

중한사전의 표제어 선정과 구성에 관한 문제

李 修 珍*

<目次>

- | | |
|-----------------------------------|---------------------|
| I. 머리말 | IV. 중한사전의 표제어 선정 과정 |
| II. 기존 중한사전의 표제어 선정
내용에 대한 문제점 | V. 중국어 코퍼스 활용의 문제점 |
| III. 중한사전의 표제어 선정
기준과 처리 | VI. 나가기 |

I. 머리말

사전의 표제어란 사전에 등재되어 있는 올림말을 가리키는 것으로, 사전의 거시 구조를 구성하는 중심요소이다. 따라서 표제어의 범위를 설정하고 구체적 어휘들을 선정하는 것은 사전 편찬의 전 과정 중에서 가장 기초가 되는 작업이며, 사전 편찬학에 있어서 표제어에 관한 연구 또한 이러한 이유로 인해 중요한 연구 분야가 된다.

홍종선 외(2009;166~167)는 사전에서 표제어가 가지는 역할 네 가지를 다음과 같이 설명한다.

- 첫째, 해당 사전의 어휘 정보를 찾을 수 있도록 하는 길잡이
- 둘째, 해당 언어의 표준적 어형 제시
- 셋째, 해당 사전의 성격을 구현

* 부산대학교 시간강사

넷째, 민족어의 어휘 전체를 드러내는 것

이 중에서 어휘 정보의 길잡이와 표준적 어형 제시라는 역할은 이중어 사전의 표제어에서 더욱 부각된다. 대부분의 이중어 사전은 외국어 학습의 기본 도구서나 참고 자료로 활용되므로, 사전 사용자는 해당 외국어에 대한 직관에 의존하여 표제어를 판단하기 힘들며 전적으로 사전 편찬자가 제공하는 정보에 의존하기 때문이다. 중한사전의 경우 표제어가 되는 중국어가 역사적으로, 지역적으로 아주 방대한 사용역을 가지고 있는 만큼 그 어휘수 또한 방대하므로, 어떤 단어를 어떤 형태로 중한사전 사용자에게 제공할 것인가 하는 것은 중요한 문제가 될 수 있다. 그럼에도 불구하고 중한사전의 표제어에 관한 국내 연구는 상당히 미미한 편으로, 이상도(1998)의 논문에서 중국어 사전의 표제자 수록 문제를 언급하고 있고, 중한사전 편찬에서 코퍼스의 활용을 논의한 박영중(2008)의 연구에서 일부 언급될 뿐이다. 이러한 연구 동향은 국내 중한사전에 대한 수요나 출판 현황에 반비례한다고 할 수 있다. 특히 최근 증보되고 출판되는 중한사전들 중 대다수가 대사전 편찬 지향으로 인해 가능한 많은 수의 (표제자를 포함하여) 표제어를 수록하려는 경향을 보인다는 것을 고려한다면, 관련 연구를 바탕으로 한 표제어 선정 문제를 점검할 필요가 있다고 여겨진다.¹⁾ 이에 본고에서는 국내에서 출간된 중한사전의 표제어 선정과 구성에 대한 기준과 원칙을 다시 한 번 점검하고, 보다 정제된 표제어의 선정과 구성을 위한 방법을 모색하고자 한다.

1) 중한사전은 중국어의 언어적 특성상 표제자와 표제어를 모두 수록하고 있으나 본고는 표제자의 대부분이 일음절 단어에 해당하므로 표제자 또한 넓은 의미에서 어휘로 본다. 이에 표제자와 표제어를 구분하지 않고 임의적으로 표제어에 포함시켜 논의하고자 한다.

II. 기존 중한사전의 표제어 선정 내용에 대한 문제점

현재 국내에서 출간된 중한사전은 그 종류와 규모에 있어 괄목할 만한 성과를 이루었다고 할 수 있다. 특히 그 중 4종의 중한사전이 문언어, 고유어, 전문용어 등 가능한 많은 어휘들을 모으고 정리했다는 점에서 확장된 정규 사전으로 볼 수 있으며, 수록하고 있는 표제자와 표제어 수에 근거해 볼 때 대사전에 속한다고 할 수 있다.²⁾ 대사전의 규모를 가진 정규사전이라면 수록하고 있는 어휘수와 어휘범주에 근거하여 다른 사전과 차별화되어야 하며, 이를 위해 실제로 필요 이상으로 표제항의 숫자를 확장시키기도 한다. 가능한 많은 양의 어휘 정보가 제공된다는 점에서 볼 때 표제어의 수가 많다는 것은 사전의 커다란 장점이 되지만, 만일 그 표제어가 모국어 화자들 사이에서 사용되지 않고 인정받지 않는 것이라면 확장된 표제어의 수는 더 이상 장점이 될 수 없다.

중한사전의 표제어는 어떠한 범주와 규모를 가지며, 그 기준은 무엇이어야 하는 지 알아보기 위해 위에서 언급한 4종의 대규모 중한사전에서는 표제어의 범주를 어떻게 규정하고 표제어의 선정 기준은 무엇인지 살펴보

2) 현재 국내에서 출간된 중한사전은 대략 17종이다. 그 중 본고에서는 규모가 가장 크며 또한 2005년 이후 개정된 사전에 한해서 선택한 결과 4종의 사전으로 추려낼 수 있었다. 나머지 사전은 동일한 출판사에서 어휘수를 줄여 소사전 혹은 포켓용으로 따로 편집해서 만든 것으로 수록하고 있는 어휘수가 4종에 비해 크게 적거나, 개정된 지 오래되어 표제어 선정과 구성을 총괄적으로 살펴보기에는 부족하다고 여겨져 검토대상에서 제외하였다. 그 4종은 모두 아래와 같다.

- i) 고대민족문화 연구원 중국어대사전 편찬실 편저, 《中韓辭典》, 2008년 2판 4쇄본, 고려대학교 민족문화 연구원 출판. 이하 《(고대)中韓辭典》이라고 약칭한다.
- ii) 한양대학교 교수 손예철 편저, 《Prime 東亞 中韓辭典》, 2쇄; 두산동아, 2008년. 이하 《Prime中韓辭典》이라고 약칭한다.
- iii) 문학박사 이용목 편저, 《民衆 옛센스 中國語辭典》 中韓·韓中 合本, 3판 10쇄; 民衆書林, 2010년. 이하 《옛센스 中國語辭典》이라고 약칭한다.
- iv) 문학박사 박영중 편저, 《現代中韓辭典》, 3쇄; 교학사, 2009년.

고, 수록 현황을 살펴 그 결과로 도출되는 문제점이 무엇인지 살펴보도록 한다.

대부분의 사전들은 머리말이나 일러두기를 통해 사전의 편찬과 구성, 지침에 관한 기초 정보들을 제공한다. 본고에서 살펴보고자 하는 4종의 중한사전 또한 마찬가지이지만, 살펴본 바에 따르면 표제어의 선정 기준이나 원칙에 대해서는 대체로 원론적인 내용만을 제시하고 있을 뿐이다. 즉, ‘표준어로 구성된 일상어휘와 전문어, 외래어, 신조어, 문언어, 방언, 속담, 성어(成語), 관용어, 험후어(歇後語)’ 등의 표제어 범주와 적개는 10만 여 개에서 많게는 18만 여 개에 이르는 표제항목의 수만을 제시하고 있을 따름이다. 다만 《Prime 中韓辭典》과 《現代中韓辭典》은 모두 기본적으로 1986년에 발표된 <簡化字總表(新版)>를 표제자의 1차 자료로 삼고 있으며, 이 외 《Prime 中韓辭典》은 중국과 대만의 대표적인 사전을 2차 자료로, 《現代中韓辭典》은 중국 정부의 발표 자료를 2차 자료로 삼고 있다는 것을 머리말과 일러두기를 통해 일러주고 있다. 그러나 중국어 사전의 어휘 범주에는 일음절 표제자뿐 아니라 이 표제자로 구성된 ‘표제어’까지 포함될 고려할 때 표제어의 선정 기준과 과정에 대해서는 구체적인 언급이 없다는 사실에 주목할 필요가 있다. 그렇다면 이 4종의 사전에서 사용 빈도수가 비교적 높은 표제자를 선택하여 하위 표제어의 유표 빈도수를 통계분석하고, 이를 통해 기존 중한사전이 가지고 있는 표제어 선정에 있어서의 문제점 여부를 살펴보자.

우선 《現代漢語常用詞表》³⁾에서 사용 빈도수 20위 안에 속하며 동시에 하위 표제어 항목이 1000여 개에 이르는 글자 중 실사인 ‘大’와 허사인 ‘不’를 대표자로 선택해 그 표제어의 수록 현황을 살펴보았다. 아래 표는 표제자 ‘大’의 하위 표제어에 해당하는 단어 목록을 추출하여 각 단어들의 실제 사용 현황을 조사하여 도출한 통계자료이다. 여기서 ‘실제 사용 현황’이라 함은 각 단어들이 코퍼스와 검색 사이트에 출현하는 빈도에 근거한

3) 國家語言文字工作委員會 編輯, 《現代漢語常用詞表》, 商務印書館出版, 2009年.

것이다.4) 비출현 표제어는 코퍼스와 검색 사이트에 한 번도 출현하지 않은 부류이고 저빈도 표제어는 코퍼스에서 3회 미만 출현하는 것이다.

i) ‘大’

	전체 표제어	비출현 표제어	저빈도 표제어
《(고대)中韓辭典》	1210	133 (10.99%)	98 (8.09%)
《Prime 中韓辭典》	1118	108 (9.66%)	85 (7.60%)
《옛센스 中國語辭典》	960	97 (10.10%)	77 (8.02%)
《現代中韓辭典》	663	23 (3.49%)	13 (1.96%)

ii) ‘不’

	전체 표제어	비출현 표제어	저빈도 표제어
《(고대)中韓辭典》	881	38 (4.31%)	42 (4.76%)
《Prime 中韓辭典》	872	34 (3.89%)	44 (5.04%)
《옛센스 中國語辭典》	795	34 (4.27%)	32 (4.02%)
《現代中韓辭典》	673	4 (0.59%)	7 (1.04%)

박동근(2011;121~122)은 이처럼 표제어로 수록되어 있으나 실제로는 거의 사용하지 않는 낱말을 유령어라고 규정한다. 이 때 실제로 사용하지 않는다는 것을 그는 두 가지 측면에서 해석하는데 첫째, 본래 언중들이 사용하던 말인데 시간이 지나면서 현재는 사용하지 않아 사어(死語)로 인정

4) 코퍼스와 웹 사이트 또한 검색 자료로 활용한 이유는 사전은 다소 잘못된 언어 형태라고 해도 그 유효 빈도수가 높으면 표제어로 수록하여 이용자들이 검색할 수 있도록 해야 하는데 코퍼스의 언어 자료는 비교적 모범적이고 규범적인 성격을 가지므로 일반 언어 사용자들이 더 많이 접근하는 웹 사이트에서 그 근거를 찾고자 하였기 때문이다. 검색에 활용한 코퍼스와 웹 사이트는 아래와 같다.

CCL : http://ccl.pku.edu.cn:8080/ccl_corpus/
 國家語委現代漢語 : <http://202.114.41.172:9090/cqs/index.gsp>

되는 경우, 둘째 현재 사전에 표제어로 수록되어 있긴 하지만 사전 외에는 사용 출처를 확인할 수 없는 말로 본래 사용되지 않았던 말이라고 한다.⁵⁾ 상술한 표에서 표제자 ‘大’의 표제어 중 평균 8%에 달하는 표제어가, ‘不’의 표제어 중 평균 3.2%에 달하는 표제어가 유령어라는 것이다. 그 중 일부가 아래의 예들이다.

- i) 大包单儿/大颠儿/大大洋洋/大草/大墩
- ii) 不登木/不育风/不幸娃娃/不分主从/不学无文

이러한 유령어들은 내용 범주와 단어 범주로 나누어서 살펴볼 수 있다. 대체로 유령어들의 내용 범주는 해당 사전에서 제공하는 석의(釋義)에 근거해볼 때 출처나 유래를 알 수 없는 단어, 고어(古語), 문언어, 방언, 고유명사, 전문 용어, 이체어인 경우가 많고, 유령어의 단위 범주는 표제어의 기본 단위인 단어의 범위를 넘어 명사나 문장으로 이루어진 경우가 많다.

표제어의 내용 범주에는 문언어, 방언, 고유명사 등도 포함된다. 이들이 비록 일상 어휘에는 속하지 않는다 해도 언어 사용자들의 다양한 언어활동 중에 여전히 활용되는 어휘군이기에 때문이다. 뿐만 아니라 중한사전의 이용자들이 일상생활용 어휘의 검색 외 고대 중국의 자료 강독 및 연구, 여러 분야의 번역 등을 위한 도구서로 중한사전을 사용하기도 함을 고려한다면 사전 사용자의 입장에서 여전히 필요한 부분이기도 하다. 문제는 문언어이기 때문에, 혹은 방언이기 때문이 아니라 이러한 어휘들 중에는 유효 빈도수가 전혀 산출되지 않는 것들도 포함되어 있다는 사실이다.

또한 단위 범주에 있어서도 마찬가지이다. 구는 고정된 어휘가 아니라 끊임없이 조합되고 분리되는 생성적 특징을 가진다. 이들이 모두 사전의 표제어로 등재될 수는 없으며, 만일 수록된다면 그것은 하나의 관용어구가

5) 이러한 유령어를 가리키는 말로 잠어(latent word)라고도 한다. 시드니 I. 랜도우(2002:99, 김영안 외 역)는 존재할 수 있지만 그것이 사용된 일이 있다는 것을 증명할 수 있는 기록이 없는 단어를 이와 같이 설명한다.

된 경우, 또는 구로 조합되면서 단어와 단어의 사전적 의미 결합이 아니라 새로운 의미를 생성해내는 경우이다. 그러나 ‘大远(아주 먼)’, ‘大肚(큰 배)’ ‘煎鸡蛋(달걀을 부치다)’, ‘讲卫生(위생을 중시하다)’⁶⁾과 같이 새로운 의미를 파생시키지 못하는 구도 표제어로 등록되어 있는 예가 많았다. 앞서 살펴 본 ‘不’의 표제어 구성에서도 이러한 특성이 많이 나타났는데 ‘不’은 허사적 성격으로 인해 단어를 구성하는 형태소로 작용하기 보다는 동사나 형용사를 수식하여 단어구를 표제어로 형성하는 경우가 훨씬 많았다. 예를 들면 ‘不见’ ‘不方便’ ‘不忙’ ‘不明白’ ‘不同’과 같이 ‘不+V’의 형태로 ‘V’의 부정형 의미를 그대로 추측할 수 있는 동사구, 성어 등이 전체 표제어의 90%를 웃돌고 있다. 이러한 ‘구’들은 유효빈도수는 높지만 두 단어의 조합으로 그 의미를 충분히 유추해낼 수 있음에도 불구하고 표제어로 선정되어 있다.

이러한 점들을 고려하여 다음 장에서는 중한사전의 표제어 선정 기준과 내용에 대해 살펴보자.

Ⅲ. 중한사전의 표제어 선정 기준과 처리

본 장에서는 중한사전의 특성을 통해 표제어 선정의 기본 원칙을 점검하고 실제 사전 편찬과정 중 표제어의 범주별 문제들을 어떻게 처리해야 할 것인지 살펴보고자 한다.

1. 중한사전의 특성과 표제어 선정의 기본 원칙

章宣華·雍和明(2007;213~219)는 중국어 사전의 표제어 선정 원칙을 9가지로 나누어 설명하는데 그 내용을 요약하면 아래와 같다.

6) ‘大远(아주 먼)’, ‘大肚(큰 배)’ ‘煎鸡蛋(달걀을 부치다)’는 2종의 중한사전에서, ‘讲卫生(위생을 중시하다)’는 1종의 중한사전에서 표제어로 수록되어 있다.

- i) 현실성 원칙[現實性原則]: 객관적인 언어 사실을 반영해야 한다.
- ii) 보편성 원칙[普遍性原則]: 언어 사회 내에서 보편적으로 인정받고 사용되는 단어들을 수록해야 한다.
- iii) 대상성 원칙[針對性原則]: 사전편찬의 목적과 대상에 따라 수록어휘 범주가 달라야 한다.
- iv) 규범성 원칙[規範性原則]: 언어 사회 내에서 표준어로 인정받는 언어 표현 형식을 수록해야 한다.
- v) 논거성 원칙[理據性原則]: 신조어의 조어 형식과 의미 형식에 신뢰할 만하고 타당한 논리와 근거가 뒷받침되어야 한다.
- vi) 고빈도 원칙[高頻次原則]: 언어사회 내에서 해당 단어가 가지는 언어적 지위와 상용 정도를 나타내는 중요한 요소로 사용빈도가 높아야 한다.
- vii) 정착성[穩定性原則]: 단어의 생명이 일정 정도의 시간 동안 지속되어 언어 사용자들 사이에 정착되어 있어야 한다.
- viii) 생산성 원칙[能產性原則]: 새로운 단어를 조어해낼 수 있는 잠재적 형태소를 수록한다.
- ix) 경제성 원칙[經濟性原則]: 간결하면서도 의미가 명확해서 경제적 언어 효과를 가질 수 있는 단어들을 선정한다.

이 9가지 원칙들은 비교적 원론적인 내용으로 중한사전 뿐 아니라 대부분의 언어 사전에 적용할 수 있을 것이다. 다만 상술한 원칙 중 고빈도 원칙에 대해서 좀 더 고려해볼 필요가 있다. 현대적 의미의 사전 편찬에서 가장 기초적인 기준은 사용자들의 사용 빈도에 있다. 이 때 빈도란 일상생활에서 실제하는 어휘의 사용 빈도수를 말하는 것으로 상술한 9가지 원칙 중 가장 중요한 선정 기준으로 작용할 수도 있다. 그러나 고빈도 원칙은 모든 표제어에 일괄 적용할 수는 없다. 중한사전과 같은 이중어 사전, 혹은 대역(對譯)사전은 외국어를 표제항으로 세우고 그에 대한 모국어 대역어를 제시하는 것이므로 그 주요 기능은 외국어 교육 및 학습, 통번역과 관련된다. 중한사전 또한 한국어 화자들이 중국어를 ‘이해’하고 중국어로 ‘표현’할 때 사용하는 도구이므로 그 ‘이해’와 ‘표현’을 위한 사전의 내용은 중국인들의 ‘이해’와 ‘표현’을 위한 것과는 다른 것이다. 또한 중국어는 문

자 및 기록 문화의 역사, 그로 인한 수많은 문헌 자료, 문어적인 서술 표현의 발달 등으로 고어, 문언어, 방언이 비교적 발달해 있고 전체 언어 사용량 중 그 비율 역시 적지 않다. 나아가 이 고어와 현대어, 문언어와 구어, 방언과 보통화의 구분이 비교적 두드러지기 때문에 이로 인한 언어 사용상의 차이가 크다고 할 수 있다. 즉, 중국어의 고어와 문언어, 방언은 ‘표현’을 위한 어휘라기보다는 주로 ‘이해’의 영역에 속하는 것으로 사용상의 유효 빈도수가 떨어질 수밖에 없다는 것이다.⁷⁾ 그러므로 중한 이중어 사전의 표제어가 현실 언어를 적극적으로 수용함으로써 실재하는 유의미한 정보를 사전에 반영하는 것 이외에도, 사전 편찬의 목적과 계획된 규모에 따라 고어와 문언어, 방언은 사용 빈도와 중요도에 근거하여 표제어로서의 올림 여부를 결정해야 한다.

이는 실제 사전 편찬 과정 중 표제어의 선정에 있어 중요한 문제가 된다. 비록 전자 사전이나 웹사전의 발달로 사전의 규모 자체가 문제시되지 않는다 해도, 지나친 정보의 범람과 정비되지 않은 정보의 내용이 반드시 유익하다고는 할 수 없기 때문이다.

다음 절에서는 상술한 원칙과 내용을 근거로 표제어의 내용 범주와 단위 범주와 관련된 세부 항목 및 처리 문제를 살펴보도록 한다.

2. 표제어의 내용 범주

언어 사전 표제어의 주된 대상은 표준어이다. 규범적인 단어와 표현을 제공하는 것이 언어사전의 기본 임무이며 그 사전이 외국어 학습의 도구로 사용되는 것이라면 정보의 정확성과 규범성은 더 많이 요구될 것이다.

7) 강현화(2010:46)는 표제어의 구성은 이해를 위한 어휘(Receptive Vocabulary)와 표현을 위한 어휘(Productive Vocabulary)로 나누어 구성하여 의미 기술의 수준을 조절해야 한다고 한다. 전자는 듣기와 읽기 등 해당 외국어를 이해하는 데 요구되는 어휘이고, 후자는 쓰기와 말하기 등을 표현을 위한 핵심적인 어휘를 말한다.

따라서 중한사전 표제어는 중국의 규범 언어, 즉 보통화여야 한다. 표제어의 내용 범주를 크게 두 가지로 분류할 경우 살아 있는 현대어로서의 보통화를 기준으로 하는 표준어와 방언이나 고어(古語), 속어(俗語) 등의 비표준어로 나눌 수 있다. 이 비표준어의 수록 여부는 사전의 규모와 관련지어 생각해야 하는데 실용위주의 중사전이 아니라 확장된 대사전의 경우 주요 표제어 대상으로 포함시킬 수 있다. 비표준어라고 해서 일괄적으로 사전의 표제어에서 제외될 수는 없다는 것이다. 실제로 비표준어라고 해도 높은 사용 빈도수를 보이는 어휘들이 있는데, 실제 사용되는 어휘들을 표제어 선정에 첫 번째 기준으로 꼽는 만큼 비표준어라고 해도 유효 빈도 이상을 보이는 경우 검토를 거쳐 표제어로 선정해야 할 것이다. 이 때 선정된 비표준어들은 문언어, 구어, 방언, 잘못된 어형 등의 관련된 미시 정보를 정확하게 제시하여 사전 이용자로 하여금 사전의 사용역 정보를 구분할 수 있도록 하는 것이 중요하다. 여기서는 사전의 내용 범주 중 (비표준어를 포함하여) 중국어의 특성에 기인하여 논의의 여지가 있는 몇 가지 항목에 한정해서 살펴보도록 한다.

1) 문언어

문언어는 ‘古語詞’라고도 불리는 만큼 고대 중국어를 어원으로 가지는, 구어에 상대되는 어휘를 말한다. 고대 중국어에서 비롯되었으나 비교적 특수한 의미를 나타내기 위해, 정중하고 격식을 갖춘 언어적 효과를 표현하기 위해, 혹은 정확하고 간결한 표현을 나타내기 위해 현대 중국어에서도 여전히 쓰여지는 어휘들이다. 이 문언어는 반드시 활자를 통해서만 전달되는 것은 아니며 학술발표, 강연 등의 구어적 텍스트에서도 찾을 수 있다. 일상에서도 이러한 문어적 환경은 쉽게 접할 수 있는데 이러한 점들은 문언어들의 유효 빈도수가 구어보다 낮다고 할 수 없음을 암시한다. 아래의 예들은 앞서 살펴본 표제어 ‘大’의 하위 표제어 중 국내 중한사전에 사용역이 문언어로 표시되어 있는 단어들로 각종 코퍼스에서 유효빈도수가 산출되지 않는다.

i) 大差/大舅太太/大两/大酺/大衾/大统/大裁/大轱辘儿车/大钧

이를 고려할 때 문언어들도 사전의 표제어로 선정될 수 있다. 다만 표제어 선정과 관련해 문언어는 두 가지 고려해야 할 점이 있다.

첫째, 중국어의 경우 문언어와 구어의 사용역이 비교적 뚜렷하게 구별되므로 표제어의 문언성을 표명해주어야 한다는 것이다. 사실 구어와 문언어에 대한 정의도 학자들마다 다르기 때문에 모종의 어휘에 대해 문언어로 사용역을 제한할 경우 그 기준도 논의되어야 한다.

둘째, 이 문언어와 관련해서 고려해보아야 할 어휘 범주로 조기 백화를 포함한 고어가 있다. 중국어는 오랜 역사적 시간 동안 많은 언어 변화를 겪으며 현대 중국어에 이르렀기 때문에 그 수량이 현대어에 버금간다. 따라서 사전의 규모상 고대 중국어와 현대 중국어를 하나의 사전에 모두 수록하기 힘들며, 또한 해당 어휘의 사용역이 다른 두 어휘군을 하나의 사전에 수록하게 되면 사전 이용자들이 이해와 사용에 혼란을 일으킬 수 있을 것이라고 추측된다. 그래서 중국에서 출간되는 한어사전류는 현대한어사전류와 고대한어사전류로 나누어 표제항의 성격을 달리하는 것이 일반적이다. 국내 출간된 중한사전의 경우 비교적 많은 양의 고어와 조기 백화들이 표제항으로 수록되어 있는데 문언어들이 고어를 어원으로 가지는 경우가 많기는 하지만, 유효 빈도수가 산출되지 않는 고어, 조기 백화들이 사전의 표제어로 선정될 수는 없을 것이다.

2) 방언

방언은 특정지역에서만 유통되는 언어이므로 규범성을 요구하는 이종어 사전에서 표제어로 선정할 수 있느냐 하는 것은 까다로운 문제가 된다. 더구나 중국어의 경우 넓은 지역적 분포에서 다양한 민족들이 다양한 방언을 사용하기 때문에 그 숫자와 내용은 다른 외국어에 비해 훨씬 방대하다. 그러나 방언이라고 해도 그 지역적 사용 범위가 확대되거나, 사용 인구가 증가되어 전국적으로 확대되는 경우 표준어로 흡수되기도 한다. 이러한 유

형의 방언은 사전의 표제어로 수록할 수 있다. 이 때 보통화와 방언의 경계를 구분하여 표제어의 기타 정보에 방언임을 분명히 표기해야 할 것이다.

그러나 해당 방언의 유효 빈도수가 낮다면, 대다수의 언어 사용자들에게 흡수 되지 못하고 있다는 것을 증명하는 것이므로 표제어로 수록될 수 없을 것이다. 앞서 제시한 국내 중한사전 4종 중 일부 사전에 등재되어 있는 아래 예들은 각각의 사전에서 방언으로 사용역 정보를 밝히고 있는 단어들인데, 각종 코퍼스와 검색 사이트에서 단 1회도 검색되지 않는 단어들이다.

- i) 大估摸儿/大家家儿/大拢意儿/大骚/大宾/大天
- ii) 不吃不斗/不得哥们儿/不分陇儿/不好价/不着家儿
- iii) 青刺嘎唧/伸把手儿/生外/守谱/死吞兒儿/讨老公/捅饥荒/爷老子/议合

방언이면서 유효 빈도수도 전무하다면 표제어로 등재될 수 없을 것이다.

3) 고유명사

고유명사는 특정 대상의 고유한 기호나 명칭을 의미하는 것으로 인명·지명·국명·책명·민족명·단체명 등으로 세분화할 수 있다. 대부분의 중국어 단일어 사전에서는 대체로 고유명사를 수록하지 않는다. 《現代漢語詞典》이나 《現代漢語規範詞典》⁸⁾의 경우 실제 표제어 항목에서도 ‘北京’과 같이 대표성을 지닌 일부 고유명사만이 수록되어 있다. 그러나 일부 고유명사의 경우 그 고유명사의 지시 범주는 해당하는 특정 대상에서 벗어나 그러한 특성을 가진 사람으로 보통명사화되어 활용되기도 한다. 동화나 신화속의 인물이 여기에 해당하는 예가 많은데 이 단어들이 동화나 신화와 같은 문화를 기반으로 하여 사람들 사이에서 또 다른 개념을 생성하기 때문이다. 중국 또한 오랜 역사와 문화를 기반으로 하여 이와 같은 형

8) 《現代漢語規範詞典》은 그 범례에서 고유명사를 수록하지 않는다고 밝히고 있다.

태로 그 의미를 공유하게 된 고유 명사가 많다. 諸葛亮과 같은 역사적 인물, 阿Q와 같이 문은 작품속의 인물, 毛澤東과 같은 당대 인물들이 그 예에 해당한다고 하겠다. 이 고유명사는 특정인을 지시함과 동시에 그러한 유형의 사람이라는 의미로 파생되어 阿Q精神, 毛澤東思想이라는 복합어를 생성하기도 한다. 특정한 지역이나 사람, 사물을 지칭하는 고유명사 자체의 기능도 중요하지만 이처럼 의미가 파생되는 경우 해당 단어는 표제어로 수록되어야 한다. 또한 표기가 번역된 고유명사들도 빈도수 및 중요도에 근거하여 표제어가 되어야 한다. 중국어의 외국어 표기 정책상 음역이나 의역 등으로 번역된 각종 고유명사들은 단어와 단어가 지시하는 내용을 일치시키기 어려운 경우가 많기 때문이다.

4) 신조어, 외래어

표제어 선정에서 가장 까다로운 작업이 될 수 있는 부분이 신조어에 관한 것이다. 새로운 사물이나 새로운 개념을 표현하게 되는 신조어는 새로운 단어로 만들어지기도 하고 원래 있던 단어에 새로운 의미가 부가되기도 한다. 또한 언어 사용자들 사이에서 자연스럽게 만들어질 수도 있고 정부의 언어 정책에 따라 계획적으로 만들어질 수도 있다. 사용자들의 사용 욕구에 부합하여 기성 어휘로 자리 잡을 수도 있으며 한시적으로 사용되다 폐기될 수도 있기 때문에 사전을 편찬하는 입장에서 이 신조어들을 표제어로 수록할 것인지에 대한 판단은 쉽지가 않다. 앞서 살펴본 표제어 선정 원칙에도 관련된 두 가지 원칙이 있는데 ‘논거성 원칙’과 ‘정착성 원칙’이다. 다양한 매체의 발달로 신조어의 조어 양상 또한 복잡하고 다양해지고 있는데, 이러한 신조어의 범람 속에서 하나의 규범적 단어로 인정받으려면 이 두 가지 원칙에 부합할 필요가 있다. 즉, 조어 형식과 의미 형식이 기존 언어의 특성을 그대로 유지하되 타당한 논리와 근거를 가져야 하며, 일정 시간동안 사용되어 언어 사회 내 정착되어야 한다는 것이다. 예를 들어 휴대폰을 의미하는 속칭으로 ‘大哥大’라는 단어가 있다. ‘大哥大’는 이동전화기 처음 등장하였던 몇 년 간 크게 유행하였던 단어이지만 지

금은 거의 사용되지 않고 ‘手机’라는 단어로 대체되었다. 그 이유로 통신 기계 자체의 변화도 들 수 있겠지만 언어학적 측면에서 볼 때 ‘大哥大’라는 단어의 유래⁹⁾ 및 조어 방식이 휴대폰이라는 물건의 특성을 제대로 반영하지 못하였기 때문이라고 여겨진다. 그에 반해 ‘手机’는 휴대폰의 여러 특성을 모두 나타냄과 동시에 기계를 나타내는 중국어의 일반적인 조어 형식을 그대로 따르기 때문에 언어 사용자들이 훨씬 더 쉽게 받아들였을 것이다. 그런데 이 신조어의 수록은 사전의 종류에 따라 다르게 적용될 수 있는데, 인터넷 사전이나 전자 사전의 경우 변화하는 언어 양상을 비교적 쉽게 적용시킬 수 있으나 종이 사전의 경우 여러 가지 출판상의 제약 사항으로 인해 그 변화들을 적절한 시간 내에 반영하기 힘들기 때문이다. 예를 들어 ‘博客(blog)’¹⁰⁾와 같은 신조어는 4종의 중한사전에 모두 수록되어 있지 않으며, ‘上座率(관객동원률)’은 3종의 중한사전에 수록되어 있지 않지만 인터넷 사전이나 전자 사전에는 이미 수록되어 있는 것과 같은 경우가 그러하다.

5) 전문용어

전문용어는 특정 분야에서 주로 활용되는 어휘들로 새로운 기술, 학문 분야가 출현하거나 발전함에 따라 전문분야 종사자들에 의해 인위적으로 만들어지는 어휘이다. 전문용어를 신조어에 포함시켜야 한다는 견해도 이

-
- 9) 周有光(1997)은 ‘大哥大’는 홍콩의 르느와르 영화 등에서 자주 출현한 조직 폭력배의 두목 혹은 무술계의 권위자들이 모두 大哥大라고 불리는데 그들의 손에 늘 이동전화기 일종의 상징처럼 들려있어 휴대폰이 ‘大哥大로 불리게 되었다고 한다.
- 10) 최근 들어 이 신조어에서 비교적 많은 비율을 차지하는 것이 외래어이다. 이 ‘博客(blog)’ 또한 외래어에 속한다고 할 수 있는데, 외래어는 신조어의 생성 경로 중 하나로서 새로운 문화, 문물의 도입과 함께 만들어지는 경우가 많으므로 넓은 의미에서 신조어에 속한다고 할 수 있다. 따라서 신조어가 표제어로 인정될 때 적용되는 원칙이 외래어에도 그대로 적용될 수 있으며 정착된 정도와 사용 기간 등은 편찬자의 기준과 그 기준을 토대로 한 고유한 지침을 세워야 할 것이다.

러한 이유에서 비롯된 것이다. 전문용어들은 특정 내용을 지칭해야 하므로 다의적이지 않으며 형태적으로 경제적이기도 하다. 한국어의 전문용어는 해당 분야의 원어를 그대로 차용하기도 하고 번역이나 조어(造語)등을 거쳐 사용하게 된다. 그러나 중국어는 음절 생성의 제한으로 인해 원어(原語)를 그대로 차용하기보다 주로 번역과 조어를 통해 만든다. 전문용어 자체가 전문적 개념을 나타내므로 일반인들은 이해하기 힘들며, 전문용어의 사용자가 한정되어 있다면 중국어 학습자라는 제한적 사용자를 가지는 중한사전에 전문용어를 표제어로 수록해야 하는 가는 고민할 필요가 있다. 그러나 전문용어도 널리 통용되어 쓰이면 일상 어휘로 취급될 수 있다. 박형익(2004;511)은 언어 사전에서는 전문용어를 표제어로 선정하지 않는 것이 원칙이긴 하되 일상 어휘로 사용되는 전문용어는 표제어가 될 수 있다고 보고 있다. 단일어 사전의 경우 대개 백과사전식으로 편찬되는 경우가 많아 전문용어에 대한 수용정도가 비교적 적극적일 수 있다. 그러나 이중어 사전의 경우 전문용어의 활용도는 전문성의 정도에 따라 차이가 날 것이다. 전문성이 강해 일반인들이 잘 사용하지 않는 전문용어들은 대체로 전문서적의 번역 등을 위해 활용될 것인데 이러한 사용자들의 편의까지 모두 고려해야 한다면 표제어 항목의 수는 엄청나게 늘어날 것이다. 일상에서도 자주 쓰이는 전문용어는 일상 어휘에 준하여 표제어로 수록하여도 전문성이 강한 용어는 배제할 필요가 있다. 이 경우 그 판단은 코퍼스에서 추출한 빈도정보에 근거할 수 있다. 앞서 2장에서 살펴보았듯 전문용어 중 코퍼스와 검색 사이트에서도 전혀 등장하지 않는 어휘들이 있었으니 이런 표제어들은 수록하지 않아도 좋을 것이다. 예를 들어 4종의 중한사전 중 3종의 사전에서 아래 예의 화학 전문용어가 표제어로 등장하고 있다.

- i) 大后方制度/大元手套/大石板瓦/大肚儿塔/白氏塑膠/寶砂/笨并非/鈣鏷磷肥/激妊酮/來瀝他唑/郎格來脫/木質醛醣/偶價元質

그런데 이들 용어가 코퍼스와 검색사이트에서 검색되지 않는 단어들이라고 해도 실제로 존재하지 않는 유행어인지, 다만 전문성이 강해 사용자가 많지 않아 검색되지 않는 용어인지 판단할 수 있어야 한다.

6) 성어

성어¹¹⁾도 중한사전의 표제어에서 등재 비율이 높은 범주이다. 표제자 ‘不’은 4종의 중한사전에서 1,000여개에 달하는 하위 표제어 중 200여 개가 성어에 속할 정도이다. 성어는 ‘정형성(定型性)’을 특징으로 하여 각 성분의 조합이 형식적 제약을 받기 때문에 성어에서 각 개별 성분은 그 독립적 지위를 상실하고 하나의 전체로 특정한 의미를 형성하게 된다. 이 경우 성어들은 한 글자 혹은 한 단어의 사전적 의미 결합으로 해석해선 안 되고, 관련 고사(故事)와 연결하여 혹은 글자의 파생 의미를 고려해야만 해석이 되는 경우가 많다. 또한 각 성어의 연어(collation) 또한 비교적 정형화되어 있기 때문에 자의적인 결합으로 다양한 의미를 만들어 내는 동사구, 형용사구, 명사구 등과는 구별된다. 사전에서 성어의 단위가 단어를 넘어 문장에 이른다고 할지라도 표제어가 되는 이유가 바로 여기에 있다. 그런데 이 성어들 중 일부는 완전히 고정되어 형식적으로도 의미적으로도 더 이상 파생되지 못하고, 다른 일부는 일정한 규율 내에서 일부 형태소나 단어를 전환하여 의미를 확대시키거나 변형시켜 활용되기도 한다. 그 과정 중 과잉 파생되는 성어들이 많으며 대체로 그들은 코퍼스를 통한 유효 빈도수가 아주 낮거나 아니면 검색되지 않는 경우가 많다.

i) 大荒大乱, 大明大摆

ii) 笨口拙舌/笨嘴笨舌 ----- 笨嘴挠腮

11) 이 때 성어(成語)는 4자 성어에만 한정하는 것이 아니라 속담, 혈후어, 관용어구 등과 같이 그 형식이 고정되어 다른 단어와의 조합이 일어나지 않는 구문에 대한 총칭으로 사용하기로 한다. 이러한 특성으로 인해 이들을 모두 중국어로 ‘固定词组’이라는 단어로 지칭하기도 한다.

상술한 예들이 바로 이에 해당하는 성어들로 i)의 예는 ‘大~大~’라는 고정 격식 속에 형태소나 단어들을 대체하여 만들어진 것이지만 유효 빈도수가 나타나지 않으며, ii)는 비슷한 의미를 가진 형태소들이 조합되어 동의어를 형성하지만 ‘笨嘴拙腮’은 나머지 두 성어와 달리 코퍼스에서 검색되지 않는다. 따라서 표제어로서 성어에 대한 조건은 빈도수 외 다른 조건, 즉 전체적인 단위가 특정한 의미를 형성해야 한다는 것이다.

7) 이체어

빈도가 낮거나 규범성을 벗어난 단어들 중 형태적으로 이체어에 속하는 단어들이 표제어로 수록되기도 한다. 이체어란 발음은 같지만 형태가 다른 한자어가 포함되는 것으로 중국어의 경우 오랜 시간 동안 가차, 통용 등으로 발음이 같은 한자들을 돌려쓰는 예가 많아 이러한 이형태 즉, 이체어가 많이 양산되어 있는 편이다. 이중어 사전의 표제어는 규범성을 가져야 하지만, 틀린 단어나 표현법이라고 해도 언어의 사용자들이 사용하는 어휘 범주에 속하여 유효 빈도수가 높은 경우 수록할 필요가 있다. 예를 들어 ‘溜黄鱼’는 ‘溜黄鱼’의 이체어로 규범적 형태가 아닌데, 이런 이체어는 표제어로 올리기 보다는 가표제어로 나타내어 그 내용 기술은 규범적 어휘 형태를 참고할 수 있도록 하는 예가 많다. 사전 편찬에 있어서 지면과 작업량 등의 경제성, 사전 이용자들에게는 검색의 용이함을 제공해주기 위해서인데, 이보다 더 중요하게 다루어야 할 문제는 어떤 단어형을 규범적 형태로 규정할 것인가 하는 문제이다. 이에 대해서는 중국의 문자 공작 위원회[国家语言文字工作委员会] 규정을 따라야 할 것이며, 규범적 형태에 따른 이체어의 선택은 우선 빈도수가 가장 높은 것, 사용 시간이 가장 오래된 것, 각 형태소의 의미가 단어의 의미에 기여하는 정도 등과 같은 기준을 마련해야 할 것이다.

3. 표제어의 단위 범주

표제어의 범위는 사전마다 다르겠지만 대부분의 언어 사전은 기본적으로 단어를 단위로 한다.¹²⁾ 중한사전의 경우 표제자와 표제어로 나누어 살필 수 있는데 다른 언어 사전과 달리 표제자가 있는 것은 하나의 한자가 하나의 의미를 나타내어 단어로 활용되기도 하고 동시에 다른 단어의 형태소로 작용하기도 하기 때문이다. 그런데 이 표제자는 글자마다 단어로, 혹은 형태소로, 혹은 단어와 형태소를 겸하여 활용되기 때문에 넓은 의미에서 다른 언어 사전의 표제어와 동일하게 취급할 수 있다.¹³⁾

이 ‘단어’가 사전의 기본 단위가 되는 것은 다른 말의 도움 없이 독립적으로 운용될 수 있는 최소 자립 단위이기 때문이다. 단어는 다시 단일어, 파생어, 합성어로 나눌 수 있는데 이들은 모두 표제어의 기본 단위가 될 수 있다. 그런데 파생어와 합성어에 대해서는 선정 원칙과 기준을 세워 표제어로 수록할 필요가 있다. 왜냐하면 파생어와 합성어는 형태소간의 결합, 혹은 형태소와 단어 간의 결합, 단어와 단어 간의 결합 등을 전제로 하므로 통사적 구와 혼동될 수 있기 때문이다. 중국어의 경우 단일어와 파생어, 합성어의 구분이 더욱 어려운데 이는 앞서 언급한 바와 같이 일음절 한자가 형태소와 단어로 동시에 역할 하는 경우가 많기 때문이다. 예를 들어 ‘草’는 단어로 쓰이면 명사로 ‘풀’을 의미하지만 ‘草案’, ‘草書’ 등의 단어 속에서는 ‘초고’ 혹은 ‘흘림체’를 나타내는 형태소가 된다. 따라서 ‘草案’, ‘草書’는 단일어가 되고, ‘풀’을 의미하는 단어가 다른 단어와 결합한 ‘花草’, ‘野草’는 합성어가 된다. 그런데 이 합성어는 ‘未伏天气’ ‘农业资本家’ 등과 같이 단어 간의 단순한 의미 조합으로 생성되어 합성성의 원리를

12) 유현경·남길임(2009)은 이를 ‘한 단어 한 표제어 원칙’이라고 한다.

13) 이는 주표제어와 부표제어의 개념과 어느 정도 상통한다고 할 수 있겠다. 주표제어는 사전에서 다른 표제어에 귀속지 않고 배열되는 말을 가리키며 부표제어는 사전 구조적 측면에서 주표제어에 부가되는 하위 표제어라고 정의할 수 있다.

그대로 지키는 경우 굳이 표제어로 올리지 않아도 되지만, ‘花草’, ‘野草’처럼 결합된 형태가 하나의 단어 못지않게 쓰인다면 단일어만큼 의미 있는 단위가 될 것이다.

이와 반대로 기존의 중한사전에서는 단어로 역활하지 못하면서 또한 형태소라고 할 수 없으며, 관용어구에 속하지도 않는 단어군들이 있다. ‘大牲’, ‘大喜功’, ‘大放’ 등이 그 예로 이들은 코퍼스에서 각각 ‘大牲畜’, ‘好大喜功’, ‘大放光彩’ 혹은 ‘大放异彩’ 등으로 확장된 형태로 검색된다. 이들 또한 잘못된 단위 형태의 표제어라고 할 수 있다.

사전의 표제어는 단어보다 큰 단위로 구나 문장으로 이루어지는 경우도 있다. ‘한 단어 한 표제어 원칙’을 따를 경우 품사적 지위를 획득하지 못한 단위의 표제어들은 기술이 어렵기 때문이다. 주로 명사구, 동사구가 해당하며 속담, 혈후어, 관용어 등을 표제어로 수록할 경우 그 문법 단위는 더욱 커지기도 한다. 이때의 구나 문장은 단순한 자의적 해석을 넘어 의미의 변화를 가지고 있는 구조 형태일 때가 많으며 이에 한정해서 표제어의 선정 여부를 판단하는 것이 좋다. 그렇지 않다면 표제어의 수가 불필요할 정도로 늘어나 경제적인 정보 제공에 부합하지 못하기 때문이다. 예를 들어 중국어의 동사가 보어와 결합된 동사구의 경우 새로운 의미를 만드는 경우가 있다.

i) 令人发指(머리털이 곤두서도록 화나게 하다)

令人喷饭(남의 실소를 자아내게 하다. 사람을 웃기다.)

이들 동사구는 글자의 조합을 통한 자의(字意)적 해석을 넘어 그 의미가 파생되어 다르게 활용되기 때문에 표제어가 될 수 있다. 그런데 같은 형식의 동사구라고 해도 새로운 의미를 만들어내지 못하는 아래의 예들은 표제어가 될 수 없다.

- ii) 令人失望(실망하게 하다) / 令人兴奋(흥분하게 하다)
- iii) 发糊涂(멍청해지다) / 分高低(높낮이를 가리다)
- iv) 犯不了(위반할 수 없다) / 给不了(주지 못하다)
- v) 大觉(크게 깨닫다) / 大毫(큰 붓)

즉, 이러한 동사구들은 합성어로 분석될 수 없으며, 그 전체가 하나의 특정한 의미를 만들어내지 못하므로 표제어로 적당하지 못하다.

IV. 중한사전의 표제어 선정 과정

상술한 표제어 선정 기준을 참고로 하여 실제 중한사전의 표제어를 선정하게 될 경우 그 과정을 살펴보자.

1. 중한사전 표제어 목록의 자료

중한사전 표제어의 목록은 크게 세 가지 자료—기존의 중한사전, 코퍼스, 기타 어휘 통계자료에서 추출할 수 있다. 기존의 사전은 일종의 선행 연구물로 표제어 선정과 관련해서 가장 많이 활용되는 자료이다. 대부분의 언어 사전들이 모두 이러한 기초 작업을 거치게 될 것이다. 기존의 사전을 자료로 활용할 경우 몇 가지 고려해야 할 사항이 있다. 첫째, 기존의 사전에서 제공되는 정보를 모두 수용할 것인가, 자체적인 점검 작업을 거쳐 선별할 것인가, 둘째, 기존의 사전을 새로운 사전을 위한 주 자료로 활용할 것인가, 보조 자료로 활용할 것인가 하는 문제이다. 기존 사전의 어휘 정보를 모두 수용한다는 것은 그 사전의 정보가 모두 옳다는 것을 전제로 해야 하기 때문에 오류나 부족한 내용까지 답습할 수 있고, 그래서 변화하는 언어 현실을 제대로 반영할 수 없다는 문제를 가진다. 자체적인 점검 작업을 거쳐 선별하는 경우에는 그 선별 기준을 정해야 하는데 가장 우려

되는 것은 이 때 편찬자의 직관에 의존하게 된다는 것이다. 이는 중한사전과 같은 이중어 사전일 때, 편찬자가 중국어 모국어 화자가 아닐 경우 더욱 유념해야 한다. 이러한 문제점을 극복하기 위한 하나의 방편으로 두 번째 어휘 자료 즉, 코퍼스를 활용할 수 있다. 중한사전의 이용자들은 표제어의 현실성에 대해 언어적 직관으로 판단할 수가 없다. 중한사전의 표제어는 객관적인 현실 언어를 수용할 수 있어야 하는데 이 때 ‘객관성’은 몇몇 중국어 모국어 화자에 의해서 확인될 수 있는 것이 아니므로 대규모의 자연언어 데이터베이스인 ‘코퍼스(corpus)’에서 추출한 언어 정보를 활용하는 것이 바람직하다. 표제어 선정과 관련하여 사전 편찬과정을 살펴보면 대체로 기존 사전에 등재된 표제어 목록을 작성하고 그 목록 가운데 사전 편찬자의 직관에 따라 어휘의 중요도를 판단하여 선정하는 경우가 많다. 그러나 이 경우 그 중요도에 대한 판단 기준을 제시하지 못하는 경우가 많고, 더구나 중한사전의 편찬이 대부분 국내에서 국내 중국 언어학자들을 위주로 이루어진다는 점을 고려할 때 그 직관과 통찰력의 한계를 인정하지 않을 수 없으며, 이는 중국어 모국어 화자와의 공동 작업이 전제된다고 해도 마찬가지다. 따라서 자연 언어 자료인 코퍼스에 근거하여 용례 및 활용과 같은 유의미한 정보를 추출할 수 있어야 하고 특히 어휘의 계량적 분석을 진행해야 한다. 앞서 표제어의 내용 범주에서, 비표준어라고 할지라도 사용 빈도수가 높으면 표제어가 될 수 있음을 살펴보았는데, 결국 이중어사전은 해당 언어의 사용자들이 그 어휘들을 어떤 의미로 어떻게 사용하는지를 보여주는 것이므로 표제어 선정에 가장 큰 기준은 객관적 유효 빈도수가 될 것이며, 이렇게 사용 빈도에 근거하여 선정된 표제어는 중국어 학습자에게 더 유용한 정보가 될 것이다. 코퍼스를 통해 얻을 수 있는 정보는 해당 어휘의 빈도 정보에만 국한되지 않는다. 다른 어휘와의 공기 관계를 나타내는 언어(collocation)정보, 구의 확장 패턴과 관련된 정보, 일반적인 결합 형태에 부합하지 않는 제약 사항 등 코퍼스를 통해 얻을 수 있는 정보는 다양하다. 이러한 내용들은 표제어 선정 뿐 아니라 표제어의 뜻풀이, 용례, 문법 환경 등의 미시 구조와도 관련한 여러 정보를 제공

할 수 있다.

2. 중한사전 표제어의 규모

이 때 표제어의 어휘 규모에 대한 논의가 전제되어야 한다. 앞서 살펴본 4종은 많게는 18만에서, 적게는 10만 여 개에 이르는 표제어를 수록하고 있다. 이중어 사전의 어휘수는 편찬 목적에 따라, 편찬하고자 하는 사전의 규모에 따라 다르겠지만 대사전이라는 전제아래 표제어의 수는 일반 모국어 화자의 어휘량을 근거로 해야 할 것이다. 김광해(2003)는 일반 모국어 화자의 어휘량을 대략 4만~5만으로 보고 국어 교육량으로 필요한 어휘량을 약 6만 단어 선으로 잡았다. 이와 중국에서 출판된 한어사전의 표제어 수를 비교해보면 대표적인 한어사전 중 《现代汉语规范词典》의 경우 68,000여 단어를, 《现代汉语词典(第5版)》의 경우 65,000단어를 수록하고 있으므로 김광해(2003)가 제시한 모국어 화자의 어휘량에 상당함을 알 수 있다. 그런데 이 2종의 사전은 모두 현대한어를 기준으로 하고 있는 점을 주의해야한다. 중한사전은 현대한어만을 기준으로 할 수 없으며 국내 중한사전 이용자의 필요와 이용 실태에 따라, 또한 사전 사용자가 대부분 중국어 학습자임을 고려하여 교육이라는 측면을 고려한다면 문어어, 방언, 지명, 외래어 등을 비롯한 고유명사 등도 표제항에 오를 것이며 이를 고려하여 표제어의 규모를 설정하도록 한다.

3. 표제어 선정 과정

(표제자를 포함하여) 표제어의 선정을 위한 어휘 구축작업은 아래의 세 가지 기본 작업을 전제로 한다.

- i) 기존사전의 표제어 검토
- ii) 한어 사전과의 대조 및 코퍼스 빈도수 조사

iii) 신조어를 비롯한 추가 단어 조사

1) 1단계: 출간된 사전의 표제어에 대한 분석을 진행하고 그 목록을 정리한다. 사전 편찬 과정 중 가장 먼저 진행되어야 하는 기초 작업으로 이를 통해 1차 표제어 목록을 작성한다.

2) 2단계: 1차 표제어 목록에는 있으나 코퍼스에서는 추출되지 않는 단어들을 점검하여 코퍼스에 출현하지 않는 이유를 분석하고, 유령어로 분류될 수 있는 단어들을 폐기한다. 동시에 단어 이상의 표제어들을 검토하여 의미적 파생을 일으키지 않는 통사적 구나 문장 범주에 속하는 표제어들을 삭제하여 2차 표제어 목록을 작성한다. 제2장에서 중한사전 4종의 표제자 ‘大’의 하위 표제어를 점검하여 코퍼스와 대조한 것이 바로 여기에 해당한다고 하겠다.

3) 3단계: 중국에서 출간된 한어 사전의 표제어 추출 및 코퍼스를 통한 빈도수 조사에 기반하여 산출해낸 표제어 목록을 2차 표제어 목록과 대조하여 3차 표제어 목록을 선정한다. 기존 사전들의 표제어를 원시 표제어로 검토하는 이 작업은 그 사전들의 표제어가 타당하고 근거가 있다는 것을 전제로 하기 때문에 선정 기준이나 표제어 목록에 있어 일종의 순환론에 빠질 위험이 있다. 이를 방지하기 위해 3단계에서는 두 가지 대조 조사 작업을 통해 3차 목록을 작성할 것이다. 우선 중국의 대표적인 한어 사전을 선정하여 2차 표제어 목록과의 대조를 통해 2차 목록에는 있으나 한어 사전에는 없는 표제어를 모아 가표제어 목록 1을 만들고, 그리고 2차 목록에는 없으나 한어 사전에는 등재되어 있는 표제어를 모아 가표제어 목록 2를 각각 산출한다. 한어 사전의 표제어를 기준으로 가표제어 목록을 만드는 것은 코퍼스 어휘 목록 추출에 관련된 대규모 작업의 한계와 코퍼스 자체의 불완전함을 보완하기 위해서이다. 가표제어 목록 1과 2에 대해 우선 내용 범주와 단위 범주에 따라 그 등재의 타당성을 검토하고 다시 코퍼스를 통해 어휘의 절대 빈도를 산출해내어야 한다. 이 과정은 원시 표제어 목록이 실제 언어 현실을 얼마나 반영하는지 점검하는 동시에, 코퍼스

출현 빈도에 따라 중요도를 판단하는 과정이 되기도 한다. 이와 같은 판단 과정을 거쳐 추출된 표제어는 1차 표제어에 첨가하여 3차 표제어 목록을 완성한다.

4) 4단계: 중국에서 출간된 신조어 사전의 표제어를 추출하여 가표제어 목록 3을 작성한다. 가표제어 목록 3의 표제어를 유효 빈도수에 근거하여 정착된 정도를 측정하여 앞서 언급한 ‘정착성 원칙’에 부합하는 지 점검한다. 또한 각 표제어를 내용 범주와 단위 범주 측면에서 등재 자격이 되는 지를 분석하여 ‘논거성 원칙’에 부합하는 지 점검한다. 이로 추출된 표제어를 3차 목록에 추가하여 4차 표제어 목록을 완성한다.

V. 중국어 코퍼스 활용의 문제점

이와 같이 사전의 표제어 선정에는 코퍼스와 같은 자연언어 데이터 베이스가 큰 몫을 한다. 컴퓨터 기술의 발달과 보조를 같이 하는 코퍼스의 발전은 사전 편찬에 있어 필수적인 시스템이 되어 가고 있음은 두말할 필요가 없을 것이다. 그런데 코퍼스를 사전 편찬에 응용할 경우, 특히 중국어코퍼스를 중한사전 편찬에 응용할 경우 유의할 점이 있다.

첫 번째, 중국어 코퍼스에 대한 신뢰도이다. 최정도(2010)는 사전편찬에 이용되는 코퍼스는 ‘균형성’과 ‘대표성’을 가지고 있어야 한다고 주장한다. 표제어 선정의 기본적 원칙이 고빈도라고 주장하는 만큼 표제어의 유표 빈도수가 아주 중요한 점으로 작용하기는 하지만 저빈도 어휘라고 해서 덜 중요하다거나 표제어로서의 가치가 없다고 단정하지는 못한다. 이는 텍스트와 장르 등의 균형성에서 요구되는 바로, 모든 이가 매일같이 사용하는 어휘만 표제어가 될 수는 없기 때문이다. 또한 코퍼스의 자료가 연구자에 의해 선택된 자료라는 점은 코퍼스의 대표성을 상실할 가능성을 제공한다. 코퍼스가 대표성을 가져야 하는 것은 코퍼스가 균형성을 가져야 하는 것과 엄밀한 의미에서 같은 맥락이다. 즉 균형성이 담보되지 않으면

대표성 또한 사라진다. 장르에 대한 비율과 규모, 텍스트에 대한 객관적 평가 등과 관련된 여러 구체적인 세부 지침을 근간으로 하는 코퍼스가 구축되어야 한다. 이러한 코퍼스에 대한 사용자의 신뢰도는 자료의 내원을 밝히고 그 종류를 세분화하며, 빈도 목록과 횡수에 대한 통계 수치를 제시함으로써 담보될 수 있을 것이다. 특히 중국어는 문언어와 구어의 사용역이 비교적 뚜렷하므로 자료 내원의 공개를 통해 해당 어휘의 사용역을 밝히는 것이 효율적이라고 여겨진다. 또한 코퍼스의 텍스트와 장르가 문헌에 치우칠 경우 구어의 어휘 산출에 소홀할 수 있으므로 코퍼스가 현실 언어를 보다 객관적으로 반영할 수 있도록, 그리고 보다 합리적인 정보를 제공할 수 있도록 지속적인 연구와 동보(同歩)적인 데이터의 구축이 아주 강력히 요구되는 바이다.

두 번째, 중국어 코퍼스의 경우 통계처리 시스템이 미약하다는 것이다.

박영중(2008)의 연구 조사에 따르면 중국어의 코퍼스는 17개에 이르고 있으며 고대중국어에서부터 현대중국어에 이르기까지, 수록하고 있는 글자 수 또한 어휘 구축의 목적에 따라 몇 십 만에서 몇 억 자에 이르기까지 그 종류가 비교적 다양하다고 할 수 있다.¹⁴⁾ 그러나 이들 코퍼스의 시스템은 검색어가 포함된 용례를 추출하는데 한정되어 있을 뿐, 검색 결과에 따른 여러 통계와 가공 시스템이 구축되어 있지 않다는 문제를 안고 있다.¹⁵⁾ 가장 큰 문제는 형태소와 의미 분석에 대한 결과가 뒷받침되어 있

14) 박영중(2008)에서 중국어 코퍼스에 대한 자세한 자료를 참고할 수 있다.

15) 이러한 문제점은 한국어의 세종 말뭉치 프로그램과 비교하면 더욱 확연하게 알 수 있다. 세종 말뭉치는 1998년부터 2007년까지 10년간 정부 지원으로 수행된 언어 자원 구축 프로젝트이다. 그러나 데이터 베이스에 대한 이해가 높지 않은 사용자들을 위해 활용 시스템을 다시 구축하였는데 그것이 꼬꼬마 세종 말뭉치이다. 이 꼬꼬마 세종 말뭉치 활용 시스템은 많은 사용자가 쉽게 말뭉치를 활용할 수 있도록 웹 응용으로 구현되어 있다.(<http://kkma.snu.ac.kr>) 꼬꼬마 시스템은 여러 가지 형태의 처리 모듈을 구현하여 적재하였고 이렇게 정형화된 폼에 조건을 입력하도록 하여 다양한 기능을 지원하는 조회 프로그램을 만들었다. 이를 통해 세종 말뭉치의 다양한 통계 정보를 표 형태로 조회할 수 있게 되었으며 용례, 굴절형, 언어 정보 등을 조회할 수 있게 되어있다. 그 외 구문 분석의 결과나 연어에 대한 시각화 기능 등도 포함되어 있어 보다

지 않다는 것이다. 하나의 단어가 하나 이상의 의미를 가질 경우 의미마다 품사, 용례, 연어(collation)에 대한 정보가 달라질 수밖에 없다. 만일 사용 빈도수가 높거나 의미가 다양한 단어라면 검색되는 용례 또한 비례할 것이므로 검색어의 여러 의미에 따른 용례를 다시 일일이 판단하고 구분해야 하는 작업을 진행해야 한다. 특히 중국어는 ‘打’ ‘火’ ‘收’ ‘外’처럼 일음절 형태소 겸 단어 중 대다수가 평균 10여개 이상의 의미항목을 가지고 있기 때문에 각 항목마다 각기 다른 정보를 가지고 있을 것이고, 그 정보들이 따로 구분되지 않고 일괄 제시된다면 각 용례별 점점이 뒤따라야 하기 때문에 이러한 코퍼스의 활용은 원시적일 수밖에 없다.

그러므로 자료의 출처와 각 자료별 추출 시스템, 품사나 용례 등의 여러 기준별 추출 및 분석 시스템의 개발과 활용이 뒤따라야 한다.

VI. 나가기

본고는 국내 중한사전들이 대사전 편찬 지향으로 인해 많은 수의 표제어를 수록하려고 하면서도 표제어 목록의 선정 방법과 그 과정에 대해 명확하게 설명하지 않음과 동시에 내용 범주와 단위 범주상의 오류를 가지거나 유령어로 분류될 수 있는 표제어까지 수록하고 있다는 문제에서 출발하였다. 이러한 경향을 가지게 된 근본적인 이유 중 하나로 이중어 사전으로서의 중한사전이 가져야 할 표제어 목록의 각종 범주와 그 선정 방법 및 과정에 대한 근본적인 연구의 부재라고 여겨진다. 이에 본고는 국내의 기존 중한사전에서 보이는 표제어 수록의 문제점을 짚어보고 중한사전의 표제어 선정과 구성에 대한 기준과 원칙을 점검하고 그 구체적인 선정 방법을 모색하고자 하였다. 다만 본고에서는 원칙과 기준 및 선정 방법에 관해 비교적 원론적인 논의를 진행하였으며 보다 구체적인 실천 방안과 중국어 코퍼스의 제반 문제 및 활용에 관해서는 다음 연구에서 논의하고자 한다.

쉽게 확인할 수 있도록 되어 있다.

<參考文獻>

1. 사전류

고대민족문화 연구원 중국어대사전 편찬실 편저, 《中韓辭典》, 고려대학교 민족문화 연구원, 2008.

손예철 편저, 《Prime 東亞 中韓辭典》, 두산 동아, 2008.

이용목 편저, 《民衆 엡센스 中國語辭典》, 民衆書林, 2010.

박영중 편저, 《現代中韓辭典》, 교학사, 2009.

2. 국내외 논저

강현화, <두말 사전 편찬(Bilingual Dictionaries): 한국어 학습용 이중어사전 역할에 대한 논의>, 2001.

강현화 외, <한국어 학습자 사전 표제어 선정을 위한 자료 구축 및 선정 방법에 관한 연구>, 《한국사전학》 16, 2010.

김광해, <국어 교육용 어휘와 한국어 교육용 어휘>, 《국어교육》 11, 2003.

도원영, <‘민연 국어 사전’(가칭)의 표제어 선정과 그 실제에 대하여>, 《한국어학》 12:1, 2000.

박동근, <국어사전의 표제어 늘리기와 ‘유령어(ghost word)’ 처리의 문제>, 《한국사전학회 학술대회 발표논문집》, 2, 2011.

박영중, <한어코퍼스를 활용한 중한사전의 개선>, 《중국어문학논집》 53, 2008.

박형익, 《한국어 사전과 사전학》, 2004.

유현경·남길임, 《한국어사전 편찬학 개론: 사전 편찬의 이론과 실제》, 2009.

이상도, <중국어사전 표제자 수록문제에 대하여>, 《중국언어연구》, 6, 1998.

정희원, <표제어 선정의 제 문제>, 《새국어생활》 5:1, 1995.

최정도, <말뭉치를 이용한 사전 편찬에서의 몇 문제에 대하여>, 《한국사

전학회 학술대회 발표논문집》 2, 2010.

홍중성 외, 《국어사전학 개론》, 2009.

周有光, 《語文閑談》(續編), 上海: 三聯書店, 1997.

章宣華·雍和明, 《當代詞典學》, 商務印書館, 2007.

Sidney I. Landau, 김영안 외 역, 《사전편찬론》, 2002.

<中文提要>

随着国内中韩辞典的大规模化发展趋势, 编纂辞典也倾向于词条数目的大量化。但对选定词条的方法及其过程, 大部分的辞典解释得不太明确, 甚至在内容范畴和单位范畴上错误的词条和死词也收入其中。本稿从这个实际问题出发, 考察国内中韩辞典的词条上存在的一些问题, 树立中韩辞典词条的选定标准和原则, 并摸索其具体的实践方案。收选词汇条目的首要工作是检讨现存辞典的词目, 其次与中国出版的汉语词典对照, 并通过权威性的语料库调查词语使用频率数, 最后调查并检定新词。选定词汇条目的整个过程在遵循这些基本原则下进行。

주제어 : 중한사전, 표제어, 코퍼스활용, 유효빈도, 유령어