

초등학생의 구어에 나타난 어휘 빈도와 분포도 조사*

이필영** · 김정선***

< 차례 >

- I. 서론
- II. 연구 방법
- III. 범주별 어휘 사용 분포
- IV. 결론

I. 서론

본 연구의 목적은 초등학생의 구어에 나타난 어휘의 사용 빈도와 분포도를 조사하는 것이다. 이를 위해 초등학생의 실제 대화 자료를 수집·분석하여 전체 어휘 목록을 추출하고 어휘별 사용 회자수와 빈도를 조사하였다.

현재 국어 어휘에 대한 조사는 대부분 학교 교육과 외국인 대상 한국어 교육에서 교육용 어휘를 선정하기 위해 이루어지고 있다(국립국어원 2002, 임지룡 1991, 김광해 2003). 국립국어원(2002)은 한국어 학습용 어휘를 선정하기 위한 사전 작업으로 현대 국어 사용 빈도 조사하였다. 이 조사에서는 1990년대 이후의 교과서, 문학, 잡지, 신문 등을 대상으로 하여 가나

* 본 연구는 한국과학재단 특정기초연구(R01-2005-000-10840-0) 지원으로 수행되었음.

** 주저자, 한양대학교 한국어언어문학과 교수

*** 교신저자, 한양대학교 국어국문학과 강사

다순 목록별 빈도와 빈도순 목록을 단어, 조사, 어미의 범주로 나누어 제시하고 있다. 임지룡(1991)에서는 객관적 방법에 의해 선정된 고빈도어와 주관적 방법에 의한 의미 분야를 절충하여 내용어 1,500개의 기초 어휘를 선정하였다. 특히, 어휘 목록을 가나다순이나 빈도순이 아닌 의미 분야를 기준으로 제시한 점은 시소러스의 배치 방식이나 머릿속 사진의 구조를 상정할 경우에 적용될 수 있는 것으로 어휘 연구에 시사하는 바가 크다(김광해 2003). 김광해(2003)에서는 이미 이루어진 선행 연구에서 제출된 어휘 목록을 하나의 DB에 넣은 뒤 그 목록에 대한 선행 연구들의 지지도를 비교하는 방법을 통해 어휘의 등급을 정하는 메타 계량 방법을 이용하였다. 그 결과 총 237,990어휘를 교육적 중요도에 따라 7등급으로 구분한 목록을 얻었다.¹⁾

이상의 연구들은 대부분 문어를 대상으로 하여 어휘 목록을 선정하고 있다. 국립국어원(2002)은 방송 대본을 비롯한 구어를 포함한다고 하였지만 전체 대상 어절 가운데 3.2%만을 차지하고 있어 구어의 특성이 반영되었다고 보기에는 무리가 있다. 교육을 목표로 하고 있는 어휘 선정은 문어만이 아니라 실제 발화된 구어도 대상에 포함하여야 한다. 그러나 지금까지 구어에 나타난 어휘는 자료 수집의 어려움, 구어 연구의 부족 등으로 인해 어휘 선정 시 구어의 필요성이 인정됨에도 불구하고 실제 목록에는 포함되지 못하는 것이 현실이다.

구어는 문어와는 다른 특성을 보이고 있다. 구어는 음성이라는 매체를 사용하는 의사 전달의 수단으로 '상호성'과 '실시간성'의 특성을 갖으며(구현정 2005), 이러한 특성으로 인해 사용 어휘, 조사/어미 목록, 문장 구조와 길이 등에서 문어와는 차이가 있다.²⁾ 또한 맥락에 따른 언어 변이나 언어의 구조 변화, 음운·어휘 등의 변이 및 분포를 알 수 있는 것은 문어보다 구어에서임은 자명한 일이다(Milroy, 1991). 구어는 일상생활에서 현존

1) 김광해(2003)에서는 국어교육용 등급과 한국어 교육용 등급을 별도로 처리하였으며, 책에서는 지면의 제약 등으로 1~4등급의 어휘만을 보이고 있다.

2) 구어는 무계획적이고 비문법적이며 자동적인 특성을 가진다(Hughes, 2002 : 134~5 지현숙 2007 재인용).

화자들에 의해 널리 사용되고 있고, 변이형의 확산은 사회적 맥락과 상황적 맥락에서 발생하는 대화를 통해 개인에서 개인으로 옮겨가기 때문이다. 따라서 구어를 체계적으로 관찰함으로써 비로소 현실 언어에 대한 타당성 있는 분석이 가능해진다.

특히 구어에 나타나는 어휘는 어휘 교육과 언어 발달 연구에서 핵심이 되는 대상임에도 아직 국어 어휘 연구에서는 조사가 활발히 이루어지지 못하는 못하였다.³⁾ 이는 구어 말뭉치 구축이 자료 수집과 가공에 많은 인력과 비용, 시간이 필요하기 때문인 것으로 생각된다. 교육용 어휘 선정을 위해서는 교육 대상자가 사용하는 어휘에 대한 조사가 선행되어야 하고, 학령 전보다 아동기 이후의 어휘 발달이 더 빠른 속도로 지속적으로 이루어지고 있기에 이에 대한 연구가 필요하다(Anglin 1993).⁴⁾ 본 조사에서는 구어에 나타난 어휘 목록을 추출하고 화자 수 및 출현 빈도를 조사하여 어휘 사용 분포를 살펴보고, 이를 문어와 비교하여 초등학생 구어의 특징적인 점들을 밝히고자 한다.⁵⁾ 특히 균형 말뭉치를 대상으로 하기 때문에 지금까지 구어에 대해 몇몇 특징적인 현상에 제한되어 이루어진 논의들이 지닌 한계를 벗어나 초등학생 구어에 나타난 전반적인 양상을 살펴볼 수 있을 것이다.

3) 임철성(1997), 장영희(2000)는 준구어인 드라마를 대상으로 어휘를 계량적으로 조사한 바 있다.

4) 어휘 크기의 증가는 초등 시절에 급격히 이루어지는데, 이를 연구하기 위해서는 대규모 말뭉치를 바탕으로 한 어휘에 대한 계량적 조사 방법이 사용되어야 한다.

5) 국내에서 이루어지는 국어교육이나, 외국인을 대상으로 하는 한국어 교육을 막론하고 피교육자의 수준에 따라 등급별로 어휘를 제공하고 있는 것이 현실이다. 또한 교육 자료, 교과서, 평가 도구 개발 등을 위한 수준을 정하는 데 도움을 줄 수 있는 가장 객관적인 자료를 찾는다면 그것은 현재로서는 등급별 어휘목록밖에 없다. 이러한 점만 보더라도 등급화된 어휘 목록의 필요성은 쉽게 확인된다.

II. 연구 방법

1. 연구 대상

본고의 연구 대상은 서울·경기 지역에 거주하는 초등학교 1학년~6학년 학생으로 학년별로 40명씩 총 240명이며, 학년별 성별 비율은 균등하게 하였다.⁶⁾

2. 분석 자료와 절차

대화 자료는 학교의 빈 공간에서 동일 학년, 동일 성별의 아동 2명씩 짝을 지어 주제의 제한을 두지 않고 40분 동안 자유롭게 대화하도록 하고 이를 녹음하여 수집하였다.⁷⁾ 녹음 자료는 음성 파일로 디지털화한 후, 장경희 외(2005)에서 마련한 전사 지침에 따라 국어학 및 국어교육학 전공자(학부, 대학원)가 3차에 걸쳐 한글로 전사하였다.⁸⁾

-
- 6) 표본 연구에서 표본의 수가 많을수록 결과의 신뢰도는 높아진다. 그러한 현실적인 면을 고려해 볼 때 한정된 표본이라도 균형성과 대표성을 갖추어 수집한다면 이 또한 의미있는 일이라 생각된다. 이런 관점에서 볼 때 본고에서 조사한 대상자는 학년별 성별 균형성을 갖추었기 때문에 논의 진행에 큰 무리가 없다고 생각된다.
 - 7) 녹음 상황이 교실에서 친구와의 대화로 한정된 것은 자료의 일관성을 확보하기 위한 것이다. 대화의 목적, 대화 상대방과의 관계, 대화 환경 등 여러 요인에 의해 대화의 주제는 무궁무진하다. 구어 자료 수집의 어려움은 이렇듯 다양한 대화 가운데 어떤 것을 대상으로 할 것이나부터 시작된다. 본고에서 분석한 결과는 초등학생 어휘의 일부분을 보여주는 것이라 볼 수 있다. 그러나 조사 자료 전체가 지닌 동일 상황에서의 자료이기 때문에 초등학생 또래 대화에 나타난 어휘의 특징을 설명하는 데에는 기여한다고 본다.
 - 8) 1차 전사는 발화자 표시와 발화 내용을 그대로 문자화한 것이다. 2차 전사에서는 1차 전사 결과물에 발화자 정보, 녹음 시간, 날짜, 전사 시간 등에 대한 정보를 표시한 헤더를 부착하였으며, 표준어 표기, 말걸침이나 군말, 비언어적 음성, 휴지 등 발화 상황을 이해하는 데 도움이 되는 정보, 발화자의 특이한 어조 등의 정보를 표시하였다. 3차 전사에서는 2차 전사 결과물에 대한 수정 보완 및 상세 정보 태그를 부착하였으며 자료의 익명성을 보장하기 위해 마크업을 실시하였다.

전사한 한글 자료를 다음의 분석 절차를 거쳐 어휘에 대한 계량 조사를 실시하였다.

① 발화 단위 추출

전사 자료에서 화자별로 50발화 이상을 추출하여 엑셀 파일에 저장하였다. 학년별 총 발화수를 보이면 다음과 같다.

<표 1> 학년별·성별 총 발화 수

단 계		성 별		합 계
		남	여	
초등학생	1학년	1,167	1,129	2,297
	2학년	1,167	1,130	2,299
	3학년	1,181	1,121	2,305
	4학년	1,177	1,127	2,308
	5학년	1,156	1,137	2,298
	6학년	1,205	1,148	2,359
합 계		7,053	6,792	13,866

9) 전사 원칙은 다음과 같다.

- (가) 발화자의 실제 발음이나 특이한 말투를 그대로 반영하여 여기에 표준어 표기를 덧붙이도록 하여 연구 대상자의 현실적인 어휘 사용 양상이 최대한 반영되도록 한다.
- (나) 전사는 음성 파일 하나당 2,000어절까지 하였으며, 화자의 발화가 끝나지 않은 상태에서 2,000어절이 되는 경우는 해당 발화의 종결 부분까지 전사한다.
- (다) 전사에 사용된 기호 목록은 다음과 같다.

전사 기호	기호 설명
?	상승 억양
!	활기에 넘치는 기운찬 어조
,	약한 상승 또는 하강, 약간의 휴지 등
.	한 발화의 끝
((-))	잘 들리지 않는 발화
[두 사람이 말이 겹칠 경우
-	의도적인 장음
< >	사람의 음성 중 비언어적인 소리
[]	발화의 표준 형태
{ }	대화 상황에 대한 설명
“ ”	인용 표시

② 품사 태깅

품사 태깅은 21세기 세종계획의 지침에 따랐으며, 글잡이 프로그램을 이용하여 자동 태깅에 의해 기본적인 형태 정보를 부착하였고, 오류는 수작업으로 수정하였다.¹⁰⁾

③ 동음이의어 구분

본고에서는 동음이의어의 의미를 국립국어원에서 개발한 Wordanal 2.0 프로그램을 이용하여 구분하였다.

④ 이형태 분류

어휘수를 계량하기 위해서 하나의 어휘에 대해 이형태가 존재하는 경우 이를 기본 어휘와 동일한 어휘로 분류하였다.¹¹⁾¹²⁾

마지막으로 태그가 완료된 결과는 형태소 단위이기 때문에 어휘 단위로 처리하는 단계를 거쳤다.¹³⁾ 최종 결과물은 대상자별로 엑셀 파일로 정리하여 사용 분포를 조사하였다.¹⁴⁾

10) 형태소 태그 목록과 약호는 <부록>에 제시하였다. 21세기 세종계획 형태소 태그 지침에서 접사와 감탄사는 제외하였다.

11) 음운적 조건에 따라 필연적이고 자동적으로 결정되는 음운론적 이형태와 특정한 형태 조건에 따르는 형태론적 이형태에 해당되는 형태 목록을 작성하고 이를 검색 프로그램에 반영하였다. 이와 같이 자동 처리 대상이 되는 이형태 목록은 형태 추출 시 이형태를 기본 형태로 자동 전환된다.

12) 이상의 어휘 태깅 작업은 국어학 전공 박사과정생 4명이 작업하고 박사급 연구원 4명이 검토하였다.

13) 본 연구에서 접사 가운데 접두사가 빠진 이유는 형태소 목록이 아닌 어휘 목록을 추출하였기 때문이다. 접두사는 어휘를 형성하는 형태소이기 때문에 제외하였고, 본 연구에서 제시한 접미사는 어휘 형성 접미사가 아닌 문법적 접미사들이다.

14) 최종 결과물은 다음의 형식으로 정리하였다.

III. 범주별 어휘 사용 분포

1. 품사별 출현 어휘 수와 출현 빈도

본 연구에서는 240명을 대상으로 하여, 총 13,866발화에서 나타난 101,167어절을 분석하였다. 그 결과, 출현 어휘는 5,982개이었고, 출현 빈도의 총합은 101,167회이었다. 품사별 출현 어휘 수와 출현 빈도, 백분율을 보이면 다음 표와 같다.

<표 2> 품사별 출현 어휘 목록 수와 출현 빈도

품 사		목록 수	(백분율)	출현 빈도	(백분율)	
체언	명사(NN)	일반명사(NNG)	2,322	(38.8%)	13,169	(13.0%)
		고유명사(NNP)	1,207	(20.2%)	2,821	(2.8%)
		의존명사(NNB)	110	(1.8%)	3,386	(3.3%)
	대명사(NP)	86	(1.4%)	7,965	(7.9%)	
	수사(NR)	39	(0.7%)	1,194	(1.2%)	
소 계		3,764	(62.9%)	28,535	(28.2%)	
용언	동사(VV)	982	(16.4%)	12,193	(12.1%)	
	형용사(VA)	263	(4.4%)	4,846	(4.8%)	
	보조용언(VX)	38	(0.6%)	2,626	(2.6%)	
	지정사(VC)	긍정지정사(VCP)	1	(0.0%)	834	(0.8%)
		부정지정사(VCN)	2	(0.0%)	252	(0.2%)
소 계		1,286	(21.5%)	20,751	(20.5%)	
수식언	관형사(MM)	88	(1.5%)	3,216	(3.2%)	
	부사(MA)	일반부사(MAG)	370	(6.2%)	7,192	(7.1%)
		접속부사(MAJ)	38	(0.6%)	2,188	(2.2%)
소 계		496	(8.3%)	12,596	(12.5%)	

대상자코드	학 년	성 별	어 휘	품 사	출현 빈도

품 사		목록 수	(백분율)	출현 빈도	(백분율)	
소 계		496	(8.3%)	12,596	(12.5%)	
관계언	격조사(JK)	주격조사(JKS)	3	(0.1%)	3,411	(3.4%)
		보격조사(JKC)	1	(0.0%)	39	(0.0%)
		관형격조사(JKG)	2	(0.0%)	61	(0.1%)
		목적격조사(JKO)	2	(0.0%)	930	(0.9%)
		부사격조사(JKB)	29	(0.5%)	3,022	(3.0%)
		호격조사(JKV)	1	(0.0%)	76	(0.1%)
		인용격조사(JKQ)	5	(0.1%)	251	(0.2%)
	보조사(JX)	35	(0.6%)	4,265	(4.2%)	
	접속조사(JC)	6	(0.1%)	333	(0.3%)	
소 계		84	(1.4%)	12,388	(12.2%)	
의존형태	어미(E)	선어말어미(EP)	12	(0.2%)	3,609	(3.6%)
		종결어미(EF)	149	(2.5%)	11,607	(11.5%)
		연결어미(EC)	139	(2.3%)	7,653	(7.6%)
		명사형전성어미(ETN)	3	(0.1%)	187	(0.2%)
		관형형전성어미(ETM)	25	(0.4%)	2,934	(2.9%)
	접미사(XS)	명사파생접미사(XSN)	16	(0.3%)	888	(0.9%)
		동사파생접미사(XSV)	3	(0.1%)	11	(0.0%)
		형용사파생접미사(XSA)	1	(0.0%)	1	(0.0%)
	어근(XR)	4	(0.1%)	7	(0.0%)	
소 계		352	(5.9%)	26,897	(26.6%)	
총합계		5,982	(100.0%)	101,167	(100.0%)	

<표 2>를 보면 어휘 목록수와 출현 빈도에서 흥미로운 점들을 발견할 수 있다. 우선 어휘 목록은 체언(62.9%)이 가장 많았고, 그 다음이 용언(21.5%), 수식언(8.3%), 의존형태(5.9%), 관계언(1.4%)의 순으로 나타났다. 특히 체언과 용언은 전체 어휘 목록 중 84.4%를 차지하고 있지만 출현 빈도에서는 48.7%를 차지하고 있었다. 체언의 경우는 어휘 목록 수의 백분율과 출현 빈도의 백분율이 각각 62.9%와 28.2%를 차지하고 있어 큰 차이를 보이고 있었지만, 용언의 경우에는 그다지 차이를 보이지 않고 있었다.

이에 반해 수식언과 관계언, 의존형태는 어휘 목록은 15.6%이지만 출현빈도는 51.3%를 차지하고 있어 체언 및 용언과는 반대 현상을 보이고 있다.¹⁵⁾ 즉, 체언은 다양한 어휘를 사용하지만, 관계언과 의존 형태와 같이 문법적 기능을 중심으로 하는 어휘는 소수의 어휘를 빈번하게 사용하는 것을 확인할 수 있었다.

다음은 전체 어휘 목록을 사용 화자 수에 따라 정리해 보기로 한다. 단순 출현 빈도순이 아닌 전체 조사 대상자 가운데 해당 어휘를 사용한 화자가 몇 명인지를 기준으로 순위를 제시한다. 종래 기준으로 삼은 단순 출현 빈도의 합은 대화의 주제 등에 크게 영향을 받기 때문에 몇 번 해당 어휘가 나왔는지보다는 전체 대상자 중 몇 명의 화자가 사용하는지가 보다 어휘의 사용 실태를 보여줄 수 있다고 생각한다. 본고의 조사 대상자는 모두 240명으로 사용 화자 수가 많은 순으로 어휘를 정리하면 다음 <표 3>과 같다.

<표 3> 사용 화자와 출현 빈도에 따른 어휘 목록

순위	어휘	품사	화자 수	출현빈도 (빈도순위)	순위	어휘	품사	화자 수	출현빈도 (빈도순위)
1	어11	EF	240	4,510 (1)	51	자24	EC	142	279 (63)
2	가11	JKS	239	3,395 (2)	52	막02	MAG	137	597 (34)
3	어06	EC	238	1,889 (4)	53	현01	MM	137	287 (62)
4	있	EP	237	3,143 (3)	54	모르	VV	137	272 (65)
5	하01	VV	236	1,773 (6)	55	주01	VX	133	301 (60)
6	나03	NP	236	1,679 (7)	56	애02	NNG	130	310 (59)
7	는01	JX	234	1,860 (5)	57	졌	EP	130	239 (69)
8	계01	NNB	229	1,472 (8)	58	구(고24)	EC	129	474 (37)
9	예04	JKB	229	1,290 (10)	59	진짜	MAG	129	346 (54)
10	고24	EC	225	1,454 (9)	60	나05	NP	128	445 (39)

15) 이러한 문법적 범주와 어휘적 범주의 어휘 목록 수와 빈도의 출현 양상은 서상규(1998 : 244)에서도 나타났다. 최고 빈도 구간인 빈도 1만 이상의 488개의 낱말들의 품사별 낱말수와 비율 구성을 조사한 결과, 낱말 수는 일반 낱말이 370개로 문법적 형태 118개보다 3배 이상이지만 빈도수의 합은 어휘적 형태가 12,039,054이고, 문법적 형태가 18,466,782로 나타났다.

순위	어휘	품사	화자 수	출현빈도 (빈도순위)		순위	어휘	품사	화자 수	출현빈도 (빈도순위)	
11	야13	EF	224	1,003	(12)	61	그렇	VA	126	406	(44)
12	논03	ETM	223	973	(15)	62	ㄴ디01	EF	125	218	(80)
13	있01	VA	222	1,220	(11)	63	개	NP	123	368	(52)
14	이03	VCP	218	834	(21)	64	이거01	NP	123	326	(57)
15	은06	ETM	217	979	(14)	65	니가	EC	123	278	(64)
16	안02	MAG	217	763	(25)	66	갈	VA	122	263	(66)
17	ㄴ02	JX	216	678	(28)	67	이렇게	MAG	121	376	(51)
18	지25	EF	215	849	(19)	68	만4	JX	121	225	(74)
19	도15	JX	215	816	(22)	69	오01	VV	118	234	(71)
20	잖아	EF	213	960	(16)	70	맞01	VV	114	254	(67)
21	뒤	NP	212	855	(18)	71	랑05	JKB	114	232	(72)
22	너01	NP	210	912	(17)	72	게10	EC	112	222	(76)
23	르03	ETM	209	753	(26)	73	가지	VX	111	407	(43)
24	다07	EF	205	842	(20)	74	그냥	MAG	109	221	(78)
25	되01	VV	198	652	(29)	75	ㄴ데01	EC	107	170	(98)
26	근데01	MAJ	194	986	(13)	76	가36	ETN	106	181	(93)
27	그01	MM	187	811	(23)	77	어어02	EC	105	177	(95)
28	논데01	EC	187	695	(27)	78	거기01	NP	103	222	(77)
29	면09	EC	187	621	(30)	79	말01하	VV	103	191	(88)
30	를	JKO	184	806	(24)	80	한테	JKB	102	203	(84)
31	그거	NP	180	615	(32)	81	르까	EF	101	166	(102)
32	보01	VX	176	457	(38)	82	여기01	NP	99	211	(81)
33	내04	NP	174	546	(35)	83	그럼01	MAJ	99	178	(94)
34	가01	VV	172	616	(31)	84	고22	JKQ	96	190	(89)
35	에서02	JKB	172	430	(40)	85	채뒀	VA	95	198	(85)
36	알	VV	172	402	(45)	86	네08	XSN	95	185	(91)
37	냐	EF	170	546	(36)	87	거든03	EF	94	238	(70)
38	우리03	NP	166	609	(33)	88	그래서	MAJ	94	219	(79)
39	없01	VA	162	380	(49)	89	몇	MM	94	185	(92)
40	보01	VV	159	416	(42)	90	어떻게	MAG	94	153	(111)
41	로07	JKB	159	386	(48)	91	무슨	MM	94	150	(115)
42	왜02	MAG	159	334	(55)	92	하01	VX	93	166	(103)

순위	어휘	품사	화자 수	출현빈도 (빈도순위)		순위	어휘	품사	화자 수	출현빈도 (빈도순위)	
43	어서03	EC	158	390	(47)	93	그때	NNG	93	160	(107)
44	있01	VX	152	329	(56)	94	러09	EF	93	152	(112)
45	때01	NNG	151	426	(41)	95	엄마	NNG	92	232	(73)
46	들09	XSN	151	350	(53)	96	나오	VV	92	171	(96)
47	다03	MAG	148	319	(58)	97	선생님	NNG	91	288	(61)
48	아니	VCN	145	251	(68)	98	자8	EF	91	167	(100)
49	그러	VV	144	399	(46)	99	또	MAG	90	140	(124)
50	좋01	VA	142	377	(50)	100	많이	MAG	90	131	(133)

<표 3>을 보면, 전체 대상자 240명 모두가 사용하는 어휘는 종결어미 ‘어’로 나타났다. 종결어미 ‘어’는 사용 화자 수도 가장 많았지만 빈도 순위에서도 가장 고빈도로 사용된 어휘이다. 다음으로 사용 화자 수가 많은 어휘는 주격 조사 ‘가’, 연결 어미 ‘어’, 선어말 어미 ‘었’의 순으로 나타나 어미, 조사와 같이 문법적 기능을 지닌 어휘가 사용 화자 수가 많은 것으로 나타났다. 1위부터 100위까지의 어휘 중, 의존형태(어미, 접사는 모두 30개, 관계언(조사)은 12개, 체언은 17개, 용언은 24개, 수식언은 17개가 사용되어, 문법적 기능을 지닌 어휘가 총 42개를 차지하고 있었다.

2. 사용 화자 수에 따른 사용 분포

이제 범주별로 사용 화자 수에 따른 어휘 사용 분포를 살펴보기로 한다.

1) 체언

명사, 대명사, 의존명사, 수사로 이루어진 체언 범주를 사용 화자 빈도가 높은 순으로 정리해 보이면 <표 4>와 같다.¹⁶⁾

16) 고유명사는 제외하였다.

<표 4> 체언의 사용 화자 수에 따른 어휘 목록

순위	어휘	품사	화자 수	순위	어휘	품사	화자 수	순위	어휘	품사	화자 수	순위	어휘	품사	화자 수
1	나03	NP	236	26	십	NR	81	51	누02	NP	44	76	수학05	NNG	31
2	거01	NNB	229	27	애기	NNG	76	52	컴퓨터	NNG	43	77	오늘	NNG	31
3	뭐	NP	212	28	아빠	NNG	68	53	동생01	NNG	42	78	한번	NNG	30
4	너01	NP	210	29	것01	NNB	67	54	진짜	NNG	42	79	저거01	NP	30
5	그거	NP	180	30	개10	NNB	64	55	학년02	NNG	42	80	백05	NR	30
6	내04	NP	174	31	줄04	NNB	64	56	옛날	NNG	41	81	머리01	NNG	29
7	우리03	NP	166	32	반11	NNG	64	57	원래01	NNG	41	82	소리01	NNG	29
8	때01	NNG	151	33	학교	NNG	59	58	전08	NNG	39	83	일01	NNG	29
9	애02	NNG	130	34	하나	NR	59	59	저기01	NP	39	84	그림01	NNG	28
10	나05	NP	128	35	이09	NR	58	60	분08	NNB	37	85	요즘	NNG	28
11	개	NP	123	36	삼06	NR	56	61	영어02	NNG	36	86	저금03	NNG	28
12	이거01	NP	123	37	학년	NNG	55	62	책01	NNG	36	87	뒤01	NNG	27
13	거기01	NP	103	38	중04	NNB	54	63	데01	NNB	35	88	물01	NNG	27
14	여기01	NP	99	39	게임	NNG	52	64	옆	NNG	35	89	생각01	NNG	27
15	그때	NNG	93	40	시10	NNB	51	65	애03	NP	35	90	정도11	NNG	27
16	엄마	NNG	92	41	여자02	NNG	51	66	일05	NR	35	91	이번01	NP	27
17	선생님	NNG	91	42	친구02	NNG	51	67	꺼 (거01)	NNB	34	92	둘01	NR	27
18	말01	NNG	89	43	오04	NR	50	68	이름	NNG	33	93	살04	NNB	26
19	다음01	NNG	88	44	명03	NNB	49	69	사11	NR	33	94	노래01	NNG	26
20	질01	NNG	88	45	돈01	NNG	48	70	적03	NNB	32	95	쿠션01	NNG	25
21	수02	NNB	85	46	모(뭐)	NP	47	71	공부01	NNG	32	96	오빠	NNG	25
22	번04	NNB	83	47	원01	NNB	44	72	언니	NNG	32	97	위01	NNG	25
23	어디01	NP	83	48	남자02	NNG	44	73	거01	NP	32	98	날01	NNG	24
24	사람	NNG	82	49	시간04	NNG	44	74	차기04	NP	32	99	흔저01	NNG	24
25	누구	NP	81	50	처음	NNG	44	75	구01	NR	32	100	그01	NP	24

체언 범주에서 가장 많은 화자가 사용한 어휘는 대명사 ‘나’(236명)로 나타났다. 그 다음은 의존명사 ‘거’(229명), 대명사 ‘뭐’(212명), 대명사 ‘너’(210명), 대명사 ‘그거’(180명), 대명사 ‘내’(174명), 대명사 ‘우리’(166명) 순으로 1위에서 7위까지가 일반명사가 아닌 대명사와 의존 명사가 차지하고 있다. 또 하나의 특징은 이들 대명사나 의존명사가 모두 구어형인 점이다.

‘것, 무엇, 그것’의 형태가 아닌 이들의 구어형을 화자들이 구어에서 실제 많이 사용하고 있었다. ‘것’은 전체 화자 240명 가운데 67명만이 사용하고 있어 구어형 ‘거’와 큰 차이를 보이고 있다.

또한 초등학생의 일상 대화에서의 특징이라고 볼 수 있는 점 중의 하나는 ‘엄마(16위, 92명), 아빠(28위, 68명), 동생(53위, 42명), 언니(72위, 32명), 오빠(96위, 25명)’ 등의 친족 명칭 어휘가 100위 안에 들었다는 것이다. 이와 함께 ‘선생님(17위, 91명), 반(32위, 64명), 게임(39위, 52명), 학교(33위, 59명), 친구(42위, 51명), 컴퓨터(52위, 43명), 학원(55위, 42명), 영어(61위, 36명), 책(62위, 36명), 수학(76위, 31명)’ 등의 어휘들을 통해서 초등학생의 대화 주제나 생활 등을 짐작할 수 있겠다.

2) 용언

용언을 사용 화자 수에 따라 정리해 보이면 다음 <표 5>와 같다.

<표 5> 용언의 사용 화자 수에 따른 어휘 목록

순위	어휘	품사	화자 수	순위	어휘	품사	화자 수	순위	어휘	품사	화자 수	순위	어휘	품사	화자 수
1	하01	VV	236	26	좋아하	VV	88	51	받01	VV	42	76	들01	VV	28
2	있01	VA	222	27	않	VX	82	52	못04하	VA	41	77	생각하	VV	28
3	이03	VCP	218	28	지04	VX	81	53	들01	VV	39	78	어떡하	VV	28
4	되01	VV	198	29	싫01	VA	75	54	오01	VX	39	79	괜찮	VA	27
5	보01	VX	176	30	주01	VV	75	55	아프	VA	38	80	어렵	VA	27
6	가01	VV	172	31	말03	VX	75	56	말03	VV	38	81	결리01	VV	27
6	알	VV	172	32	싶	VX	72	57	살01	VV	38	82	보이01	VV	27
8	없01	VA	162	33	얘기하	VV	67	58	웃기	VV	38	83	쓰01	VV	27
9	보01	VV	159	34	니01	VV	66	59	짜증나	VV	38	84	어쩐01	VV	27
10	있01	VX	152	35	질02하	VV	65	60	쓰03	VV	37	85	이기01	VV	27
11	아니	VCN	145	36	무섭	VA	60	61	내02	VV	36	86	키우	VV	27
12	그러	VV	144	37	죽01	VV	59	62	들어가01	VV	34	87	어떻	VA	26
13	좋01	VA	142	38	놀01	VV	57	63	버리01	VX	34	88	맞03	VV	26
14	모르	VV	137	39	이러	VV	57	64	생기	VV	33	89	올라가	VV	26
15	주01	VX	133	40	사	VV	52	65	자01	VV	33	90	똑같	VA	25

순위	어휘	품사	화자 수	순위	어휘	품사	화자 수	순위	어휘	품사	화자 수	순위	어휘	품사	화자 수
16	그렇	VA	126	41	많	VA	51	66	잡01	VV	33	91	끝나	VV	25
17	갈	VA	122	42	나가	VV	49	67	가지	VV	32	92	만나	VV	25
18	오01	VV	118	43	가01	VX	48	68	때리01	VV	32	93	빼01	VV	25
19	맞01	VV	114	44	크01	VA	47	69	치02	VV	32	94	싸우	VV	25
20	가지	VX	111	45	다니	VV	46	70	재미있	VA	31	95	앉	VV	24
21	말01하	VV	103	46	이상하	VA	45	71	그리02	VV	30	96	남01	VV	23
22	재밋	VA	95	47	놓01	VX	44	72	이렇	VA	29	97	넋02	VA	22
23	히01	VX	93	48	갖01	VV	43	73	재미없	VA	29	98	맛있	VA	22
24	나오	VV	92	49	갖01	VX	43	74	만지	VV	29	99	들어오	VV	22
25	떡02	VV	88	50	만들	VV	42	75	타02	VV	29	100	읽	VV	22

용언 가운데 가장 사용 화자 수가 많은 어휘는 ‘하다’로 총 236명이 사용하였다. 그 다음으로 ‘있다’(236명, 본용언, 이다(218명), 되다(198명), 보다(176명, 보조용언), 가다(172명), 알다(172명), 없다(162명), 보다(159, 본용언), 있다(152, 보조용언)가 비교적 높은 순위를 차지하고 있다. 목록 가운데 ‘재밋다’(22위, 95명)와 같은 구어형 어휘의 순위가 높은 것이 주목할 만하다.

그러나 용언의 경우 체언과 달리 구어형의 출현보다는 출현 어휘 목록이 지닌 의미면에서의 특징이 눈에 띈다. 우선, 보조 용언 가운데 ‘보다(5위, 176명), 있다(10위, 152명), 주다(15위, 133명), 가지다(20위, 111명)’가 20위 안에 들었다는 점이다. ‘알다(6위, 172명), 모르다(14위, 137명)’와 같이 인지 상태를 설명하는 어휘가 상위를 차지한 것도 흥미롭다. 마지막으로 목록 가운데 ‘짜증나다(59위, 38명)’와 같은 어휘를 통해서도 학생들의 습관적 말 버릇을 확인할 수 있었다.

3) 수식언

관형사와 부사가 포함된 수식언의 경우, 사용 화자 수가 많은 순서로 정리해 보이면 다음과 같다.

<표 6> 수식언의 사용 화자 수에 따른 어휘 목록

순위	어휘	품사	화자 수	순위	어휘	품사	화자 수	순위	어휘	품사	화자 수	순위	어휘	품사	화자 수
1	안2	MAG	217	26	딱3	MAG	75	51	별로1	MAG	39	76	정말1	MAG	22
2	근데1	MAJ	194	27	내13	MM	73	52	좁 (좁2)	MAG	39	77	다섯	MM	22
3	그1	MM	187	28	어떤	MM	73	53	사1	MM	38	78	육2	MM	22
4	왜2	MAG	159	29	또	MAJ	68	54	삼6	MM	38	79	천3	MM	22
5	다3	MAG	148	30	지금3	MAG	62	55	오늘	MAG	37	80	인제1	MAG	21
6	막2	MAG	137	31	그러면	MAJ	62	56	다른	MM	37	81	맨1	MM	21
6	한1	MM	137	32	이제1	MAG	60	57	딱2	MAG	36	82	이케	MAG	20
8	진짜	MAG	129	33	제일4	MAG	60	58	언제1	MAG	34	83	요즘	MAG	19
9	이렇게	MAG	121	34	갑자기	MAG	55	59	그래도	MAJ	33	84	쪼금	MAG	19
10	그냥	MAG	109	35	되게	MAG	53	60	그리구	MAJ	33	85	쨍2	MAG	17
11	그럼1	MAJ	99	36	그런데	MAJ	52	61	거여1	MAG	31	86	하여튼	MAG	17
12	어떻게	MAG	94	37	빨리	MAG	51	62	오4	MM	31	87	얌3	MM	17
12	그래서	MAJ	94	38	일5	MM	51	63	아까	MAG	30	88	얼마나	MAG	16
12	뵈	MM	94	39	맨날	MAG	49	64	어제1	MAG	30	89	디게 (되게)	MAG	15
12	무슨	MM	94	40	그렇게	MAG	48	65	네2	MM	30	90	웁3	MM	15
16	또	MAG	90	41	계속4	MAG	47	66	아주1	MAG	29	91	바로2	MAG	14
16	많이	MAG	90	42	십	MM	47	67	아작1	MAG	29	92	벌써	MAG	14
18	그리고	MAJ	88	43	이5	MM	46	68	그니까	MAJ	29	93	자꾸1	MAG	14
19	못4	MAG	87	44	이런1	MM	46	69	저4	MM	27	94	조금1	MAG	14
20	좁2	MAG	87	45	같이	MAG	44	70	엄청	MAG	26	95	금 (그럼1)	MAJ	14
21	너무1	MAG	86	46	먼저	MAG	44	71	백5	MM	26	96	맨3	MM	14
22	잘2	MAG	86	47	그니까	MAJ	44	72	어느1	MM	25	97	존나	MAG	13
23	두1	MM	81	48	이9	MM	44	73	젤	MAG	23	98	참1	MAG	13
24	더1	MAG	79	49	다시1	MAG	41	74	따른 (다른)	MM	23	99	여덟	MM	13
25	그럼1	MM	79	50	세1	MM	40	75	만6	MM	23	100	꼭3	MAG	12

<표 6>을 보면, 수식언 가운데 가장 많은 화자가 사용한 어휘는 부정 부사 ‘안(217명)’으로 조사되었다. 그 다음이 ‘근데(194명), 그(187명), 왜(159명), 다(148명), 막(137명), 한(137명), 진짜(129명), 이렇게(121명), 그냥(109명)’ 등

의 어휘가 높은 순위를 차지하고 있다.

수식언 어휘 목록 100위까지 중에서 부사는 68개, 관형사는 32개로, 관형사에 비해 부사가 훨씬 많이 사용되었다. 부사 가운데에서도 ‘근데(2위, 194명), 왜(4위, 159명), 막(6위, 137명), 진짜(8위, 129명), 그럼(11위, 99명), 딱(26위, 75명), 이제(32위, 60명)’와 같이 담화 표지로 사용된 어휘가 높은 순위를 차지하고 있어, 구어가 지닌 특징적인 점을 확인할 수 있었다. 또한, ‘좀(20위, 87명), 되게(35위, 53명), 쯤(52위, 39명), 그리구(60위, 33명), 그니까(68위, 29명), 쟈(73위, 23명), 이케(82위, 20명), 짱(85위, 17명), 하여튼(85위, 17명), 존나(97위, 13명)’ 등의 구어 형태들도 초등학교생의 일상 대화의 단면을 볼 수 있는 어휘들이라 생각된다.

32개의 관형사를 자세히 살펴보면, 지시 관형사가 13개, 수 관형사가 17개, 성상 관형사가 1개 나타나고 있다.¹⁷⁾ 13개의 지시 관형사는 다시 ‘그(3위, 187명), 그런(25위, 79), 이(43위, 46명)’와 같은 일반 지시 관형사가 9개, ‘땃(12위, 94명), 무슨(12위, 94명)’ 등과 같은 의문 지시 관형사 4개 나타나고 있어 현장성이 중요한 구어의 특징을 확인해 볼 수 있다. 성상 관형사는 ‘맨(81위, 21명)’ 하나만이 유일하게 100위 안에 나타났는데, 국어 어휘에서 관형사의 숫자가 그다지 많지 않지만 실제 체언의 성질이나 상태를 제한하는 기능을 관형사가 크게 담당하지 않는다고 추측해 볼 수 있겠다.

4) 관계언

관계언의 사용 화자 수에 따른 어휘 목록을 보이면 다음과 같다.

17) 성상 관형사 : 맨01(81위, 21명) / 일반 지시 관형사 : 그01(3위, 187명), 그런01(25위, 79명), 내13(27위, 73명), 이05(43위, 46명), 이런01(44위, 46명), 다른(56위, 37명), 저04(69위, 27명), 따른(다른)(74위, 23명), 단03(96위, 14명) / 의문 지시 관형사 : 땃(12위, 94명), 무슨(12위, 94명), 어떤(28위, 73명), 어느01(72위, 25명) / 수 관형사 : 한01(7위, 137명), 두01(23위, 81명), 일05(38위, 51명), 십(42위, 47명), 이09(48위, 44명), 세01(50위, 40명), 사11(53위, 38명), 삼06(54위, 38명), 오04(62위, 31명), 네02(65위, 30명), 백05(71위, 26명), 만06(75위, 23명), 다섯(77위, 22명), 육02(78위, 22명), 천03(79위, 22명), 열03(87위, 17명), 팔03(90위, 15명), 여덟(99위, 13명)

<표 7> 관계언의 사용 화자 수에 따른 어휘 목록

순위	어휘	품사	화자 수	순위	어휘	품사	화자 수
1	가11	JKS	239	26	처럼	JKB	33
2	논01	JX	234	27	요17	JX	30
3	예04	JKB	229	28	가11	JKC	27
4	ㄴ02	JX	216	29	에다05	JKB	21
5	도15	JX	215	30	루(로07)	JKB	20
6	를	JKO	184	31	구(고22)	JKQ	19
7	에서02	JKB	172	32	마다04	JX	16
8	로07	JKB	159	33	라06	JKQ	15
9	만4	JX	121	34	나10	JC	14
10	랑05	JKB	114	35	르(로07)	JKB	14
11	한테	JKB	102	36	에다가	JKB	13
12	고22	JKQ	96	37	다06	JX	13
13	르02	JKO	88	38	하구(하고05)	JKB	11
14	두10	JX	86	39	라고01	JKQ	10
15	까지03	JX	81	40	대토10	JKB	9
16	랑05	JC	75	41	다기02	JX	9
17	서16	JKB	66	42	과12	JC	8
18	밖에	JX	59	43	가11	JX	8
19	보다04	JKB	55	44	라도01	JX	8
20	나10	JX	54	45	에게	JKB	7
21	하고05	JC	46	46	같이	JKB	6
21	하고05	JKB	46	47	보고01	JKB	6
23	아09	JKV	45	48	야11	JX	6
24	부터	JX	44	49	과12	JKB	5
25	의10	JKG	43	50	으루(으로01)	JKB	5

<표 7>을 보면, 관계언 가운데 가장 많은 화자가 사용한 어휘는 주격 조사 ‘가(1위, 239명)’이다. 관형격 조사 ‘의(25위, 43명)’은 주격 조사나 목적격 조사에 비해 사용 화자가 적은 것을 볼 수 있었다. 관계언 목록을 보면, 목적격 조사 ‘을/를’의 구어적 형태인 ‘르(13위, 88명)’의 순위가 높은 편

이고, ‘랑(부사격, 10위, 114명), 랑(접속, 16위, 75명), 한테(11위, 102명), 하고(접속/부사격, 21위, 46명), 보고(47위, 6명)’와 같이 구어형이 높은 순위를 차지하고 있고, ‘두(14위, 86명), 루(로07 ; 30위, 20명), 구(고22 ; 31위, 19명), 르(로07 ; 35위, 14명), 하구(하05 ; 38위, 11명), 으루(으01 ; 50위, 5명)’와 같이 모음 변이 형이 많이 나타나고 있어, 구어에서의 조사 사용상 특징을 확인할 수 있다. 또한, 접속 조사 가운데 ‘과(42위, 8명)’에 비해 ‘랑(16위, 75명), 하고(21위, 54명)’가 높은 순위를 나타내고 있는 점도 구어 자료의 특징적인 점이라 볼 수 있다.

5) 의존형태

다음은 의존형태를 사용 화자 수에 따라 정리해 보기로 한다.

<표 8> 의존형태의 사용 화자 수에 따른 어휘 목록

순위	어휘	품사	화자 수	순위	어휘	품사	화자 수	순위	어휘	품사	화자 수	순위	어휘	품사	화자 수
1	어11	EF	240	26	르가	EF	101	51	사23	EP	41	76	ㄴ다고01	EC	13
2	어06	EC	238	27	네08	XSN	95	52	ㄴ대03	EF	40	77	냐면	EC	13
3	였	EP	237	28	거든03	EF	94	53	야저12	EF	39	78	래04	EF	13
4	고24	EC	225	29	러09	EF	93	54	ㄴ자01	EC	36	79	비니다	EF	13
5	야3	EF	224	30	자28	EF	91	55	고25	EF	36	80	르려고 (려고02)	EC	12
6	느03	ETM	223	31	이30	XSN	86	56	다고03	EF	36	80	드라 (더라)	EF	12
7	은06	ETM	217	32	는데02	EF	77	57	짜리02	XSN	36	80	ㄴ디는	ETM	12
8	자25	EF	215	33	ㄴ데02	EF	76	58	러01	EC	34	80	책02	XSN	12
9	잖아	EF	213	34	대22	EF	75	59	테나01	EC	31	84	르라고 (려고02)	EC	11
10	르03	ETM	209	35	ㄴ가01	EF	73	60	던02	ETM	31	85	대16	EF	10
11	다07	EF	205	36	네07	EF	70	61	다는	ETM	30	85	케요	EF	10
12	는데01	EC	187	37	애16	EF	70	62	다고02	EC	25	85	ㄴ됐	EP	10
13	면09	EC	187	38	나12	EF	65	63	랬	EP	24	88	깨(크까)	EF	9
14	냐	EF	170	39	다카03	EC	63	64	구 (고25)	EF	22	88	잖어	EF	9

순위	어휘	품사	화자 수	순위	어휘	품사	화자 수	순위	어휘	품사	화자 수	순위	어휘	품사	화자 수
15	어서03	EC	158	40	다08	EC	59	65	구나03	EF	21	88	단21	ETM	9
16	들09	XSN	151	41	르게	EF	57	66	라2	EC	19	88	라느02	ETM	9
17	자24	EC	142	42	면서	EC	55	67	려고02	EC	19	92	다면02	EC	8
18	졌	EP	130	43	어노02	EC	55	68	냐고01	EF	19	93	어야지01	EC	8
19	구 (고24)	EC	129	44	였었	EP	55	69	르걸	EF	19	94	나고02	EF	8
20	나01	EF	125	45	니10	EF	52	70	씩03	XSN	18	95	다구 (다고03)	EF	8
21	니까	EC	123	46	더라	EF	51	71	지만05	EC	17	96	다니까01	EF	8
22	게10	EC	112	47	라08	EF	48	72	게11	EF	17	97	길래02	EC	7
23	나테01	EC	107	48	나	ETM	47	73	쯤02	XSN	17	98	단테01	EC	7
24	가36	ETN	106	49	느자01	EC	43	74	단테02	EF	15	99	라세02	EC	7
25	어아02	EC	105	50	르래	EF	41	75	냐구 (냐고)	EF	14	100	군09	EF	7

의존 형태 중, 조사 대상자 240명이 모두 사용한 어휘는 종결 어미 ‘어’이다. 그 다음이 연결어미 ‘어’(2위, 238명), 과거 시제 선어말 어미 ‘았(3위, 237명)’, 연결어미 ‘고’(4위, 225명)’의 순으로 나타났다. 종결 어미 ‘다(11위, 205명)’에 비해 ‘아’(5위, 224명), ‘지’(8위, 215명), ‘잖아’(9위, 213명)’ 등의 형태가 더 높은 순위를 차지한 것, 연결어미 ‘거든’은 100위 안에 들지 못했지만 종결어미 ‘거든’(28위, 94명)’은 높은 순위를 차지한 것 등은 구어의 특징이라고 볼 수 있겠다. 특히 ‘구(고24 ; 19위, 129명), 르려고(려고02 ; 80위, 12명), 드라(더라 ; 80위, 12명), 르라고(려고02 ; 84위, 11명), 까(르까 ; 88위, 9명), 다구(다고03 ; 95위, 8명)’와 같은 비표준 형태 목록이 많다는 점은 주목할 만하다.

3. 출현 빈도에 따른 사용 분포¹⁸⁾

지금까지 사용 화자 수에 따른 어휘 목록에 나타난 특징들을 살펴보

18) 연령 등에서 차이가 나는 다른 구어 담화와의 출현 빈도 비교가 필요하다든 의견도 있었는 데 현재 국내에서 순수 구어만을 대상으로 하여 계량적 분석을 한 연구는 일부 품사, 예를 들어 조사·어미(박석준·남길임·서상규 2003)에 제한되어 비교가 불가능하다.

있는데, 구어 자료의 특징은 문어와의 비교를 통해서 보다 분명해질 것이다. 본고에서는 국립국어원(2002)에 조사한 ‘교육용 어휘 선정을 위한 현대 국어 사용 빈도 조사’와 비교해 보고자 한다. 국립국어원(2002)은 총 1,531,966어절을 대상으로 빈도 조사를 실시하였는데, 조사 대상 자료 가운데 구어가 차지하는 비율은 3.2%로 아주 낮은 편이고, 방송 대본이 포함되어 있어 실제 생활에서 수집한 구어라고 보기에는 무리가 있어 문어 자료로 간주하기로 한다.¹⁹⁾ 문어와 비교를 위해 본고에서 조사한 자료를 다시 출현 빈도를 중심으로 하여 고찰하기로 한다. 그러나 절대적인 수치인 출현 빈도를 비교하는 것은 두 코퍼스 규모 간의 차이로 인해 불가능하기 때문에 전체 어휘 총합 가운데 출현 빈도가 차지하는 비율을 함께 제시하여 살펴보기로 한다.

1) 체언

출현 빈도를 기준으로 하여 구어와 문어에 나타난 체언 어휘 목록을 고빈도 순으로 보이면 다음 <표 9>와 같다.

<표 9> 출현 빈도에 따른 구어와 문어의 체언 어휘 목록

순위	구 어				순위	문 어			
	어 휘	품 사	빈 도	출현 비율 ²⁰⁾		어 휘	품 사	빈 도	출현 비율
1	나03	NP	1,679	6.53%	1	것01	의	25,567	3.12%
2	거01	NNB	1,472	5.72%	2	수02	의	10,915	1.33%
3	너01	NP	912	3.55%	3	나03	대	10,564	1.29%
4	뭐	NP	855	3.33%	4	사람	명	7,004	0.85%

19)

	한국어 교재	교과서	교양	문학	신문	잡지	대본	구어	기타	총합계
자료 종수	20	31	26	20	39	14	12	8	6	176
어절 수	81,781	100,164	374,125	278,028	303,911	209,279	48,844	48,337	78,497	1,531,966
비율 (%)	5.3	7.1	24.4	18.1	19.8	13.7	3.2	3.2	5.1	100

순위	구 어				순위	문 어			
	어 휘	품 사	빈 도	출현 비율 ²⁾		어 휘	품 사	빈 도	출현 비율
5	그거	NP	615	2.39%	5	우리03	대	6,583	0.80%
6	우리03	NP	609	2.37%	6	그01	대	6,305	0.77%
7	내04	NP	546	2.12%	7	등05	의	5,967	0.73%
8	니05	NP	445	1.73%	8	때01	명	5,813	0.71%
9	때01	NNG	426	1.66%	9	거01	의	5,678	0.69%
10	개	NP	368	1.43%	10	년02	의	4,630	0.56%
11	이거01	NP	326	1.27%	11	말01	명	4,286	0.52%
12	애02	NNG	310	1.21%	12	일01	명	4,271	0.52%
13	선생님	NNG	288	1.12%	13	이05	대	3,918	0.48%
14	엄마	NNG	232	0.90%	14	때문	의	3,889	0.47%
15	십	NR	225	0.88%	15	씨07	의	2,920	0.36%
16	거기01	NP	222	0.86%	16	일07	의	2,808	0.34%
17	여기01	NP	211	0.82%	17	사회07	명	2,770	0.34%
18	다음01	NNG	190	0.74%	18	그것	대	2,542	0.31%
19	집01	NNG	168	0.65%	19	집01	명	2,457	0.30%
20	학년	NNG	167	0.65%	20	문제06	명	2,370	0.29%
21	사람	NNG	162	0.63%	21	저03	대	2,281	0.28%
22	그때	NNG	160	0.62%	22	속01	명	2,168	0.26%
23	얘기	NNG	156	0.61%	23	테01	의	2,104	0.26%
24	원01	NNB	146	0.57%	24	앞	명	1,979	0.24%
25	말01	NNG	144	0.56%	25	경우03	명	1,976	0.24%
26	반11	NNG	143	0.56%	26	중04	의	1,950	0.24%
27	누구	NP	142	0.55%	27	그녀	대	1,884	0.23%
28	수02	NNB	141	0.55%	28	자신01	명	1,799	0.22%
29	번04	NNB	139	0.54%	29	문화01	명	1,790	0.22%
30	아빠	NNG	138	0.54%	30	원01	의	1,776	0.22%
31	어디01	NP	129	0.50%	31	생각01	명	1,750	0.21%
32	이09	NR	126	0.49%	32	명03	의	1,743	0.21%
33	개10	NNB	124	0.48%	33	소리01	명	1,660	0.20%

20) 총 어휘 빈도 가운데 해당 어휘의 출현빈도가 차지하는 비율을 나타낸다.

순위	구 어				순위	문 어			
	어 휘	품 사	빈 도	출현 비율		어 휘	품 사	빈 도	출현 비율
34	것01	NNB	120	0.47%	34	여지02	명	1,645	0.20%
35	모(뒤)	NP	118	0.46%	35	개10	의	1,622	0.20%
36	언니	NNG	110	0.43%	36	정도11	명	1,619	0.20%
37	학교	NNG	106	0.41%	37	뒤01	명	1,608	0.20%
38	오04	NR	100	0.39%	38	아이01	명	1,495	0.18%
39	시10	NNB	100	0.39%	39	시간04	명	1,491	0.18%
40	삼06	NR	98	0.38%	40	너01	대	1,489	0.18%
41	돈01	NNG	97	0.38%	41	인간01	명	1,488	0.18%
42	게임	NNG	96	0.37%	42	사실04	명	1,459	0.18%
43	친구02	NNG	96	0.37%	43	어머니01	명	1,436	0.18%
44	중04	NNB	96	0.37%	44	눈01	명	1,425	0.17%
45	줄04	NNB	90	0.35%	45	뒤	대	1,420	0.17%
46	학원02	NNG	89	0.35%	46	점10	명	1,417	0.17%
47	하나	NR	88	0.34%	47	시대02	명	1,396	0.17%
48	여자02	NNG	85	0.33%	48	다음01	명	1,386	0.17%
49	명03	NNB	85	0.33%	49	누구	대	1,384	0.17%
50	동생01	NNG	84	0.33%	50	전08	명	1,378	0.17%
합계			13,774	53.57%	합계			171,275	20.89%

<표 9>를 보면, 상위에 출현한 어휘의 세부 품사가 구어와 문어가 다르다는 것을 확인할 수 있다. 구어는 ‘나(1위, 1,679회), 거(2위, 1,472회), 너(3위, 912회), 뒤(4위, 855회), 그거(5위, 615회)’ 등 10위까지 대명사 8개, 의존명사 1개, 일반명사 1개가 출현하고 있는 반면에 문어는 ‘것(1위, 25,567회), 수(2위, 10,915회), 나(3위, 10,564회), 사람(4위, 7,004회), 우리(5위, 6,583회)’ 등 대명사 3개, 의존명사 5개, 일반명사 2개가 출현해 차이가 나고 있다. 특히 대명사의 출현은 문어 목록에는 없는 ‘이거(11위, 326회), 거기(16위, 222회), 여기(17위, 211회)’ 등의 어휘가 구어에는 나타나는 것, 구어에는 없는 ‘그녀

(27위, 1,884회)가 문어 목록에는 나타나고 있어 구어의 문어의 차이를 볼 수 있다. 또한 같은 대명사라 할지라도 구어에서는 ‘그거(5위, 615회)’가, 문어에서는 ‘그것(18위, 2,542회)’이 출현하고 있다.

구어와 문어 목록의 비교에서 나타나는 또 다른 특징은 문어 목록에서 의존명사 ‘수, 등, 줄’이 고빈도로 나타나고, 구어에서는 수사가 고빈도로 출현하는 것이다. 그리고 50위까지의 어휘가 체언 중 총 53.57%의 비율을 차지하며 출현하는데 비해 문어는 20.89%의 비율을 차지해 문어에서 보다 다양한 어휘를 사용하고 있다고 생각된다.²²⁾

2) 용언

구어와 문어에 나타난 용언 어휘 목록을 고빈도 순으로 정리해 보면 다음과 같다.

<표 10> 출현 빈도에 따른 구어와 문어의 용언 어휘 목록

순위	구 어				순위	문 어			
	어 회	품 사	빈 도	출현 비율		어 회	품 사	빈 도	출현 비율
1	하01	VV	1,773	8.54%	1	이다	지	49,507	9.04%
2	있01	VA	1,220	5.88%	2	하다01	동	22,064	4.03%
3	이03	VCP	834	4.02%	3	있다01	보	18,553	3.39%
4	되01	VV	652	3.14%	4	있다01	형	18,202	3.32%
5	가01	VV	616	2.97%	5	되다01	동	11,506	2.10%
6	보01	VX	457	2.20%	6	하다01	보	10,758	1.96%
7	보01	VV	416	2.00%	7	없다01	형	8,969	1.64%
8	가지	VX	407	1.96%	8	않다	보	8,852	1.62%
9	그렇	VA	406	1.96%	9	아니다	지	6,245	1.14%
10	알	VV	402	1.94%	10	보다01	동	6,045	1.10%
11	그러	VV	399	1.92%	11	보다01	보	5,231	0.96%

22) 이러한 차이가 코퍼스 규모에 의한 것이라 생각할 수도 있지만 또 한편으로는 구어에서의 표현 어휘가 문어보다 적기 때문이라고 볼 수도 있겠다.

순위	구 어				순위	문 어			
	어 휘	품 사	빈 도	출현 비율		어 휘	품 사	빈 도	출현 비율
12	없01	VA	380	1.83%	12	같다	형	5,199	0.95%
13	좋01	VA	377	1.82%	13	주다01	보	4,929	0.90%
14	있01	VX	329	1.59%	14	대하다02	동	4,871	0.89%
15	주01	VX	301	1.45%	15	가다01	동	4,670	0.85%
16	모르	VV	272	1.31%	16	말하다	동	3,738	0.68%
17	같	VA	263	1.27%	17	위하다01	동	3,497	0.64%
18	맞01	VV	254	1.22%	18	오다01	동	3,350	0.61%
19	아니	VCN	251	1.21%	19	알다	동	2,933	0.54%
20	오01	VV	234	1.13%	20	그렇다	형	2,905	0.53%
21	좋아하	VV	206	0.99%	21	크다01	형	2,835	0.52%
22	재밌	VA	198	0.95%	22	많다	형	2,697	0.49%
23	말이하	VV	191	0.92%	23	좋다01	형	2,661	0.49%
24	나오	VV	171	0.82%	24	받다01	동	2,566	0.47%
25	히01	VX	166	0.80%	25	나오다	동	2,435	0.44%
26	먹02	VV	164	0.79%	26	따르다01	동	2,388	0.44%
27	싫01	VA	145	0.70%	27	살다01	동	2,297	0.42%
28	주01	VV	141	0.68%	28	못하다	보	2,249	0.41%
29	않	VX	133	0.64%	29	생각하다	동	2,237	0.41%
30	지04	VX	132	0.64%	30	모르다	동	2,218	0.41%
31	무섭	VA	130	0.63%	31	만들다	동	2,151	0.39%
32	사	VV	129	0.62%	32	먹다02	동	1,874	0.34%
33	싫	VX	128	0.62%	33	오다01	보	1,854	0.34%
34	놀01	VV	120	0.58%	34	어떻다	형	1,748	0.32%
35	얘기하	VV	116	0.56%	35	통하다	동	1,674	0.31%
36	말03	VX	113	0.54%	36	그러다	동	1,662	0.30%
37	갖01	VX	113	0.54%	37	듣다01	동	1,593	0.29%
38	죽01	VV	100	0.48%	38	들다01	동	1,535	0.28%
39	잘02하	VV	99	0.48%	39	싶다	보	1,527	0.28%
40	나01	VV	97	0.47%	40	보이다01	동	1,522	0.28%
41	이러	VV	88	0.42%	41	가지다	동	1,512	0.28%
42	그리02	VV	85	0.41%	42	지나다	동	1,494	0.27%

순위	구 어				순위	문 어			
	어 휘	품 사	빈 도	출현 비율		어 휘	품 사	빈 도	출현 비율
43	많	VA	83	0.40%	43	주다01	동	1,489	0.27%
44	다니	VV	78	0.38%	44	나다01	동	1,451	0.27%
45	만들	VV	77	0.37%	45	이렇다	형	1,448	0.26%
46	내02	VV	74	0.36%	46	의하다01	동	1,412	0.26%
47	크01	VA	74	0.36%	47	이러하다	형	1,386	0.25%
48	이상하	VA	71	0.34%	48	버리다01	보	1,330	0.24%
49	가01	VX	69	0.33%	49	들다04	동	1,272	0.23%
50	나가	VV	67	0.32%	50	시작하다01	동	1,264	0.23%
합계			13,801	66.51%	합계			257,805	47.09%

<표 10>을 보면, 구어에서는 ‘하다(1위, 1,773회)’가 가장 높은 빈도를 차지하였고, 문어에서는 ‘이다(1위, 49,507회)’가 차지하였다. 구어와 문어 목록에서 드러나는 몇 가지 차이를 보면, 구어에서는 ‘그렇다(9위, 406회), 그러다(11위, 399회), 이러다(41위, 88회)’가 비교적 높은 빈도로 나타나고 있는데, 문어에서는 ‘그렇다(20위, 2,905회), 그러다(36위, 1,662회)’로 구어보다 상대적으로 낮은 빈도로 나타나고 있다. 또 하나 문어에서는 ‘대하다(14위, 4,871회), 위하다(17위, 3,497회), 통하다(35위, 1,674회), 의하다(46위, 1,412회)’의 어휘가 나타나는데 구어에서는 이런 어휘가 목록 안에 들어 있지 않다. 그밖에 문어에는 없는 ‘맞다(18위, 254회)’가 구어에는 있는데, 이는 대화하면서 상대방의 말에 대한 동의 표현인 ‘맞어’ 등이 빈번하게 사용되기 때 문인 것으로 보인다.

50위까지의 어휘가 출현하는 비율을 보면 구어는 66.51%, 문어는 44.02%를 차지하고 있어 22% 정도의 차이가 나고 있다.

3) 수식언

구어와 문어에 나타난 수식언의 차이를 살펴보기로 한다. 출현 빈도에 따라 어휘 목록을 정리해 보면 다음과 같다.

<표 11> 출현 빈도에 따른 구어와 문어의 수식언 어휘 목록

순위	구 어				순위	문 어			
	어 휘	품 사	빈 도	출현 비율		어 휘	품 사	빈 도	출현 비율
1	근테01	MAJ	986	7.83%	1	그01	관	9,413	5.97%
2	그01	MM	811	6.44%	2	이05	관	6,332	4.02%
3	인02	MAG	763	6.06%	3	헌01	관	4,485	2.85%
4	막02	MAG	597	4.74%	4	그러나	접	3,457	2.19%
5	이렇게	MAG	376	2.99%	5	또	접	2,814	1.79%
6	진짜	MAG	346	2.75%	6	인02	부	2,676	1.70%
7	왜02	MAG	334	2.65%	7	더01	부	2,567	1.63%
8	다03	MAG	319	2.53%	8	그리고	접	2,384	1.51%
9	헌01	MM	287	2.28%	9	그런01	관	2,353	1.49%
10	그냥	MAG	221	1.75%	10	두01	관	2,082	1.32%
11	그래서	MAJ	219	1.74%	11	어떤	관	1,917	1.22%
12	딱03	MAG	194	1.54%	12	질02	부	1,890	1.20%
13	몇	MM	185	1.47%	13	다시01	부	1,659	1.05%
14	그럼01	MAJ	178	1.41%	14	다른	관	1,653	1.05%
15	그리고	MAJ	163	1.29%	15	이런01	관	1,653	1.05%
16	좀02	MAG	159	1.26%	16	다03	부	1,564	0.99%
17	너무01	MAG	155	1.23%	17	좀02	부	1,555	0.99%
18	어떻게	MAG	153	1.21%	18	함께	부	1,501	0.95%
19	그런01	MM	152	1.21%	19	많이	부	1,491	0.95%
20	무슨	MM	150	1.19%	20	여러	관	1,348	0.86%
21	못04	MAG	147	1.17%	21	가장01	부	1,278	0.81%
22	또	MAG	140	1.11%	22	바로02	부	1,260	0.80%
23	어떤	MM	137	1.09%	23	어느01	관	1,255	0.80%
24	질02	MAG	136	1.08%	24	그래서	접	1,254	0.80%
25	제일04	MAG	134	1.06%	25	모든	관	1,239	0.79%
26	많이	MAG	131	1.04%	26	그런데	접	1,227	0.78%
27	이제01	MAG	127	1.01%	27	모두01	부	1,215	0.77%
28	내13	MM	124	0.98%	28	왜02	부	1,139	0.72%
29	두01	MM	120	0.95%	29	없이	부	1,103	0.70%
30	또	MAJ	118	0.94%	30	몇	관	1,024	0.65%

순위	구 어				순위	문 어			
	어 휘	품 사	빈 도	출현 비율		어 휘	품 사	빈 도	출현 비율
31	더01	MAG	115	0.91%	31	특히	부	1,004	0.64%
32	그러면	MAJ	114	0.91%	32	지금03	부	981	0.62%
33	지금03	MAG	105	0.83%	33	아직01	부	956	0.61%
34	딱02	MAG	105	0.83%	34	같이	부	952	0.60%
35	일05	MM	90	0.71%	35	못04	부	931	0.59%
36	되게	MAG	85	0.67%	36	이제01	부	908	0.57%
37	갑자기	MAG	82	0.65%	37	물론01	부	894	0.57%
38	계속04	MAG	82	0.65%	38	무슨	관	869	0.55%
39	그런데	MAJ	81	0.64%	39	너무01	부	859	0.55%
40	빨리	MAG	81	0.64%	40	이미01	부	842	0.53%
41	그렇게	MAG	73	0.58%	41	및	접	792	0.50%
42	같이	MAG	73	0.58%	42	또한	접	788	0.50%
43	맨날(만날)	MAG	72	0.57%	43	또는	접	778	0.49%
44	십	MM	69	0.55%	44	즉01	부	715	0.45%
45	이런01	MM	67	0.53%	45	따라서	접	712	0.45%
46	다시01	MAG	67	0.53%	46	더욱	부	700	0.44%
47	그러니까	MAJ	66	0.52%	47	아주01	부	697	0.44%
48	이09	MM	64	0.51%	48	먼저	부	669	0.42%
49	이05	MM	62	0.49%	49	얼마나	부	668	0.42%
50	삼06	MM	59	0.47%	50	그냥	부	628	0.40%
합계			9674	76.80%	합계			83,126	52.76%

<표 11>을 보면, 수식언 50위까지의 목록에서 구어는 ‘근데(1위, 986회), 안(3위, 763회), 막(4위, 597회), 이렇게(5위, 376회), 진짜(6위, 346회)’와 같이 부사가 높은 빈도로 출현하였는데 문어는 ‘그(1위, 9,413회), 이(2위, 6,332회), 한(3위, 4,485회)’의 관형사가 가장 높은 빈도로 출현하여 차이를 보이고 있다. 구어 목록 중 ‘근데(1위, 986회), 막(4위, 597회), 진짜(6위, 346회), 그러면(32위, 114회), 그러니까(47위, 66회), 되게(36위, 85회)’ 등은 문어 목록에는 보이지 않는 어휘들이고, 반대로 ‘그러나(4위, 3,457회), 및(41위, 792회), 즉(44위, 715회)’은 구어 목록에는 없는 어휘들이다.

앞 절에서 구어 목록 100위까지 중 관형사의 수가 적다는 지적을 한 바 있는데, 50위까지의 수식언 목록에서도 구어는 14개, 문어는 13개 출현하고 있다. 특히 성상 관형사는 구어와 문어 모두에서 하나도 나타나지 않아 국어에서의 성상 관형사의 출현 빈도가 높지 않다는 것을 확인할 수 있었다.

수식언 50위까지의 어휘 목록의 출현 비율을 보면, 구어는 76.80%, 문어는 52.76%로 나타나 약 24% 정도의 차이가 있어, 용언에서의 차이와 비슷한 정도의 차이가 나고 있다.

4) 관계언

다음은 관계언의 구어에서의 출현 빈도와 문어에서의 출현 빈도를 비교해 보기로 한다. 어휘 목록을 고빈도 순으로 정리하면 다음 <표 12>와 같다.

<표 12> 출현 빈도에 따른 구어와 문어의 관계언 어휘 목록

순위	구 어				순위	문 어			
	어 휘	품 사	빈 도	출현 비율		어 휘	품 사	빈 도	출현 비율
1	가11	JKS	3395	27.41%	1	의	관	72,437	11.37%
2	는01	JX	1860	15.01%	2	을	목	71,370	11.20%
3	에04	JKB	1290	10.41%	3	에	부	57,750	9.06%
4	도15	JX	816	6.59%	4	이	주	50,894	7.99%
5	를	JKO	806	6.51%	5	는	조	45,348	7.12%
6	ㄴ02	JX	678	5.47%	6	를	목	41,429	6.50%
7	에서02	JKB	430	3.47%	7	은	조	35,606	5.59%
8	로07	JKB	386	3.12%	8	가	주	32,530	5.10%
9	랑05	JKB	232	1.87%	9	도	조	23,812	3.74%
10	만14	JX	225	1.82%	10	으로	부	21,559	3.38%
11	한테	JKB	203	1.64%	11	에서	부	19,663	3.09%
12	랑05	JC	192	1.55%	12	로	부	15,511	2.43%

순위	구 어				순위	문 어			
	어 휘	품 사	빈 도	출현 비율		어 휘	품 사	빈 도	출현 비율
13	고22	JKQ	190	1.53%	13	과	접	9,914	1.56%
14	두10	JX	171	1.38%	14	고	인	9,732	1.53%
15	까지03	JX	127	1.03%	15	와	접	6,471	1.02%
16	르02	JKO	124	1.00%	16	만	조	6,057	0.95%
17	서16	JKB	123	0.99%	17	이	보	4,949	0.78%
18	하고05	JC	110	0.89%	18	에게	부	4,769	0.75%
19	밖에	JX	93	0.75%	19	과	부	4,435	0.70%
20	하고05	JKB	86	0.69%	20	까지	조	4,283	0.67%
21	아09	JKV	76	0.61%	21	ㄴ	조	3,906	0.61%
22	보다04	JKB	75	0.61%	22	요	조	3,439	0.54%
23	나10	JX	73	0.59%	23	와	부	3,414	0.54%
24	부터	JX	61	0.49%	24	부터	조	2,548	0.40%
25	의10	JKG	59	0.48%	25	가	보	2,539	0.40%
26	요17	JX	55	0.44%	26	보다	부	2,429	0.38%
27	처럼	JKB	41	0.33%	27	처럼	부	2,379	0.37%
28	가11	JKC	39	0.31%	28	이나	접	2,373	0.37%
29	구(고22)	JKQ	29	0.23%	29	나	접	1,630	0.26%
30	에다05	JKB	25	0.20%	30	르	목	1,597	0.25%
31	마다04	JX	24	0.19%	31	서	부	1,417	0.22%
32	루(로07)	JKB	23	0.19%	32	나	조	1,254	0.20%
33	라06	JKQ	19	0.15%	33	로서	부	1,059	0.17%
34	하구 (하고05)	JKB	19	0.15%	34	이나	조	886	0.14%
35	르로 (로07)	JKB	18	0.15%	35	밖에	조	876	0.14%
36	다06	JX	16	0.13%	36	으로써	부	863	0.14%
37	나10	JC	16	0.13%	37	라	인	787	0.12%
38	에다가	JKB	14	0.11%	38	한테	부	776	0.12%
39	다가02	JX	11	0.09%	39	마다	조	688	0.11%
40	께서	JKS	11	0.09%	40	으로서	부	671	0.11%
41	라고01	JKQ	11	0.09%	41	이라	인	653	0.10%

순위	구 어				순위	문 어			
	어 휘	품 사	빈 도	출현 비율		어 휘	품 사	빈 도	출현 비율
42	대로10	JKB	11	0.09%	42	로부터	부	603	0.09%
43	과12	JC	9	0.07%	43	뿐	조	600	0.09%
44	가11	JX	8	0.06%	44	라고	인	561	0.09%
45	라도01	JX	8	0.06%	45	라도	조	551	0.09%
46	에게	JKB	7	0.06%	46	께서	주	535	0.08%
47	같이	JKB	7	0.06%	47	대로	조	500	0.08%
48	야11	JX	6	0.05%	48	이라고	인	480	0.08%
49	보고01	JKB	6	0.05%	49	이란	조	479	0.08%
50	과12	JKB	6	0.05%	50	이라고	조	466	0.07%
합계			12320	99.45%	합계			579,508	90.93%

관계언, 즉 조사 어휘 목록을 보면, 구어와 문어의 큰 차이는 우선 구어의 고빈도 어휘가 주격 조사 ‘가(1위, 3,395회)’이고, 문어는 관형격 조사 ‘의(1위, 72,437회)’라는 사실이다. 구어에서 ‘의’는 25위로 총 59회 출현하고 있지만, 문어에서는 가장 높은 빈도로 출현하고 있다. 그리고 앞서 사용 화자 수 순으로 살펴본 바와 같이 ‘랑(9위, 232), 하고(20위, 86회), 예다(30위, 25회), 다가(39위, 11회), 보고(49위, 6회)’ 등은 구어 목록에 있지만 문어 목록에는 나타나지 않는다. 구어 목록에는 ‘두(도 ; 14위, 171회), 구(고 ; 29위, 29회), 루(로 ; 32위, 23회), 하(구하고 ; 34위, 19회), 르(로 ; 35위, 18회)’ 등과 같이 모음 변이형이나 비표준형이 나타나고 있지만 문어 목록에는 이들이 나타나지 않고 있다.

구어와 문어 목록 모두에 출현하고 있지만 출현 순위에서 차이를 보이는 어휘도 있다. 구어에서 ‘한테(11위, 203회)’는 비교적 높은 순위에 출현하고 있지만 문어에서는 38위에 나타나고 있어 구어에서 보다 활발하게 사용하는 것으로 보인다. 반대로 ‘과’는 구어에서는 43위로 낮은 빈도로 출현하고 있지만 문어에서는 ‘과(13위, 9,914회), 와(15위, 6,471회)’로 구어보다 높은 빈도로 출현하고 있다.

관계언의 경우 50개 어휘 목록의 총 출현 비율을 보면, 구어는 99.45%

이고 문어는 90.93%로 구어와 문어 모두 이들 어휘가 대부분 사용된다고 볼 수 있겠다.

5) 의존형태²³⁾

마지막으로 구어와 문어에 나타난 의존형태 어휘를 출현 빈도에 따라 정리해 보면 다음과 같다.

<표 13> 출현 빈도에 따른 구어와 문어의 의존형태 어휘 목록

순위	구 어				순위	문 어			
	어 휘	품 사	빈 도	출현비율		어 휘	품 사	빈 도	출현비율
1	어11	EF	4510	17.35%	1	다	종	76,710	11.83%
2	었	EP	3,143	12.09%	2	ㄴ	관	69,661	10.74%
3	어06	EC	1,889	7.27%	3	는	관	60,551	9.34%
4	고24	EC	1,454	5.59%	4	고	연	46,689	7.20%
5	어13	EF	1,003	3.86%	5	었	선	35,239	5.43%
6	은06	ETM	979	3.77%	6	았	선	30,743	4.74%
7	는03	ETM	973	3.74%	7	아	연	26,904	4.15%
8	잖아	EF	960	3.69%	8	어	연	26,266	4.05%
9	지25	EF	849	3.27%	9	르	관	24,098	3.72%
10	디07	EF	842	3.24%	10	ㄴ다	종	20,056	3.09%
11	르03	ETM	753	2.90%	11	계	연	18,406	2.84%
12	는데01	EC	695	2.67%	12	은	관	17,749	2.74%
13	면09	EC	621	2.39%	13	기	명	13,807	2.13%
14	냐	EF	546	2.10%	14	지	연	12,144	1.87%
15	구(고24)	EC	474	1.82%	15	라	종	8,233	1.27%
16	어서03	EC	390	1.50%	16	면	연	8,036	1.24%
17	지24	EC	279	1.07%	17	아서	연	7,411	1.14%
18	니까	EC	278	1.07%	18	던	관	6,609	1.02%
19	겠	EP	239	0.92%	19	을	관	6,447	0.99%

23) 국립국어원에서 조사한 목록은 의존형태가 아닌 어미만을 따로 분류하였기 때문에 문어 목록과의 비교를 위해 접시는 제외하고 어미만을 대상으로 하여 비교하였다.

순위	구 어				순위	문 어			
	어 회	품 사	빈 도	출현비율		어 회	품 사	빈 도	출현비율
20	거든03	EF	238	0.92%	20	며	연	6,126	0.94%
21	계10	EC	222	0.85%	21	습니다	종	5,790	0.89%
22	나디01	EF	218	0.84%	22	겠	선	4,374	0.67%
23	기36	ETN	181	0.70%	23	면서	연	4,274	0.66%
24	어어02	EC	177	0.68%	24	어서	연	4,202	0.65%
25	나테01	EC	170	0.65%	25	지만	연	4,075	0.63%
26	자28	EF	167	0.64%	26	는데	연	3,881	0.60%
27	르까	EF	166	0.64%	27	버니다	종	3,734	0.58%
28	리09	EF	152	0.58%	28	시	선	3,613	0.56%
29	니10	EF	152	0.58%	29	아아	연	3,454	0.53%
30	대22	EF	138	0.53%	30	ㅁ	명	2,929	0.45%
31	는데02	EF	129	0.50%	31	어아	연	2,827	0.44%
32	나테02	EF	127	0.49%	32	어요	종	2,782	0.43%
33	나가01	EF	114	0.44%	33	나가	종	2,594	0.40%
34	다가03	EC	113	0.43%	34	아	종	2,407	0.37%
35	애16	EF	102	0.39%	35	어	종	2,317	0.36%
36	르계	EF	99	0.38%	36	으며	연	2,168	0.33%
37	면서	EC	94	0.36%	37	는다	종	2,085	0.32%
38	네07	EF	94	0.36%	38	자	종	2,071	0.32%
39	나12	EF	92	0.35%	39	라	연	1,890	0.29%
40	있었	EP	89	0.34%	40	다가	연	1,882	0.29%
41	디08	EC	82	0.32%	41	야	종	1,833	0.28%
42	시23	EP	82	0.32%	42	으면	연	1,796	0.28%
43	ㄴ	ETM	79	0.30%	43	아요	종	1,737	0.27%
44	리08	EF	73	0.28%	44	지	종	1,718	0.26%
45	더라	EF	71	0.27%	45	도록	연	1,573	0.24%
46	어도02	EC	70	0.27%	46	르까	종	1,538	0.24%
47	는지01	EC	56	0.22%	47	거나	연	1,464	0.23%
48	더니01	EC	52	0.20%	48	니	연	1,432	0.22%
49	르래	EF	52	0.20%	49	다	연	1,406	0.22%
50	나대03	EF	52	0.20%	50	아도	연	1,375	0.21%
합계			24,580	94.57%	합계			601,086	92.69%

<표 13>을 보면, 구어에서는 ‘어(1위, 4,510회)가 가장 고빈도어인 반면에 문어에서는 ‘다(1위, 76,710회)’가 가장 고빈도어로 나타나고 있다. 구어에서는 ‘잡아(8위, 960회), 거든(20위, 238회), 네(38위, 94회)’ 등의 종결어미가 출현하고 있지만 문어에서는 50위까지의 목록에는 나타나지 않고 있다. ‘습니다(21위, 5,790회), 버니다(27위, 3,734회)’ 등의 높임체 어미가 문어에는 보이지만 구어에는 없는 것인데, 분석 자료가 또래 대화이기 때문에 기인한 결과로 보인다. 연결 어미는 구어와 문어 모두 ‘고’가 가장 많이 출현하고 있다. ‘며(20위, 6,126회)’는 문어에는 있지만 구어에서는 없는 어휘이고, ‘니까(18위, 278회)’는 구어에는 있지만 문어에는 없는 어휘이다. 전성어미에서는 ‘ㅁ(30위, 2,929회)’이 문어에서만 나타나고 있었다.

관계언와 마찬가지로 의존형태도 50위까지의 목록이 차지하는 비율이 의존형태 총 출현비율의 90% 이상 나타나 이들 목록이 대부분을 차지하고 있었고, 구어가 94.57%, 문어가 92.69%로 별 차이가 없는 것으로 나타났다.

IV. 결론

본 연구에서는 초등학생의 실제 대화 자료를 수집·분석하여 어휘의 사용 빈도와 분포도를 조사하고, 문어와의 비교를 통해 구어 어휘의 특징을 살펴보았다. 초등학생의 대화에 나타난 전체 어휘 목록을 추출하고 어휘별 사용 화자수와 빈도를 분석한 결과, 총 5,982개의 어휘가 101,167회 출현하였으며, 전체 어휘 중 사용 화자 수가 많은 어휘는 문법적 형태소들이 차지하였다.

범주별로 정리해보면 다음과 같다. 1) 체언은 구어에서는 대명사, 구어형들이 높은 순위를 차지하였고, 문어에서는 의존명사가 높은 순위를 차지하였으며, 어휘 목록에서도 차이를 보였다. 2) 용언은 다른 범주에 비해서 고빈도 어휘 목록에서는 차이가 크지 않았지만, ‘대하다, 워하다, 통하다,

의하다'와 같은 어휘들이 문어 목록에서만 나타나고 있었다. 3) 수식언은 구어에서는 부사가, 문어에서는 관형사가 높은 순위를 차지하였으며, 성상 관형사는 구어와 문어 모두에서 출현하는 빈도가 낮았다. 4) 관계언은 구어에서는 주격 조사 '가'가, 문어에서는 관형격 조사 '의'가 고빈도로 나타나 차이를 보였다. 특히, 관계언은 모음변이형이나 비표준형이 구어에 많이 나타나고 있었다. 5) 의존 형태는 구어에서는 종결어미 '어'가, 문어에서는 종결어미 '다'가 가장 높은 순위를 차지하였으며, 출현 어휘 목록에서도 구어와 문어의 차이가 있었다.

국어 교육에서 기초 어휘 및 교육용 어휘 선정은 매우 중요하다. 현재 국내에서 진행되고 있는 기초 어휘 선정은 메타 계량과 코퍼스에서의 출현 빈도 등을 기준으로 이루어지고 있는데, 두 가지 방법 모두 문어에 집중되어 이루어지고 있어 구어의 어휘 조사가 절실한 실정이다.

본고의 결과는 교실이라는 한정된 상황에서 도출한 것이어서 초등학교의 어휘 사용 전모를 밝히기에는 무리가 있다. 앞으로 보다 다양한 상황에서의 구어 조사가 이루어지고 이들이 축적된다면 구어에 사용되는 기초 어휘 선정도 가능할 것이라 생각한다. 구어에 사용된 어휘를 수집하고 분석하는 일은 지난한 일이지만 본 연구를 계기로 이 분야의 연구가 활발히 이루어지기를 기대한다.*

* 본 논문은 2008. 10. 31. 투고되었으며, 2008. 11. 3. 심사가 시작되어 2008. 11. 22. 심사가 종료되었음.

▣ 참고문헌

- 구현정(2005), “말뭉치 바탕 구어 연구”, 『언어과학 연구』, 32, 1~20.
- 국립국어원(2002), 『현대국어 사용빈도 조사』, 국립국어원.
- 김광해(2003), 『등급별 국어교육용 어휘』, 박이정.
- 박석준·남길입·서상규(2003), “대학생 구어 텍스트에서의 조사·어미의 분포와 사용 양상에 대한 연구”, 『텍스트언어학회』, 14, 139~167.
- 서상규(1998), “말뭉치 분석에 기반을 둔 낱말 빈도의 조사와 그 응용”, 『한글』, 242.
- 이성연(2007), “남북한 중학교 1학년 국어 교과서의 어휘 분포”, 『새국어교육』, 77, 553~582.
- 임지룡(1991), “국어의 기초 어휘에 대한 연구”, 『국어교육연구』, 23, 87~132.
- 임칠성(1997), “연령층별 성별 어휘 사용의 계량적 고찰”, 『새국어교육』, 54, 한국국어교육학회, 253~276.
- 장영희(2000), “20대 남녀 사용 어휘의 대비적 고찰”, 『화법연구』, 93~115.
- 지현숙(2007), “한국어 구어 문법 교육을 위한 과제 기반 교수법”, 『국어교육연구』, 20, 247~270.
- Anglin, J.M.(1993), “Vocabulary development: A morphological analysis”, Monographs of the Society for Research in Child Development, 58.(No.10)
- Hughes, R.(2002), Teaching and Researching Speaking, Longman.

<초록>

초등학생의 구어에 나타난 어휘 빈도와 분포도 조사

이필영 · 김정선

본 연구에서는 초등학생의 실제 대화 자료를 수집·분석하여 어휘의 사용 빈도와 분포도를 조사하고, 문어와의 비교를 통해 구어 어휘의 특징을 살펴보았다. 초등학생의 대화에 나타난 전체 어휘 목록을 추출하고 어휘별 사용 화자수와 빈도를 분석한 결과, 총 5,982개의 어휘가 101,167회 출현하였으며, 전체 어휘 중 사용 화자 수가 많은 어휘는 문법적 형태소들이 차지하였다.

범주별로 정리해보면 다음과 같다. 1) 체언은 구어에서는 대명사, 구어형들이 높은 순위를 차지하였고, 문어에서는 의존명사가 높은 순위를 차지하였으며, 어휘 목록에서도 차이를 보였다. 2) 용언은 다른 범주에 비해 고빈도 어휘 목록에서는 차이가 크지 않았지만, ‘대하다, 위하다, 통하다, 의하다’와 같은 어휘들이 문어 목록에서만 나타나고 있었다. 3) 수식언은 구어에서는 부사가, 문어에서는 관형사가 높은 순위를 차지하였으며, 정상 관형사는 구어와 문어 모두에서 출현하는 빈도가 낮았다. 4) 관계언은 구어에서는 주격 조사 ‘가’가, 문어에서는 관형격 조사 ‘의’가 고빈도로 나타나 차이를 보였다. 특히, 관계언은 모음변이형이나 비표준형이 구어에 많이 나타나고 있었다. 5) 의존 형태는 구어에서는 종결어미 ‘어’가, 문어에서는 종결어미 ‘다’가 가장 높은 순위를 차지하였으며, 출현 어휘 목록에서도 구어와 문어의 차이가 있었다.

【핵심어】 어휘, 사용화자수, 빈도, 체언, 용언, 수식언, 관계언, 의존형태, 구어, 문어, 초등학생

<Abstract>

A Study on the Vocabulary Frequency and Distribution
in Spoken Language of Elementary School Students

Lee, Phil-young · Kim, Jung-sun

This study collected and analyzed actual conversational data of elementary school students to examine the frequency and distribution of the vocabulary used by them and compared the findings with those of written language to understand the characteristics of spoken vocabulary. After making up an overall vocabulary list based on the conversations among elementary school students, the investigator analyzed the number of speakers and frequency according to vocabulary.

The results were summarized by the categories : 1) As for uninflected words, while pronouns and spoken words ranked a high place in spoken language, dependent nouns did in written language, which means there was a difference in the vocabulary list ; 2) The category of declinable words didn't vary much in the list of high-frequency vocabulary compared to other categories. In written language, the list included such words as '대하다, 위하다, 통하다, and 의하다.' ; 3) As for modifiers, while adverbs were placed at a high place in spoken language, pre-nouns did in written language. Pre-nouns that present nature or state were low in frequency both in spoken and written language ; 4) As for connective words, the subjective postposition "가" and the modifier postposition "의" demonstrated a high frequency in spoken and written language, respectively. ; and 5) As for dependent forms, the final endings "어" and "다" recorded the top place in spoken and written language, respectively.

[Key words] vocabulary, number of speakers, uninflected words, declinable words, modifiers, connective words, dependent forms, spoken language, written language, elementary school students

【부 록】

<형태소 분류 체계 및 약호>*

대분류	소분류	세분류
(1) 체언	명사NN	일반명사\NNG
		고유명사\NNP
		의존명사\NNB
	대명사NP	
	수사NR	
(2) 용언	동사VV	
	형용사VA	
	보조용언VX	
	지정사VC	긍정지정사\VCP 부정지정사\VCN
(3) 수식언	관형사MM	
	부사MA	일반부사\MAG 접속부사\MAJ
(4) 관계언	격조사JK	주격조사\JKS
		보격조사\JKC
		관형격조사\JKG
		목적격조사\JKO
		부사격조사\JKB
		호격조사\JKV
		인용격조사\JKQ
	보조사JX	
	접속조사JC	

* 이러한 분류 체계는 21세기 세종계획의 안을 따른 것이다.

대분류	소분류	세분류
(5) 의존형태	어미E	선어말어미EP
		종결어미EF
		연결어미EC
		명사형전성어미ETN
		관형형전성어미ETM
	접두사XP	체인접두사XPN
	접미사 XS	명사파생접미사XSN
		동사파생접미사XSV
		형용사파생접미사XSA
	어근XR	

<국립국어원 약호>

범 주	약 호	품 사	약 호	품 사
체 인	명	일반명사(NNG)	대	대명사(NP)
	의	의존명사(NNB)	수	수사(NR)
용 언	동	동사(VV)	형	형용사(VA)
	보	보조 용언(VX)	지	긍정 지정사(VCP)
	지	부정 지정사(VCN)		
수식언	관	관형사(MM)	부	일반 부사(MAG)
	접	접속 부사(MAJ)		
조 사	주	주격 조사(JKS)	보	보격 조사(JKC)
	관	관형격 조사(JKG)	목	목적격 조사(JKO)
	부	부사격 조사(JKB)	호	호격 조사(JKV)
	인	인용격 조사(JKQ)	조	보조사(JX)
	접	접속 조사(JC)		
어 미	선	선어말 어미(EP)	종	종결 어미(EF)
	연	연결 어미(EC)	명	명사형 전성 어미(ETN)
	관	관형형 전성 어미(ETM)		