는 동사가 약간 많다.

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

### Abstract

This paper studied the statistics such as the total number of syllables, the kinds of syllables, the frequency of syllables, the number of eojeols(word phrases unique in Korean language), the kinds of eojeols, average length of eojeols, the frequency of eojeols and the parts of speech in four different Korean textbooks for 6th-grade students(6-1 Korean Reading, 6-1 Korean Speaking · Listening · Writing, 6-2 Korean Reading and 6-2 Korean Speaking · Listening · Writing). The results of the statistical survey are as follows: the number of Hangeul syllables was 194,683; the kinds of syllables were 1,290; the average frequency of syllables was 150.9; the number of eojeol was 70,185; the kinds of eojeol were 22,647; the average frequency of eojeol was 3.1; the average length of eojeols was 2.8 syllables, the longest one consist of 10 syllables. In parts of speech, nouns are used more in the Korean Reading textbook, and verbs are used more in Korean Speaking · Listening · Writing.

■ keyword : | Vocabulary | Statistical-survey | Syllable | Eojeol | Parts of Speech | Frequency |

## 1. 서론

국어교과서는 국어 교육의 가장 중요한 기본 자료이자 대표적인 자료로써, 정규 교육과정을 이수한 모든

사람들은 국어교과의 교육과정을 거친다. 국어교과서는 국어교육의 집대성이라 할 수 있으며, 체계적인 국어교육을 위한 어휘 선택은 국어교과서 편찬에 중요한 요소가 된다. 특히, 초등 국어교과서에 수록된 어휘는

의도적인 국어교육의 기준이며 첫 단계라는 점에서 매우 중요하다.

초·중등교육법 제38조에 “초등학교는 국민생활에 필요한 기초적인 초등교육을 하는 것을 목적으로 한다.”고 명시되어 있다. 초등 교과서에 수록된 대부분의 어휘들은 일상생활에 꼭 필요한 기본 어휘들로 국어 생활과 밀접한 관계가 있는 기초적이고 중요한 의미를 가지는 필수 어휘들의 집합체이다. 국어교과서의 어휘들에 대한 분석 자료는 새로운 국어교과서 편찬의 중요한 기초 자료가 된다. 특히 중등교육 과정으로 전환되기 직전인 초등교육 과정의 마지막 단계인 6학년 국어교과서에 수록된 어휘들의 조사는 또 다른 의미를 가진다. 교과서 수록된 어휘들에 대한 연구들은 주로 국어학회 및 한글학회를 중심으로 이루어지고 있으며 IT 분야의 도움으로 우리말 시소러스 등이 개발되어 사용되는 등 많은 발전이 있었다[1][2].

본 연구는 언어생활의 기초 단계가 정립되는 초등학교 6학년 국어교과서를 기준으로 하였으며 7차 교육과정 6학년 1학기의 '읽기'[3], '말하기·듣기·쓰기'[4]와 6학년 2학기의 '읽기'[5], '말하기·듣기·쓰기'[6] 4종 교과서에 수록된 전체 내용들에 대한 음절 개수, 음절 빈도수, 음절 종류, 어절 개수, 어절 종류, 어절 빈도수, 어절 길이, 어절에 대한 품사 빈도수 등에 대하여 조사하였다.

## 2. 관련 연구

국어교과서의 어휘들에 관한 연구는 서정국[7]을 비롯하여 80년대에 들어오면서 어휘력을 향상시키고자 하는 많은 노력들이 활발히 진행되어 왔다.

김광해[8]는 그 동안의 연구를 통합해 어휘 분석과 관련된 용어들을 정의하였고, 현재까지 진행되어 온 국어 어휘와 관련한 많은 연구가 김광해의 정의를 활용하고 있다. 김광해는 “조사대상이 되는 자료에 출현하는 모든 조사 단위의 어휘 수”를 언어휘(N)라 정의하였고, “중복되어 나타난 것을 헤아리지 않은 개별적인 어휘 수”를 개별어휘(V)로 정의하였다. 빈도수(P)는 “하나의 조사 대상 단위가 등장한 횟수”로 정의하고 있다. 연령별 어휘량에서 13세(6학년)의 어휘량은 남자는 28,809

개, 여자는 31,035개로 기술하고 있다.

진태경[9]은 제6차 교육과정의 6학년 국어교과서의 언어휘 개수는 65,825개, 개별어휘 개수는 36,073개로 기술하고 있다.

조남호 연구[10]에 따르면 현대 국어에서 사용되는 단어 빈도 조사에서 1회 이상 출현하는 단어는 약 59,000개이다. 이 가운데 출현빈도에 따른 단어 순위에서 1~15순위 사이의 단어 빈도수 합이 10,352회이다. 59,000개의 단어 중 일반인들이 사용하는 단어 개수는 1만 안팎이며, 이 가운데 절반 정도를 초등학교 6학년 과정에서 습득하는 것으로 나타났다.

본 연구의 분석 대상인 제7차 교육과정의 초등학교 6학년 교과서는 2002년도부터 사용되었으며 2004년, 2005년, 2006년 3회에 걸쳐 부분 수정이 있었고, 2009년에는 4차 수정된 교과서가 사용되었다. 전면적으로 수정된 2007 개정교육과정이 2009년에 초등학교 1~2학년을 대상으로 시행되었고 2010년에는 초등학교 3~4학년을 대상으로 시행되었다.

제6차 교육과정의 초등학교 국어교과서를 대상으로 연구[11]가 있으며, 2학년 국어 읽기 교과서를 대상으로 한 연구[12]가 있다.

초등학교 국어교과서를 대상으로 한 어휘 분석 관련 연구들은 주로 어휘들의 출현빈도, 품사별 구성, 어원 분류 등에 관한 연구가 주를 이루며 어절을 구성하는 음절에 관한 분석은 미미하다.

## 3. 어휘 분석

### 3.1 분석 기준

어휘 분석 결과는 분석 범위 및 분석 기준에 따라 달라진다. 본 연구의 어휘 분석 기준은 다음과 같다.

- ① 음절 분석은 한글로 표기된 음절을 대상으로 하되, 숫자의 경우 아라비아 숫자 1개를 1음절로 한다. 특수문자는 제외하였다.
- ② 어휘(어절)는 띄어 쓴 문자열을 하나의 어휘로 함을 원칙으로 한다. 성명은 띄어쓰기 되어 있어도 붙여서 하나의 어휘로 한다.
- ③ 품사는 국어 문법에서 보편적으로 사용하는 9품사 체계를 따랐으며 고유명사는 명사에 포함시켰다.

하나의 어휘는 1개의 대표 품사를 가지는 것으로 하였으며 조사는 명사나 명사 상당어구와 결합하는 특성이 있어 명사와 결합된 조사는 제외하였으며 독립된 어절이 조사인 경우만을 조사하였다.

- ④ 한글과 외국어, 한글과 특수문자, 한글과 숫자가 혼합된 어휘는 하나의 어절 및 품사로 한다.
- ⑤ 접사가 붙은 파생어는 별개의 어휘로 한다.
- ⑥ 사투리와 외국어도 독립된 단위로 한다.
- ⑦ 본용언과 분리된 어절의 보조용언은 별개의 단위로 한다. 단 본용언과 보조용언이 합하여져 굳어진 것은 하나의 단위로 한다.
- ⑧ 각 교과서에 대한 분석 범위는 교과 내용이 실제 시작되는 쪽에서 종료되는 쪽까지 하였다.

### 3.2 분석 방법

각 교과서에 수록된 어휘들에 대한 음절 관련 통계 정보, 어절 관련 통계 정보, 품사 관련 통계 정보를 조사하기 위해서는 연속된 텍스트 데이터를 어절 및 음절로 분리하는 과정이 선행되어야 한다. 본 논문에서는 연속된 텍스트 데이터를 각각의 어절로 분리하거나 음절 개수 및 개별음절 개수를 자동으로 조사하기 위하여 C 언어를 이용하여 자동화 프로그램을 제작하였다. 분리된 어절들에 대한 품사 분류는 수작업으로 하였으며 3차례의 확인 과정을 거쳤다.

### 3.3 음절 분석

6학년 국어 교과서에 출현하는 어휘들의 숫자를 포함한 전체 음절수는 196,838개, 읽기 교과에 출현하는 음절수는 150,152개, 말하기·듣기·쓰기 교과에 출현하는 음절수는 46,686개로 교과별 출현 음절수는 [표 1]과 같다.

표 1. 출현 음절수

교과구분	학기	음절수	소계	계
읽기	1	78,093	150,152	196,838
	2	72,059		
말하기·듣기·쓰기	1	22,621	46,686	
	2	24,065		

각 교과에 출현하는 음절수를 비교해 보면, 읽기 교

과는 1학기 교과가 2학기 교과보다 6,000여 개 많지만, 말하기·듣기·쓰기 교과는 1학기 교과가 2학기 교과보다 1,400여 개 적다. 이는 읽기 교과는 1학기 교과 내용이 많고, 말하기·듣기·쓰기 교과는 2학기 교과 내용이 많음을 의미한다.

실제 내용이 담겨 있는 각 교과와 전체 쪽 수는 751쪽이고, 출현하는 전체 음절수는 196,838개이므로 쪽 당 출현하는 평균 음절수는 262개이다[표 2].

표 2. 쪽 당 평균 음절수

교과구분	학기	음절수	쪽수	쪽당 평균음절수
읽기	1	78,093	236	330.9
	2	72,059	220	327.5
말하기·듣기·쓰기	1	22,621	153	147.8
	2	24,065	142	169.5
합계		196,838	751	262.1

읽기 교과에 출현하는 쪽 당 평균 음절수는 1학기 교과가 2학기 교과보다 3.4음절 많다. 그러나 말하기·듣기·쓰기 교과는 2학기 교과가 1학기 교과보다 쪽 당 11.7음절이 더 많다. 196,838개의 음절 개수는 2,155개의 숫자 음절을 포함한 것으로 각 교과에 출현하는 숫자 음절수는 [표 3]과 같다.

표 3. 숫자 음절수

교과구분	학기	음절수	소계	계
읽기	1	533	1,359	2,155
	2	826		
말하기·듣기·쓰기	1	422	796	
	2	374		

196,838개의 전체 음절에서 2,155개의 숫자 음절을 제외한 194,683개의 한글 음절들이 각 교과에 출현하는 한글 음절수는 [표 4]와 같다.

표 4. 한글 음절수

교과구분	학기	음절수	소계	계
읽기	1	77,560	148,793	194,683
	2	71,233		
말하기·듣기·쓰기	1	22,199	45,890	
	2	23,691		

6학년 국어교과에 출현하는 194,683개의 음절들에 대한 음절종류(개별음절)는 1,290개이다. 읽기 교과에 출



표 7. 음절 출현순위 (말뜨쓰: 말하기·듣기·쓰기)

순 위	전체교과		6-1 읽기		6-2 읽기		6-1 말뜨쓰		6-2 말뜨쓰	
	음절	빈도	음절	빈도	음절	빈도	음절	빈도	음절	빈도
1	이	7,106	이	2,957	이	2,689	다	867	다	839
2	다	6,947	다	2,804	다	2,437	이	705	이	755
3	는	4,883	는	2,076	는	1,860	을	589	을	714
4	을	4,815	을	1,755	을	1,757	어	546	하	642
5	하	4,273	하	1,634	에	1,477	하	543	는	488
6	에	3,866	에	1,441	하	1,454	에	472	에	476
7	고	3,526	고	1,435	고	1,235	는	459	어	455
8	가	3,331	가	1,388	가	1,221	고	415	고	441
9	어	3,052	지	1,204	지	1,118	니	385	시	412
10	지	2,941	나	1,138	의	993	시	374	가	409
11	나	2,712	어	1,133	니	990	의	369	나	348
12	니	2,707	니	1,033	나	955	가	313	지	325
13	의	2,577	은	1,010	어	918	지	294	아	307
14	은	2,400	아	982	은	914	여	286	니	299
15	아	2,356	의	930	서	845	를	273	여	287
16	서	2,194	서	887	기	841	나	271	의	285
17	를	2,183	리	867	아	833	불	265	한	285
18	리	2,167	그	862	를	788	리	257	를	271
19	시	2,094	를	851	리	774	한	252	리	269
20	기	2,077	기	790	그	716	로	236	불	269

를 비교해보면 20위까지 비슷하였다. 20위 안에 존재하지 않은 음절은 ‘시’(19위→31위), ‘리’(18위→22위), ‘를’(17위→21위), ‘니’(12위→45위)이다. 음절 출현순위에 따른 출현빈도와 누적빈도는 [표 8]과 같고, [표 8]에 대한 음절 출현비율과 누적비율을 그래프로 나타내면 [그림 1]과 같다.

표 8. 음절 출현순위 구간별 출현빈도와 누적빈도

출현순위	출현		누적	
	빈도	비율(%)	빈도	비율(%)
1 ~ 50	104,899	53.9	104,899	53.9
51 ~ 100	29,987	15.4	134,028	68.9
101 ~ 200	26,832	13.8	160,485	82.4
201 ~ 500	26,758	13.7	187,056	96.1
501 ~ 290	7,627	3.9	194,647	100.0

[표 8]과 [그림 1]에서 음절 출현 순위에 따른 누적 비율을 보면 50순위까지의 음절이 전체 음절의 53.9%를 차지하고, 100순위까지가 68.9%, 200순위까지가 82.4%, 500순위까지가 96.1%를 차지한다. 이는 6학년 국어교과

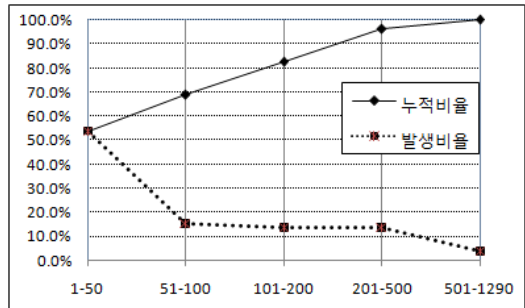


그림 1. 음절 출현순위 구간별 발생비율과 누적비율

과의 절반 정도가 50개의 음절로 구성되고, 82.4%가 200개의 음절로, 96.1%가 500개의 음절로 구성되어 있음을 의미한다. 이는 50개의 음절이 집중적으로 사용되어 50개의 음절만 알고 있어도 50%정도는 읽을 수 있으며 500개의 음절을 알고 있으면 대부분(96.1%) 읽을 수 있음을 의미한다. 이러한 음절 출현순위에 따른 누적비율은 6학년 국어교과서를 읽기 위해 습득해야 하는 음절 범위를 잘 제시하고 있다.

### 3.4 어절

어절은 하나 이상의 음절들이 결합된 띄어 쓴 단위로 6학년 국어교과서에 출현하는 어절 개수는 읽기 교과가 52,996개, 말하기·듣기·쓰기 교과가 17,189개로 전체 어절 개수는 70,185개 이다. 각 교과에 출현하는 어절 개수는 [표 9]와 같다.

표 9. 어절 개수(언어휘)

교과서 구분	학기	개별교과	단위교과	전체교과
읽기	1	27,616	52,996	70,185
	2	25,380		
말하기·듣기·쓰기	1	8,249	17,189	
	2	8,940		

어절 개수 관점에서 각 교과를 비교해 보면, 읽기 교과는 1학기 교과가 2학기 교과보다 2,200여 개 많지만, 말하기·듣기·쓰기 교과는 1학기 교과가 2학기 교과보다 700여 개 적다. 읽기 교과와 말하기·듣기·쓰기 교과의 어절 개수를 합한 학기별 어절 개수를 비교해 보면 1학기 교과가 35,865개, 2학기 교과가 34,320개로 1학기 교과가 1,500여개 많다.

70,185개의 어절들에 대한 어절 다양성을 확인하기 위해 어절 종류(개별어절)를 조사하였다. 70,185개의 어절에 대한 전체 교과서의 개별어절 개수는 22,657개, 읽기 교과서의 개별어절 개수는 19,531개, 말하기·듣기·쓰기 교과서의 개별어절 개수는 6,148개다[표 10].

표 10. 개별어절 개수

교과서 구분	학기	개별교과	단위교과	전체교과
읽기	1	11,731	19,531	22,657
	2	11,005		
말하기·듣기·쓰기	1	3,418	6,148	
	2	3,639		

교과별 개별어절 개수를 비교해 보면, 읽기 교과서는 1학기 교과서가 2학기 교과서보다 700여 개 많지만 말하기·듣기·쓰기 교과서는 1학기 교과서가 2학기 교과서보다 200여 개 적다. 이는 읽기 교과서는 1학기 교과서가 2학기 교과서보다 다양하지만 말하기·듣기·쓰기 교과서는 2학기 교과서가 1학기 교과서보다 다양함을 의미한다.

[표 10]을 통계학 관점에서 분석해 보면 1학기과 2학기의 읽기 교과서에 공통으로 출현하는 개별어절은 3,205개이고, 말하기·듣기·쓰기 교과서에 공통으로 출현하는 개별어절은 909개이다. 6학년 읽기 교과서와 말하기·듣기·쓰기 교과서에서 공통으로 출현하는 개별어절은 3,022개이다. 말하기·듣기·쓰기 교과서에 출현하는 개별어절 개수가 6,148개임을 고려할 때 49%(3,022개)의 개별어절은 읽기 교과서에서도 출현하지만 51%(3,126개)의 개별어절은 말하기·듣기·쓰기 교과서에만 출현함을 의미한다. 그렇지만, 읽기 교과서에 출현하는 19,531개의 개별어절 중 말하기·듣기·쓰기 교과서와 공통으로 출현하는 3,022개의 개별어절을 제외한 16,329개의 어절은 읽기 교과서에만 출현하므로 84.4%의 개별어절은 읽기 교과서에만 출현하고 15.6%(3,022개)의 개별어절은 말하기·듣기·쓰기 교과서와 공통으로 나타난다. 이는 읽기 교과서와 말하기·듣기·쓰기 교과서가 상호보완적인 관계임을 보여주고 있다.

각 교과서에 출현하는 음절수, 어절 개수, 개별어절 개수를 종합해 보면, 읽기 교과서는 1학기에 더 치중하고, 말하기·듣기·쓰기 교과서는 2학기 교과서에 더 치중되었을

을 알 수 있다.

[표 9]와 [표 10]을 종합해보면 6학년 국어 교과에 22,657개의 개별어절이 70,185회 출현하여 개별어절들의 평균 출현빈도는 3.1회이다. 각 개별 교과서를 독립적으로 고려하면 2.4회, 읽기 교과서만을 고려하면 2.7회, 말하기·듣기·쓰기교과서를 고려하면 2.8회, 6학년 국어 교과 전체를 대상으로 하면 3.1회 출현한다[표 11].

표 11. 개별어절 평균 출현빈도

교과서 구분	학기	개별교과	단위교과	전체교과
읽기	1	2.4	2.7	3.1
	2	2.3		
말하기듣기쓰기	1	2.4	2.8	
	2	2.5		

22,657개의 개별어절들에 대한 출현빈도를 조사하였다. 최고빈도를 가지는 어절은 ‘봄시다’(847회), 2위는 ‘수’(461회), 3위는 ‘그’(438회)이다. 출현빈도에 따른 1위에서 20위까지 개별어절들은 [표 12]와 같다.

표 12. 개별어절 출현빈도 순위

순위	전체교과		읽기		말하기·듣기·쓰기	
	어절	빈도	어절	빈도	어절	빈도
1	봄시다	847	수	371	봄시다	483
2	수	461	그	368	한	150
3	그	438	봄시다	364	내용을	110
4	한	380	있는	267	써	103
5	있는	320	한	230	바꾸어	95
6	우리	283	우리	225	수	90
7	이	243	이	188	어떤	81
8	어떤	217	나는	173	잘	80
9	그리고	209	더	144	대하여	77
10	내용을	207	그리고	141	내	75
11	잘	202	내가	141	글을	73
12	나는	199	어떤	136	어떻게	71
13	더	199	있다	130	그	70
14	내가	187	것은	123	그리고	68
15	읽고	173	잘	122	읽고	62
16	있다	163	있습니다	118	말하여	62
17	내	162	하고	112	생각이나	60
18	글을	161	것이	112	우리	58
19	를	160	있었다	112	를	57
20	대하여	154	읽고	111	다른	57

출현 순위 20위 안에 존재하는 대부분의 어절들은 읽기 교과와 말하기·듣기·쓰기 교과 모두에서 높은 순위를 나타내고 있다. 전체 교과에서 출현순위 20위 안에 존재하는 어절 중 읽기 교과에서 20위 안에 포함되지 않은 어절들은 {내용을, 내, 글을, 를, 대하여} 5개이고, 포함되지 않은 5개의 어절 대신 {것은, 있습니다, 것이, 있었다, 하고} 5개의 어절이 추가되었다. 또한 전체 교과의 출현순위 20위 안에 존재하는 어절 중 말하기·듣기·쓰기 교과의 20위 안에 포함되지 않은 어절들은 {있는, 이, 나는, 더, 내가, 있다} 6개의 어절이고, 포함되지 않은 6개의 어절 대신 {써, 바꾸어, 어떻게, 말하여, 생각이나, 다른} 6개의 어절이 포함되었다.

전체 교과의 출현순위 20위 안에 존재하는 어절 중 읽기 교과와 말하기·듣기·쓰기 교과의 어절들을 비교해 보면 9개의 어절은 공통으로 존재하지만 11개의 어절은 읽기 교과에는 있으나 말하기·듣기·쓰기 교과에는 존재하지 않으며, 반대로 11개의 어절은 말하기·듣기·쓰기 교과에는 존재하지만 읽기 교과에는 존재하지 않는다.

출현 순위 50위까지 어절 범위를 확장하면 38개의 어절은 공통으로 존재하지만 12개의 어절은 상호 배타적으로 존재한다. 이는 읽기 교과와 말하기·듣기·쓰기 교과가 상호 보완적인 역할을 하고 있음을 잘 보여주고 있다. 개별어절들의 출현빈도에 따른 출현 순위를 그래프로 나타내면 [그림 2]와 같다.

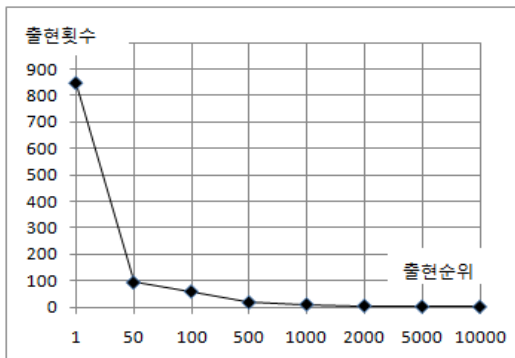


그림 2. 출현빈도에 따른 개별어절 순위

출현빈도에 따른 개별어절들의 출현 순위에서 100순

위까지는 60회 이상 출현하고 1,100순위까지는 10회 이상 출현하지만 2,000순위 이후는 5회 이하로 출현빈도가 감소하며 8,170순위 이상의 어절은 단 1번씩 출현한다. 이는 일부 어절이 집중적으로 사용되고 있기 때문이다. 즉 전체 어절 중 1,100여 개의 개별어절이 10회 이상 자주 출현하고 나머지 1,300여 개의 개별어절이 5회 이상 출현하며, 나머지 2,300여 개의 개별어절이 3회 이상 출현하며, 나머지 3,500여 개의 개별어절은 2회 정도 출현하고 나머지 14,400여 개의 개별어절은 1회 출현한다. 출현빈도에 따른 개별어절들의 순위 구간별 출현비율 및 누적비율은 [표 13]과 같다.

표 13. 어절 출현빈도에 의한 순위 구간별 구성 비율

순위구간	출현		누적	
	빈도	비율(%)	빈도	비율(%)
1~1,000	31,585	45.0	31,585	45.0
1,001~2,000	7,411	10.6	38,996	55.6
2,001~4,000	7,672	10.9	46,668	66.5
4,001~6,000	4,691	6.7	51,359	73.2
6,001~8,000	4,000	5.8	55,359	78.9
8,001~10,000	2,168	3.2	57,527	82.0
10,001~12,000	2,000	2.8	59,527	84.8
12,001~14,000	2,000	2.8	61,527	87.7
10,001~16,000	2,000	2.8	63,527	90.5
16,001~18,000	2,000	2.8	65,527	93.4
18,001~20,000	2,000	2.8	67,527	96.2
20,001~22,658	2,658	3.8	70,185	100.0
합계	70,185	100.0	70,185	100.0

[표 13]에서 보는 것처럼 1,000개의 개별어절이 6학년 국어교과서 전체의 45%를 차지하고, 2,000개의 개별어절이 55.6%를 차지한다. 2회 이상 출현하는 8,200여 개의 개별어절은 국어 교과의 79.4%를 차지하며 나머지 14,500여 개의 개별어절은 단 한 번씩 출현하여 20.6%를 차지한다. 개별어절들의 출현빈도에 의한 순위에서 구간별 발생비율과 누적비율을 그래프로 나타내면 [그림 3]과 같다.

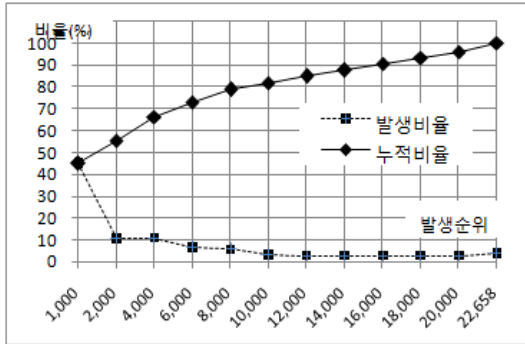


그림 3. 어절 출현 순위 구간별 발생비율과 누적비율

[그림 3]에서 보는 것처럼 8,000순위까지는 출현비율의 변화가 있지만 8,000순위 이상의 어절 출현비율은 일정함을 볼 수 있다. 이는 8,000순위 이상의 어절들은 1번씩 출현하기 때문이다.

6학년 국어 교과에 출현하는 70,185개 어절들에 대한 어절 길이를 조사하였다. 전체 어절 중 3음절로 구성된 어절이 23,895개로 가장 많고, 2음절로 구성된 어절이 21,898개로 2음절과 3음절로 구성된 어절이 전체 어절의 65.3%로 절반 이상을 차지한다. 교과별 어절 길이 구성은 [표 14]와 같다.

표 14. 어절 길이 구성

어절 길이	읽기		말하기 듣기 쓰기				전체			
	1학기		2학기		1학기		2학기		빈도	비율
	빈도	비율	빈도	비율	빈도	비율	빈도	비율		
1	2,897	10.5	2,438	9.6	1,131	13.7	1,229	13.7	7,695	10.96
2	8,595	31.1	8,131	32.0	2,464	29.9	2,708	30.3	21,898	31.20
3	9,433	34.2	8,554	33.7	2,739	33.2	3,169	35.4	23,895	34.05
4	4,473	16.2	4,323	17.0	1,364	16.5	1,349	15.1	11,509	16.40
5	1,647	6.0	1,410	5.6	451	5.5	376	4.2	3,884	5.53
6	442	1.6	353	1.4	70	0.8	83	0.9	948	1.35
7	107	0.4	139	0.5	27	0.3	24	0.3	297	0.42
8	19	0.1	27	0.1	2	0.0	2	0.0	50	0.07
9	3	0.0	3	0.0	1	0.0	0	0.0	7	0.01
10	0	0.0	2	0.0	0	0.0	0	0.0	2	0.00
합계	27616	100.0	25380	100.0	8249	100.0	8940	100.0	70185	100.00

가장 긴 어절은 2학기 읽기 교과의 10음절로 된 ‘북방 거꾸로여덟팔나비’이다. 두 번째로 긴 어절은 9음절로 된 ‘멀티미디어방이었다’(1학기 읽기)와 ‘안타까워하였습니다’(1학기 말하기·듣기·쓰기)이다. 읽기 교과와 말하기·듣기·쓰기 교과에 나타난 어절 길이 구성 비율을 그래프로 나타내면 [그림 4]와 같다.

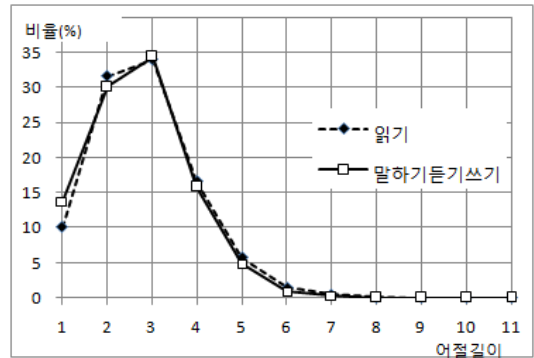


그림 4. 어절 길이 구성 비율

[표 14]와 [그림 4]에서 보는 것처럼 어절 길이별 출현 순위는 3음절, 2음절, 4음절, 1음절, 5음절 순서이며 6음절 이상은 1.8% 이다. 각 교과의 음절 길이 구성 비율은 거의 비슷하였다.

### 3.5 품사

본 절에서는 6학년 국어 교과에 출현하는 어절들에 대한 품사를 조사하였다. 품사 분류체계는 1963년에 공포된 9품사 체계를 따랐다. 9품사는 대명사, 명사, 수사, 관형사, 형용사, 동사, 부사, 감탄사, 조사이다. 조사는 명사와 결합하여 출현하는 경향이 있어 많이 출현한다. 어절에 대한 품사 분류에서 하나의 어절은 하나의 대표 품사만을 고려하였으며 공용과 활용에 의해 2개 이상의 품사가 결합된 경우 첫 번째 품사만을 고려하였다. 이에 따라 조사의 경우 독립적으로 분리된 경우와 인용 부호에 결합하여 나타난 경우만을 고려한 까닭에 발생 빈도가 낮게 나타났다. 고유명사는 명사에 포함하였다. 각 교과의 품사 출현빈도 및 구성 비율은 [표 15]와 같다.

전체 어절에 대한 품사별 구성 비율은 명사 48.7%,



표 15. 품사별 출현빈도 및 구성비율

품사	읽기				말하기·듣기·쓰기				전체	
	1학기		2학기		1학기		2학기		빈도	비율
	빈도	비율	빈도	비율	빈도	비율	빈도	비율		
명사	12,956	46.9	12,304	48.5	3,693	44.8	3,828	42.8	32,781	46.7
대명사	979	3.5	840	3.3	237	2.9	375	4.2	2,431	3.5
수사	405	1.5	357	1.4	380	4.6	329	3.7	1,471	2.1
관형사	950	3.4	783	3.1	256	3.1	281	3.1	2,270	3.2
형용사	2,549	9.2	1,815	7.2	540	6.5	599	6.7	5,503	7.8
동사	7,345	26.6	7,292	28.7	2,572	31.2	2,896	32.4	20,105	28.6
부사	2,121	7.7	1,686	6.6	413	5.0	477	5.3	4,697	6.7
감탄사	87	0.3	91	0.4	27	0.3	26	0.3	231	0.3
조사	224	0.8	212	0.8	131	1.6	129	1.4	696	1.0
합계	27,616	100	25,380	100	8,249	100	8,940	100	70,185	100

동사 28.6%, 형용사 7.3%, 부사 6.7%, 대명사 3.5%, 관형사 3.2%, 수사 2.1%, 조사 1%, 감탄사 0.3%이다. 읽기 교과, 말하기·듣기·쓰기 교과 그리고 전체 교과의 품사별 출현비율을 그래프로 나타내면 [그림 5]와 같다.

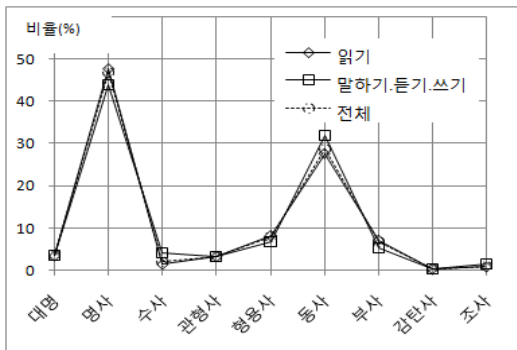


그림 5. 품사 구성 비율

읽기 교과와 말하기·듣기·쓰기 교과를 비교했을 때, 동사는 말하기·듣기·쓰기 교과가 읽기 교과보다 약 4% 높게 나타났다. 이는 교과서의 특성상 지시어가 많기 때문이다. 명사는 읽기 교과가 말하기·듣기·쓰기 교과보다 4% 높게 나타났다.

#### 4. 결론

본 연구는 제7차 교육과정에 따라 편찬된 초등학교 6학년 국어교과서 4종에 출현하는 어휘들에 대한 전체 음절 개수, 개별음절 개수, 음절들의 평균 출현횟수, 전체 어절 개수, 개별어절 개수, 어절들의 평균 출현횟수, 음절과 어절들의 출현 순위, 어절의 평균길이, 어절의 품사 등에 대하여 조사하였다.

6학년 1,2학기의 읽기 교과와 말하기·듣기·쓰기 교과에 출현하는 전체 음절 개수는 196,838개이고, 이 가운데 숫자 음절 개수는 2,155개, 한글 음절 개수는 194,683개이다. 194,683개의 한글 음절들에 대한 개별음절 개수는 1,290개이다. 초성, 중성, 종성이 결합하여 컴퓨터상에서 표현 가능한 완전 음절 개수는 11,172개이지만 85%정도의 음절은 거의 사용되지 않고 일부 음절만 집중적으로 사용된다. 국어대사전의 표제어에 실제로 출현하는 개별음절 개수가 2,463개이므로 6학년 국어 교과에 출현하는 개별음절 개수는 국어대사전의 52.4%이다. 1,290개의 개별음절은 평균적으로 150.9회 출현한다.

194,683개 음절들의 출현빈도에 따른 개별음절들의 출현 순위별 누적비율은 50순위까지가 53.9%, 100순위까지가 68.9%, 200순위까지가 82.4%, 500순위까지가 전체의 96.1%를 차지하였다. 이는 500개의 음절만 알아도 국어 교과의 96.1%를 읽을 수 있음을 의미하며 6학년 교과서를 통하여 저학년 학생들이 읽혀야 하는 어휘들에 출현하는 음절 범위 자료를 제시하고 있다. 최고 출현빈도를 나타내는 음절은 ‘이’ 음절로 7,106회이다. 교과서의 한 쪽에 출현하는 평균 음절 개수는 읽기 교과는 329음절, 말하기·듣기·쓰기 교과는 158음절, 전체 교과는 262 음절이다. 읽기 교과의 1학기, 2학기 음절수 구성 비율은 53.6 : 46.4로 1학기가 2학기보다 7.2% 높지만, 말하기·듣기·쓰기 교과의 1학기, 2학기의 음절수 구성 비율은 48.5:51.5로 2학기가 3% 높다.

전체 어절 개수는 70,185개이고 개별어절 개수는 22,657개로 개별어절들의 평균 출현빈도는 3.1회이다. 이는 6차 교육과정 국어 교과[9]의 전체 어절 개수 65,825개, 개별어절 개수 36,073보다 상당히 증가한 것이다. 이는 교과 내용이 확대되었음을 의미한다.

70,185개의 어절은 196,838개의 음절로 구성되므로 한 어절을 구성하는 평균 음절수는 2.8이다. 어절 길이 구성 비율은 3음절(34.1%), 2음절(31.2%), 4음절(16.4%), 1음절(10.9%), 5음절 5.6%, 6음절 순서이다. 가장 긴 어절은 10음절로 이루어진 어절 ‘북방거꾸로여덟팔나비’이다.

품사별 구성 비율은 명사 46.7%, 동사 28.6%, 형용사 7.8%, 부사 6.7%, 대명사 3.5%, 관형사 3.2%, 수사 2.1%, 조사 1.0%, 감탄사 0.3% 순으로 나타났다. 읽기 교과과 말하기·듣기·쓰기 교과와 품사 구성 비율을 비교하면 읽기 교과는 명사가 말하기·듣기·쓰기 교과는 동사가 상대적으로 많이 나타났다.

본 연구는 초등학교 6학년 국어교과서를 대상으로 분석하였으나 초등학교 과정에서 임해야하는 한글 음절의 범위, 사용음절에 따른 관련 어휘 등 국어교과서의 편찬에 있어 어휘 선정에 기초 자료로 활용될 수 있을 것이다.

참 고 문 헌

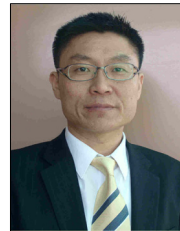
[1] 정인삼, 컴퓨터 한글자판에 관한 연구: 자모결합 빈도조사, 국립국어연구원, 1992.  
 [2] 이상섭, 송만섭, “전자 사전 구현을 위한 어휘 데이터베이스의 설계와 국어 정보 분석 도구의 개발”, 한국 통신공사 후원 연구과제, 2004.  
 [3] 한국교육과정평가원, 국어 읽기 6-1, 대한교과서 주식회사, 2010.  
 [4] 한국교육과정평가원, 국어 말하기·듣기·쓰기 6-1, 대한교과서주식회사, 2010.  
 [5] 한국교육과정평가원, 국어 읽기 6-2, 대한교과서 주식회사, 2010.  
 [6] 한국교육과정평가원, 국어 말하기·듣기·쓰기 6-2, 대한교과서주식회사, 2010.  
 [7] 서정국, 國語 基本語彙의 研究: 國民學校 國語科 教科書의 語彙調査를 中心으로, 고려대학교 석사 학위논문, 1969.  
 [8] 김광해, 국어 어휘론 개설, 집문당, 1993.

[9] 진태경, 초등학교 국어 교과서 어휘 연구-제6차 읽기교과를 중심으로, 진주교육대학교 교육대학원 석사학위논문, 2001.  
 [10] 조남호, 한국어 학습용 어휘선정 결과보고, 국립국어연구원 2003.  
 [11] 변은주, “초등학교 국어교과서 어휘 분석”, 우리말교육8집, 2003.  
 [12] 김경선, “초등 학교 2학년 국어 읽기 교과서의 어휘 조사”, 초등 국어교육, No.8, pp.151-188, 1998.  
 [13] 국립국어연구원, 표준국어대사전, 두산동아출판사, 1999.  
 [14] 김철수, 김양범, “대용량 전자사전 구축을 위한 국어 대사전의 통계정보”, 한국콘텐츠학회논문지 (B), Vol.7, No.6, pp.60-68, 2007.

저 자 소 개

김 중 영(Jong-Young Kim)

정희원



- 1993년 2월 : 전주교육대학교(학사)
- 2002년 7월 : 서남대학교 교육대학원(석사)
- 2007년 8월 : 서남대학교대학원 박사과정 수료
- 2011년 현재 : 남원주생초등학교 교사
- <관심분야> : 교육, IT

김 철 수(Cheol-Su Kim)

정희원



- 1987년 2월 전북대학교 졸업(이학사)
- 1989년 2월 전북대학교 졸업(이학석사)
- 1998년 8월 전북대학교 졸업(이학박사)
- 1995년 3월 ~ : 서남대학교 컴퓨터정보통신학과
- <관심분야> : 정보검색, 음성인식, 한국어처리