東洋學 第88輯(2022年 7月) 檀國大學校 東洋學硏究院

한자 어휘 의미망 개발을 위한 한자어 사전 표제어의 성격 고찰
-『漢韓大辭典』, 『韓國漢字語辭典』, 『한국고전용어사전』, 『옛편지낱말사전』을 중심으로-*

손혜옥** · 김다미*** · 차명희**** · 김우정*****

▌국문초록▐

이 연구는 한자 어휘 의미망 개발의 기본 자원이 되는 기구축된 디지털 한자어 사전 표제어의 성격을 고찰하여 한자 어휘 의미망 개발에 앞서 먼저 논의해야 할 점이 무엇인지를 확인하는 것을 목적으로 하였다. 본 연구에서 모든 디지털 한자어 사전의 표제어들의 성격을 고찰하는 것은 현실적으로 불가능하므로 가장 많은 표제어를 싣고 있는 『漢韓大辭典』, 한국 문헌에만 등장하는 어휘를 모아 놓은 『韓國漢字語辭典』, 특수 목적 사전이라 할 수 있는 『한국고전용어사전』, 『옛편지낱말사전』을 연구 대상으로 선정하였다. 이들 사전 표제어의 성격을 표제어 범위 면에서 살펴본 결과, 『한한대사전』과 『한국한자어사전』은 한국 문헌에만 나타나는 것을 포함하여 한·중·일 고전 한문 자료에 나타난 모든 자와 어휘를 포괄하고자 했고 표제어 규모도 커 한문 어휘에 대한 포괄성은 가장 넓다. 한편 『한국고전용어사전』과 『옛편지낱말사전』도 고전 한문 자료를 바탕으로 한 사전이나 전자는 주로 기관에서 발행한 공적 문서를 바탕으로 하였다면 후자는 사적 문서인 간찰을 바탕으로 표제어를 선정하였으므로 특정 장르적 성격을 반영하고 있었다. 언어적 특성 면에서는, 네 사전 모두 표제어의 언어적 성격이 사전의 거시 구조 차원에서 구분되어 반영되어 있지 않았다. 따라서 단어로 볼 수 있는 『한한대사전』의 표제자를 제외하고 『한한대사전』과 『한국한자어사전』의 어휘 표제어와 『한국고전용어사전』, 『옛편지낱말사전』의 표제어들은 모두 단어에서부터 새로운 의미를 획득한 관용구, 의미 합성성을 지키고 있는 통사적 구성에 이르기까지 꽤 다양한 언어적 특성을 가진 것들이 혼재된 것으로 보인다. 이상의 표제어 성격의 연구 결과를 바탕으로 어휘의미망 개발을 위한 다음의 논점을 확인할 수 있었다. 첫째, 개발하려는 어휘의 미망의 성격을 확정하고 그에 합당한 데이터가 무엇인지 골라내는 논의가 필요하다. 둘째, 보통 어휘의미망 개발은 유사한 언어적 속성을 가진 것을 모아 개발되는 편이므로 『한한대사전』의 표제자를 제외하고 다양한 언어적 속성이 혼재된 한자 어휘 데이터들에서 균질한 언어적 특성을 가진 데이터를 어떻게 골라낼 수 있을지, 혹은 이러한 데이터의 성격을 고려할 때 현실적으로 개발 가능한 어휘 의미망은 무엇이 될 수 있을지 논의가 필요하다.

[주제어] 한자 어휘 의미망, 표제어, 『한한대사전』, 『한국한자어사전』, 『한국고전용어사전』, 『옛편지낱말사전』

* 이 논문은 2021년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2021S1A5C2A02086984)
** 제1저자, 연세대학교 강사 / darmal@hanmail.net
*** 단국대학교 한문교육연구소 연구원 / muchbeauty_418@naver.com
**** 단국대학교 강사 /myunghee812@naver.com
***** 연결저자, 단국대학교 교수 / rtoran@dankook.ac.kr

| 목 차 |

Ⅰ. 서 론
Ⅱ. 국어 사전의 표제어 성격
Ⅲ. 한자어 사전 표제어의 성격
Ⅳ. 한자 어휘 의미망 개발을 위한 전제적 논의점
Ⅴ. 결 론

Ⅰ. 서 론

컴퓨터의 등장으로 인한 자료의 구축과 처리 기술의 발달은 그간 인문학 연구의 방법론에 큰 영향을 주었고, 이로 인해 인문학 연구자들은 그 이전에 보지 못했던 새로운 인문학적 현상을 발견하거나 새로운 인문적 해석을 할 수 있게 되었다. 현재 디지털 자료와 기술을 활용한 연구 방법론의 개발과 이를 이용한 해석 등은 인문학 연구의 큰 흐름 중 하나라고 할 수 있다. 이에 국내 동아시아 한문 고전 연구에서도 한문 고전 디지털 자료의 구축과 그것을 활용한 연구 방법론이나 연구 주제 발굴에 대한 다양한 탐색과 시도가 이루어지고 있다고 하겠다.

그간 한자·한문 고전 디지털 자료들은 상당수 구축된 바 있으나 주로 검색과 그 내용의 확인이라는 측면에서 개발되고 서비스되는 경우가 대부분이었다. 그러나 이제 기존의 디지털화된 한자·한문 자료를 가지고 더 다양한 연구에 활용하기 위해 원시 말뭉치를 넘어선 주석, 병렬 말뭉치 등의 구축, 한문 텍스트 처리에 활용되는 형태소 분석기, 구문 분석기, 어휘 의미망 개발 등에 있어서도 관심이 고조되고 있다. 이 중 본 연구에서 논의할 어휘 의미망이란 다양한 의미 정보를 컴퓨터가 처리할 수 있도록 체계화시킨 다양한 의미 관계 데이터베이스를 말한다.[1] 한자 어휘를 대상으로 하는 어휘 의미망이 개발된다면 한문의 어휘 연구의 데이터로서 기능할 뿐만 아니라 정보 검색 및 기계번역, 주제별 텍스트의 유형 분류나 내용 요약의 자동화 등 다방면으로 활용이 가능하다. 그리고 이를 바탕으로 한문학적, 또한 한문 텍스트를 다루고 있는 관련 연구에 디지털 인문학적 방법론으로의 연구 성과를 내는 데 활용될 수 있을 것이다.

한자 어휘 의미망 구축을 위해서는 한국 고전 자료의 어휘를 대표할 수 있는 어휘 데이터가 필수인데 대량의 어휘 데이터로 참조할 수 있는 대표적인 것이 바로 사전이라고 할 수 있다. 현재 가장 많은 표제어 수를 포함하는 『한한대사전』을 비롯하여 다양한 특수 목적 한자 어휘·용어 사전들이 데이터화되어 있으며 그 중 상당수는 웹상으로 서비스를 하고 있다.[2] 이들 데이터들은 모두 한자 어휘 의미망 구축을 위한 기초 자료

1) 영어를 대상으로 하는 대표적인 어휘 의미망으로는 PWN이 있다. 일본어를 대상으로 한 것으로는 NTT LN, PWN을 바탕으로 동유럽어에 적용한 것으로는 EWN 등이 있으며, 한국어를 대상으로 한 것으로는 세종전자사전, U-Win, KorLex, CoreNet 등이 있다. 국내외 어휘의미망 구축 현황에 대한 더 자세한 내용은 윤애선(2007)을 참조할 수 있다.

2) 공개된 데이터로는 웹기반 동양고전번역 용어/용례 사전(경북대), 조선시대 유서 지식DB 및 지식지도(고려대 민족문화연구원), 조선시대 유서(類書) 기반 지식 DB 및 지식지도 구축(한국학중앙연구원), 조선 후기 물명 집성 DB(한국학중앙연구원), 옛 편지 낱말사전 DB(한림대), 한중일 고대 한자 자전 통합DB(한국연구재단), 고전용어 시소러스(한국고전번역원), 한국민족대백과 사전(한국학중앙연구원), 문화재용어 사전(한국국제교류재단), 한국 전통용어 사전 정보(특허청), 한국복식사전(한

로서 기능할 수 있을 것이나, 이들 한자 어휘 데이터를 수집하여 의미망 구축을 위한 설계를 하기 전에 먼저 각 한자 어휘 사전 데이터의 표제어의 성격을 이해하는 것이 필요하다. 각 사전들은 그 편찬 목적에 따라 표제어 선정 기준이 다르고, 그에 따라 수집된 표제어의 성격에도 차이가 있으므로 어휘 의미망 개발 전에 해당 데이터들의 표제어가 어떤 언어적 지위를 가지고 있는지, 주로 어떤 의미 부류에 속하는 것인지를 파악해야 어휘의미망 개발의 목적에 따라 어휘를 선정하고 정제하는 작업을 할 수 있다.

이에 이 연구는 한자 어휘 의미망 개발의 기본 자원이 되는 기구축된 디지털 한자어 사전 표제어의 성격을 고찰하여 한자 어휘 의미망 개발에 앞서 먼저 논의해야 할 점이 무엇인지를 확인하는 데 목적을 둔다. 이를 위해 디지털 한자어 데이터베이스의 표제어들이 어떻게 수집되고 선정된 어휘인지 조사하려고 한다. 그러나 후술하겠지만 디지털 한자어 사전의 표제어의 성격을 파악하는 데 있어서 표제어 선정에 대해 자세히 기술해 놓은 자료를 찾기 어려워 수집 기준을 명확히 파악하기 어려운 면이 있다. 따라서 현대적 방식으로 편찬된 몇몇 국어사전에서의 표제어 수록 범위와 선정 기준, 해당 표제어들의 언어학적 특성에 기대어 한자어 사전 표제어의 성격을 가늠해 보려고 한다.[3] 본 연구에서 모든 디지털 한자어 사전의 표제어들의 성격을 고찰하면 좋겠지만 그러한 작업은 현재로서는 불가능하므로 가장 많은 표제어를 싣고 있어 고전 한자 어휘의 상당수를 포괄할 수 있는 『漢韓大辭典』과 『韓國漢字語辭典』,[4] 특수 목적 사전이라 할 수 있는 『한국고전용어사전』, 『옛편지낱말사전』을 선정하여 이들의 표제어 성격에 대해 살펴보도록 하겠다.

연구의 구성은 다음과 같다. 2장에서는 현대적 방식으로 편찬되어 웹으로 서비스되고 있는 3종의 국어사전의 표제어 선정 기준에 대해 살펴 수록 범위와 규모 면에서의 표제어의 성격과 언어학적 성격에 대해 살펴보도록 하겠다. 보통 표제어의 언어학적 성격은 사전 구조나 기술상에 변별되어 반영되기 마련이므로 이에 대해서도 함께 살피도록 한다. 3장에서는 한자어 사전의 표제어들의 선정 과정과 기준 등을 살펴보고 수록 범위와 규모 면에서의 표제어의 성격에 대해 살핀다. 한자어 사전 표제어의 언어학적 성격에 대해서는 관련 서술을 찾기 어려운 면이 있어, 표제어 항목을 어떻게 구조화하여 기술하는지 살펴 국어사전의 기술에 빗대어 가늠하기도 하며, 국어사전의 표제어의 언어학적 특성에 빗대어 일부 표제어에 대해 실제 분석을 하기로

국연구재단), 한국 전통 약재 정보(특허청), 한국전통 공계 정보(특허청). 한국전통 식품 정보(특허청) 등이 있다.

3) 본 연구에서 현대식 국어사전 편찬 방식을 따른 사전의 표제어 선정 기준과 그 성격을 확인하여 한자어 사전을 바라보려는 것은 두 사전을 단순히 비교하려는 것이 아니다. 국어사전과 한자어 사전은 구축의 제반 여건이 다를 수 있기 때문이다. 그럼에도 현대식 국어사전 편찬 방식에 대해 확인하고 한자어 사전 표제어의 성격을 확인하려는 것은 기구축된 디지털 사전은 디지털 인문학적 방법론에 따른 한문 고전 연구를 위한 다양한 자료 개발을 위한 중요한 데이터이기 때문이다. 한자어 사전의 데이터를 활용하여 다른 유형의 디지털 데이터를 개발하려면 실제적으로 현대식 사전에서 얻을 수 있는 정보의 수준이 요구된다. 그것을 충족시키지 못한다고 해도 무엇이 충족되지 않는지 확인해야 이것이 해당 언어 자체의 특성인지, 아니면 필수적으로 확인되어야 하는 정보인지를 판단하고, 어떻게 가공해야 하는지, 어떤 방법으로 자료의 한계를 보완할지 다음 작업을 진행할 수 있다. 다만 국어사전과 한자어 사전은 대상 언어가 다르기 때문에 국어사전을 기준으로 한자어 사전의 표제어의 성격을 살피려는 것이 타당한가라는 의문이 제기될 수 있을 것이다. 그러나 본 연구에서 살피는 국어 사전들이 일반적인 현대적 방식의 언어 사전과 표제어 선정 기준이나 선정된 표제어의 성격이 크게 다르지 않고, 대상 언어가 서로 다르기는 해도 언어마다 표제어 선정 기준은 유사하므로 가까이 접할 수 있는 국어 사전을 비교 대상으로 선정하였다.

4) 단국대학교 동양학연구원에서는 기 간행한 『한한대사전』과 『한국한자어사전』을 통합한 『통합디지털한한대사전』을 구축 중이며, 대부분의 공정이 완료된 것으로 알려져 있다. 그러나 아직 일반에 공개되지 않고 있으므로 본 연구에서는 기 간행된 종이 사전을 참조하였다.

한다. 4장에서는 3장의 논의를 바탕으로 한자 어휘 의미망 개발을 위한 전제적 논의점을 확인하고, 5장에서는 논의 내용을 요약하고 한계를 밝힌다.

Ⅱ. 국어 사전의 표제어 성격

국어사전의 편찬은 말뭉치를 이용하게 되면서 획기적으로 변화하게 되었다고 할 수 있겠다. 1990년대 컴퓨터와 말뭉치를 이용하여 사전 편찬 작업을 하게 되면서 사전 편찬 방법론에 큰 변화가 발생하며 사전과 관련된 논문이 활발하게 나오게 되었고, 사전에 대한 기대도 이전에 학습, 참조는 물론이고 온라인 서비스나 기계 번역 등의 자연언어처리에 활용될 수 있는 언어 데이터베이스로 확장되는 추세이다.[5] 말뭉치를 활용한 사전으로는『연세한국어사전(1998)』,『표준국어대사전(1999)』,『고려대한국어대사전(2009)』등을 비롯하여 초등생 학습자를 위한『연세 초등국어사전(2002)』, 한국어를 공부하는 외국인 학습자를 위한『한국어기초사전(웹용)』, 한국어 기본 어휘의 빈도 정보를 담고 있는『한국어기본어휘의미빈도사전(2014)』등의 몇몇 특수 목적 사전들이 개발되어 있다. 이 장에서는 웹으로 서비스하고 있는 범용적 현대 국어사전인『연세한국어사전』,『표준국어대사전』,『고려대한국어대사전』[6]의 표제어 선정 기준이나 과정을 소개하고[7] 이들 사전에서의 표제어의 성격을 확인하여 한자어 사전 표제어 성격의 이해를 위한 참조로 삼도록 하겠다.

먼저 표제어의 수록 범위와 규모에 대해서 살펴보도록 하겠다.『연세』는 최초의 말뭉치 중심 사전이라고 할 수 있다. 말뭉치를 활용한 최초의 사전은『표준』이겠지만,『표준』에서의 말뭉치의 활용은 제한적이므로 사전의 거시구조와 미시구조[8] 전체의 구성과 기술을 말뭉치를 중심으로 하여 만들어진 사전은『연세』가 최초라 하겠다.[9]『연세』는 대략 4300만 어절의 말뭉치[10]에서 빈도 14 이상의 빈도를 보이는 대략 5만 개의

5) 남길임,「말뭉치 기반 사전 편찬의 동향과 지향 방향: 최근 30년간의 사전편찬 방법론을 중심으로」,『한말연구』16, 한말연구학회, 2005, 76, 80쪽.

6) 이하 각각『연세』,『표준』,『고려』로 지칭하도록 하겠다. 이들 사전은 모두 웹으로 내용을 서비스하고 있으며, 각각 https://ilis.yonsei.ac.kr/ysdic/, https://stdict.korean.go.kr/main/main.do, https://dic.daum.net/index.do?dic=kor에서 그 내용을 확인할 수 있다.

7) 이들 사전의 표제어 선정과 그 과정, 표제어의 성격에 대한 기술은 이들 사전의 일러두기와 관련 논문을 바탕으로 기술하도록 한다.『연세』는 남기심 · 김한샘(2017),『표준』은 윤용선(1995), 정희원(1995), 최혜원(2000), 송길룡 · 민경모 · 서상규(2003),『고려』는 도원영(2000), 도원영 · 차준경(2009)을 주로 참고하였으며, 그 외 이들 사전에 대한 비교 분석이 있는 안의정 · 황은하(2010), 유현경(2012), 남길임(2007), 유현경 · 남길임(2008)도 참고로 하였다.

8) 사전의 거시구조는 표제어 구성과 관련한 것을 가리키고, 미시구조는 하나의 표제어를 구성하는 하위 정보 항목, 즉 뜻풀이, 용례 등과 관련한 것을 가리킨다.

9) 남기심 · 김한샘(2012: 26)에서는 사전의 거시구조와 미시구조 전반에 걸쳐서 말뭉치의 분석을 기반으로 한 사전 편찬을 Sinclair(1987)의 개념을 가져와 말뭉치 중심 사전 편찬(corpus-driven lexicography)이라 하고,『연세』,『고려』의 편찬이 이에 해당한다고 언급하였다.

10)『연세』를 위한 말뭉치의 구축은 초반 400만 어절의 균형 말뭉치를 시작으로 1980, 1990 등의 연도별 말뭉치, 교과서 말뭉치, 구어 말뭉치 등이 추가되었다.『연세』를 위한 연세말뭉치 제작 과정과 구성은 남기심 · 김한샘(2017: 14-21)을 참고할 수 있다.

어휘만을 표제어로 선정하였다. 빈도를 기준으로 표제어를 선정하였으므로 비표준어나 순화의 대상이 되는 어휘라고 해도 14 이상의 빈도를 보이는 것이라면 널리 쓰이는 것의 경우 선별하여 표제어로 선정하였다. 또한『연세』는 일러두기에서 현대 한국어의 학습을 위한 언어사전[11]임을 밝히고 있는 것처럼 기존의 사전들이 고유명사나 전문어를 다수 수록하고 있는 것에 반해 이들을 표제어에서 제외하였다.[12] 일정한 빈도 이상의 어휘만 수록하며, 고유명사나 전문어도 표제어 선정에서 제외하고 있으므로『연세』의 표제어 수는 그리 많은 편은 아니다. 대사전에 속하는『표준』의 표제어 수가 대략 50만이라는 것을 생각하면 상대적으로『연세』의 표제어 수는 현대 한국어의 언어생활을 망라하여 반영할 만한 규모는 아님을 알 수 있다.

『표준』은 정부 기관인 국립국어원에서 만든 사전으로 강인선(1995)에 따르면 언어 생활의 표준을 제시하고, 민족의 언어적 동질성을 회복하고, 오천 년 문화 민족의 자긍심을 고취하기 위한 목적을 가지고 제작되었다 한다.[13]『표준』에서는 이러한 목적에 따라 개별 어휘 하나하나에 어문 규정을 반영하고 널리 쓰는 비표준어를 표제어로 수록하여 이에 대응되는 표준어 정보를 제공하였다. 북한의 문화어를 수용하기 위해『조선말대사전(1992)』의 어휘를 일부 선별하여 수록하였고, 우리 민족이 사용했던 모든 말들을 시공간의 구애 없이 수용하기 위해 전문 용어, 고유어, 방언, 옛말도 표제어로 수용하였다.『표준』은 표제어 수가 대략 50만 개로 현재 편찬된 사전 중 표제어 수가 가장 많은데 이는『표준』의 표제어 선정 범위가 넓기 때문으로 보인다.

『표준』의 표제어 선정은 6종의 기존 사전[14]의 총 63만개의 표제어 목록을 바탕으로 규범에 맞지 않는 것, 언어학적으로 잘못 판단된 것, 사전의 성격과 맞지 않는 것은 제외하여 선별하고[15] '소설, 수필, 신문, 잡지' 등으로 구성된 대략 5000만 어절의 말뭉치를 참조하여 신어를 일부 수용하는 방식으로 이루어졌다.[16]『표준』에서 전문어의 선정은 6종의 사전을 바탕으로 하되 고등학교 교과 과정에서 다룰 수 있는 정도의 난이도에 해당하는 것으로 하였고, 영역별 전문어 사전, 전문 서적에서 폭넓게 나타나는 전문어를 일부 수집하였다.[17]

---

11) 언어사전은 단어들에 대한 참고서로서 언어에 대한 책인 반면 백과사전은 언어에 대한 것이 아니라 사물, 사람, 장소, 관념 등 실제 세계에 대한 책이다(Jackson, 2002: 21). 언어사전이 단어에 대한 언어적 정보를 다루고 있다면 백과사전은 단어들이 가리키는 실세계의 정보를 다룬다. 이러한 특성의 차이로 언어 사전은 표제어 구성에 있어서 언어사전은 어휘적 요소와 문법 형태를 모두 포함하나, 백과사전은 어휘적 요소 가운데 주로 명사만이 표제어가 된다. 언어사전과 백과사전은 여러 측면에서 차이가 있지만 대부분의 국어사전에는 이 두 가지 요소가 섞여 있어 엄격하게 구분되기 어렵다. 대사전류는 대부분 고유어와 전문어를 표제어로 받아들여 기술하고 있고, 언어사전을 표방하는『연세』에서도 일부 뜻풀이에는 백과사전적 정보를 실을 수밖에 없다(유현경, 2012: 38).

12)『연세』에서도 고유명사 중에 일반어적인 용법을 가지고 있는 경우에는 표제어로 삼고 있다.

13) 강인선, 종합국어대사전 편찬 사업,『새국어생활』5: 1, 국립국어원, 1995, 4쪽.

14)『표준』에서 검토 대상이 된 사전은『국어대사전(금성출판사, 1991)』,『우리말큰사전(한글학회, 1992)』,『새우리말큰사전(삼성출판사, 1986)』,『국어대사전(민중서림, 1982)』,『조선말대사전(북한 사회과학원, 1992)』,『조선말사전(북한 사회과학원, 1962)』이다.

15) 사실 6종의 사전을 참조하여 적절한 표제어를 선정할 경우 명확한 기준을 세우는 것이 쉽지는 않다. 윤용선(1995: 22)에서는 6종의 사전에 모두 실려 있는 말이면 일단 수록하였는데, 편찬 과정에서 비록 낯선 단어라 할지라도 기존의 6개 대사전에서 모두 등재된 표제어의 경우 제외시키기는 힘든 일이라고 언급한 바 있다.

16)『표준』에서 다루고 있는 표제어의 수가 매우 많기 때문에 5000어절의 말뭉치에서 모든 표제어 후보의 용례를 찾는 것이 불가능했을 것이고, 따라서『표준』의 표제어 선정에서 말뭉치의 활용은 제한적일 수밖에 없었을 것이다.

17) 고유어 선정에 대한 것은 지침이나 관련 연구에서도 명확한 기술을 찾아보기 어렵다. 다만 고유명의 경우 전문어와 관련성이 있으므로 정희원(1995: 82)에서 전문어와 고유어를 어느 정도까지 사전 표제어로 수록하는가의 문제가 있으며, 이러한

『표준』의 표제어는 구성이나 규모를 보면 『연세』에 비해 백과사전적 특성을 보이고 있으며, 현대 언어 생활의 관점에서는 일부 유령어가 있을 수도 있겠으나 한국인의 언어 생활을 전반적으로 반영할 수 있는 성격을 보인다고 할 수 있겠다.[18]

『고려』는 『연세』와 마찬가지로 말뭉치를 적극 활용하여 편찬한, 대략 38만 표제어를 담고 있는 대사전이다. 『고려』는 사전 이용자를 위한 실용성을 강조하여 북한어를 제외한 전문어, 고유어, 방언, 빈도 높은 비표준어를 모두 포괄하였다는 점에서 『표준』과 유사하나 앞서 출간된 사전에서 추출한 원시 표제어 목록에서 1억 어절 규모의 말뭉치에서 출현하지 않는 어휘를 제외함으로써 모어 화자의 언어 생활에서 사용되지 않는 유령어를 배제하려고 노력하였다.[19] 또한 말뭉치에서는 확인되나 기존 사전의 표제어에 등재되지 않은 어휘에 대해 사전적 단어로서의 판별을 거쳐 표제어 등재 여부를 결정하였다.[20] 『고려』는 『표준』과 백과사전적 특성을 가진 대사전이라는 면에서는 유사하나 현대 국어 화자의 언어 생활을 망라하는 어휘를 비교적 객관적 근거에 기반하여 선정하여 싣고 있다는 면에서 『표준』과 차별화된다고 볼 수 있다.

『연세』, 『표준』, 『고려』의 수록 범위와 규모면에서의 표제어의 특성을 종합하여 보면 『표준』이 수집의 객관성은 다소 떨어지나 수록 표제어 수가 가장 많고 그 대상이 넓기 때문에 가장 한국어 어휘에 대한 포괄성이 가장 큰 편이다. 『연세』는 표제어 선정 기준이 가장 분명하므로 수집의 객관성은 높지만 포괄성이나 현대 국어 일상 어휘로서의 대표성은 다소 떨어진다고 할 수 있다. 『고려』의 표제어는 『표준』에 비해 포괄성은 떨어지나 수집의 객관성은 더 높고, 말뭉치 빈도를 활용해 유령어를 제외하려 하였으므로 현대 국어 일상 어휘로서의 대표성은 더 높다고 할 수 있겠다.

다음으로 이들 세 사전의 표제어의 언어학적 성격과 그에 따른 사전 구조·기술상의 특성을 살펴보도록 하겠다. 언어사전에서 표제어는 주로 단어로 구성되지만 국어는 교착어이므로 단어 이하 단위인 파생접사, 어미에 대한 정보도 중요하여 이들을 표제어로 싣는 것이 기본이다. 단어 이상의 단위 중에서 관용구, 연어, 속담 등의 표현들도 사전에 등재되어야 할 중요 요소들이므로 사전마다 이들도 역시 그 방식은 조금씩 차이가 있지만 표제어에 포함하고 있다. 표제어의 구성이 사전마다 조금씩 차이가 있을 수 있지만 일반적으로 단어, 파생접사, 어미 등은 주표제어로 제시되고, 구 이상의 단위는 부표제어나 주표제항 아래의 뜻풀이나 용례로 제시되기도 한다.

성격의 표제어는 전문어 분과에서 영역별 담당자가 심의하여 판단한다고 밝히고 있다.

18) 『표준』은 종이 사전으로 출판된 이후에도 웹으로 서비스하면서 부분적인 영역에서 수정과 업데이트를 통해 미시항목의 수정과 표제항의 추가나 삭제 등을 지속하고 있다. 이에 대해서는 https://stdict.korean.go.kr/notice/noticeList.do에서 확인할 수 있다. 『표준』은 규범적 성격을 가지고 있으므로 신어에 대한 수용이 보수적일 수 있으며, 방언이나 전문어에 대한 수용이 제한적일 수 있다. 이에 국립국어원에서는 신어, 전문어, 방언에 대한 수용성을 높이고, 사전 이용자의 요구에 대한 만족을 높이기 위해 100만 어휘로 시작한 개방형 사전인 『우리말샘』, https://opendict.korean.go.kr/main을 서비스하고 있다. 『우리말샘』은 『표준』을 기반으로 하되 『표준』의 제한성을 보완하기 위한 사전이라고 할 수 있겠다.

19) 도원영(2000)에서는 말뭉치에서 확인되지 않았다고 기존 사전의 어휘를 무조건 배제하는 것이 아니라 말뭉치의 한계로 확인되지 않는 것일 수도 있으므로 이들 어휘에 대해서는 구어에서 주로 사용되는 것인지, 우리 어휘 체계에 정착한 단어인지 등을 확인하여 표제어 여부에 대한 선별 작업을 거쳤다고 밝히고 있다.

20) 비표준어, 외래어, 전문어, 고유명사, 신어 등도 코퍼스를 활용하여 일정한 빈도 이상의 것에 대해 전문가의 검토를 거쳐 수록하였는데 그 자세한 과정과 기준에 대한 자세한 내용은 도원영(2000: 46-56)에서 확인할 수 있다.

『연세』는 1단어 1표제어 원칙에 따라 단어, 예외적으로 접사, 어미의 경우만 주표제어로 기술하고, 구 이상의 단위는 부표제어로 분류한다. 부표제어는 주표제항 아래에서 기술된다. 『연세』의 일러두기에 따르면 부표제어에 해당하는 것은 관용표현이나 연어를 이루어 쓰이는 구에 해당한다. 구체적인 예시로는 '가격 파괴'처럼 명사와 명사로 이루어진 구이나 보통 한 단어처럼 많이 쓰이는 구, '비행기를 태우다'와 같은 관용 표현, '뾰족한 수'와 같은 특수 표현들이다.

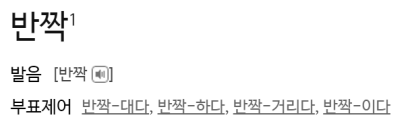
〈그림 1〉『연세』의 '미역국' 표제어 항목 기술

『표준』은 『연세』와 달리 전문어나 고유어의 경우에 한해 구 이상의 단위도 주표제어로 다루고 있다. 다만 구 이상의 단위를 주표제어로 다룰 때에는 품사 정보를 주지 않고 '진급^시험'과 같이 '^'로 구임을 밝히고 있다. 『연세』에서 부표제어로 분류하였던 관용 표현이나 특수 표현은 『표준』에서는 주표제어의 미시구조 아래에서 '관용구/속담' 항목이나 예문으로 제시된다.

〈그림 2〉『표준』의 '수' 표제어 항목 기술

〈그림 3〉『표준』의 '미역국' 항목 기술

『표준』의 부표제어는 '-거리다, -대다, -되다, -이/히, -이다, -적(的), -하다'가 결합한 파생어이다. 이들 접사가 결합한 파생어는 주표제어로 선정된 인접 어휘 아래 제시된다.



〈그림 4〉『표준』의 '반짝[1]' 부표제어

〈그림 5〉『표준』의 '반짝[1]'의 표제어와 부표제어 제시 예

『고려』에서는 표제어의 유형을 구분하고 있지 않다. 『고려』에서도 『표준』과 마찬가지로 구이상의 단위도 표제어로 삼고 있는데 고유어나 전문어에 속하는 구 단위뿐만 아니라 관용 표현과 속담도 포함한다. 다음은 『고려』에서 표제어로 삼고 있는 예들이다.

(1) 가. 가격 파괴, 복지 사업
나. 스승의 날, 플레밍의 왼손 법칙
다. 미역국을 먹다, 제 코가 석자, 가는 말이 고와야 오는 말이 곱다, 무슨 뾰족한 수가 있나?

이상의 내용을 종합해 보면 『연세』, 『표준』, 『고려』 모두에서 단어 이하 단위를 표제어로 선정하는 것을 원칙으로 하는 것을 확인할 수 있다. 다만 예외적으로 구 이상의 단위라도 전문어나 고유어, 사전에서 기술 대상으로 삼을 만한 다중 어휘 표현들[21]에 대해서는 사전마다 서로 다른 판단을 하고, 서로 다른 방식으로 변별하여 싣고 있다. 『연세』의 경우 구 이상의 연어, 관용 표현, 특수 표현에 대해 부표제어라 하여 주표제어와 구분함으로써 표제어의 문법 단위의 다름을 사전의 거시구조 차원에서 드러내고 있다. 『표준』은 주표제어에도 일부 단어도, 일부 구 구성도 포함하고 있으므로 표제어의 문법적 특성이 거시구조에서 엄격하게 반영되었다고 보기는 어렵다. 그러나 표제어에 속하는 구 구성이 전문어와 고유명사에 한한 것이며, 이들이 단

21) 남길임(2007: 148)에서는 사전에서 관용 표현, 연어 및 패턴, 속담, 상투 표현, 고빈도 자유 표현 등의 구 구성의 다중어휘 표현들이 사전에서 주로 부표제어로서 기술되는 대상이 된다고 하였다. 각 표현들의 예시는 다음과 같다(남길임, 2007: 148-149).
예) 관용표현 : 비행기를 태우다, 미역국을 먹다
연어 및 패턴 : 눈을 감다, 감을 잡다, NP1뿐만 아니라 NP2
속담 : 가는 날이 장날이다
상투 표현 : 안녕하십니까? 반갑습니다. 어서 오십시오.
고빈도 자유 표현 : 과일 가게, 대기 오염, 수질 오염

어가 아님을 ^로 표기상에서 변별하고 있다. 『고려』는 『표준』보다 더 넓은 범위의 다중 어휘 표현을 표제어 유형 구분 없이 제시하고 있으므로 마찬가지로 표제어의 문법적 특성이 사전의 거시구조에 반영되지는 않는다. 그러나 구 구성 표제어의 대상이 무엇인지 밝히고 있으며, 띄어쓰기를 통해 표기상으로 변별된다. 『연세』, 『표준』, 『고려』가 단어 이상의 단위들에 대한 선정 여부나, 사전 구조 내에서의 기술 방법에서는 차이가 있으나 이들을 선정하여 사전에서 기술해야 할 필요성에 대해 인식하고, 또 그것을 기술할 때 단어 표제어와 구 표제어를 변별하여 기술했다는 점에서는 공통점을 확인할 수 있다.

Ⅲ. 한자어 사전 표제어의 성격

이 장에서는 『한한대사전』, 『한국한자어사전』, 『한국고전용어사전』, 『옛편지낱말사전』의 표제어의 성격에 대하여 살펴보도록 하겠다.[22] 먼저는 이들 사전의 일러두기와 관련 연구를 통해 이들 사전 표제어의 수록 범위 면에서의 성격을 살피고 그 이후 언어학적 면에서의 성격에 대해 살펴보도록 하겠다.

『한한대사전』은 일러두기에서 밝히고 있는 바와 같이 자전과 사전을 겸한 책으로 한·중·일 삼국에 현존하는 모든 한자를 싣고, 한자 아래에 해당 한자가 첫 글자가 되는 어휘들을 함께 수록하고 있다. 『한국한자어사전』은 한국에만 존재하는 한자나 어휘를 담은 사전이다. 두 사전은 한문 고전을 해석하는 데 활용될 목적으로 제작되었으므로 표제자와 표제 어휘의 채록은 고전 문헌을 기반으로 하고 있으며, 전문어나 고유명사까지 포괄하는 백과사전적 특성을 지니고 있다. 『한한대사전』은 표제자 약 5만 5천자, 표제 어휘 약 42만 규모의 대사전이며, 『한국한자어사전』의 한자 어휘도 9만 개에 가깝다. 두 사전의 표제자와 표제 어휘는 말뭉치를 바탕으로 출현 빈도를 참조하여 선정된 것은 아니나 철저히 문헌을 바탕으로 하였고, 19세기 이전까지의 동아시아의 주요 문헌에 나타난 어휘를 통시적으로 망라한 자료라고 할 수 있다.

『고전』은 역사서, 문집, 법전, 운서 등 각 분야 한국 고전 126개 문헌에서 발췌한 용어 5만 여 개를 표제어로 삼아 뜻을 풀이하고 용례와 출전을 밝힌 용어 사전이다. 『고전』의 표제어 추출 문헌은 다음과 같다.

(2) 가. 1차 사료 : 『삼국사기(三國史記)』, 『삼국유사(三國遺事)≫, 『고려사(高麗史)』, 『고려사절요(高麗史節要)』, 『조선왕조실록(朝鮮王朝實錄)』, 『승정원일기(承政院日記)』 등
나. 법전 : 『대명률직해(大明律直解)』, 『경국대전(經國大典)』, 『속대전(續大典)』, 『대전통편(大典通編)』, 『대전회통(大典會通)』 등
다. 일반 고전 : 『훈민정음(訓民正音)』, 『용비어천가(龍飛御天歌)』, 『석보상절(釋譜詳節)』, 『월인천강지곡(月印千江之曲)』, 『훈몽자회(訓蒙字會)』, 『악학궤범(樂學軌範)』, 『국조오례의(國朝五禮儀)』, 『의방유취(醫方類聚)』, 『향약채취월령(鄕藥採取月令)』, 『칠정산내외편(七政算

22) 이하 기술의 편의상 『한한대사전』과 『한국한자어사전』은 통합하여 지칭할 경우 『한한』, 『한국고전용어사전』과 『옛편지낱말사전』은 각각 『고전』, 『옛편지』라 하겠다.

內外篇)』, 『서운관지(書雲觀志)』, 『증보문헌비고(增補文獻備考)』, 『신증동국여지승람(新增東國輿地勝覽)』 등

라. 기타 : 문집 및 야사류, 금석문 등

『고전』은 한국의 문헌에서 수록된 용어를 발췌한 것이므로 '세사미두(歲賜米豆), 도승지(都承旨), 시명지보(施命之寶)' 등과 같이 한국 한문 문헌에만 나타나는 표현들을 수록하고 있는데, 이런 특성은 한국 문헌에만 등장하는 어휘만을 수록하고 있는 『한국한자어사전』과 유사하며, 실제 이런 성격의 어휘들은 두 사전에서 모두 표제어로 등장한다.[23] 『고전』의 표제어 선정의 기준은 자세히 알려져 있지 않으나 용어 사전이라는 명칭 그대로 사건명, 관직명 등을 다수 포함하고 있으며, (2라)나 『삼국유사』를 제외하고는 정부 주도로 편찬된 문헌들이므로 수록된 표제어들이 공적 영역에서 주로 사용되는 표현들이 많이 수록되었다고 볼 수 있겠다.

『옛편지』는 한문으로 쓰여진 옛편지에 쓰인 한자어를 해설하는 사전이다. 총 56개의 자료집을 바탕으로 7600여 개의 낱말 표제어를 추출했는데 고려말 정몽주(鄭夢周)의 간찰부터 일제 강점기 독립운동가의 옥중서한까지가 이 사전의 연구 대상이 되었다.[24] 『옛편지』에서 밝히고 있는 표제어 선정 유형은 아래와 같다.

(3) 가. 수신자 및 발신자 상호간의 호칭, 인사말, 편지에만 나오는 상투적인 낱말, 봉투의 서식 등에 나오는 낱말
나. 관혼상제 때의 축하 및 위로와 관련된 낱말
다. 질병 및 건강과 관련된 낱말 예) 숙아(夙疴), 소신(所愼), 흠상(欠常), 위섭(違攝)
라. 간찰과 함께 보낸 문방구, 음식물, 약재 등 각종 선물의 물명(物名)
마. 날씨, 농사, 신간, 집기 등 일상생활과 관련된 낱말
바. 과거, 조세, 부역, 흉년, 풍년, 구휼 등 행정과 관련된 낱말

간찰은 일정한 장르적 특성을 가진 사적 장르의 특수한 글이기 때문에 편지 봉투나, 간찰 본문에만 쓰이는 상투적인 낱말들이 다수이다. 따라서 『옛편지』의 표제어는 간찰의 장르적 특성을 나타내는 '균후(勻候), 정리(政履), 사리(仕履), 시학(侍學), 만만(萬萬), 불비식(不備式)' 등과 같이 상투적인 표현과 (3나-다)에서 확인할 수 있듯 편지에 자주 등장할 법한 내용의 의미 영역에 해당하는 단어들을 다수 포함하고 있다.

『한한』은 말뭉치 분석 결과에 기반한 것이 아니라 사전 편찬자의 직관에 의존하기는 했으나 한국 문헌에만 나타나는 것을 포함하여 한·중·일 고전 한문 자료에 나타난 모든 자와 어휘를 포괄하고자 했고 실제 표제어 수도 방대하므로 사실상 한문 어휘에 대한 포괄성은 가장 넓다고 할 수 있다. 『고전』과 『옛편지』도 고전 한문 자료를 바탕으로 한 사전이나 이들은 『한한』과 표제어의 규모에서 큰 차이가 있으며 『고전』은 주로 기관에서 발행한 공적 문서를 바탕으로 하였다면 『옛편지』는 사적 문서를 바탕으로 표제어를 선정하였으

23) 간단한 예로 세사미두(歲賜米豆), 도승지(都承旨), 시명지보(施命之寶) 등은 『고전』과 『한국한자어사전』에 모두 등재된 표현들이다.

24) 『옛편지』에서 표제어를 추출한 간찰 영인본 목록은 하영휘 외(2011)의 부록편을 참조할 수 있다.

므로『한한』보다는 상대적으로 표제어들이 특정 의미 영역을 다룬 특수사전의 성격을 지닌다.[25)]

『한한』,『고전』,『옛편지』에 대한 사전적, 언어적 특성에 대해서는 사전의 거시 구조나 미시구조 차원에서 변별되어 있지 않고 선정 기준에 대한 자료를 찾기 어려워 더욱 판단이 어려운 면이 있다. 보통 국어사전을 포함하여 언어 사전에서 표제어의 대상으로 삼는 것은 단어와 구 구성의 경우 고유어나 전문어 혹은 긴밀한 공기 관계를 보이거나 긴밀한 결합뿐만 아니라 의미 합성성으로 설명하기 어려운 다중 어휘 표현들이다. 그러므로 이에 미루어 한자어 사전의 표제어들이 단어인지 구인지, 혹은 관용적 의미를 획득한 것인지 등의 언어적 특성에 대해서 임의로 몇몇 표제어를 골라 살펴보도록 하겠다.

『한한』은 거시구조를 보면 표제자와 표제자 밑에 딸린 어휘들로 구성된다. 구성상의 특성으로 본다면 표제자는 오늘날 국어사전에서의 주표제어, 어휘들은 부표제어로 볼 수 있다.『한한』에서는 자(字)를 기준으로 사전의 거시구조를 설계했으므로 자의 언어적 단위가 무엇인지에 대해서는 명시적으로 설명하고 있지 않지만 중국어의 경우 형태론적으로 고립어에 해당하므로 표제자가 곧 단어 표제어인 경우가 상당하다.[26)] 그렇다면『한한』의 주표제어에 해당하는 표제자는 꽤 일관된 언어적 특성을 가진 것이라 할 수 있다. 다만『한한』에서 어휘를 선정할 때 어떤 기준으로 선정하였는지 명확하게 밝히고 있지 않았으므로 '어휘' 표제어의 경우에는 2자 이상의 표현이라는 것 외에는 그 언어적 특성에 대해서 명확히 알기는 어려운 면이 있다. 따라서『한한』의 일부 표제자에 딸린 어휘 표제어들을 분석해 보았다.

다음은『한한』의 '귤(**橘**)' 표제자에 딸린 어휘들을 그 언어적 특성에 따라 분류한 것이다.[27)]

(4) 가. 명사 : 귤림(橘林), 귤음당(橘蔭堂), 귤포기(橘浦記), 귤감(橘監), 귤관(橘官), 귤록(橘錄), 귤산(橘山), 귤산사륙(橘山四六), 귤생(橘生), 귤송(橘頌), 귤수(橘叟), 귤원(橘苑), 귤원(橘園), 귤은(橘隱), 귤자주(橘子州), 귤재도인(橘齋道人), 귤주(橘洲), 귤지사(橘枝詞), 귤산(橘山), 귤산문고(橘山文稿), 귤실(橘實), 귤옥(橘屋), 귤옥집(橘屋集), 귤우(橘于), 귤은재(橘隱齋), 귤은재집(橘隱齋集), 귤정(橘亭), 귤정(橘井),[28)] 귤주(橘洲), 귤포(橘浦)[29)]

나. 명사 or 명사구 : 귤정(橘井), 귤노(橘奴), 귤동(橘童), 귤홍(橘紅), 귤향고(橘香餻), 귤중수(橘中叟), 귤중희(橘中戲),[30)] 귤적(橘籍), 귤도(橘涂), 귤여(橘如) / 귤과(橘果), 귤병(橘

25) 사실『한한』을 포함하여,『고전』,『옛편지』의 표제어 선정 기준이나 과정에 대한 기록이나 연구들이 많이 없어 현재로서는 제한된 자료만을 가지고 이들 표제어의 수록 범위 면에서의 특성을 추정할 뿐이다.

26) 다만 '포도(葡萄)'와 같이 '포'와 '도'가 문장 안에서 단독으로 잘 쓰이지 아니하고 언제나 '포'는 '도'와 '도'는 '포'와만 결합해서 사용한다면 '포'와 '도'는 각각 형태소로서의 지위만 가지고 단어로서의 지위는 가지지 않는다.

27) 본래 어떤 표현이 합성어인지 구 구성(관용구, 연어, 패턴을 포함하여)인지의 구분은 언제나 논리적인 것만은 아니다. 따라서 본 연구에서 한자어의 문법적 지위에 대한 판단은 가능성의 차원에서 분석된 것이다.

28) '귤정'은 특정 우물의 이름을 가리키기도 하고, '양약(良藥)'을 가리키기도 한다.『한한』에서 의미항목을 두 개로 기술하고 있는데 본 연구에서 (4가)에서 고유명사로 분류한 '귤정'은 첫 번째 의미항목을 나타낼 때의 경우이다.

29)『한한』에서 '귤포'의 의미는 다음과 같이 기록되었다. "1. 귤주(橘洲)의 1번 뜻. 2. 청 과원(戈源)의 호. 3. 전설상 유의(柳毅)가 용녀(龍女)에게 책을 전하였다는 곳. 뒤에 남녀가 언약하는 곳을 이른다." 1번 '귤주'는 호남성 장사시 서쪽 상강에 있는 모래톱의 이름을 뜻한다. 따라서 '귤포'의 모든 의미는 '귤포'가 고유명사임을 말해 준다. 그러나 3.의 의미항목 중 '남녀가 언약하는 곳을 이른다'라는 기술이 있는데 이는 고유명사였던 귤포가 보통명사로서의 확장을 보인다는 기술이다.

30) 귤중희(橘中戲)는 '橘中之戲'에서 '之'가 탈락한 표현이다. '之'는 명사구를 나타내는 표지인데 고대 중국어에서 선진시기에는 다소 규칙적으로 나타났던 '之'가 명사구 내부 구성에서 동한 시기 이후에는 아예 보이지 않게 되는 변화를 보이는데,

餠), 귤성(橘性), 귤심(橘心), 귤실(橘實), 귤엽(橘葉), 귤자(橘子), 귤주(橘酒), 귤피(橘皮), 귤핵(橘核), 귤환(橘丸), 귤락(橘絡), 귤로(橘露)

다. 명사구 : 귤중지락(橘中之樂)

라. 절 : 귤황(橘黃), 귤화위지(橘化爲枳), 귤록등황(橘綠橙黃)

위의 예에서 고유명으로 쓰이는 (4가)의 경우는 2음절 명사로 볼 만하고, '之'가 개입된 (4다)는 명사구, (4라)는 절로 판단할 수 있겠다. (4다, 라)는 주로 고사에서 유래하며 축자적 의미가 아니라 '귤중지락'이 '바둑의 즐거움'을 가리키는 것처럼 새로운 의미를 가지고 사용된다. 한편, (4나)는 이것이 단어인지 명사구인지 판단이 쉽지 않아 그 문법적 지위가 어떠한지 명확히 알기 어려운 편이다. (4나)의 예들은 대부분 명사와 명사의 결합 구조를 보이는데 축자적 의미의 결합이 아닌 새로운 의미를 획득하여 사용되는 것도 있고 축자적 의미의 결합으로 보이는 경우도 보인다. '귤정(橘井)'류는 '귤정'이 '양약'이라는 의미로 사용되는 것에서 알 수 있듯 축자적 의미 외에 새로운 의미를 획득한 것으로 보이는데 이때 이것이 합성어일 수도 있고, 관용적 의미를 획득한 명사구일 수도 있다. '귤과(橘果)'류는 투명한 의미 결합을 보이고 있는 예들로 '귤정'류보다는 명사구의 가능성이 높은 예들이라고 할 수 있다. 그러나 표현의 의미 투명성이 높아도 패턴과 같이 높은 공기관계를 보이고 사용되는 연속 표현은 합성어의 지위를 획득하기도 하므로 일단 (4나)의 유형으로 분류하였다. 이상의 분석 결과를 보면 『한한』의 '귤'의 표제항에 실려 있는 표제 어휘들은 명사부터, 절까지 꽤 다양한 문법적 지위를 가지고 있는 것들임을 알 수 있다.

다음은 『한한』에서 '남(覽)' 표제자에 딸린 어휘들을 그 언어적 특성에 따라 분류한 것이다.

(5) 가. 명사 : 남주(覽州), 남화사(覽華寺)

나. 명사 or 명사구 : 남규(覽揆), 남견(覽見), 남침(覽寢), 남침연(覽寢宴)

다. 동사 or 동사구 : 남경(覽鏡), 남고(覽古), 남물(覽物), 남승(覽勝), 남시(覽試), 남정(覽政), 남핵(覽核) / 남견(覽見), 남관(覽觀), 남구(覽究), 남독(覽讀), 남력(覽歷), 남성(覽省), 남시(覽示), 남열(覽閱), 남조(覽照), 남찰(覽察), 남촉(覽矚), 남총(覽總), 남취(覽取),

라. 동사구 : 남서구실(覽書具悉), 남문변견(覽聞辯見)

(5가)의 경우 고유명으로 사용되는 것이므로 명사로 볼 수 있겠다. (5라)의 '남서구실'은 '세자가 대리청정할 때 상소문을 읽어 보고 그 내용을 다 안다'는 뜻으로 사용되는데 이때 '남서구실' 안에 주어에 해당하는 명사는 확인할 수 없으므로 동사구로 볼 수 있다. '남문변견'은 '식견이 넓고 말솜씨가 좋다'는 의미로 사용되는데 이 역시 동사구이다.

(5나), (5다)는 각각 명사구나 동사구로 쓰일 수 있는 표현들이다. (5나)의 '남규(覽揆)'는 고사에서 유래한 말로 '생일의 별칭', '남침(覽寢)', '남침연(覽寢宴)'은 '첫날밤을 보낸 신랑신부의 잠자리를 보는 풍습'을 가리

'귤중희(橘中戱)'도 그렇게 해석할 가능성이 있다. 그러나 '귤중희(橘中戱)'는 출전(『幽怪錄』 또는 『太平廣記』)로 볼 때 당송시기 이후에 생성된 단어이므로 일단은 합성어 가능성을 열어두고 분류하였다.

키는 말이므로 합성명사 혹은 관용적 쓰임을 보이는 명사구로 볼 수 있겠다. '남견(覽見)'은 (5다)로도 분류했듯 유사한 의미의 동사적 쓰임이 있는데, 명사적 쓰임도 『한한』에서 별도의 의미항목으로 용례와 함께 제시했으므로 (5나)로도 분류하였다.

(5다)에서 '남경(覽鏡)'은 결합에 따른 새로운 의미의 획득도 잘 확인되지 않아 '경'이 '남'의 논항으로 해석될 가능성이 있다. 그렇게 본다면 '남경'은 동사구로 볼 수 있겠다. 그러나 VN의 결합이 동사화했다고 볼 가능성도 있으므로 일단 동사로의 분석도 열어 둔다. '남경'과 함께 제시한 표현들은 '남경'과 같이 VN 구성을 이루며 비교적 의미 합성성을 잘 지키고 있는 편이다. (5다)에서 '남견(覽見)'류는 합성 동사인지 동사 연속 구성인지 판단이 쉽지 않다. 보통 고립어의 경우 한 절 안에서 동일 논항을 취하는 동사 연속 구성이 자주 나타나며 또 이들은 흔히 합성동사로 문법화되기도 한다. (5다)의 표현들은 비교적 의미 합성성이 잘 지켜지고 있는 편이어서 그 판단이 더욱 어렵다고 하겠다.

『한한』의 '어휘' 표제어들은 국어사전에서 부표제어로 기술되거나 주표제항의 미시구조 안에서 관용어, 속담(절의 구조를 가지므로)으로 기술될 만한 것과 『고려』에서 사전적 단어로 그 성격을 규정하여 주표제어로 선정했던 유형을 포괄하는 것으로 보인다. 그러나 (4나)의 '귤과'류, (5다)의 표현들은 전형적인 통사적 구성으로 해석할 만한 가능성이 있어 『한한』의 어휘 표제어의 다른 항목과 동일하게 다룰 수 있는 것인지 의심스러운 면이 있다. 또한 고유명과 제3의 의미를 가진 '귤정(橘井)'류는 단어로 볼 만한데, 그렇다면 일부 어휘들은 '표제자'와 동일한 언어적 성격을 가진다고 볼 수 있다. 정리하면 『한한』의 어휘 표제어들은 합성어로 추정되는 것부터, 연어, 관용구, 전형적인 통사적 구성에 이르기까지 꽤 다양한 언어적 특성을 보이는 단위들이 혼재되어 있는 것으로 추정된다.

다음은 『고전』의 일부 표제어를 언어적 특성에 따라 분류한 것이다.

(6) 가. 명사 : 도승지(都承旨), 숙의(淑儀), 요동(遼東), 염락관민(濂洛關閩)
나. 명사 or 명사구 : 세자빈인(世子嬪印), 관현맹인(管絃盲人), 임오년수립(壬午年樹立)
다. 명사구 : 시명지보(施命之寶)
라. 동사구 : 제세안민(濟世安民), 무본억말(務本抑末), 세사미두(歲賜米豆), 종성부용초성(終聲復用初聲)

(6가)는 관직명, 지명에 해당하는 것으로 명사로 볼 수도 있다. 다만 '염락관민'은 '송(宋)나라의 유학자인 주돈이(周敦頤)·정호(程顥)·정이(程頤)·장재(張載)·주희(朱熹)를 말하는데 이들이 살던 지역의 명칭의 첫 글자를 따서 부르는 것'으로 지금의 두음절어와 유사한 성격을 가진 단어로 보인다. (6나)의 예들은 명사구이다. '세자빈인', '관현맹인', '임오년수립'은 '세자빈', '관현', '임오년'이 수식어 명사구, '인', '맹인', '수립'이 피수식어 명사나 명사구이다.[31] (6다) '시명지보'의 경우 명사구 내부 구성에서 출현하는 '之'가 확인되므로

31) '임오년수립'은 사도세자의 죽음이라는 특정 사건을 가리키는 표현이다. '임오년수립'은 고유명으로 볼 수 있으나 명사구의 가능성이 있어 (6가)와는 분류를 달리 하였다.

명사구로 볼 만한다. (6라)의 '제세안민', '무본억말', '세사미두', '종성부용초성'은 동사구의 구조를 가지고 있다. '세사미두'의 경우 『고전』에서는 '해마다 대마도주(對馬島主)에게 내리는 쌀과 콩'으로 뜻풀이가 명사구로 되어 있으나, 용례를 보면 '해마다 쌀, 콩 등의 물건을 하사한다.'로 해석되어 뜻풀이와 달리 동사구로 분류할 수 있겠다.[32] 매우 소략한 분석이기는 하나 『고전』도 표제어가 꽤 다양한 언어적 특성을 가진 것을 확인할 수 있다.

다음은 『옛편지』의 일부 표제어를 언어적 특성에 따라 분류한 것이다.

(7) 가. 명사 : 칠황(柒黃)
나. 명사 or 명사구 : 균후(勻候), 정리(政履), 사리(仕履), 시학(侍學) / 하사(下駟), 혜향(惠餉), 존당(尊堂), 영전(令前)
다. 명사구 : 서하지척(西河之慽)
라. 동사 or 동사구 : 찬하(攢賀), 만만(萬萬), 최운(摧隕)
마. 동사구 : 불비식(不備式), 가건(加健), 취중(就中), 엄연관사(奄捐館舍)

(7가)의 '칠황'은 '황칠나무의 수액'으로 '황'에 '수액'이라는 의미를 확인하기 어려워 합성명사로 보아야 할 것이다. (7나)의 '균후'는 '정승의 안부를 물을 때 사용하는 말인데 '균'은 안부를 묻는 상대가 정승이기 때문에 사용된 것이며, '후'는 '체후'를 나타내는 것이다. '균'과 '후'의 결합에 따른 의미의 합성으로만은 '균후'의 의미를 확인하기 어려우므로 합성어로 볼 수도 있고, 관용적 의미를 가진 명사구로 분석할 수도 있다. 안부를 물을 때 자주 사용되는 '정리, 사리, 시학' 등도 '균후'와 유사한 특성을 가진다. '하사(下駟), 혜향(惠餉), 존당(尊堂), 영전(令前)'은 의미합성성이 지켜지고 있는 표현들이라 명사구로 해석될 가능성이 있다. (7다)는 고사에서 유래한 말로 축자적 의미 외에 추가적인 의미를 가지고 있지만 명사구 내부 표지인 之가 있으므로 관용적 의미를 획득한 명사구라고 할 수 있겠다. (7라)의 '찬하', '만만', '최운'은 고립어에서 자주 나타나는 동사 연속 구성으로 해석될 수도 있고, 만약 이들이 결합하여 자주 사용되는 예들이라면 합성어로 해석할 수도 있겠다. (7마)는 편지글에서 상투적으로 자주 등장하는 표현이기는 하나 비교적 의미합성성이 잘 지켜지고 있어 동사구의 구조를 가지는 것으로 볼 수 있다. 『옛편지』가 낱말 사전을 표방하고 있기는 하나 몇몇 표제어를 분석한 결과 단어뿐만 아니라 구로 볼 만한 표제어들도 꽤 나타나는 것을 확인할 수 있었다.

32) 『고전』의 '세사미두'의 다음의 용례에서는 동사구로 해석된다. "我朝以貴島 世效忠勤 鎭守海服 勞績可嘉 故**歲賜米豆**等物 恩數優渥 無以復加… [명종실록 권제14, 36장 뒤쪽, 명종 8년 윤3월 10일(병진)]" 그러나 "當此機會, 歲賜米豆及歲遣船等, 斟酌加給何如[조선왕조실록, 중종실록, 중종 36년]"에서는 '세사미두'가 명사구로 해석된다. 본 연구에서는 『고전』의 표제어를 분석하는 것이므로 일단 『고전』에서 제시한 용례를 바탕으로 판단하였고, 언어 맥락이 달라지면 '세사미두'의 언어적 특성이 달라질 수 있다는 것을 밝힌다.

Ⅳ. 한자 어휘 의미망 개발을 위한 전제적 논의점

이 장에서는 3장에서 논의된 내용을 바탕으로 어휘 의미망 개발과 관련하여 논의해야 할 점에 대해서 언급해 보려고 한다.

첫째, 어휘의미망 개발을 위한 디지털 데이터의 선정에 대한 것이다. 3장의 분석을 통해 『한한』, 『고전』, 『옛편지』의 표제어의 성격에 대해 살펴보았다. 『한한』은 표제어의 범위가 넓고 규모도 매우 크므로 한문 고전 문헌의 여러 단어와 표현들을 포괄할 만하다. 『고전』과 『옛편지』도 고전 한문 자료를 바탕으로 한 사전이나 이들은 『한한』과 표제어의 규모에서 큰 차이가 있으며 특정 영역의 문헌만을 바탕으로 선정된 표제어를 수록하고 있다. 『고전』과 『옛편지』에서 바탕으로 한 문헌들이 장르적 특정성이 있으며, 특히 『고전』의 경우 용어 사전의 성격을 가지고 있어 표현의 언어적 제한에서 더 자유로울 수 있으므로 『한한』에서 포착하지 못한 표현들도 있을 수 있겠다. 그러나 『한한』의 표제어 수가 워낙 크기 때문에, 더 자세한 조사를 해 보아야겠지만 『고전』, 『옛편지』의 표제어가 『한한』과 일부 겹칠 가능성이 있다.[33]

만약 한문 문헌을 바탕으로 한 범용 어휘의미망을 개발한다고 하면 『한한』의 표제자들만을 대상으로 개발하되 관련어 정보를 『고전』, 『옛편지』와 같은 특수 목적 데이터나 고전용어 시소러스 등의 데이터에서 추가로 추출할 수도 있을 것이다.[34] 그러나 하위 온톨로지 수준까지의 장르 특정적인 표현들까지 포괄한 어휘의 미망을 개발한다고 하면 개발된 다양한 한자 용어, 어휘 데이터들이 어떤 의미나 장르적 성격을 보이는 것인지 파악하고, 『한한』과 겹치지 않는 어휘들을 걸러내어 포함하는 작업이 필요할 것이다.[35]

둘째, 어휘 의미망의 대상으로 삼을 어휘의 선정에 대한 것이다. 3장에서 살펴본 바와 같이 『한한』, 『고전』, 『옛편지』의 표제어들은 그 언어적 특성이 사전 구조상으로 구분되어 있지 않다. 단어에서부터 새로운 의미를 획득한 관용구, 의미 합성성을 지키고 있는 통사적 구성에 이르기까지 꽤 다양한 언어적 특성을 가진 것들이 혼재되어 있다. 보통 어휘 의미망은 명사 어휘 의미망, 동사 어휘 의미망 등과 같이 유사한 언어적 특성을 가진 것들을 모아서 개발되는데, 본 연구에서 살핀 사전 표제어들은 여러 단위의 것이 혼재되어 있으므로 이들 중 어떤 것을 한자 어휘의미망의 대상으로 할 것인지 골라내는 작업이 필요할 것으로 보인다. 단어에 해당하는 것만을 고를 것인지, 구이지만 제3의 의미를 가지고 사용되는 것을 포괄할 것인지, 의미 합성성을 지니고 있는 구는 제외할 것인지 등등의 고민이 필요해 보인다.[36]

---

33) 본 연구에서 살펴본 『고전』의 표제어 중 '도승지(都承旨), 숙의(淑儀), 관현맹인(管絃盲人), 시명지보(施命之寶), 세사미두(歲賜米豆)', 『옛편지』의 표제어 중 '균후(勻候), 정리(政履), 찬하(攢賀), 최운(摧隕), 취중(就中)' 등이 『한국한자어사전』에서도 발견된다. 다만 『한국한자어사전』에서는 '관현맹인'은 '관현맹', '세사미두'는 '세사미'로 실려 있다는 차이가 있다.

34) 현실적으로 기구축된 한자 어휘 데이터들의 언어적 성격을 파악하고 그 성격에 따라 유형을 분류하는 것은 쉬운 작업이 아니다. 따라서 범용 차원의 온톨로지를 개발한다고 하면 『한한』의 표제자만을 대상으로 하는 것도 하나의 대안이 될 수 있다. 『한한』의 표제어 규모가 꽤 크므로 한국 한문 텍스트의 상당수의 어휘를 포괄할 수 있을 것이며, 한국 한문 텍스트가 반영하고 있는 언어는 이전 시기 중국어이므로 음절 1단어 1자의 특성을 근, 현대 시기 중국어보다 더 현저히 보이고 있기 때문이다. 또한 『한한』의 2음절 이상의 단어 중에는 고유명이 상당수인데 어차피 이들은 온톨로지 개발 시 어휘 관계 설정에서 보통명과 동일하게 처리될 수 없으므로 필요하다면 별도의 온톨로지로 개발해야 한다.

35) 혹시 동일한 표현이라고 해도 한자 어휘 사전 데이터의 특성에 따라 다르게 뜻풀이될 가능성이 있으므로 이에 대한 고려는 필요할 수도 있다.

V. 결 론

본 연구에서는 고전 한자 어휘의미망 개발을 위해 디지털 한자 사전 3종을 임의로 선택하여 그 성격에 대해서 분석하고 이를 바탕으로 어휘 의미망 개발 시 전제적 논의점에 대해 확인해 보았다. 연구의 결과는 각 장의 후반부와 4장에서 종합하여 언급된 바 있으므로 연구의 한계를 언급하고 한문 자료 정보화를 위한 제안을 하는 것으로 논의를 마무리 짓도록 하겠다.

본 연구는 연구 대상으로 삼은 사전의 표제어 선정의 기준이나 각 과정에 대한 선행연구를 충분히 찾기가 어려워 제한된 자료를 바탕으로 성격을 파악한 면이 있다. 또한 추가적으로 국어사전의 표제어의 성격에 기대어 한자어 사전 표제어의 언어적 특성에 대해 분석하는 방법을 취하였다. 각 사전 표제어 선정에 대한 선행연구의 부족을 뛰어넘으려면 사실상 각 사전의 모든 표제어들을 분석의 대상으로 삼아, 의미별, 추출 문헌별 등으로 분류해 보고 계량하는 방법이 있겠지만, 현재 이 연구는 거기에까지는 미치지 못하였다. 그리고 한자 어휘 표제어의 언어적 특성에 대해서는 고대 중국어 문법 전문가의 연구가 더 있어야 할 것으로 보인다.

어휘 의미망을 포함하여 한문 자료의 정보화를 위한 각종 자료의 개발을 위해서는 기구축된 자료에 대한 언어적, 의미적 특성을 포함하여 다양한 관점에서의 분석이 필요할 것이다. 본 연구에서는 표제어의 성격에 대해서만 분석하였으나 뜻풀이나 각종 약물로 주어진 정보, 관련어 정보 등도 분석의 대상이 될 수 있다. 이러한 분석이 정교해졌을 때 데이터마다 무엇이 없는지, 어떤 것을 골라내야 하는지, 필요한 자료를 어떻게 데이터 마이닝(data mining)할 수 있을지 등의 다음 단계로 나아갈 수 있다. 따라서 앞으로도 기구축된 한자 어휘 자료의 특성에 대한 다양한 관점의 논문과 또 디지털 자료의 개발 과정에서 도출되는 다양한 문제들과 그 해결 과정에 대한 논문이 지속적으로 학계에 발표되고 공유될 필요가 있을 것이다.

36) 여기서는 자세히 분석하지 않았으므로 본문 안에서 기술하지는 않으나 어휘의미망은 품사별로 개발되는 사례가 많아 현재 사전에 품사 정보가 주어지지 않은 것에 대해서도 어휘의미망 개발시 고민이 필요할 것으로 보인다. 중국어의 경우 명사와 동사의 전성(활용)이 매우 일반적 현상이므로 품사 구분 없이 어휘들을 모아 어휘의미망을 개발할 수도 있겠으나 전형적으로 동사로만 쓰이고, 명사로만 쓰이는 예들도 있을 것이므로 이들을 품사 구분 없이 어휘의미망을 개발하는 것이 적절한 것인지에 대해서도 논의해야 할 부분이다.

〈참고문헌〉

강인선, 종합국어대사전 편찬 사업,『새국어생활』5: 1, 국립국어원, 1995.

고려대학교 민족문화연구원,『고려대 한국어대사전』, 고려대학교 민족문화연구원, 2009.

국립국어연구원,『표준국어대사전』, 두산동아, 1999.

국립국어원, 한국어기초사전, https://krdict.korean.go.kr/, 2016.

남길임,「말뭉치 기반 사전 편찬의 동향과 지향 방향: 최근 30년간의 사전편찬 방법론을 중심으로」,『한말연구』16, 한말연구학회, 2005.

남길임,「부표제어의 범위와 유형-속담·관용표현·연어·패턴·상투표현·자유표현의 기술」,『한국사전학』9, 한국사전학회, 2007.

남기심·김한샘,「말뭉치 활용 한국어사전 편찬의 실제: 연세한국어사전 사례를 중심으로」,『한국사전학』30, 한국사전학회, 2017.

단국대학교 동양학연구소 편,『한국한자어사전』, 단국대학교 출판부, 1996.

단국대학교 동양학연구소 편,『한한대사전』, 단국대학교 출판부, 2008.

도원영,「'민연 국어사전(가칭)'의 표제어 선정과 그 실제에 대하여」,『한국어학』12, 한국어학회, 2000.

도원영·차준경,「고려대 한국어대사전』의 종합적 고찰」,『민족문화연구』51, 고려대 민족문화연구원, 2009.

배은한·박용범·허철,「한문고전 인공지능 번역 연구의 필요성과 선결 과제」,『한문학논집』53, 근역한문학회, 2019.

서상규,『한국어기본어휘의미빈도사전』, 한국문화사, 2017.

송길룡·민경모·서상규,「『표준국어대사전』의 표제어 구성에 대하여」,『한국사전학회 학술대회 발표논문집』, 한국사전학회, 2003.

안의정·황은하,「『연세 현대한국어사전』의 표제어 목록 구성의 이론과 실제」,『한국사전학』15, 한국사전학회, 2010.

연세대학교 언어정보개발연구원,『연세한국어사전』, 두산동아, 1998.

연세대학교 언어정보개발연구원,『연세 초등국어사전』, 두산동아, 2002.

유현경,「연세 한국어사전과 고려대 한국어대사전 비교 연구: 언어사전으로서의 특성을 중심으로」,『한글』295, 한글학회, 2012.

유현경·남길임,『한국어사전 편찬학 개론』, 도서출판 역락, 2008.

윤애선,「국·내외 어휘의미망 구축과 활용」,『새국어생활』17: 3, 국립국어원, 2007.

윤용선,「종합국어대사전의 표제어」,『새국어생활』5: 1, 국립국어원, 1995.

이수진,「중한사전의 표제어 선정과 구성에 관한 문제」,『중국어문학』61, 영남중국어문학회, 2012.

정희원,「표제어 선정의 제문제」,『새국어생활』5: 1, 국립국어원, 1995.

최혜원, 「표준국어대사전의 특징」, 『새국어생활』 10: 1, 국립국어원, 2000.
하영휘 외, 『옛편지낱말사전』, 돌베게, 2011.
한국고전용어사전 편찬위원회, 『한국고전용어사전』, 세종대왕기념사업회, 2001.
허철, 「고전적(古典籍) 전산화에서 말뭉치 언어학 적용의 필요성과 방향에 대한 일견」, 『민족문화』 47, 한국고전번역원, 2016.
Jackson, H., *Lexicography: An Introduction*, New York: Routledge, 2000.
Sinclair, J. M.(Ed.), *Looking up: An account of the COBUILD project in lexical computing and the development of the Collins COBUILD English language dictionary*. Collins Elt, 1987.

* 이 논문은 2022년 5월 6일에 투고되어,
2022년 6월 21일에 심사위원을 확정하고,
2022년 7월 7일까지 심사하고,
2022년 7월 11일에 게재가 확정되었음.

▌Abstract ▌

A Study on the Characteristics of Headwords in Chinese Dictionaries for a Development of a Classical Chinese Lexical Semantic Network

- Focused on 『Great Chinese-Korean Dictionary』, 『Dictionary of Korean-Chinese Characters』, 『Dictionary of Korean Literature's Term』, 『Word Dictionary of Traditional Letter』 -

Son, Hyeok* · Kim, Dami** · Cha, Myunghee*** · Kim, Woojeong****

The aim of this present study is to examine the characteristics of headwords in digital Chinese dictionaries, which are the basic resources for developing a classical Chinese lexical semantic network, and to find out what needs to be discussed before developing it. It is practically impossible to examine all of the digital Chinese dictionaries, so we select 『Great Chinese-Korean Dictionary』 with largest number of headwords, the special dictionaries, 『Dictionary of Korean-Chinese Characters』, 『Dictionary of Korean Literature's Term』, 『Word Dictionary of Traditional Letter』 as the research subjects. The results of examining the characteristics of these dictionary headwrods are as follows. First, 『Great Chinese-Korean Dictionary』 and 『Dictionary of Korean-Chinese Characters』 have the greatest coverage of classical Chinese vocabulary, while 『Dictionary of Korean Literature's Term』, 『Word Dictionary of Traditional Letter』 contain vocabulary of specific genres. Second, in all four dictionaries, the linguistic features of the headwords were not reflected in the macrostructure of them. Therefore, except for the head-characters of 『Great Chinese-Korean Dictionary』, which can be seen as a word, it seems that the four dictionaries contain headwords with quite a variety of linguistic characteristics, such as a word, idioms that acquire new meanings, patterns, typical syntactic constructions. Through the above research results on the characteristics of headwords, we were able to find out several points to discuss before developing a classical Chinese lexical semantic network. First, the selection of appropriate data according to the purpose of developing lexical semantic networks should be discussed. Second, it is necessary to discuss whether it is possible to select data with homogeneous linguistic characteristics from digital

* Lecturer, Department of Korean Language & Literature, Yonsei University
** Ph.D Program Researcher, Han-Character Education Research Center, Dankook University
*** Lecturer, Department of Korean Language & Literature, Dankook University
**** Professor, Department of Sino-Korean Education, Dankook University

Chinese vocabulary data, and how to select it if possible.

[Keywords] classical Chinese lexical semantic network, headword, 『Great Chinese-Korean Dictionary』, 『Dictionary of Korean-Chinese Characters』, 『Dictionary of Korean Literature's Term』, 『Word Dictionary of Traditional Letter』