

<http://dx.doi.org/10.20880/kler.2018.53.2.205>

『조선어문』 7학년 어휘의 계량 연구

임칠성 전남대학교 국어교육과 교수

- * 이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2017S1A5A2A01024318)

- I. 연구의 배경
- II. 계량 자료와 계량 방법
- III. 어휘의 분포
- IV. 결론

I. 연구의 배경

어휘 계량 연구는 대상 어휘의 빈도별, 품사별, 어종별 분포 실태와 그 실태가 보여 주는 어휘 분포적 특성을 파악하는 연구이다. 어휘 분포 실태는 사전 편찬의 기초 자료일 뿐만 아니라 교과서나 출판물 등의 어휘 수준을 결정하는 기초 자료를 제공한다. 빈도별 분포를 통해 어휘의 수준을 결정하기도 하고 교육용 어휘나 기초 어휘¹⁾를 추출하는 자료로 활용된다.

이런 이유 때문에 많은 나라들이 대대적인 어휘 조사를 한다.²⁾ 우리나라 는 1956년 당시 문교부 주관으로 교과서, 신문, 잡지 등을 대상으로 대대적인 어휘 계량을(개별 어휘가 56,485개) 실시하여 그 결과를 『우리말 말수 사용의 잣기 조사』라는 보고서로 펴낸 적이 있다. 이후 2002년에 이르러서야

-
- 1) 기초 어휘란 사용 빈도가 높을 뿐만 아니라 어휘 생산력이 높은 어휘로서 일상 소통의 기본이 되는 어휘이면서 해당 언어 자료의 바탕이 되는 어휘를 가리킨다.
 - 2) 어휘 계량과 그 연구가 활발한 나라는 일본이다. 일본에서는 일본의 국어연구소 주관으로 잡지 90종의 어휘를 대대적으로 조사하여 그 결과를 1962년 『現代 雜誌 九十種の用語用字』로 펴냈다. 이후 일본의 계량국어학회가 활발하게 계량 연구를 하고 있다.

국립국어원에 의해 대대적인(계량 대상 어절 1,531,966개) 어휘 계량 연구가 있었고, 그 결과물이 『현대 국어 사용 빈도 조사』로 나왔다. 북한에서도 1993년 대대적인 어휘 계량을 하여(개별 어휘 39,389개) 그 결과를 『조선어 빈도 수 사전』으로 펴냈다.

어휘 계량 연구는 특정한 언어 자료의 어휘 분포를 통해 언어 사용 집단들의 의식을 파악하게 해 준다. 임칠성·노시훈(2003)에서 5·18 민주화운동 전, 중, 후에 나온 성명서와 유인물의 어휘를 분석하여 당시 관련 집단들의 사태 인식을 조사 연구한 것이 어휘 계량을 통한 언중의 인식 파악 연구의 한 예라고 할 수 있다.

그런데 국어 어휘의 계량 연구사를 살펴보면, 어휘 계량 연구가 가장 관심을 보이는 언어 자료는 국어 교과서이다. 국어 교과서 어휘의 계량 연구는 다른 계량 연구에 비해 대대적인 작업이 아니지만 해당 언어의 표본적인 언어 실태를 파악할 수 있는 단서를 제공해 준다. 그리고 국어 교과서 어휘 계량 연구는 현재는 물론 미래 우리말 사용 양상을 파악할 수 있는 잣대가 된다.³⁾ 외국어로서의 한국어 교재의 어휘를 계량 연구하는 것이 같은 맥락이다.

우리나라에서 대대적인 어휘 계량이 시작된 1956년 이후 국어 교과서의 어휘 계량 연구가 간헐적으로 이어졌다. 우리나라의 광범위한 어휘 조사의 역사는 김한샘(2012)에 잘 정리되어 있다.⁴⁾

이 연구에서는 중국 조선족자치구민족학교에서 현재 사용되고 있는 『조선어문』 7학년의 어휘를 계량하여 『조선어문』 어휘의 분포 실태를 밝히

3) 현재 우리나라에서는 어휘 계량 조사에 의거하여 국어 교과서를 편찬하지 못하고 있다. 그러나 국어 교과서를 편찬하는 이들과 심사하는 이들은 글을 고르거나 쓸 때, 혹은 심사를 할 때, 외국어나 외래어가 지나치게 사용되었는지, 학령의 수준에 비해 어려운 어휘가 사용되었는지를 고려할 수밖에 없기 때문에 항상 어휘의 분포(어종별, 빈도별 등)를 고려하지 않을 수 없게 된다.

4) 이충우(1994)는 어휘 계량을 통해 교육용 어휘를 연구하여 책으로 펴낸 바 있다. 임칠성·水野俊平·北山一雄(1997)은 어휘 계량 연구를 종합적으로 정리하였다. 김광해(2003)에서는 어휘 계량을 바탕으로 등급별 교육용 어휘를 분류하여 제시하고 있다.

고자 한다. 그리고 이 어휘 분포를 남한의 국어 교과서 어휘 계량 연구 결과(이성연, 2007)와 북한의 국어 교과서 어휘 계량 연구 결과(민병곤, 1999)와 비교하고, 또 조남호(2002)와 비교하여 조선어 어휘 분포의 특성을 살펴보자 한다.『조선어문』의 어휘 계량을 통해 조선족 언어의 현실을 살펴볼 수 있을 것이다. 또 이를 남한과 북한의 교과서와 비교함으로써 조선족의 조선어 정책의 일상을 살펴볼 수 있는 단서를 제공할 수 있을 것이다.

II. 계량 자료와 계량 방법

중국 조선족자치구는 민족학교에서 조선어로 조선족의 언어, 사회, 문화, 역사 등을 의무 교육한다.『조선어문』은 이 민족학교의 국어 교과서이다. 중국 조선족자치구는 기존에 중국의 교육과정을 조선족 민족학교에 적용하던 것에서 벗어나 민족의식을 강화하면서 2010년 『의무교육조선족학교 조선어문과정표준』을 공포하였다. 그 이후 조선어 교육의 방향을 크게 바꾸고, 이에 의거하여 ‘전국조선문교재심사위원회 2014 심사통과’ 『의무교육조선족 학교교과서 조선어문』을 출판하여 현재 사용하고 있다(임칠성, 2016). 이 연구는 이렇게 출판된 『조선어문』 7학년 상권과 하권의 어휘를 대상으로 한다.

『조선어문』 어휘의 계량을 위해서 몇 가지 단계를 거쳤다. 우선, 교과서 상권과 하권의 내용을 컴퓨터로 입력하여 한글 워드 파일로 변환하였다. 교과서의 구성 중 ‘제목, 목차, 학습 안내’는 계량 대상에서 제외하였다. 다만 학습활동은 교과서의 주요 내용으로 인식하여 계량에 포함하였다. 부록과 각주도 계량 대상에서 제외하였다.

연구 결과를 비교하기 위해서는 주요한 계량 단위화 방법이 차이가 나지 않아야 한다. 컴퓨터로 입력된 교과서의 어휘들을 기본적으로 임칠성(2003)과 민병곤(1999)의 계량 단위화 방법에 따라 계량 단위화 하면서 필

요한 정보를 태깅하였다.⁵⁾

계량 단위화의 기본 원칙은 국립국어원의 『표준국어대사전』의 표제어이다. 표제어로 등록된 단위로 계량 단위화를 하였다.⁶⁾ 다만, 보조 동사와 보조 형용사는 본용언에 통합하여 계량하였다. 지명이나 인명과 같은 고유명사는 표제어로 제시되지 않은 경우 계량에서 제외하였다. ‘나, 너, 저’, ‘내, 네, 제’는 주격형과 속격형을 구분하여 별도로 계량하였다. 다만, 하나의 표제어일지라도 예를 들어, ‘조금’이 명사와 부사로 구분되어 나와 있는 경우처럼 품사가 다른 경우에는 따로 단위화 하였다.

『조선어문』에 ‘사기꾼, 벤취, 아빠트, 맙소서’ 등과 같이 표기 방식의 차이로 인해 『표준국어대사전』에 등재되지 않은 단어라 판단되는 경우에는 ‘사기꾼, 벤치, 아파트, 맑소사’ 등 『표준국어대사전』 수록 단어로 수정하여 단위화 하였다. 표기 방식의 차이가 아니라 어휘 자체가 미등재 되어 있는 경우는 따로 단위화 하였다.

『표준국어대사전』에 북한말로 제시된 경우에는 이를 따로 분류하였다. 다만 사전에 북한말로 분류되어 있더라도 한국말의 의미가 함께 있는 경우에는 한국말로 분류하였다. 예를 들면, ‘애미, 여직, 열적다, 제끼다’와 같은 경우가 그것이다. 이들은 이에 상응하는 한국어가 있어서 해당 한국어로 분류하였다.

계량 단위화가 끝난 자료를 어휘 계량 프로그램인 ‘깜짝새’를 사용하여 빙도를 측정하였다. ‘깜짝새’의 결과를 엑셀 파일로 변환하고, 이 엑셀 파일에서 어종과 품사 등 필요한 정보를 분류하여 기입한 후 필요한 결과를 도출

-
- 5) 이성연(2007)과 조남호(2002)는 계량 단위화 방법이 이 연구와 다른 경우가 있었다. 『조선어문』과 이들을 비교하면서 이 점을 염두에 두었다.
 - 6) 계량 단위화란 해당 어휘가 하나의 개별 어휘로 확정되도록 사전의 표제어로 고치는 작업을 의미한다. 여기에는 단위 구분을 위한 태깅 작업이 수반된다. 계량 단위화 방법은 계량 대상과 목적에 따라 달라질 수 있다. 그러나 다른 계량 결과와 비교하기 위해서는 단위화 방법이 가능하면 동일해야 한다.

하였다. 상권과 하권을 각각 계량한 다음 상권과 하권의 데이터를 엑셀 파일로 합해서 7학년 전체의 결과를 얻었다.

『조선어문』의 어휘 규모는 〈표 1〉과 같다.⁷⁾

〈표 1〉 대상 어휘의 규모

교과서	개별 어휘	운용 어휘	운용 비율
상권	5065	22666	4.48
하권	4964	22645	4.56
계	7554	45311	6.00

개별 어휘는 단위 어휘이고, 운용 어휘는 이 단위 어휘가 사용된 어휘이다. 그러므로 개별 어휘의 수는 단위 어휘의 개수이고, 운용 어휘는 이 단위 어휘가 사용된 총 개수이다. 일본의 계량학에서는 이들을 각각 ‘이어(異語), 연어(延語)’라고 부른다. 임지룡(1991)에서는 ‘이어’에 대해 ‘개별 어휘’라는 용어를 사용하고, ‘연어’에 대해 ‘연어휘’라는 용어를 사용하였다. 임칠성(1993)에서는 ‘이어’ 대신 ‘개별 어휘’를, ‘연어’ 대신 ‘운용 어휘’를 사용하였다. 그리고 하나의 개별 어휘가 몇 번이나 사용되었는지를 나타내는 평균 운용 개수를 ‘운용 비율’이라 불렀다. 이 글의 용어는 임칠성(1993)의 용어에 따랐다.

『조선어문』의 어휘 분포를 민병곤(1999), 이성연(2007), 조남호(2002)의 연구 결과를 통해 남한과 북한의 분포와 비교하였다. 그러나 『조선어문』과 민병곤(1999), 이성연(2007), 조남호(2002)은 어휘 조사의 시기와 대상이 조금씩 다르다.⁸⁾

민병곤(1999)에서는 1989년, 1990년에 북한에서 간행한 『국어: 고등중

7) 이 글의 모든 통계는 소수점 셋째 자리에서 반올림하였다.

8) 2014년 이후 남한과 북한의 국어교과서의 어휘를 계량한 결과를 찾아볼 수 없었다. 이러한 차이는 이 논문의 한계이다.

학교』 1, 2, 3권의 어휘를 계량하였다. 교과서 내 본문과 학습활동을 모두 계량 대상에 포함하였다. 이 연구에서는 북한의 고등중학교 1학년부터 3학년 까지의 국어교과서 고빈도 어휘를 각각 300위까지 부록으로 제시하고 있어서 서로 직접적인 비교가 가능하였다. 북한의 고등중학교 1학년은 7학년에 해당한다. 이성연(2007)은 북한에서 2002년에 간행한 『고등중학교 국어』 1, 2, 3과 남한에서 2005년에 간행한 『중학교 국어』 1-1에서 3-2 교과서를 대상으로 계량하였다. 남한 교과서는 각 단원 앞뒤에 나오는 학습목표와 학습 활동을 포함하지 않았고, 북한 교과서는 각 단원 앞과 뒤에 나오는 속담 등 의 보충 자료를 포함하지 않았다.

민병곤(1999)의 계량 결과가 이 연구와 계량 방법이 거의 동일하였기 때문에 『조선어문』을 북한 교과서와 비교할 때는 민병곤(1999)의 결과를 이용하였고, 이성연(2007)은 고빈도어가 따로 제시되어 있지 않고, 계량 방법이 명확하게 제시되어 있지 않아서 북한 교과서의 경우는 참고하지 않고 남한 교과서의 경우만 비교하였다. 『조선어문』의 고빈도어를 남한의 고빈도어와 비교할 때는 국립국어원에서 연구보고서로 나온 조남호(2002)의 자료를 참고하였다. 조남호(2002)는 국립국어원에서 2000년부터 2002년까지 대대적인 어휘 계량을 통해 5,965개의 학습용 어휘를 추출한 결과이다.

어휘 계량의 시기와 대상, 그리고 일부 방법이 조금씩 다름에도 계량 방법이 기본적으로 같고, 민병곤(1999)과 이성연(2007)이 중학교 국어교과서의 어휘를 계량하였다는 점과 조남호(2002)도 한국어 학습용 어휘 선정을 목적으로 하였다는 점에서 이들 연구들의 결과를 비교 대상으로 선택하였다.

III. 어휘의 분포

1. 빈도별

고빈도 어휘의 분포는 연구 대상이 어떤 것에 관심을 두고 있는지를 보여준다. 빈도 순 50위까지의 어휘는 <표 2>와 같다.

<표 2> 『조선어문』의 고빈도 어휘 50개

순위	단어	빈도	순위	단어	빈도	순위	단어	빈도
1	하다	1345	18	때	229	35	못하다	125
2	것	693	19	가다	221	36	뜻	120
3	나	676	20	말하다	217	37	속	117
4	있디0형	597	21	어머니	213	38	알아보다	108
5	보다	490	22	한01	211	39	보이다	107
6	되다	412	23	않다	202	40	잘	107
7	수0의명	369	24	같다	201	41	듣다	104
8	있디0동	368	25	쓰디0글	184	42	미음	104
9	그	353	26	선생님	183	43	아니다	103
10	없다	344	27	대하다	175	44	알다	102
11	이0관	325	28	그0대	168	45	먹다	101
12	주다	305	29	다음	163	46	사실	97
13	말	304	30	일	150	47	두0관	94
14	사람	267	31	쓰다	148	48	집	93
15	우리	254	32	오다	134	49	더	89
16	글	247	33	어떤	131	50	또, 시, 지다	88
17	내0대	235	34	아이	128			

대부분의 어휘 계량에서 고빈도 어휘들이 차지하는 비중은 매우 높다. 그리고 특정 어휘들이 최상위 빈도를 차지한다. 『조선어문』7학년의 경우 1위부터 50위까지 52개 어휘의 운용 어휘 개수는 총 12477개로 전체 운용 어휘 45311개의 27.54%에 해당한다. 한국어의 경우 ‘하다(1위), 것(2위), 있다(4위), 보다(5위), 되다(6위), 수(7위), 있다(8위)’는 어휘 계량에서 최상위를 차지한다.

『조선어문』에서는 남한이나 북한과 다르게 ‘나’가 3위라는 최상위에 들어 있다. ‘나’는 민병곤(1999)에서는 18위이고, 조남호(2002)에서도 8위이다. ‘내’의 경우 민병곤(1999)은 61위인데 반해 『조선어문』은 17위이다. 조남호(2002)에서 ‘내’는 계량 단위 어휘에 포함되어 있지 않다. 이는 ‘내’를 ‘나’에 포함하여 계량한 까닭이다. 따라서 ‘나’와 ‘내’를 분리하면 조남호(2002)의 ‘나’의 순위는 8위 이하로 떨어질 것이다. 반면에 ‘우리’라는 어휘는 『조선어문』은 15위이지만 민병곤(1999)에서 8위이다. 조남호(2002)에서 ‘우리’는 13위이다.

이로만 미루어 보면 북한 교과서는 집단의식을, 조선족 교과서는 개인의식을 더 강하게 보여주고 있다는 것을 추론할 수 있다. 『조선어문』 남한보다도 더 강한 개인주의적 성향을 보인다.

『조선어문』의 어휘들은 오래된 시대의 어휘라는 인상을 준다. 『조선어문』에서는 ‘엄마 : 어머니 : 어머님’의 빈도 순위가 483 : 22 : 399이다. 7학년의 경우를 보자면 ‘엄마’는 최근에 사용량이 많아진 어휘이지만 ‘어머니’나 ‘어머님’은 엄마보다 오래된 시기에 많이 사용된 어휘이다. 조남호(2002)에서는 이들 비율이 240 : 106 : 2902로 나타난다. ‘어머니’라는 어휘가 민병곤(1999)에서 조차 300위권 밖에 있다.⁹⁾ ‘아버지’라는 표현을 살펴보면 『조선어

9) 이성연(2007)은 중학 국어 교과서 1~3학년 모두를 계량한 결과이다. 그리고 논문에 고빈도 자료를 제시하지 않고 품사별로 고빈도어를 따로 제시하였다. 이성연(2007)의 남한 국어 교과서를 살펴보면 ‘나, 우리, 그’가 대명사 내에서는 1위, 2위, 4위이고, ‘어머니’와 ‘아버지’가 명사 내에서만 4위와 6위이다. 그러나 전체 빈도 순위가 아니기 때문에 『조선

문』에서는 74위이지만 조남호(2002)에서 '아버지'는 153위이다.

『조선어문』에서는 '스승, 스승님'이 각각 9회와 7회 사용되고, '애비, 예미, 육친, 집사람, 집식구' 등과 같은 어휘들이 사용되고 있다. 이런 어휘들은 남한의 중학교 교과서에서 찾아보기 어려운 어휘들이다.

이런 어휘 분포는 조선족 사회의 인간관계 어휘가 남한에 비해 비교적 오래된 시대에 사용된 어휘라는 점을 확인시켜 준다. 『조선어문』에 '사양餉養, 설(누에설), 양봉원, 양봉장, 연자간, 오령, 잠실蠶室, 쟁기, 풋옥수수' 등과 같이 남한의 중학생들에게는 생소한 농업 어휘들이 많이 사용된 것도 오래된 시대의 어휘라는 느낌을 더해준다.

『조선어문』에는 『표준국어대사전』 미등재 어휘가 개별 어휘로 55개 있다. 미등재 어휘가 여전히 많이 남아있다는 것은 우리에게 조선족 언어인 조선어가 생경하다는 느낌을 줄 뿐만 아니라, 『조선어문』이 조선족이 중국에 정착 당시에 사용하였던 어휘들을 계승어로 교육하고 있다는 증거라고 본다.

미등재 어휘는 다음과 같다. 제시된 어휘 중에서 '0' 이하는 '0' 앞의 어휘를 구분하기 위한 태깅이다. 예를 들어, '썩하면0툭하면'에서 '0툭하면'은 '썩하면'의 의미를 더해 표시한 것이다.

분수못(4), 기윽자(3), 병급표(3), 쯔(3), 갑급표(2), 막대사탕(2), 문예야회(2), 산수세트(2), 스키신(2), 주공부(2), 지식성(2), 진공서(2), 찬해나다(2), 체육장(2), 폐지(2), 포고판(2), 포병련(2), 한오리한오리(2), 노벨문학상(1), 농악기(1), 문치무력(1), 바이올린수(1), 방적선(1), 별통문(1), 복합선(1), 부족점(1), 삼각모(1), 서면어(1), 선전란(1), 선전화(1), 솜배자(1), 수완군(1), 시체멋(1), 쑥대불(1), 아버지절(1), 암니(1), 야영대오(1), 어머니절(1), 옴니(1), 용송거리다(1), 운율성(1), 조련관(1), 중군막(1), 집거미(1), 썹하면0툭하면(1), 쪼로롱(1), 채색기(1), 청년절(1), 팔간0집(1), 팔괘진(1), 포도장선(1), 하많다(1), 허가론0許可論(1), 헐망하다(1)

어문』의 결과와 단순 비교하기 어렵다.

‘분수못, 스키신, 기윽자, 폐지, 방적선, 찍하면, 청년절, 하많다, 헐망하다’는 『조선말대사전』에 표제어로 등재되어 있다.¹⁰⁾ ‘시체멋’과 ‘운율성’의 경우 『조선말대사전』에 ‘시체, 운율’ 자체가 표제어에서 빠져 있다. ‘쑥대불, 아버지절, 어머니절, 채색기’ 역시 표제어에서 빠져 있다. 이들은 모두 『표준국어대사전』에 등재되어야 할 표제어들이다.

2. 품사별

『조선어문』의 품사별 분포는 <표 3>과 같다.

<표 3> 『조선어문』의 품사별 분포

품사	명사	대명사	수사	동사	형용사	관형사	부사	감탄사	계
개별어휘 (%)	4064 (53.80)	41 (0.54)	23 (0.30)	1834 (24.28)	790 (10.46)	84 (1.11)	681 (9.02)	37 (0.49)	7,554 (100)
운용어휘 (%)	20631 (45.53)	1997 (4.41)	201 (0.44)	12991 (28.67)	4598 (10.15)	1691 (3.73)	3100 (6.84)	102 (0.23)	45,311 (100)
운용 비율	5.08	48.71	8.74	7.08	5.82	20.13	4.55	2.76	6.00

다른 한국어 어휘 계량과 마찬가지로 품사의 비율이 가장 높은 것은 명사(53.80%)이다. 이어 동사 > 형용사 > 부사 > 관형사 > 대명사 > 수사의 순서이다. 『조선어문』에서는 대명사가 운용 비율이 48.71로 가장 높다. 이성연(2007)의 경우 대명사의 운용비율은 25.88이다. 『조선어문』에서 감탄사는 37개의 개별어휘로 102회 사용되어 운용 비율이 2.76로 가장 낮다.

『조선어문』의 품사별 분포를 북한 국어교과서인 민병곤(1999), 남한 국어교과서인 이성연(2007)과 비교하면 <표 4>, <표 5>와 같다.

10) 『조선말대사전』에 ‘급표’, ‘서면질문, 서면토론’이 표제어로 등재되어 있다. ‘선전대, 선전망, 선전문, 선전물, 선전실...’ 등 매우 많은 표제어가 있지만 ‘선전란’은 빠져 있다. ‘노벨상, 벌통, 용송그리다, 수완가, 중군’은 『표준국어대사전』에 표제어로 등재되어 있다.

〈표 4〉 민병곤(1999: 475)의 운용 어휘 품사별 분포(1학년)

품사	명사	대명사	수사	동사	형용사	관형사	부사	감동사	계
운용 어휘	12326 (48.15)	945 (3.69)	192 (0.75)	7576 (29.59)	2148 (8.39)	315 (1.23)	1993 (7.79)	105 (0.41)	25600 (100)

〈표 5〉 이성연(2007: 558)의 품사별 분포

품사	명사	대명사	수사	동사	형용사	관형사	부사	감동사	계
개별어휘 (%)	3273 (51.86)	50 (0.79)	18 (0.29)	1644 (26.05)	526 (8.33)	76 (1.20)	636 (10.08)	88 (1.39)	6311 (100)
운용어휘 (%)	14123 (45.12)	1294 (4.13)	128 (0.41)	9081 (29.01)	2647 (8.46)	1127 (3.60)	2604 (8.32)	205 (0.65)	31299 (100)

전체 어휘 중에서 명사가 차지하는 비율은 개념적, 혹은 지식적인 내용이 어느 정도인지를 가늠할 수 있게 해 준다. 『조선어문』의 경우 운용어휘에서 명사가 차지하는 비율은 45.53%이고, 민병곤(1999)의 경우 48.15%이고, 이성연(2007)의 경우 45.12%이다. 지식의 양으로 볼 때 『조선어문』은 북한 교과서보다 적고 남한 교과서와 비슷한 수준이다.

3. 어종별

『조선어문』의 어종별 분포는 〈표 6〉과 같다.¹¹⁾

〈표 6〉 『조선어문』의 어종별 분포

구분	고유	한자	외래	고유어+ 한자어	외래어+ 고유어/한자어	계
개별어휘 (%)	3707 (48.88)	2485 (32.77)	97 (1.23)	1241 (16.36)	24 (0.32)	7584 (100)
운용어휘 (%)	30955 (68.32)	10201 (22.51)	257 (0.57)	3852 (8.50)	46 (0.10)	45311 (100)

11) ‘외래어+고유어/한자어’는 ‘외래어+고유어’와 ‘외래어+한자어’를 합한 것이다.

이성연(2007), 민병곤(1999)의 어종별 분포는 〈표 7〉, 〈표 8〉과 같다.¹²⁾

〈표 7〉 이성연(2007: 561)의 어종별 분포

구분	고유	한자	외래	혼종어	계
개별어휘 (%)	3396 (53.81)	1901 (30.12)	89 (1.41)	925 (14.66)	6311 (100.000)
운용어휘 (%)	22916 (73.22)	6208 (19.83)	262 (0.84)	1913 (6.11)	31299 (100)

〈표 8〉 민병곤(1999: 472)의 어종별 분포

구분	고유	한자	외래	고유어+ 한자어	고유어+ 외래어	계
개별어휘 (%)	2728 (62.20)	1133 (25.83)	30 (0.68)	493 (11.24)	1 (0.02)	4386 (100)
운용어휘 (%)	18709 (73.08)	4949 (19.33)	48 (0.19)	1892 (7.39)	2 (0.004)	25600 (100)

고유어의 비율을 『조선어문』: 이성연(2007) : 민병곤(1999)'으로 살펴보면 개별 어휘의 경우 48.88 : 53.81 : 62.20이고 운용 어휘의 비율은 68.32 : 73.22 : 73.08이다. 『조선어문』이 남한과 북한에 비해 낮다. 반면, 한자어의 비율을 살펴보면 개별 어휘의 경우 32.77 : 30.12 : 25.83이고 운용 어휘의 경우 22.51 : 19.83 : 19.33이다. 외래어의 비율은 이성연(2007)의 남한이 가장 높고 『조선어문』과 민병곤(1999)이 함께 매우 낮다.

민병곤(1999)에서 고유어 비율이 높고 외래어 비율이 낮은 것은 강한 민족의식의 발로라고 할 수 있다. 『조선어문』은 이성연(2007)과 비슷하다. 이는 조선족이 강한 민족의식 교육을 내세우지만(임칠성, 2016) 어휘 분포에서는 그것을 찾아보기 어렵다는 것을 방증한다. 이러한 어종 분포로 볼 때,

12) 민병곤(1999)의 중1에서 중3까지 자료 중 중1 자료만 편집했다. 이성연(2007)의 남한어와 북한어 중 남한어만 편집했다.

북한은 문화어로 언어를 정비하면서 한자어 대신 고유어를 살려 쓰고자 노력했지만 조선족은 한자를 사용하는 중국어의 영향을 받으며 정착 당시의 언어들을 그대로 계승하여 사용하고 있다는 추론을 가능하게 한다.

『조선어문』은 한자어의 비중이 높을 뿐만 아니라 ‘특점(48회), 표달하다(4회), 방애0妨礙(3), 시다림0尸荼林(3) 사로0思路(2), 사양0飼養(2)’ 등과 같이 우리 중학생들은 물론 성인들도 이해하기 어려운 한자어(한자어+고유어 포함)도 많이 사용되고 있다.¹³⁾〈부록 1〉 이런 한자어들은 『표준국어대사전』에 북한말이 아니라 우리말로 등재된 어휘들로서 지금은 일상적으로 사용되지 않는 표현들이다. 조선족이 정착 당시에는 남한이나 북한과 같은 말을 사용했다는 점을 고려해 보면 남한의 성인조차 이해하기 어려운 한자어가 많다는 것은 조선족이 정착 당시의 언어를 계승하여 사용하고 있다는 방증이라고 본다.

『조선어문』에는 우리 중학생들이 쉽게 이해하기 어려운 고유어들도 많이 사용되었다. 〈부록 2〉 ‘틀어쥐다(26회), 말마디(5회), 불칼(5회), 셀(4회), 소슬하다(4회), 노상(3), 적삼(3)’ 등이 그런 어휘들이다. 이런 고유어들은 남한에서도 살려쓸 만한 것들이기는 하지만 지금은 일상적으로 사용되지 않는 어휘들이다. 이 또한 어려운 한자어와 마찬가지로 조선족이 정착 당시의 언어를 계승적으로 사용하고 있다는 방증이다.

『조선어문』에는 ‘차요하다(9회), 모대기다(3회), 분주소(3회), 굽닐다(2회), 려지꿀(2회)’ 등과 같이 북한어의 사용이 높다.¹⁴⁾〈부록 3〉 북한어는 운용 어휘로 볼 때 800회 사용되었다. 800회는 전체 운용어휘 45311개의 1.77%에 해당한다. 개별어휘로 보면 303개 어휘가 사용되어, 전체 개별어휘 7554

-
- 13) 필자의 주관적인 판단으로 중학생이 이해하기 어려울 만한 고유어나 한자어와 지금은 일상적으로 사용되지 않는 어휘를 선택하였다. 통계적 기준을 제시하지 못한 것은 이 논문의 한계이다.
- 14) ‘북한어’는 어종별 분포에 포함되지 않는다. 그러나 조선족 언어의 역사적 배경으로 볼 때 조선족 언어의 모습을 살피기 위해서는 북한어의 분포를 들여다 볼 필요가 있다.

개의 4.01%에 해당한다. 북한어 가운데 음운 상의 차이 등으로 인해 차이를 보이는 ‘페그나, 엉뎅이, 해여지다, 로부인, 메타’와 같은 어휘를 제외하여 살펴보더라도 『조선어문』에는 남한에서는 이해하기 어려운 150개의 운용 어휘가 사용되었다.

IV. 결론

조선족 민족학교 국어 교과서인 7학년 『조선어문』의 어휘를 계량하고 이를 남한 국어 교과서의 어휘와 북한 국어 교과서의 어휘와 비교하여 살펴보았다. 빈도 분포로 볼 때 『조선어문』은 민족정신 함양이라는 조선족 교육의 목표와 달리 개인주의적 성향을 보인다. 어종별 분포로 볼 때 남한이나 북한보다 고유어의 사용이 적고 한자어의 사용이 많았다. 또, 『조선어문』에는 많은 북한어가 사용되고 있다. 『표준국어대사전』이나 조선말대사전에 등재되지 않은 어휘들을 사용하고 있으며, 남한의 중학생들이 이해하기 어려운 한자어나 고유어들을 사용하기도 한다. 이러한 분포는 조선족이 정착 당시의 언어들을 계승적으로 사용하고 있다는 방증이다.

조선족의 언어는 한국어, 북한어에 이어 세 번째로 많이 사용하는 한국어이다. 그리고 조선족의 언어는 북한어를 기반으로 하고 있다. 이런 상황에서 보자면 조선족의 언어에 대한 연구는 통일 한국을 대비하는 차원이나, 한국어를 발전시키는 차원에서 적극적으로 진작되어야 할 것이다.

* 본 논문은 2018. 4. 26. 투고되었으며, 2018. 5. 8. 심사가 시작되어 2018. 6. 1. 심사가 종료되었음.

참고문헌

- 김광해(1993),『국어어휘론개설』, 서울: 집문당.
- 김한샘(2012),「한국어 어휘 계량 연구의 성과」,『한민족문화연구』41, 39-74.
- 민병곤(1999),「북한 고등중학교 1~3학년 국어 교과서 어휘의 계량적 고찰」,『남천 박갑수 교수 정년퇴임 기념 논문집』, 455-500.
- 이성연(2007),「남북한 중학교 1학년 국어 교과서의 어휘 분포」,『새국어교육』77, 553-582.
- 이충우(1994),『한국어 교육용어휘 연구』, 서울: 국학자료원.
- 임지룡(1991),「국어의 기초어휘에 대한 연구」,『국어교육연구』23, 87-120.
- 임칠성(2003),「기초어휘 선정 방법론」,『새국어생활』13(3), 91-118.
- 임칠성(2016),「중국 조선족의 화법 교육 연구:『조선어문』을 중심으로」,『화법연구』32, 153-183.
- 임칠성·水野俊平·北山一雄(1997),『한국어 계량 연구』, 광주: 전남대출판부.
- 임칠성·노시훈(2003),「5·18 항쟁 관련 유인물과 성명서 어휘의 계량 연구」,『민주주의와 인권』4(1), 179-237.
- 조남호(2002),『현대 국어 사용 빈도 조사: 한국어 학습용 어휘 선정을 위한 기초조사(연구보고서 2002-1-17)』, 서울: 국립국어연구원.
- 國語研究所(1962),『現代雑誌 九十種の用語用字』, 東京: 秀英出版.

〈부록 1〉 중학생에게 난해한 『조선어문』의 한자어(+ 고유어)

특점(48), 성구成句(29), 소학교(26), 형상적(12), 과문(11), 사시절(8), 수립(7), 교편(4), 표달하다(3), 방애0妨礙(3), 부동하다(3), 생령(3), 시다림0茶林(3), 심산(3), 어언간(3), 영양물0營養物(3), 잡박(3), 중약(3), 총집(3), 합리하다(3), 경상(2), 명문0明文(2), 미구未久(2), 백화만발하다(2), 부절히(2), 사로0思路(2), 사양하다0飼養 - (2), 성상(2), 습자첩(2), 신근하다(2), 완정하다(2), 왕가물0(2), 원립(2), 유관(2), 유화하다(2), 절주(2), 정거하다0정차(2), 중약재(2), 천야만야(2), 철리(2), 충차(2), 포가0砲架(2), 허장성세(2), 가액질하다(1), 교양하다(1), 교오(1), 군별(1), 대상성(1), 대통로(1), 등지(1), 무뢰배(1), 미구하다(1), 민항(1), 방사기(1), 버선본(1), 변0재앙(1), 비단술(1), 사모0관복모자(1), 사생0師生(1), 사시장철(1), 사양0飼養(1), 사처0사방(1), 산쥐(1), 산지사방(1), 삼밭(1), 삽시0삽시간(1), 상학(1), 상학하다0上學 - (1), 생신하다0형(1), 생초(1), 서찰0書札(1), 서한0書翰(1), 설명성(1), 설복(1), 설복력(1), 소학생(1), 수만금(1), 수태(1), 수행승0修行僧(1), 식솔(1), 신고0고생(1), 심입(1), 애간(1), 양봉원(1), 양봉장(1), 언명하다(1), 연자간(1), 영걸(1), 오령(1), 육친(1), 응방(1), 의론(1), 자신심(1), 잔광(1), 전도0앞길(1), 좀도적(1), 종시0부(1), 주렴0珠簾(1), 직소0直所(1), 진감하다(1), 진장(1), 진종일(1), 징모(1), 창성(1), 천공0天空(1), 천애0天涯(1), 청청하다0青青 - (1), 총체0總體(1), 통고(1), 편폭(1), 표달되다0북(1), 표시기(1), 표연히(1), 필경(1), 필득(1), 하학하다(1), 하학0下學(1), 한담(1), 허실(1), 허심히(1), 현성(1), 혼곤히(1), 화미조0새이름(1), 화밀(1), 화필(1), 회억(1), 회억하다(1), 후비대(1), 흐름선(1)

〈부록 2〉 중학생에게 난해한 『조선어문』의 고유어

틀어쥐다(26), 두텁다(6), 말마디(5), 불칼0벼락(5), 셨0누에셨(4), 소슬하다(4), 노상0부(3), 알심(3), 적삼(3), 깍정이(2), 비라리(2), 새되다(2), 앵돌아지다(2), 열쏭덜쏭하다(2), 홀제0부(2), 내포하다(1), 노0꼬아만든 줄(1), 뉴잦히다(1), 도리반거리다(1), 보리밭머리(1), 살바람(1), 소소리(1), 시툭하다(1), 아지미(1), 안반침(1), 안반침하다(1), 앞내(1), 앞별(1), 야드르르하다(1), 어섯눈(1), 여름비(1), 여름살이(1), 여직(1), 오줌자루(1), 우레0천둥(1), 옥이다(1), 울바자(1), 이리루하다(1), 이태(1), 잠비(1), 저저마다(1), 지중지종(1), 차하다(1), 꾀뜩(1), 한뉘(1), 한옆(1), 혜치다(1), 흰서리(1),

〈부록 3〉 『조선어문』의 북한어

안해(11), 차요하다(9), 인차(8), 배워주다(5), 어금이(5), 바위돌(4), 반명제(4), 지팽이(4), 표달하다(3), 모대기다(3), 분주소(3), 짓쫓다(3), 총적(3), 표달(3), 가제천(2), 구름본(2), 굽널다(2), 귀맛(2), 꿩지다(2), 려지꿀(2), 로동별(2), 목책(2), 몰키다(2), 비양거리다(2), 소조(2), 오돌차다(2), 장벼(2), 줄창(2), 터치다(2), 한별(2), 호함지다(2), 가녁(1), 가득누에(1), 고꽤(1), 곱대거리(1), 굽인돌이(1), 근로인민(1), 단설기(1), 단장0 -醬(1), 대가정(1), 도로표식(1), 되거리(1), 두점(1), 딱친구(1), 뻃하다(1), 뛰기(1), 록포도(1), 망짝(1), 모대김(1), 몰붓다(1), 무져놓다(1), 뭉청(1), 박산(1), 방죽0防築(1), 사려물다(1), 사탕웃(1), 선코(1), 섭쓸리다(1), 소리마루(1), 소리빛갈(1), 소털(1), 애어리다(1), 역사질0役事 - (1), 올리막길(1), 올방자(1), 운수대(1), 원주필(1), 의력(1), 이상분(1), 이윽하다(1), 입귀(1), 정지간(1), 중동무(1), 지지우다(1), 짹자그르르(1), 탈리다(1), 편역(1), 표달되다(1), 푸르싱싱(1), 함박꽃웃음(1), 호상(1), 혼자소(1), 흙구뎅이(1), 희슥희슥하다(1)

『조선어문』 7학년 어휘의 계량 연구

임칠성

이 연구는 『조선어문』 7학년의 어휘 분포를 빈도별, 품사별, 어종별로 살피고 그 결과를 남한과 북한의 그것과 비교했다.

빈도별 분포에서 볼 때, 『조선어문』은 ‘나’의 빈도가 상대적으로 높고 ‘우리’의 빈도가 상대적으로 낮았다. 이는 『조선어문』이 개인적인 성향이 강하다는 것을 방증한다. ‘엄마 : 어머니 : 어머님’의 빈도 비교와 오래된 농업 어휘의 사용 등을 통해 조선족이 정착 당시의 언어를 계승하고 있다는 것을 확인할 수 있었다. 품사별 분포로 볼 때 『조선어문』의 교육 내용이 북한보다는 남한의 교과서와 비교적 비슷한 수준이라고 확인하였다. 어종별 분포로 볼 때 『조선어문』은 남한이나 북한보다 한자어의 비율이 높은 대신 고유어의 비율이 낮고, 북한어와 함께 오래된 고유어, 한자어 어휘들이 많다. 이런 어휘 분포는 조선족이 한자를 사용하는 중국어권 내에서 정착 당시의 언어들을 그대로 계승하여 사용하고 있다는 것을 확인해 준다.

핵심어 『조선어문』, 어휘 계량, 고빈도 어휘, 품사별 어휘 분포, 어종별 어휘 분포, 조선족의 언어

ABSTRACT

Quantitative Study of the 7th Grade “Joseon Eomun” Vocabulary

Im Chilseong

This study examines the Joseon Eomun vocabulary distribution of the seventh grade by frequency, parts of speech, and lexical type, and compared the results with those in the South and North.

In its frequency distribution, Joseon Eomun shows a strong tendency toward ‘individualism’. By comparing the frequency distribution of three words for mother “eom-ma : eo-meo-ni : eo-meo-nim” we can see that Joseon Eomun is the language of past times. Unlike South Korean or North Korean textbooks, the frequency distribution of “eo-meo-mi” is higher than that of “eom-ma” in Joseon Eomun. In terms of distribution of the parts of speech, it was confirmed that the educational content of Joseon Eomun is relatively similar to that of South Korean textbook rather than that of North Korea. For each lexical type distribution, Joseon Eomun has a higher ratio of Chinese characters than that of South Korea or North Korea, a lower ratio of native languages, and higher number of old Chinese words and phrases that are common with North Korean. This distribution of vocabulary shows that the Chinese ethnic Koreans have inherited the language from the time they settled down.

KEYWORDS Joseon Eomun, Vocabulary Measurement, High Frequency Vocabulary, Vocabulary Distribution by Parts of Speech, Vocabulary Distribution by Lexical Types, Language of Chinese Ethnic Koreans