

중국 초등 어문 교과서 수록 고전시가의 말뭉치 기반 분석

김 응*

<目次>

I. 서론	IV. 분석 및 논의
II. 이론적 배경	V. 결론
III. 연구 범위 및 방법	

I. 서론

고전시가는 중국문학의 꽃이라 할 만큼 예로부터 문화적으로 중요한 위치를 차지해 왔으며 현대 중국인들의 언어생활에도 깊숙이 침투해 있다. 중국인들은 초등학교 또는 그 이전부터 당시(唐詩)를 암송하기 시작하고 성인이 되어서도 즐겨 읽고 활용하는데, 고전문학과는 거의 관계가 없어 보이는 경제 관련 기사 제목에도 다음과 같이 고전시가 구절들이 인용되는 것을 흔히 볼 수 있다.

- (1) 柳暗花明又一村, 恐慌性拋售讓你錯過投資最佳時機
버들 길고 꽃 환하더니 또 마을 하나, 패닉셀은 최적의 투자 타이밍을 놓치게 한다 (2020. 3. 9. 智通財經)
- (2) 本是同根生, 相煎何太急, 育碧和EA兩個“孤島”的多年相爭
본래 같은 뿌리에서 났거늘 들들 볶는 게 어찌 이리 급할까, 유비소프

* 대구중학교 교사

트와 EA 두 ‘외판섬’의 해묵은 다툼

(2020.5.11. 騰訊新聞)

하지만 금세기 초부터 여러 연구자들이 지적하였듯 오늘날 우리나라 사설학원, 중·고등학교는 물론 대학에서도 HSK로 대변되는 이른바 ‘실용 중국어’ 교육에 비하여 고전시가는 실용적이지 못한 것으로 여겨져 다소 등한시되고 있는 것이 사실이다.(이동향, 2000; 신하운, 2001; 류성준, 2002; 김정동, 2005)

다행히 이러한 상황에 대한 문제의식은 계속되어 고전시가 교육에 관한 연구 성과들이 꾸준히 나오고 있는데, 이 가운데 정혜영(2008), 황여옥(2009), 최정은(2010), 이혜천(2014) 등은 중국어 교육에 고전시가를 활용하는 방안을 모색한 시도로서 주목할 만하다. 다만 어휘교육적 측면에서 기존 연구들은 고전시가 구절 그 자체에 주목하기보다는 구절 내부에 쓰인 일부 어휘들을 현대중국어 어휘와 관련짓고자 하였다는 점이 아쉬움으로 남는다.

본 연구는 고전시가 구절들이 그 자체로서 하나의 어휘 단위처럼 기능하는 면이 있다고 간주하여, 중국인들의 언어생활에 자주 쓰이는 구절들은 속담, 험후어 등 숙어와 마찬가지로 중국어 교육에서 중요하게 다루어져야 한다고 본다. 이를 위한 초보적 연구로서 본 연구는 중국 초등학교 어문 교과서에 실린 고전시가를 분석 대상으로 삼고자 한다. 초등학교 어문 교과서는 대부분의 중국인들이 의무교육을 통해 학습한 경험이 있으므로 이를 대상으로 하면 그들의 실제 언어생활에 비교적 가까이 갈 수 있을 것이다. 본 연구는 이를 통해 고빈도 시구(詩句) 목록을 추출하고 그것들이 어떠한 특성을 지니는지 분석하며 어떤 작품들을 우선적으로 학습하면 좋을지 알아보는 데 목적을 두고 있다.

그런데 어떠한 어휘 단위가 자주 쓰이는지 아닌지는 연구자의 주관 또는 원어민 몇 명의 어감으로 판단하기 어려운 문제이다. 이에 본 연구는 말뭉치 언어학적 방법을 이용하여 통계적인 분석을 통해 연구의 객관성을 확보하고자 하였다.

II. 이론적 배경

1. 고전문학 작품과 속어

고전시가를 비롯한 중국 고전문학 작품은 속어(熟語)와 밀접한 관계를 가지는데 그 중에서도 특히 성어(成語)와 불가분의 관계에 있다. 徐盛桓(2006a)은 성어사전에 수록된 사자성어 5100개의 출처를 조사하였는데 그 중 옛 시문에서 온 것이 3800개에 이르는 것을 발견하였다. 대부분의 성어가 형식이 고정되고 문언적인 특성을 가지는 것은 바로 이처럼 고전에 출처를 두고 있기 때문이다.

고전문학 작품의 내용이 성어로 굳어지는 방식에는 몇 가지가 있는데, 본 연구와 관련이 가장 깊은 것은 문구를 그대로 가져온 경우이다. 예를 들면 다음과 같은 것들이 있다.

- (3) **心有靈犀一點通** (李商隱, 〈無題〉)
 마음은 무소의 뿔과 같이 통한다
- (4) **柳暗花明又一村** (陸游, 〈遊山西村〉)
 버들 짙고 꽃 환하더니 또 마을 하나

성어 ‘心有靈犀’(마음이 서로 통함)는 (3)에서 나왔고 ‘柳暗花明’(어려운 가운데 전환점이 되는 기회를 만남)은 (4)에서 나왔다. ‘每逢佳節倍思親’(王維, 〈九月九日憶山東兄弟〉)과 같은 시구 전체를 성어로 간주하는 경우도 있는데(子規, 2020), 이것을 성어로 볼 것인지에 대해서는 신중한 논의가 필요하겠지만 고전시가 구절과 성어 사이의 경계가 모호한 지점이 존재하는 것은 확실해 보인다.

오늘날 쓰이는 상당수의 성어가 고전에 출처를 두고 있음을 고려하면, 현대 중국인들이 즐겨 인용하는 고전시가 구절은 그 자체로서 성어와 유

사한 기능을 지니며 차후 성어로 발전할 가능성도 있음을 짐작해 볼 수 있다.

2. 말뭉치 언어학

말뭉치 언어학(Corpus Linguistics)은 응용언어학의 한 분야로서 ‘실제 쓰인 언어 자료’를 모아 놓은 집합, 즉 말뭉치(코퍼스)를 컴퓨터로 구축하고 그것을 기반으로 언어를 연구한다. 최초의 말뭉치는 1960년대 미국 브라운대학교에서 구축한 100만 어절의 브라운 말뭉치(Brown Corpus)로 알려져 있으며, 우리말 말뭉치로는 국립국어원의 세종 코퍼스, 연세대학교 언어정보연구원의 연세말뭉치, 고려대학교 민족문화연구원의 물결21 코퍼스¹⁾ 등이 운영되고 있다.

대표적인 중국어 말뭉치로는 북경대학교 중국 언어학 연구센터(北京大學中國語言學研究中心)의 CCL말뭉치(CCL語料庫)와 북경언어대학교 빅데이터와 언어교육 연구소(北京語言大學大數據與語言教育研究所)의 BCC말뭉치(BCC語料庫)²⁾ 등이 있는데, 이것들은 모두 균형성과 대표성을 높이기 위해 다양한 장르와 영역을 망라한 대규모의 범용 말뭉치(general corpus)이다.

범용 말뭉치에 상대되는 것은 특정한 장르 또는 영역의 텍스트로 이루어진 특수목적 말뭉치(specialized corpus)이다. 고전시가와 관련하여 말뭉치를 활용한 기존 연구 성과들은 대체로 당시(唐詩) 말뭉치, 송사(宋詞) 말뭉치 등 특수목적 말뭉치를 기반으로 하였다. 高博(2015), 呂榮等(2019)

1) 이상 3개 말뭉치의 주소는 아래와 같다.

① 세종코퍼스: <https://ithub.korean.go.kr/>

② 연세말뭉치: <https://ilis.yonsei.ac.kr/corpus/>

③ 물결21 코퍼스: <http://corpus.korea.ac.kr/>

2) 이상 2개 말뭉치의 주소는 아래와 같다.

① CCL말뭉치: http://ccl.pku.edu.cn:8080/ccl_corpus/

② BCC말뭉치: <http://bcc.blcu.edu.cn/>

등은 고전시가 작품의 영어 번역 연구에 고전시가 영역(英譯) 말뭉치를 활용하였으며, 周昌樂等(2012), 張帥(2012), 袁慧(2016)와 같이 고전시가 말뭉치를 어휘 분석, 격률 연구, 작품의 자동 생성 연구 등에 활용하는 시도 또한 이루어지고 있다.

반면 CCL말뭉치, BCC말뭉치와 같은 대규모 범용 말뭉치를 활용하여 현대 중국어의 시각에서 고전시가를 연구한 사례는 찾아보기가 매우 어렵다. 현대 중국인의 언어생활에서 여전히 고전시가가 차지하는 비중이 작지 않음을 생각하면 앞으로 이와 관련하여 많은 연구가 이루어져야 할 것으로 보인다.

Ⅲ. 연구 범위 및 방법

1. 교과서 분석 대상과 범위

본 연구에서 분석 대상으로 삼은 교과서는 2001년 출판된 인민교육출판사(人民教育出版社) 초등 어문(語文) 교과서 1~6학년 상·하권, 총 12권이다. 교육부에서 편찬한 국정교과서, 이른바 부편본(部編本)이 2016년부터 보급되기 시작하여 2019년부터는 중국 전역에서 통일된 교과서를 사용하고 있지만, 2001년부터 2016년까지는 인민교육출판사, 강소교육출판사(江蘇教育出版社), 북경사범대학출판사(北京師範大學出版社), 호남교육출판사(湖南教育出版社), 서남사범대학출판사(西南師範大學出版社), 호북교육출판사(湖北教育出版社), 어문출판사(語文出版社) 등 지역, 학교에 따라 다양한 출판사의 교재들을 활용하였다.

현재 널리 쓰이고 있는 부편본을 대상으로 하지 않은 것은 보급된 지 몇 년 되지 않아 오늘날 중국인의 언어생활에 영향을 준 것이 크다고 보기 어렵기 때문이다. 그리고 인민교육출판사의 교과서를 대상으로 한 것은 해당 출판사가 1951년부터 국정 교과서를 출판해 온 중국의 가장 권위 있

는 교과서 출판 기구일 뿐만 아니라, 교과서가 북경(北京)은 물론 하북(河北), 하남(河南), 산서(山西), 길림(吉林), 흑룡강(黑龍江), 운남(雲南), 귀주(貴州), 감숙(甘肅), 청해(青海), 각 자치구 등 다양한 지역에서 널리 사용되어 대표성을 지니기 때문이다.

인민교육출판사 초등 어문 교과서의 중국 고전시가 수록 양상은 학년별로 다소 차이가 있다. 작품 전문을 실은 경우가 대부분이지만 읽기 훈련용으로 일부 유명한 구절만을 수록하기도 하였다. 1~6학년 상·하권 각 교과서에 수록된 고전시가 작품 수와 시구(詩句) 수를 표로 나타내면 다음과 같다.

〈표 1〉 초등 어문 교과서의 학년별 중국 고전시가 수록 양상

구분	1-상	1-하	2-상	2-하	3-상	3-하	4-상	4-하	5-상	5-하	6-상	6-하	계
작품	5	4	4	5	10	16	6	17	12	6	11	14	110
시구	20	16	16	23	30	44	28	48	34	23	41	74	397

고학년으로 갈수록 작품 수와 시구 수가 대체로 증가하는 것을 볼 수 있는데, 통계 처리 면에서 설명이 필요한 것이 몇 가지 있다.

첫째, 육유(陸游)의 〈遊山西村〉은 4학년 상권에 전문이 수록되었지만 3학년 상권 ‘日積月累’ 부분에도 그 일부인 ‘山重水復疑無路, 柳暗花明又一村’이 실려 있다. 그러므로 〈표 1〉의 단순 통계와는 달리 실제 수록된 작품 수는 109개, 시구 수는 395개이다.

둘째, 고전시가의 범위는 시경(詩經), 초사(楚辭)로부터 시작하여 1919년 5·4운동 이전까지 창작된 작품에 한정하였다. 5학년 상권에 수록된 모택동(毛澤東)의 〈七律·長征〉, 〈卜算子·詠梅〉와 같은 작품들은 고전시가의 형식을 취하고 있지만 통계에서 제외하였다.

셋째, 앞서 언급하였듯 학년별 작품 수록 수, 교과서 내 위치, 전문 수록 여부에는 모두 차이가 있다. 그 자체로서 하나의 단원을 이루는 ‘古詩兩首’, ‘古詩三首’, ‘古詩詞三首’, ‘古詩詞背誦’ 등은 대체로 작품 전문을 수

록하고 있으며 학년이 올라감에 따라³⁾ 작품 수가 늘어나고 형식도 좀 더 다양해진다. 한편 읽기 훈련을 위한 코너인 ‘日積月累’의 ‘讀讀背背’ 부분에도 고전시가 작품이 많이 포함되어 있는데, 여기에는 전문이 수록되어 있지 않고 특히 유명한 구절만이 실려 있다. 이밖에 ‘閱讀材料’, ‘閱讀鏈接’, ‘趣味語文’ 등에도 고전시가로 보이는 구절들이 실려 있는 경우가 많다. 이 가운데 작자의 이름이 명기되어 있거나 출처가 확실한 것들은 전문 수록 여부를 막론하고 모두 통계 범위에 포함하였다.

2. 연구 방법

말뭉치 언어학을 기반으로 하는 어휘 연구에서는 빈도 분석이 기본이 된다. 본 연구에서는 중국 북경어언대학교에서 구축하고 운영하는 BCC말뭉치⁴⁾를 활용하여 중국 고전시가 구절들이 일종의 어휘 단위로서 어느 정도의 빈도를 보이는지 알아보고자 한다. 이를 위해서는 먼저 앞서 추출한 109개 작품의 395개 구절을 일일이 BCC말뭉치에서 검색하고 각각의 출현 횟수를 기록하는 작업이 필요하다.

그런데 빈도는 상대적인 것이므로 비교 대상이 필요하다. 앞서 언급하였듯 고전문학 작품은 숙어, 특히 성어와 밀접한 관계에 있으며 어느 정도 유사한 성질을 가지고 있다. 만약 일부 고전시가 구절이 상용(常用) 숙어와 비슷하거나 크게 떨어지지 않는 빈도를 보인다면 이는 중국어 학습자

3) ‘古詩兩首’는 4학년 하권에서 ‘古詩三首’로 바뀌고 5학년 상권부터는 ‘古詩詞三首’로 바뀐다. 가장 높은 단계인 6학년 하권에서는 ‘古詩詞背誦’을 일반 단원과 분리하여 교과서 뒷부분에 배치하고 고전시가 작품 10편을 수록하였다.

4) BCC말뭉치(BCC漢語語料庫)는 총 글자 수가 150억 자에 달하는 대규모 범용 말뭉치로 신문·잡지, 문학, 웨이보, IT, 고대중국어 등 여러 영역의 텍스트를 고루 포함하고 있다. 검색 시 여러 영역 가운데 하나를 선택할 수도 있고 각 영역이 종합된 다영역(多領域)을 고를 수도 있는데, 본 연구는 현대 중국인의 전반적인 언어생활을 알아보고자 하는 연구 목적에 맞게 후자를 선택하였다. 구어체와 문어체의 사용 빈도에 차이가 있을 것으로 예상되나 이와 관련한 문제는 후속 연구에서 다루기로 한다.

들이 고전시가를 학습해야 할 필요성을 보여 주는 강력한 증거가 될 수 있을 것이다.

화어교학출판사(華語教學出版社)에서는 외국인의 중국어 학습을 위한 숙어 교재 시리즈를 출판하였는데, 《成語100: 漢英對照》, 《諺語100: 漢英對照》, 《歇後語100: 漢英對照》⁵⁾는 성어, 속담, 헐후어(歇後語)를 각 100개씩 수록하고 있다. 본 연구는 해당 교재에 수록된 숙어들이 상용 숙어로서 비교적 대표성이 있다고 간주하여 고전시가 구절과 비교 분석을 시도해 보고자 한다.

IV. 분석 및 논의

1. 교과서 수록 고전시가의 빈도 분석

인민교육출판사 초등 어문 교과서에서 추출한 395개 고전시가 구절을 BCC말뭉치에서 검색한 결과 시구(詩句)에 따라 0에서 1024까지 매우 다른 출현 횟수를 보임을 알 수 있었다. 다음은 전체 시구의 출현 횟수를 8개 구간으로 나누고 각 구간에 해당하는 시구의 개수와 비율을 정리한 표이다.

〈표 2〉 말뭉치 출현 횟수 구간별 시구 개수 및 비율

횟수	0	1~50	51~100	101~150	151~200	201~250	251~300	301~	계
개수	83	226	47	14	9	4	7	5	395
비율	21.01%	57.22%	11.90%	3.54%	2.28%	1.01%	1.77%	1.27%	100%

말뭉치 검색 결과가 0인 경우가 약 21%이고, 누적해서 계산할 경우 50

5) 이밖에 《典故100: 漢英對照》도 있으나 전고(典故)는 일반적으로 숙어의 범주에 들어가지 않고 성어와 겹치는 부분도 있어 제외하였다.

이하는 약 78%, 100 이하는 90%에 달하는 것을 볼 수 있다. 언뜻 보아서는 고전시가 구절의 사용 빈도가 그리 높지 않지만, 숙어를 비롯한 다른 어휘군과의 상대적 비교를 거치기 전에 이러한 결론을 내리는 것은 아직 이르다. 어쨌든 확실한 것은 초등 어문 교과서에 실린 시구라고 해서 모두 널리 쓰이는 것은 아니라는 점이다.

101 이상부터는 해당하는 시구 개수가 대체로 줄어드는 양상을 보이는 데, 더 정밀한 분석을 위해서는 자료 처리 방식에 약간의 수정이 필요하다. 출현 횟수가 1024로 가장 높은 시구는 ‘靑山綠水’인데 이 수많은 검색 결과가 모두 백박(白樸)의 <天淨沙秋>를 인용했다고 보기는 어렵다. 작품의 나머지 부분인 ‘孤村落日殘霞’, ‘輕煙老樹寒鴉’ 등의 출현 횟수가 0 또는 1이라는 점이 이를 뒷받침한다. ‘靑山綠水’라는 표현은 당나라 이관(李觀)의 <弔漢武帝文>에 이미 나왔으며 송나라 때에는 전송사(全宋詞)에만 6회 출현할 만큼 널리 쓰였다.⁶⁾

이러한 문제는 ‘靑山綠水’뿐만 아니라 ‘楊柳依依’(《詩經·小雅·采薇》), ‘江南好’(白居易, <憶江南>), ‘鵝鵝鵝’(駱賓王, <詠鵝>) 등 삼언·사언 시구 전반에 걸쳐 존재한다. 이를 해결하기 위해 다음 절부터 진행할 상용 숙어와의 비교, 고빈도 시구 목록 추출 및 분석 등에서는 해당 시구들을 일률적으로 배제하도록 하겠다. 다음은 오언 이상 시구만을 대상으로 다시 진행한 통계의 결과이다.

<표 3> 말뭉치 출현 횟수 구간별 오언 이상 시구 개수 및 비율

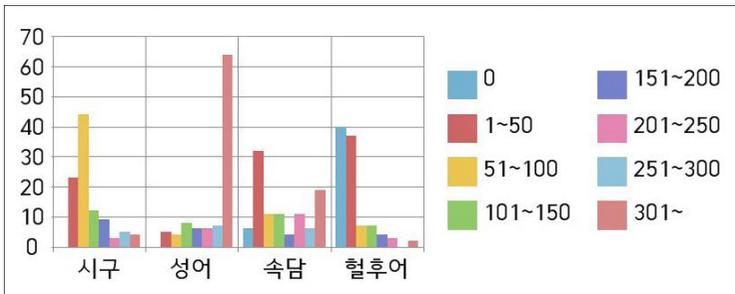
횟수	0	1~50	51~100	101~150	151~200	201~250	251~300	301~	계
개수	83	213	44	12	9	3	5	4	373
비율	22.25%	57.10%	11.80%	3.22%	2.41%	0.80%	1.35%	1.07%	100%

6) 중국 교육부 ‘온라인 말뭉치(語料庫在線)’(<http://corpus.zhonghuayuwen.org/>) 참조.

2. 상용 속어와의 비교

고전시가 구절의 말뭉치 출현 횟수와 구간별 비율이 의미를 가지려면 다른 어휘군과의 비교를 거쳐야 한다. 화어교학출판사(華語教學出版社)에서 출판된 외국인의 중국어 학습을 위한 속어 교재 시리즈는 자주 쓰이는 성어, 속담, 혈후어(歇後語)를 각 100개씩 수록하고 있다. 앞서 초등 어문 교과서에서 추출한 오언 이상 시구 373개 가운데 말뭉치 출현 횟수 상위 100개를 추출하여 속어와 상대적인 비교를 진행해 보았는데, 아래 표4는 그 결과이다.

〈표 4〉 어휘 단위 유형별 말뭉치 출현 횟수 비교



시구의 경우 출현 횟수 100회 이하에 집중되고 그 위로는 점차적으로 수가 줄어드는 반면, 성어는 301회 이상이 60%를 넘는 것을 볼 수 있다. ‘柳暗花明又一村’(419회), ‘每逢佳節倍思親’(367회), ‘舉頭望明月’(346회), ‘床前明月光’(345회) 등 성어에 뒤지지 않는 빈도를 보이는 시구들도 분명 존재하지만 전체적으로 보면 시구는 성어에 한참 못 미친다. 성어 가운데 출현 횟수가 가장 높은 것은 ‘天涯海角’(2484회)였다.

속담은 성어만큼은 아니지만 대체로 시구보다 높은 빈도를 보였다. 출현 횟수 301회 이상이 20% 가까이 되며 ‘見怪不怪, 其怪自敗’(1480회)⁷⁾와

7) 속담과 혈후어는 성어에 비해 형식이 비교적 자유로운 까닭에 변이형이 존재

같이 1000회가 넘는 것들도 있다. 흥미로운 점은 1~50 구간에 속하는 속담이 30%를 차지하고 출현횟수가 0인 경우도 일부 존재한다는 것이다. 1~50 구간의 시구가 약 20%이고 출현횟수가 0인 경우가 없는 것과 대조된다.⁸⁾ 이는 속담 못지않게 널리 쓰이는 시구가 적지 않음을 의미한다.

성어, 속담과 달리 혈후어(歇後語)는 전체적으로 시구에 못 미치는 모습을 보인다. 말뭉치 출현 횟수 0과 1~50 구간이 각각 약 40%로 합치면 80%에 달하며, 301회를 넘는 것이 ‘八仙過海——各顯神通’(664회), ‘熱鍋上的螞蟻——團團轉’(621회) 2개에 불과하다.

상용 속어와의 비교 결과 고전시가 구절의 말뭉치 출현 횟수는 성어와 속담에는 못 미치지만 혈후어보다는 높음을 알 수 있었다. 그리고 시구 가운데 일부는 성어와 속담 못지않은 빈도를 보였다. 중국어 학습에서 고전시가가 더 중시되어야 할 필요성은 명백해 보인다.

3. 중국어 학습용 고전시가 작품의 선정

위에서 고빈도 시구 100개를 추출한 것은 속어와의 비교를 위한 것이며, 중국어 학습용 고전시가 작품 선정을 위해서는 범위를 좀 더 축소할 필요가 있다. <표 4>에서 볼 수 있듯 1~50 구간은 적지 않은 수의 속담과 혈후어가 분포해 있는 구간으로 빈도가 아주 높다고는 볼 수 없다. 그렇다고 해서 학습할 가치가 없는 것은 결코 아니지만, 최대한 고빈도에 논의를 집중하기 위해서는 해당 구간을 제외할 필요가 있어 보인다. 아래 표 5는 말뭉치 출현 횟수 51회 이상 시구 77개의 목록이다.

하는 경우가 많은데, ‘見怪不怪, 其怪自敗’ 같은 경우 뒷 구절을 ‘其怪自壞’라 쓰기도 한다. 본 연구는 이러한 경우가 누락되는 것을 방지하기 위해 형식이 비교적 고정된 앞 구절을 검색 대상으로 하였다.

- 8) 물론 이것은 연구 대상 373개 시구 중 출현 횟수 상위 100개와 비교한 것이다. 하지만 화어교학출판사 교재에 실린 100개의 속담들이 상용 속담으로서 대표성을 가진다고 본다면 이러한 비교는 무리가 아니다.

〈표 5〉 말뭉치 출현 횟수 51회 이상 시구

시구	횟수	시구	횟수	시구	횟수
柳暗花明又一村	419	低頭思故鄉	120	誰言寸草心	74
每逢佳節倍思親	367	春江水暖鴨先知	117	相看兩不厭	74
舉頭望明月	346	只緣身在此山中	114	爆竹聲中一歲除	74
床前明月光	345	春風吹又生	113	淡妝濃抹總相宜	72
潤物細無聲	288	桃花潭水深千尺	111	春風又綠江南岸	72
路漫漫其修遠兮	276	讀書破萬卷	110	處處聞啼鳥	68
花落知多少	275	天涯若比鄰	107	長河落日圓	68
洛陽親友如相問	268	誰知盤中餐	100	清明時節雨紛紛	65
本是同根生	251	兩個黃鸝鳴翠柳	97	報得三春暉	62
相煎何太急	231	小荷才露尖尖角	95	大漠孤煙直	61
春眠不覺曉	215	下筆如有神	93	夜來風雨聲	58
咬定青山不放鬆	205	少小離家老大回	91	臨行密密縫	58
他鄉遇故知	200	明日複明日	91	好雨知時節	58
野火燒不盡	192	霜葉紅于二月花	90	遠上寒山石徑斜	56
悠然見南山	186	疑是地上霜	89	疑是銀河落九天	56
鋤禾日當午	181	西出陽關無故人	89	歲月不待人	56
風吹草低見牛羊	181	海上生明月	88	落紅不是無情物	56
天涯共此時	164	不拘一格降人才	85	借問酒家何處有	55
采菊東籬下	163	我勸天公重抖擻	84	一行白鷺上青天	53
吾將上下而求索	159	橫看成嶺側成峰	82	化作春泥更護花	53
不識廬山真面目	158	慈母手中線	80	煙花三月下揚州	52
山重水複疑無路	147	月是故鄉明	80	任爾東西南北風	52
粒粒皆辛苦	136	隨風潛入夜	80	明月松間照	51
飛流直下三千尺	136	遠近高低各不同	79	日出江花紅勝火	51
停車坐愛楓林晚	128	獨在異鄉爲異客	78	春來江水綠如藍	51
海內存知己	124	白了少年頭	75		

위 77개 시구는 42개 작품에서 나온 것으로 다음과 같은 특성을 지니고 있다.

첫째, 당나라 때 작품에서 나온 것이 가장 많다. 77개 중 49개로 약 64%를 차지하여 이백(李白)과 두보(杜甫)로 대표되는 당시(唐詩)의 위상을 보여 준다. 그 다음은 송(13개), 위진남북조와 청(각 6개), 선진(2개), 명(1개)의 순서이다.

둘째, 작자 24명 중 이백의 작품에 나온 구절이 가장 많다. 9개 구절이 〈靜夜思〉를 비롯한 5개 작품에서 나왔다. 그 다음은 두보로 8개이며 왕유(王維)와 소식(蘇軾)은 6개, 두목(杜牧)은 5개, 맹호연(孟浩然), 맹교(孟郊), 백거이(白居易), 공자진(龔自珍)은 4개, 도잠(陶潛)과 이신(李紳)은 3개가 포함되었다. 역시 당나라 시인이 가장 두드러지는 가운데 송나라의 소식이 공동 3위에 오른 것이 눈에 띈다. 이밖에 굴원(屈原), 조식(曹植), 왕발(王勃), 장구령(張九齡), 왕안석(王安石), 육유(陸游), 정섭(鄭燮)은 2개, 하지장(賀知章), 왕창령(王昌齡), 왕수(汪洙), 악비(岳飛), 양만리(楊萬里), 전학탄(錢鶴灘)은 1개씩 포함되었으며, 작자 미상의 작품(〈敕勒歌〉)에서 나온 것도 1개 있었다.

셋째, 〈獨坐敬亭山〉, 〈滿江紅·寫懷〉, 〈明日歌〉, 〈泊船瓜洲〉, 〈芙蓉樓送辛漸〉, 〈山居秋暝〉, 〈小池〉, 〈送元二使西安〉, 〈神童詩〉, 〈元日〉, 〈月夜憶舍弟〉, 〈飲湖上初晴後雨〉, 〈雜詩〉, 〈贈汪倫〉, 〈敕勒歌〉, 〈惠崇春江晚景二首〉, 〈黃鶴樓送孟浩然之廣陵〉, 〈回鄉偶書〉 등 많은 작품들이 한 구절씩만 포함되었다. 반면 〈春曉〉, 〈靜夜思〉, 〈題西林壁〉은 전 구절이 모두 포함되었다.

어떤 작품 내 구절들의 말뭉치 출현 횟수가 대체로 높다는 것은 바꾸어 말하면 그 작품이 사람들의 입에 널리 오르내린다는 뜻이다. 이것이 작품 자체의 가치와는 별개의 문제이지만, 중국어 학습을 위한 교육적 가치에 한정한다면 우선적으로 학습해야 할 작품을 선정하는 기준으로 삼을 수 있다.

다음은 작품 내 고빈도 시구의 비율에 근거하여 42개 작품을 다섯 그룹으로 나누는 것이다.

〈표 6〉 중국어 학습용 고전시가 작품 목록

그룹	고빈도	작품명	작품수
1그룹	100%	〈靜夜思〉, 〈春曉〉, 〈題西林壁〉	3
2그룹	51~99%	〈憫農〉, 〈山行〉, 〈憶江南〉, 〈遊子吟〉	4
3그룹	50%	〈九月九日憶山東兄弟〉, 〈己亥雜詩·其五〉, 〈己亥雜詩·其二百二十〉, 〈望廬山瀑布〉, 〈絕句·其三〉, 〈竹石〉, 〈清明〉	7
4그룹	25~49%	〈春夜喜雨〉, 〈獨坐敬亭山〉, 〈望月懷遠〉, 〈泊船瓜洲〉, 〈芙蓉樓送辛漸〉, 〈使至塞上〉, 〈小池〉, 〈送杜少府之任蜀州〉, 〈送元二使西安〉, 〈神童詩·四喜〉, 〈元日〉, 〈遊山西村〉, 〈飲湖上初晴後雨〉, 〈贈汪倫〉, 〈賦得古原草送別〉, 〈七步詩〉, 〈惠崇春江晚景二首〉, 〈黃鶴樓送孟浩然之廣陵〉, 〈回鄉偶書〉	19
5그룹	0~24%	〈離騷〉, 〈滿江紅·寫懷〉, 〈明日歌〉, 〈奉贈韋左丞丈二十二韻〉, 〈山居秋暝〉, 〈月夜憶舍弟〉, 〈飲酒·其五〉, 〈雜詩〉, 〈敕勒歌〉	9

1그룹의 세 작품은 기승전결 네 구가 모두 고빈도이므로 가장 우선적으로 학습될 필요가 있다.⁹⁾ 특히 〈靜夜思〉는 ‘擧頭望明月’과 ‘床前明月光’의 출현 횟수가 각각 346회, 345회로서 77개 시구 가운데 3, 4위를 차지하였다.

2그룹의 네 작품 가운데 〈憫農〉과 〈山行〉은 네 구 가운데 세 구가 고빈도이며, 〈憶江南〉은 다섯 구 가운데 세 구¹⁰⁾, 〈遊子吟〉은 여섯 구 가운데 네 구가 고빈도이다.

9) 세 작품은 모두 우리나라 2015 개정 중·고등학교 한문 교과서에 수록되어 있기도 하다. 이밖에 2그룹의 〈憫農〉, 〈山行〉, 4그룹의 〈春夜喜雨〉, 〈送元二使西安〉, 〈黃鶴樓送孟浩然之廣陵〉도 수록되어 있다.

10) 〈憶江南〉 중 ‘江南好’(119회)는 오언이 되지 않아 77개 목록에서 제외되었지만 여기서는 포함하였다.

3그룹의 일곱 작품은 모두 칠언절구로 네 구 가운데 절반이 고빈도이다. 1~3그룹은 작품 길이가 짧고 고빈도 시구의 비중이 높으므로 작품 전문의 학습과 암송을 권장할 만하다.

4그룹은 대부분 네 구 가운데 한 구, 여덟 구 가운데 두 구가 고빈도인 작품들이다. 다만 〈春夜喜雨〉는 여덟 구 가운데 세 구, 〈七步詩〉는 여섯 구 가운데 두 구로 비중이 다소 높다. 4그룹은 율시가 다수 포함되고 고빈도 시구의 비중이 낮은 편으로, 1~3그룹을 학습한 뒤 다음 단계에 활용하는 것이 적합해 보인다.

5그룹의 아홉 작품은 고빈도 시구의 비중이 낮고 작품 분량도 대체로 많다. 〈奉贈韋左丞丈二十二韻〉은 44구로 이루어져 있으며 〈離騷〉는 무려 372구에 달한다. 물론 〈離騷〉 전문을 학습하거나 암송한다면 중국과 중국어에 대한 이해가 상당히 깊어질 것은 자명하다. 하지만 현대 중국어 학습이 목적인 학습자를 위해서는 빈도가 높고 유명한 구절을 적절히 발췌하여 제공하는 것도 고려할 만하다. 인민교육출판사의 초등 어문 교과서는 읽기 훈련 코너 ‘讀讀背背’에서 유명한 고전시가 한두 구절을 실어 외우도록 하고 있는데, 외국인 대상 중국어 교재에서도 이러한 방식을 적용해 볼 수 있겠다.

1~4그룹에서 말뭉치 출현 횟수가 특히 높은 시구 역시 전문을 학습하는 것과는 별도로 한 번 더 중요하게 다루는 것을 고려할 만하다. 예를 들어 〈遊山西村〉과 〈九月九日憶山東兄弟〉는 각각 4그룹, 3그룹에 속하여 상대적으로 학습 우선순위가 떨어지지만, ‘柳暗花明又一村’(419회)과 ‘每逢佳節倍思親’(367회)의 출현 횟수는 전체 시구 가운데 1~2위를 차지한다. 해당 구절들을 발췌하여 1그룹에 속한 시구 못지않은 비중으로 다룰 필요가 있다.

V. 결론

인민교육출판사 초등 어문 교과서에 수록된 중국 고전시가 작품의 오언 이상 구절 373개를 BCC말뭉치에서 검색한 결과 출현 횟수가 0회에서 419회로 다양함을 발견할 수 있었다. 상위 100개 시구를 상용 속어와 비교해 보면, 비록 성어와 속담에는 미치지 못하지만 혈후어보다는 출현 횟수가 높았다. 또 일부 시구는 성어와 속담에 뒤지지 않는 높은 빈도를 보이기도 했다.

말뭉치 출현 횟수 51회 이상의 고빈도 시구 77개는 24명의 시인이 지은 42개 작품에서 나왔다. 당나라 때 작품에서 나온 것이 64%로 가장 많았으며 특히 이백과 두보의 작품에서 나온 것이 많았다. 한 작품의 모든 구절이 고빈도 시구 77개에 다 포함된 경우도 있었지만, 한 구절만 포함되어 있는 경우가 적지 않았다.

작품 내 고빈도 시구의 비율에 근거하여 42개 작품을 다섯 그룹으로 나누었는데, 1~3그룹은 작품 길이가 짧고 고빈도 시구의 비중이 높아 작품 전문의 학습과 암송이 권장된다. 4그룹은 그 다음 단계의 학습에 활용하는 것이 적절해 보이며, 5그룹은 고빈도 시구의 비중이 매우 낮고 작품 분량도 많아, 고빈도 시구가 포함된 부분만을 발췌하여 학습자에 제공하는 것이 학습 효율을 높이는 방안이 될 듯하다. 물론 이것은 일반적인 현대 중국어 학습자의 경우에 한한 것이다. 중국 관련 학과에 재학 중인 학생에게는 더욱 심도 있는 학습이 요구된다.

본 연구는 중국 고전시가 구절을 하나의 어휘 단위로 간주하고 통계적인 분석을 활용하여 어휘교육적 측면에서 고전시가 구절이 속어 못지않게 중요함을 증명하려 하였다. 아울러 중국어 학습용 고전시가 작품의 선정을 시도하였는데, 이것은 일종의 방법론적 제시이며 현대 중국인들이 보편적으로 인지하고 있는 작품을 모두 망라하였다고 보기는 어렵다. 교수·학습 현장에 실제로 적용하기 위해서는 초등 어문 교과서의 범위를 넘어서 《唐

詩三百首》, 《宋詞鑑賞辭典》과 같은 자료들을 활용하여 대규모 조사 작업을 진행해야 할 것으로 보인다.

<참고문헌>

- 강병규, 〈중국어 언어자료-말뭉치 구축 현황과 활용 방법 고찰〉, 《중국어 문논총》 제35집, 2007.
- 강병규, 〈AI의 중국 古典詩歌 창작——詩語의 학습과 생성〉, 《중국문학》 제100집, 2019.
- 권석환, 〈中國 名言의 초기 형성과정과 ‘引用’의 관계〉, 《중국어문논총》 제77집, 2016.
- 김경동, 〈中國古典詩 教育의 諸問題〉, 《중어중문학》 제37집, 2005.
- 류성준, 〈중국 고전시가의 향후 연구모색과 그 교육〉, 《외국문학연구》 제11집, 2002.
- 병선, 〈성구, 속담, 격언, 리언, 경구, 명언〉, 《중국조선어문》 제4호, 1986.
- 신하윤, 〈중국고전시가교육의 의의와 방안〉, 《중국어문학지》 제10집, 2001.
- 이동향, 〈中國 古典詩歌의 研究와 教學〉, 《중국문학》 제34집, 2000.
- 이혜천, 〈우리나라 고등학교 중국어교육에 있어서의 중국고전시가 활용 방안 연구〉, 울산대학교 석사학위논문, 2014.
- 정혜영, 《중국고전시가를 통한 중국어교육연구: 중국교육부 <소학어문교육대강>에서 추천한 시를 중심으로》, 순천향대학교 석사학위논문, 2008.
- 최정은, 〈중국어 교육에서의 중국 고전시가 활용방안 연구: 초·중급 학습자를 중심으로〉, 한국외국어대학교 석사학위논문, 2010.
- 황여옥, 《시 텍스트를 활용한 중국어 교육과 학습지도안: 唐人 絶句를 중심으로》, 이화여자대학교 석사학위논문, 2009.
- 황정봉, 《韓國과 中國의 教科書 收錄 漢詩 比較 研究》, 한국교원대학교

- 석사학위논문, 2019.
- 符淮青, 《現代漢語詞彙(增訂本)》, 北京大學出版社, 2004.
- 高博, 〈中國詩歌英譯語言特征探析——基於語料庫的研究〉, 《外文研究》第3期, 2015.
- 呂榮、洪揚、馬文麗, 〈基於對比語料庫的詩詞翻譯語言特征研究——以許淵沖宋詞英譯為個案〉, 《成都師範學院學報》第6期, 2019.
- 王旗、張瑜、彭光源, 〈成語的重要來源: 古代典籍與文學作品〉, 《大眾文藝》第12期, 2019.
- 徐盛桓, 〈相隣與補足——成語形成的認知研究之一〉, 《四川外語學院學報》第2期, 2006.
- 徐盛桓, 〈相隣和相似——漢語成語形成的認知研究之二〉, 《暨南大學華文學院學報》第3期, 2006.
- 徐耀民, 〈成語的劃界、定型和釋義問題〉, 《中國語文》第1期, 1997.
- 荀恩東、饒高琦、肖曉悅、臧嬌嬌, 〈大數據背景下BCC語料庫的研制〉, 《語料庫語言學》第1期, 2016.
- 尹斌庸編著, 《成語100: 漢英對照》, 華語教學出版社, 1999.
- 尹斌庸編著, 《歇後語100: 漢英對照》, 華語教學出版社, 1999.
- 尹斌庸編著, 《諺語100: 漢英對照》, 華語教學出版社, 1999.
- 袁慧, 《輔助古詩詞研究的唐詩語料庫的構建》, 河北大學碩士學位論文, 2016.
- 張帥, 〈計算機輔助宋詞研究〉, 北京大學碩士學位論文, 2012.
- 周昌樂、鄭旭玲、何中市、李良炎, 〈中國古典詩詞楹聯的計算化研究〉, 《心智與計算》第2期, 2012.
- 周薦, 〈熟語的經典性和非經典性〉, 《語文研究》第3期, 1994.
- 子規, 〈由唐詩派生的成語典故〉, 《文史雜誌》第3期, 2020.
- Grant, L. E. "Frequency of 'core idioms' in the British National Corpus (BNC)". *International Journal of Corpus Linguistics* 10, (2005).
- Simpson, R. & Mendis, D. "A Corpus-Based Study of Idioms in Academic Speech". *TESOL Quarterly* 37.3, (2003).

<Abstract>

This study examines how often Chinese classical poetry verses are used in modern Chinese language life, and also extracts a list of frequently used poetry and analyzes its characteristics. And this paper examined which poems Chinese learners should learn first.

The corpus linguistic method was used to find out how often the classical poetry verses are used. The BCC corpus was searched for 373 verses of five-character verses contained in Chinese elementary language textbooks, and the top 100 were extracted and compared with frequently used idioms. As a result of comparison, it was found that the frequency was less than that of the idioms and common sayings, but higher than that of the wisecracks.

77 high-frequency verses with 51 or more corpus appearances came from 42 poems, and 42 poem were divided into five groups according to the proportion of high-frequency verses. Groups 1~3 have a short length of poem and high proportion of high-frequency verses, so it is recommended to study and memorize the whole poems, but groups 4~5 are encouraged to study only partially.

Key Words : 고전문학(Classical literature), 고전시가(Classical poetry), 중국어교육(Chinese teaching), 어문(Chinese language), 교과서(Textbook), 말뭉치(Corpus).

