

## 《우리말샘》 편찬 경과

이현주

국립국어원 학예연구관

### 1. 들어가기

국립국어원에서는 ‘개방형 한국어 지식 대사전 구축’ 사업을 2009년부터 준비하여 2010년 본격적으로 시작하였고, 6년 뒤인 2016년 10월 5일 그 결과물로 세 종류의 사전을 개통하였다. 이 중에서 《우리말샘》(2016, 이하 《말샘》)은 ‘개방형 한국어 지식 대사전 구축’ 사업의 대표 사전으로, 사업 이름을 통해서도 알 수 있듯이 일반 국민이 사전 편찬에 참여할 수 있도록 만들어진 개방형 국어사전이다.

1999년에 완간된 《표준국어대사전》(1999, 이하 《표준》)은 “어문 규정을 구체적으로 단어 하나하나에 적용하여 국민들의 올바른 언어생활에 표준을 제공하”(조남호, 2000)는 것에 역점을 둔 사전으로, 편찬 당시 어문 규정을 정비한 지 10여 년이 지난 시점에서 세부 적용에서 혼란을 겪고 있는 어문 규범의 해설서 역할을 충실히 했고, 지금도 언어생활의 기준을 제공하는 규범 사전으로서의 역할을 해 오고 있다.

반면, 《말샘》의 목표는 우리말이 우리 문화와 지식을 담은 정보와 지식의 집합체라는 점을 전제로 “지금까지 우리가 사용해 왔던 우리말 정보를 모두 집대성하여 이를 국가 차원에서 관리하되 이를 일반 언중들이 들어

와 다양하게 우리말 정보를 깊고 보탬 수 있는 국가 언어 지식 관리 체계”(이 승재, 2012)의 역할을 하는 사전<sup>1)</sup>이다. 이를 위해 일반에 공개하기 전까지 일반 언중들이 들어와 다양하게 우리말 정보를 깊고 보탬 수 있도록 ‘개방형’ 편찬 체계를 갖추었으며, 실생활에서 쓰이지만 기존 사전에 수록되지 않은 실생활어들을 대폭 추가하여 100만 항목을 마중물로 부어 넣었다. 그리고 미시 구조에서 기존 사전에는 없는 다양한 정보 항목들을 마련하였다.

이 글에서는 《말샘》 편찬의 경과를 소개하고자 한다. 《말샘》은 편찬이 끝나지 않는 사전이다. 개통이 되고 두 달이 넘은 지금도 누군가에 의해 《말샘》은 계속해서 더해지고 수정되고 있다. 그러나 여기서는 《말샘》을 일반에 공개하기 전까지의 경과만 기술하기로 한다. 먼저 단계별 사업 추진 경과를 개괄한 후 자료 구축, 미시 정보 보완, 교정·교열, 사전 시스템 구축의 내용과 경과를 기술하고 앞으로의 과제를 살펴볼 것이다.

## 2. 단계별 사업 추진 경과

‘개방형 한국어 지식 대사전 편찬’ 사업은 2009년 국가경쟁력강화위원회에서 세종사업 추진 방안의 하나로 국민들의 언어생활을 반영하고 수요자 기반의 ‘쉽고 살아 있는 사전 편찬’ 추진을 보고(2009년 6월 24일)하면서 시작되었다. 이후 사전 편찬 계획과 방향을 제시하고 진행되는 과정에서의 자문 기구로 사전편찬추진위원회가 구성되었는데, 사전편찬추진위원회는 문화체육관광부 제1차관과 민간 대표가 공동으로 위원장을 맡고 사전 편찬

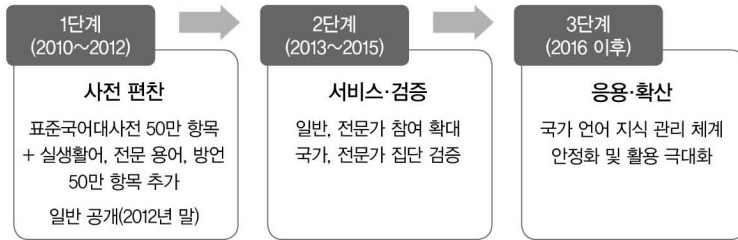
1) 다른 사전과 구별되는 《말샘》의 특징은 다음과 같이 정리할 수 있다.

실용성	일상어, 지역어, 전문어 등 실생활 언어 정보를 제공할 수 있는 사전
개방성	일반 국민이 함께 만들고, 누구나 자유롭게 이용할 수 있는 사전
기술성	규범 제시 기능에서 나아가 다양한 언어 정보와 지식, 문화를 담은 사전
현실성	종이 사전의 한계를 벗어나 사회, 언어 변화를 실시간 반영할 수 있는 웹 사전

일반 분야 5명, 전자 학습 사전 분야 4명, 정보·기술 분야 3명 등 총 12명의 분야별 전문 위원을 두었다. 사전편찬추진위원회는 《말샘》<sup>2)</sup> 개통 전까지 총 6차례 회의를 개최하여 단계별, 연차별 사업 계획을 검토하고 추진 내용에 대한 의견을 제시하였다. 2009년 두 차례의 사전편찬추진위원회 회의를 거쳐 ‘개방형 한국어 지식 대사전 구축’ 계획이 확정되고 2010년 1월 기본 계획이 수립되면서 본격적으로 편찬 사업이 시작되었다.

《말샘》 편찬 단계는 기본 계획이 수립될 당시에는 다음과 같이 세 단계로 설정하였다.

**그림 1** 변경 전 단계별 추진 절차



2010년부터 2012년까지 1단계 사업을 추진하고 2012년 말에 일반에 공개할 예정이었으나 2013년 10월 9일 정식 개통으로 일반 공개 일정이 늦춰졌다. 이에 따라 2012년 10월 9일부터 11월 30일까지 두 달간 일반인 편집 참여 절차상의 문제점과 사전 검색과 사용상의 개선점을 발굴하여 보완하기 위해 150명의 체험단을 모집하여 시범 운영을 실시하였다.

2) 사전 명칭이 《우리말샘》으로 정해진 과정은 다음과 같다. 제3차 사전편찬추진위원회(2010년 6월 1일)에서 ‘개방형 한국어 지식 대사전’이라는 목표가 포함되어 있는 사업 명칭이 아닌 친근하면서도 간결한 사전 명칭을 정할 필요가 있다는 의견이 제시되었다. 이에 따라 2010년 9월에 일반인 대상으로 명칭 공모를 하였고, 그 결과 ‘우리말누리샘’, ‘온누리말모이’, ‘열린말누리’, ‘생생한국어사전’이 당선작으로 결정되었다. 그 후 2012년 내부 검토를 거쳐 대상인 ‘우리말누리샘’을 네 글자로 간결하게 줄인 《우리말샘》으로 확정되었다.

2013년부터 일반 공개 일정을 포함한 2단계 사업에 착수하였으나 국가 주도로 구축한 사전으로서의 질적 수준을 확보하기 위해 1단계에서 구축한 자료들을 전체적으로 검토하고 정비할 필요가 있으며, 세계적으로 처음 시도되는 국민 참여형 사전의 운영을 위한 운영 체계 안정화 기간이 필요하다는 점이 제기되었다. 이에 따라 2013년 6월 3일부터 7월 5일까지 한 달간 1단계에서 구축한 결과물에 대해 분야별 전문가에 의한 진단 평가를 실시하여 문제점을 분석하고 개선 방향을 수립하였다. 진단 평가에서는 《말샘》 1단계 구축 자료에 대해 표제어 선정의 균형성 및 일관성 보안을 위한 검토가 필요하다는 점과 사전 뜻풀이의 오류 및 정보 제공의 비균질성을 개선할 필요가 있다는 점, 전문 용어 범주의 적절성을 검토할 필요가 있다는 점 등이 지적되었으며 이에 대한 단기 개선 방안으로 집필 오류 유형 분석을 통한 구축 지침 보완 및 지침 상세화, 뜻풀이 오류 및 정보 제공 비균질성 해소를 위한 전문가 전수 검토, 전문 용어 분야 간 균형성 확보 등이 제시되었다. 또 개통 이후 중·장기적 개선 방안으로는 표제어 일관성 확보, 일반 사용자의 편의성 및 활성화를 위한 기능 및 정보 제공 확대, 규범 정보 제공 사전과 《말샘》의 상호 연계, 영역별 전문가 등의 참여를 통한 지속적인 기술 내용의 엄밀성 확보, 다른 사전 및 관련 콘텐츠와의 연계 체계 구축을 통한 서비스 수준 향상, 《말샘》 내용의 직접 관리 체계 확립을 위한 국립국어원 편찬 인력 증원과 국립국어원 내 사전 구축 체계 강화 및 외부 교정·교열 체계 정비를 제시하였다.

이러한 진단 평가 결과를 바탕으로 ‘개방형 한국어 지식 대사전’ 사업의 개선 추진 계획을 수립하고 2013년 10월 공개하기로 했던 계획을 다시 변경하여 1단계 결과물 교정·교열 및 정비를 진행한 후 2016년에 공개하기로 하였다. 이에 따라 단계별 추진 일정은 다음과 같이 조정되었다.

**표 1** 변경 후 단계별 추진 절차

단계	구분	추진 내용
1단계 (‘10~‘12)	사전 편찬	<ul style="list-style-type: none"> <li>• 한국어 자료 집대성(100만 항목 사전 구축)</li> <li>• 위키피디아형 사전 편찬 및 관리 시스템 개발</li> <li>• 언어 정보의 유통·검색 지원 시스템 구축</li> </ul>
2단계 (‘13~‘16)	정비 및 안정화	<ul style="list-style-type: none"> <li>• 구축 자료의 교정·교열 및 정비, 감수</li> <li>• 사전 정보의 다각적 보완</li> <li>• 일반인 개방에 대비한 사전 정보 관리 운영 체계 확립</li> </ul>
3단계 (‘16 개통 이후)	운영 및 확산	<ul style="list-style-type: none"> <li>• 사전 개통·운영, 일반인 참여 활성화</li> <li>• 일반인 구축 자료의 전문가 감수</li> <li>• 한국어 지식 정보의 민간 활용 지원</li> </ul>

수정된 계획에 따라 2013년 후반부터 2016년 10월 개통 전까지 1단계 구축한 자료에 대해 외부 전문 감수단을 활용한 교정·교열 및 감수를 진행하고 내부 인력을 활용하여 사전 자료 간의 일관성 검토 등의 정비를 하였으며, 사용자 편의성 제고를 위한 시스템 점검 및 개선을 진행하여 2016년 10월 5일 개통 행사를 기점으로 일반에 공개되었다.

### 3. 사전 자료 구축

《말샘》은 수록된 어휘의 규모 면에서 지금까지 편찬된 국어사전 중에서 가장 크며, 일반인이 집필에 참여하면서 앞으로 그 수는 더 늘어날 것이다. 이렇게 많은 표제어를 수록하는 것은 《말샘》의 목적이 예전에 사용되었거나 현재 사용하고 있는 ‘모든 한국어 정보를 집대성’하는 것이기 때문이다. 《말샘》은 일반에 공개하여 일반인이 집필에 참여하기 전까지 100만여 항목을 구축하는 것이 목표였다. 이에 따라 《표준》의 50만여 항목을 가져 오고, 2010년부터 2013년까지 신어와 생활용어 7만 5천여 항목, 전문 용어 41만여 항목, 방언 9만여 항목을 새로 구축하였다. 연도별 구축 내용을 표로 정리하면 다음과 같다.

**표 2 연도별 신규 구축 자료**

분과	2010년		2011년	2012년	2013년	합계	
신어	1만		2만	1만	-	4만	
생활 용어	0.5만		1만	1.5만	0.5	3.5만	
방언	2만		4만	3만	-	9만	
전문 용어	신규 구축	통신·전기·컴퓨터	5만	4만	-	9만	
		의학·약학·생물	4만	3만	-	7만	
		경제	2.9만	-	-	2.9만	
		농업	-	2.6만	-	2.6만	
		수학·물리·화학	-	2만	-	2만	
		환경	-	1만	1만	2만	
		법률·사회·정치	-	1.7만	-	1.7만	
		패션	-	1만	7천	1.7만	
		역사	-	1만	5천	1.5만	
		운동·언어·문학	-	-	1.5만	1.5만	
		예술	-	1.4만	-	1.4만	
		전자 음악·음향	-	1.4만	-	1.4만	
		건축·건설	-	-	1.3만	1.3만	
	선박·항공·우주 통신	-	-	1만	1만		
			11.9만	19.1만	6만	37만	
	외부 기관 이관 자료 <sup>3)</sup>	국방 과학 기술	-	-	8,498	-	8,498
		산림·임업	-	-	4,510	-	4,510
		문화재	-	-	281	-	281
		행정 표준	-	-	2,262	-	2,262
		-	-	1.56만	-	1.56만	

집필진은 매년 분과별로 공모를 통해 선정하였다. 집필진은 표제어 선정 기준을 정하고 집필 지침을 마련하여 국립국어원과 협의한 후 집필을 진행하였다. 2010년에는 편찬 시스템의 개발이 완료되지 않아 집필 원고를 파일로 작성하였는데, 2011년 편찬 시스템이 개방된 후부터 편찬 시스템에 완성

3) 외부 기관 이관 자료는 외부 기관에서 사전의 형식으로 구축되어 있는 자료를 제공받아 《말샘》의 지침에 맞게 교열 및 정비한 자료이다. 국방 과학 기술 용어는 '국방기술품질원', 산림·임업 용어는 '산림청', 문화재 용어는 '문화재청', 행정 표준 용어는 '행정안전부'에서 제공받았다.

된 원고를 등록하고 직접 입력하는 방식으로 집필하였다.

신어 분과에서는 2010년부터 2012년까지 3년 동안 총 4만 항목을 구축하는 것을 목표로 하여 2010년에 1만, 2011년에 2만, 2012년에 1만 항목을 구축하였다. 신어 분과에서는 집필 대상 신어를 ‘일상생활에서 자주 사용되지만 《표준》에 등재되지 않은 단어나 단순한 단어 의미의 합 이상의 의미를 갖는 구’로 정하고, 현실 언어에서 활발하게 쓰이는 말, 국어로 정착한 말, 현실 언어에서 자주 나타나는 비표준어, 쓰임이 활발한 순화어, 《표준》 등재 어휘의 계열어들 중 빈자리를 보이는 어휘를 선정하여 집필하였다.

생활 용어 분과에서는 2010년부터 2013년까지 4년 동안 총 3만 5천 항목을 구축하였다. 생활 용어를 ‘전문 영역에 특화되어 쓰이지 않고 일상생활에서 두루 쓰이는 말, 또는 전통적으로 전해져 내려오는 생활어 가운데 보존의 가치가 있는 말’로 정의하고, 현대 생활 어휘로 《표준》 미등재어 가운데 말뭉치에서 일정 빈도 이상 나타나는 어휘, 생산성이 높은 접사에 의한 파생어, 자주 쓰이는 구 단위 이상의 표제어와 실사용에서는 저빈도어로 나타나지만 민족 생활어나 문학 작품 등에서 살려 쓸 만한 고유어를 선정하여 집필하였다.

방언 분과에서는 2010년부터 2012년까지 3년 동안 총 9만여 항목을 구축하였다. 국립국어원의 21세기 세종계획 <한민족 언어정보화> 한국 방언 검색 시스템에 등재된 방언 자료와 ‘국립국어원(2005~2011), <지역어 조사 보고서>’, ‘전광현(1978/2003), <동해안 방언의 어휘>’, ‘박성중·전혜숙(2009), 《강릉방언사전》, 태학사’, ‘이희순(2004), 《방언사전: 여수편》, 어드북스’ 등의 방언사전 및 방언 자료집 등의 방언 자료를 조사 정리하여 《표준》에는 등재되지 않았지만 일반 국민들이 일상생활에서 접하는 방언 표제어를 선정하여 집필하였다.

《말샘》에서 새로 구축한 자료 중에 가장 큰 비중을 차지하는 것이 전문 용어이다. 2010년에 전문 용어 구축을 위해 ‘학술단체총연합회’의 학술

전문 용어를 비롯하여 기존에 구축된 여러 가지 전문 용어 자료를 통합하여 정비하고 전문 용어 집필 지침을 작성하여 2011년부터 2013년까지 3년 동안 37만여 항목을 집필하였다. 표제어로는 국립국어원에서 제공한 2010년도 통합 정비 자료에 속한 전문 용어를 우선적으로 선정하는 것을 원칙으로 하고, 해당 분야의 기초 개념을 나타내는 전문 용어, 일반어이지만 전문적인 쓰임을 보이는 전문 용어<sup>4)</sup>, 일상생활에서 널리 쓰여 일반인들이 흔히 접하는 전문 용어, 여러 영역에서 널리 쓰여 다의어 양상을 보이는 전문 용어를 선정하여 집필하도록 하였다. 《표준》은 전문 용어를 집필할 때 국어국문학 전공자가 집필하고 해당 영역 전공자가 감수한 반면, 《말샘》에서는 해당 영역 전문가가 집필하고 국어학 및 사전학 전문가가 사전 형식에 맞게 교열 및 교정을 하여 일반 사용자가 이해하기 쉽도록 보완하였으며 이렇게 교열한 원고에 대해서는 해당 영역 전문가가 최종적으로 확인하도록 하였다.

#### 4. 미시 정보 보완

《말샘》은 《표준》에 더하여 신어, 전문 용어, 방언 등 실생활에서 쓰이고 있는 어휘들을 대폭 등재하여 표제어의 범위가 크게 늘어났을 뿐만 아니라 미시 정보도 다각적으로 보완하여 기존 사전에서는 제시되지 않은 다양하고 유용한 정보들이 제시되었다.

《말샘》은 ‘쉽고 살아 있는 사전’을 표방하였기 때문에 사전 정보의 보완 중 가장 먼저 시작한 것이 어렵고 비체계적인 뜻풀이를 수정하여 쉽게 하고 실생활 용례를 보완하는 일이었다. ‘쉬운 풀이’와 ‘용례 보완’ 작업은

4) 일반어이지만 전문적인 쓰임을 보이는 예로 ‘재킷(jacket)’을 들 수 있다.

재킷(jacket) 『001』 앞이 터지고 소매가 달린 짧은 상의. 보통 털실 따위의 모직물로 만든다.  
재킷(jacket) 『003』 『기계』 보일러, 파이프 따위를 싸서 열의 방산(放散)이나 기관이 지나치게 뜨거워지는 것을 방지하는 피복물(被覆物).



2010년부터 2012년까지 진행되었다. 《표준》에서 이관한 자료 중 풀이 없이 올라 있는 82,177항목의 동의어를 《말샘》에서는 직접 풀이하였고, 뜻풀이에 쓰인 어려운 용어 3,228 항목을 쉬운 용어로 수정하였으며, 전문 용어의 뜻풀이 2,800항목을 쉽게 풀이하였다. 이와 함께 일반인이 실생활에서 많이 사용하는 실용 용례와 다양한 쓰임을 보여 주는 용례 14,670항목을 추가하였다.

- 동의어 직접 풀이의 예[표제어: 국제^습지^조약(國際濕地條約)]
  - 기존 풀이: =람사르 협약
  - 수정 풀이: 물새의 서식지로서 국제적으로 중요한 습지의 보호에 관한 국제 조약. 1971년 이란의 람사르에서 열린 국제회의에서 채택되었다. 우리나라는 1997년에 가입하였다. 「동의어」 람사르 협약.
- 뜻풀이에 쓰인 어려운 용어 수정 예[표제어: 임지-촌(林地村)]
  - 기존 풀이: 임지를 개척하여 만든 촌락.
  - 수정 풀이: 나무가 많이 자라고 있는 땅을 개척하여 만든 촌락.
- 전문어 쉬운 풀이 예[표제어: 글라이더(glider)]
  - 기존 풀이: 비행기와 같은 고정 날개를 가진 항공기이지만, 자체에 엔진과 프로펠러나 제트 같은 추진 장치를 가지고 있지 않고 바람의 에너지나 자신의 중력의 전진 성분을 추력으로 삼아 비행하는 항공기.
  - 수정 풀이: 엔진 없이 바람만을 이용하여 나는, 날개가 달린 비행기.

뜻풀이만으로 충분한 설명이 어려울 때 이를 보충하기 위해 《표준》에도 사진과 삽화를 수록하였는데, 《말샘》에서는 이를 확대하여 2010년부터 2013년까지 사진과 삽화, 애니메이션 동영상 등의 다중 매체(멀티미디어) 자료 2,073항목을 추가로 제작하여 수록하였다.

**그림 2** 다중 매체(멀티미디어) 자료 구축 예



<그림자놀이> 애니메이션



<꺾꽂이> 동영상

《말샘》에서 의미를 두어 구축했던 정보 중 하나가 어휘 역사 정보이다. 2010년부터 2012년까지 총 5,000항목에 대한 어휘 역사 정보를 구축하였는데, 역사 정보에서는 어형 변화의 모습과 그 시기, 역사적 변천에 대한 설명, 이형태/이표기와 세기별 용례를 집필하여 제시하였다.

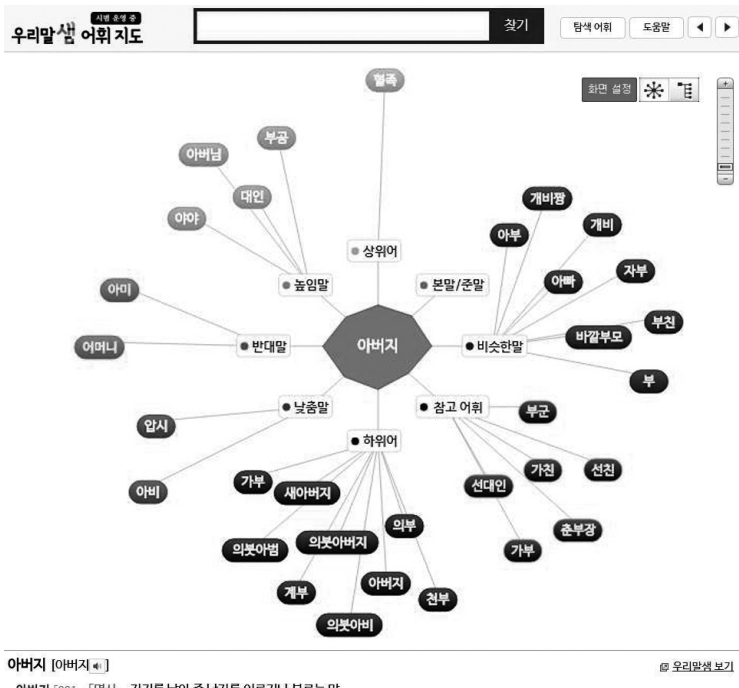
사전에서 비슷한말, 반대말, 본말, 준말, 높임말 등과 같은 표제어와 관련된 어휘들에 대한 정보는 어휘 의미 관계를 통해 표제어의 의미를 보다 잘 이해할 수 있도록 하며, 어휘 교육이나 글쓰기에 유용하게 활용될 수 있는 정보이다. 《말샘》에서는 《표준》에서 이관된 관련 어휘 정보와 《말샘》에서 새로 구축된 어휘의 관련 어휘 정보, 그리고 ‘어휘 의미 체계 구축’ 사업의 결과물을 《말샘》의 체제에 맞춰 통합하기 위해 2013년부터 2015년까지 관련어 정비 구축 사업을 진행하였다. 우선 관련 어휘 정보의 통합 정비에 앞서 《말샘》의 관련 어휘 유형과 범위를 재설정하고<sup>5)</sup> 통합 정비 지침을 마련하여 뜻풀이 단위 표제어 기준으로 2013년에 3만여 항목,

5) 《말샘》의 관련 어휘 유형과 범위는 다음과 같다.

- 비슷한말 뜻풀이가 동일한 비슷한말(동의어), 전문어-일반어, 달리 이르는 말(별칭어, 이칭어 등), 뜻풀이에 차이가 있는 비슷한말
- 반대말, 본말, 준말, 높임말, 낮춤말
- 참고 어휘: 이형태, 동류 개념, 대립어, 큰말, 작은말, 거센말, 센말, 여린말, 기타
- 상위어, 하위어

2014년에 5만여 항목, 2015년에 5만여 항목을 통합 정비하였다. 《표준》에서 제시한 관련 어휘 정보 유형의 큰 틀을 유지하면서 개념적으로 구분이 쉽지 않은 동의어와 비슷한말은 통합하고, 비슷한말과 참고 어휘는 세부적으로 하위분류를 함으로써 관련 어휘 정보를 더 상세하게 제공할 수 있도록 하였다. 《표준》에서 동형어 수준에서 제시되던 참고 어휘 정보는 《말샘》에서 뜻풀이 수준으로 정비되어 제시되었으며, 체제가 다른 자료들을 통합하는 과정에서 생기는 오류와 비일관성을 정비하였고, 뜻풀이를 중심으로 관련 어휘 정보를 살펴 누락된 관련 어휘 정보를 추가하였다. 이렇게 통합 정비된 관련 어휘 정보는 《말샘》의 ‘어휘 지도’를 통해 보기 쉽게 그림 형태로 제시되고 있다.

**그림 3** 어휘 지도



아버지 [아버지 ㅈ]

© 우리말샘 보기

아버지 「001」, 「명사」, 자기를 낳아 준 남자를 이르거나 부르는 말.

《표준》에는 22만여 항목의 전문 용어가 수록되어 있고, 이 전문 용어를 더 쉽게 검색하고 활용할 수 있도록 분야별 쓰임에 맞게 58개 분야로 분류하여 그 표지를 달고 있다. 《표준》 이후 《말샘》에서 새로 구축한 전문 용어들은 그간 전문화, 다변화된 새로운 분야의 단어들이 많이 포함되어 기존의 전문 용어 분류 체계로는 적절하게 분류하기가 어려웠다. 이에 따라 2013년에 전문 용어 분류 체계 재설정 사업을 추진하였으며, 그 연구 결과를 토대로 2014년 국립국어원에서 《말샘》의 새로운 분류 체계를 다음과 같이 확정하였다.

**표 2 전문 분야 분류 체계**

대분야	중분야
1. 인문	교육, 문학, 민속, 언어, 역사, 철학, 인문 일반
2. 사회	경영, 경제, 군사, 매체, 법률, 복지, 심리, 정치, 행정, 사회 일반
3. 자연	동물, 물리, 생명, 수학, 식물, 지구, 지리, 천문, 천연자원, 해양, 화학, 환경, 자연 일반
4. 산업	공업, 광업, 농업, 서비스업, 수산업, 임업, 산업 일반
5. 보건	수의, 식품, 약학, 의학, 한의, 보건 일반
6. 공학	건설, 교통, 기계, 재료, 전기·전자, 정보·통신, 공학 일반
7. 예체능	공예, 무용, 미술, 복식, 연기, 영상, 음악, 체육, 예체능 일반
8. 종교	가톨릭, 기독교, 불교, 종교 일반
9. 명칭	인명, 지명, 책명, 고유명 일반

2단계 분류 체계인 9개 대분야와 67개 중분야로 설정하였는데, 사전 내용을 제시할 때는 67개 중분야 표지로만 제시하고 있다. 이러한 분류 체계에 따라 재분류 지침을 마련하여 2014년부터 2016년까지 《표준》에서 이관된 전문 용어와 《말샘》에서 새로 구축한 전문 용어 60만여 개에 대해 분야를 재설정하였다.

《말샘》에서는 교육 및 언론, 출판 등의 요구와 수요에 부합하는 규범

지식과 정보 구축 사업을 진행하였다. 2013년에 ‘국립국어원(2004), 《어휘별 규범 해설》’, ‘국립국어원·한국어문교육연구소(2010), 《어문 규범의 단계별 교재 집필 지침서》’, 국립국어원의 ‘가나다 질의응답 자료’를 검토, 분석하여 규범 정보를 제시할 1만 개의 표제어를 선정하고 집필 지침을 마련하여 500개의 표제어에 대해 시범적으로 규범 정보를 집필하였다. 2차 연도인 2014년에 1만 개의 표제어를 추가 선정하여 2013년에 선정된 표제어와 함께 총 2만 개 표제어에 대한 규범 정보 집필을 완료하였다. 규범 정보를 제시할 표제어로는 국어 어문 규범과 교육 관련 어휘 및 용어, 고빈도 어휘를 선정하였다. 규범 정보의 유형을 ‘한글 맞춤법, 띄어쓰기, 표준어, 문장 부호, 외래어 표기, 로마자 표기, 표준 발음, 문법, 어휘, 언어 예절, 기타’로 나누고 유형별로 관련 조항과 간결하고 정확한 설명을 제시하였다. 이렇게 집필된 규범 정보는 2015년과 2016년에 교정·교열을 거쳐 《말샘》에 입력한 후 입력된 규범 정보를 다시 검토하였다.

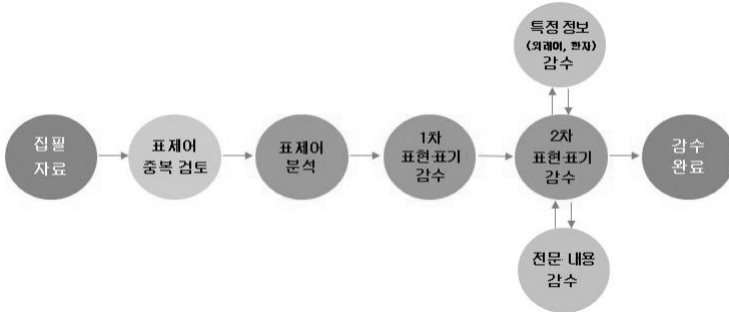
2014년에는 《표준》에서 이관된 옛말 정보에 대한 연구자 및 일반인의 접근성을 향상하고 이해도를 높이기 위해 옛말 표제어의 용례 5,972항목에 한문 원문 정보를 추가하였고, 9,276항목의 옛말 표제어의 용례는 현대 국어로 번역하였다.

## 5. 구축 자료의 교정교열

앞서 2에서 언급한 대로 계획 수립 당시와 사업 초기에는 별도의 교정·교열 단계를 설정하지 않았다. 그 이유는 《말샘》은 일반인이 사전 편찬과 수정에 참여하여 함께 만들어 가는 사전이라는 점을 표방했기 때문에 마중물로서 1차적으로 자료를 구축하고 난 후 개통을 하여 함께 수정하고 보완해 나가는 것으로 계획했기 때문이다. 그러나 2012년 말부터 새로 구축된

자료들을 내부 검토하는 과정에서 사전의 체계에 맞지 않거나 뜻풀이의 내용이 부실하거나 어려운 뜻풀이가 되어 있는 자료들이 발견되면서 새로 구축된 자료에 대해서는 일반에 공개하기 전에 최소한 한 차례의 교정·교열이 필요하다는 점이 지적되었다. 이는 1단계(2010년~2012년) 동안 기존 《표준》의 50만 항목을 가공하고 50만 항목을 새로 집필하여 총 100항목이 넘는 내용을 검토하는 데에는 그 기간과 인력이 턱없이 부족했고, 새롭게 구축된 어휘의 대다수가 다양한 분야의 전문 용어인데 전문 용어 집필자들이 해당 분야의 전문가이기는 하지만 사전 구축 경험은 부족하였던 데서 그 원인을 찾을 수 있을 것이다. 《말샘》이 지속적으로 보완을 해 나가는 사전이기는 하지만 국가가 관리하는 사전이고 구축된 자료는 일반인이 참여하여 새로운 정보를 추가하거나 수정을 할 때 표본이 되는 것이므로 최소한의 신뢰성은 확보할 필요가 있다는 지적에 무게가 실렸다. 이에 따라 외부 전문가를 구성하여 진단 평가를 실시하였고 그 결과를 바탕으로 개선 계획을 수립하였다. 2013년에 개통하기로 하였던 계획을 2016년 개통으로 변경하여 2013년 하반기부터 2016년 개통 전까지 새로 집필된 자료를 교열·교정하고 《표준》에서 이관된 자료와 새로 구축한 자료 간의 일관성을 보완하기로 한 것이다.

《말샘》에 구축된 자료 중 《표준》 이관 자료와 방언 자료를 제외한 신어·생활용어, 전문 용어 전체를 교정·교열 및 감수의 대상으로 삼았으며 그 절차는 다음 그림의 순서로 진행하였다.

**그림 4** 구축 자료 감수 절차

표제어 중복 검토는 같은 의미, 같은 형태의 표제어가 전문 용어 분과, 신어 분과, 생활 용어 분과에서 중복 구축된 것을 가려내어 삭제하는 것으로 국립국어원 내부에서 진행하였다. 표제어 분석은 표제어의 성분 분석을 검토하는 것으로 외부 감수자가 검토를 하되, 일관성 있는 분석을 위해 국립국어원 담당자와 표제어 분석 감수자가 일주일에 한 번 작업 내용을 함께 검토하여 결정한 뒤 반영하도록 하였다. 표현·표기 감수는 품사, 원어, 뜻풀이, 용례 등의 미시 정보들을 살펴 감수 지침에 맞지 않거나 표현·표기에 문제가 있는 정보들을 교정·교열하는 것으로 2차에 걸쳐 진행하였는데, 1차 감수는 외부 감수자를 활용하고 2차 감수는 사전 편찬 경력이 많은 외부 감수자와 국립국어원 내부 감수자가 나누어 맡아 진행하였다. 표현·표기 감수자가 정확한 정보를 확인할 수 없는 전문 내용이나 원어에 외래어, 한자 등 전문적인 내용이 있을 때에는 해당 분야 전문 감수자에게 의뢰하여 감수 결과를 받은 뒤 다시 표현·표기 감수를 하여 마무리하도록 하였다.

2013년 7월과 9월 2차례에 걸쳐 국립국어원 내부의 부족한 감수 인력을 보충하여 감수를 진행할 전문 감수단을 모집하였다. 선정된 감수단에게 감수 지침 교육을 실시한 후 감수를 진행하였는데, 1차 표현·표기 감수를 기준으로 2013년 하반기에 전문 용어 8천 항목, 2014년에는 전문 용어 18만 3천 항목, 2015년에 전문 용어 19만 4천 항목과 신어·생활용어 6만

8천 항목에 대한 표현·표기 감수가 이루어졌다. 2016년에는 남아 있던 2차 감수 항목들과 전문 내용을 파악할 수 없어서 표현·표기 감수를 완료하지 못한 항목들에 대한 전문 내용 감수, 세부 미시 정보 항목별 감수 작업을 진행하여 원어, 뜻풀이 내에 있는 외래어 표기, 신어·생활 용어의 용례 교정·교열, 대역어, 규범 정보 등을 검토하여 정비하였다. 또 2016년 4월에는 제주 지역어에 대해 아래아(·) 표기를 반영할 것인지에 대한 문제가 제기되어 외부 전문가 자문 회의를 통한 의견 수렴을 거쳐 표제어의 표기는 한글맞춤법의 현용 24자모로 한정하되, 아래아 포함 표제어 표기를 뜻풀이의 일부로 제시하기로 결정함에 따라 제주 지역어 표제어와 뜻풀이 수정 작업을 하기도 하였다. 이외 에도 개통 전까지 일일이 열거하기 힘들 정도의 세부적인 항목들에 대한 정비 작업이 진행되었다.

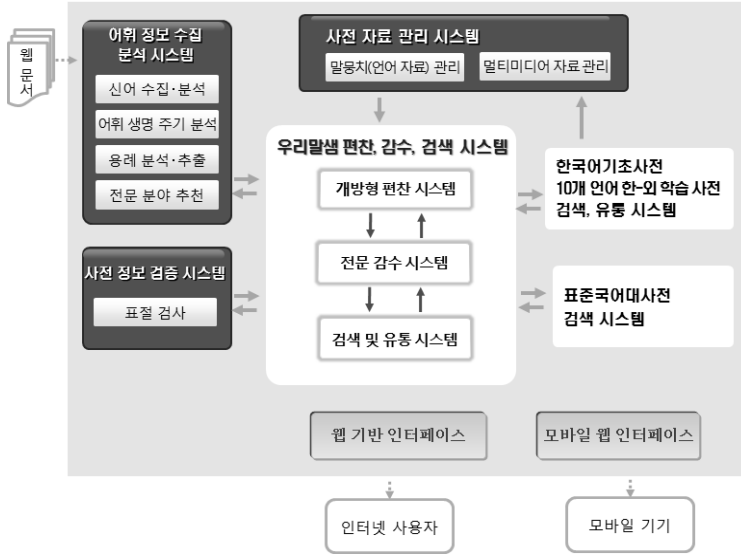
## 6. 개방형 언어 지식 관리 시스템 구축

지금까지 사전 시스템의 역할은 대체로 편찬을 보조하거나 소수 전문가에 의해 편찬된 사전 정보를 일반인이 검색할 수 있도록 하는 데에 머물렀다. 1999년에 종이 사전으로 발간된 《표준》은 2008년에 웹 사전으로 전환되었는데, 종이 사전 형식을 단순히 웹으로 옮긴 것으로 사전 편찬자나 관리자만 접근하는 폐쇄적 관리 시스템이며, 사전 정보의 검색이나 활용의 지원에 있어서도 제한적이다.

개방형 언어 지식 관리 시스템은 발달된 정보 기술을 활용하여 누구나 참여할 수 있는 개방형 구조로 사전 편찬과 활용의 전 과정을 처리할 수 있는 온라인 사전 편찬, 검색, 유통, 활용 시스템을 구축하고, 실시간으로 언어 정보를 수집·분석하고 관리하며 활용할 수 있도록 하는, 단순한 사전 편찬 및 검색 지원 도구가 아니라 국가 언어 지식 관리 도구를 목표로 하여 개발되었다.



그림 5 개방형 언어 지식 관리 시스템 구성 개념도



개방형 언어 지식 관리 시스템 개발 사업은 매년 공모를 통해 개발 사업자를 선정하여 진행하였다. 2010년에는 ‘편찬 지원 체계 구축’ 사업을 통해 언어 지식 관리 체계 구축을 위한 중장기 전략 계획(ISP)을 수립하고 기본 설계를 한 후 편찬 지원 시스템을 개발하였다. 2011년에는 ‘검색 유통 지원 체계 구축’ 사업을 통해 어휘 정보 수집 및 추출 시스템을 개발하고 검색 및 유통 시스템을 개발하였다. 2012년에는 ‘활용 유통 지원 시스템 구축’ 사업을 통해 미등록 어휘 관리 기능과 문장 인식 기반 용례 추천 기능을 설계하고 개발하였으며 기관 간 언어 정보를 연계할 수 있도록 기능을 개발하고 모바일 웹 인터페이스를 구축하였다. 2013년에는 ‘협업 지원 및 어휘 정보 분석 시스템 구축’ 사업을 통해 일반 사용자 편집 항목에 대한 감수 지원 및 관리 기능, 표제어별 토론 지원 기능을 개발하고 말뭉치 관리 기능과 멀티미디어 자료 등록, 공유, 검색 및 관리 시스템, 사용자 정보 분석 및 접속 로그 추출 기능, 어휘별 관리 이력 추출 기능 등의 관리 기능을 개발하였

다. 2014년에는 ‘분류 검증 지원 시스템 개발’ 사업을 통해 전문 용어 분류 체계 편집 및 검증 기능을 개발하고 일반 사용자 집필 요청 관리 기능 등 사용자 참여 지원 기능을 개발하였으며, 원어 및 출전 정보 제시 방법을 개선하고, 어휘 통계 추출 및 제시 기능 등을 개발하는 등 검색 및 활용 기능을 개선하였다. 2015년에는 ‘시스템 고도화 및 운영 지원’ 사업을 통해 시스템의 사용성을 평가하고 이 결과를 바탕으로 편찬, 관리, 검색, 활용 시스템을 고도화하였다. 개통 해인 2016년에는 ‘연계 활용 기능 개발 및 시스템 운영 통합 점검’ 사업을 통해 《말샘》과 《한국어기초사전》 간 정보 연계를 하고 사전 개통을 위해 지금까지 개발된 모든 기능과 보안 등 시스템 전반에 대한 통합 점검을 실시하였으며 시스템 운영을 위한 매뉴얼을 작성하는 등의 운영 준비를 하였다.

## 7. 앞으로의 과제

사전을 편찬하는 일은 끝내자마자 곧바로 수정 작업에 착수해야 한다고 들 하는데 여기에는 두 가지 이유가 있는 것으로 보인다. 첫 번째는 사전의 정보량이 워낙 방대해서 몇 차례에 걸쳐 수정을 해도 오타나 내용의 오류 등의 문제가 없을 수가 없기 때문이다. 두 번째는 늘 새로운 말이 생겨나고, 그 의미가 변하기도 하며, 없어지기도 하는 등 언어는 계속 변해서 ‘아무리 말을 모으고 뜻풀이를 하고 정의를 내려도 사전에 진정한 의미의 완성은 없<sup>6)</sup>기 때문일 것이다. 이런 두 가지 측면 모두에서 《말샘》은 수정에 가장 최적화된 사전이라고 할 수 있다. 《말샘》은 지금까지 나온 국어사전 중에 정보량이 가장 많은 사전이고, 시간과 인력의 문제로 교열·교정 작업이

---

6) 미우라 시온 저·권남희 역(2013), 《배를 워다》, 은행나무.

충분히 이루어지지 못했지만 계속 변하는 말을 지속적으로 담아낼 수 있는 ‘개방형’이라는 틀을 가지고 있기 때문이다. 이렇게 본다면 《말샘》은 개통 전까지의 편찬 작업보다는 오히려 한국어를 사용하는 우리 모두가 주체가 되어 지금부터 함께 해 나가는 수정 작업이 더 중요한 사전이라고 할 수 있을 것이다.

2016년 10월 5일 개통식에서 처음 등록된 ‘재능 나눔’을 시작으로 하여 12월 16일까지 약 2개월 반 동안 일반 이용자가 참여하여 집필한 것을 전문가가 감수하여 《말샘》에 등록된 어휘는 1,320개였다. 실제로 일반 이용자가 ‘집필하기’를 통해 제안한 어휘 수는 2,581개였으나 등록 요건에 맞지 않아 올라가지 못한 것들이 꽤 많이 있었다. 많은 사람들의 의미 있는 참여가 《말샘》의 성공 조건인 만큼 의미 있는 많은 참여를 이끌어 내는 일이 앞으로 해야 할 가장 중요한 과제일 것이다. 아직은 운영 초기라 미흡한 점이 많이 있다. 앞으로 일반 이용자가 쉽고 재미있게 《말샘》 집필 활동에 참여할 수 있도록 운영 체계와 시스템 개선을 위해 지속적으로 노력할 계획이다.

한편, 《말샘》을 풍요롭게 하는 것을 일반 이용자의 참여에만 의존할 수는 없다. 일반 이용자가 집필하기 어려운 정보들을 찾아 채우고 국립국어 원뿐만 아니라 여러 기관들과 학계에서 생산되는 다양하고 정확한 언어 정보와 지식을 《말샘》으로 모아 우리 민족 문화의 중요한 자산인 우리말 정보를 축적하고 다듬어 나갈 수 있도록 해야 할 것이다.

《말샘》은 사전이자 언어 정보의 집합체이다. 《말샘》 정보와 자료가 또 다른 사전과 교육 자료와 문화 콘텐츠를 만들고 언어 처리 기술을 발전시키는 씨앗이 될 수 있도록 《말샘》 자료는 저작권을 공개하였고 오픈 에이피아이(Open API)를 통해 다양한 서비스를 개발할 수 있는 개발 지원을 하고 있다. 그러나 단순히 저작권을 공개한 것에서 더 나아가 다양한 분야에서 언어 정보 콘텐츠로 잘 활용될 수 있도록 정보를 다듬고 활용

지원 기능들을 개선하고 추가하며 홍보를 해 나가야 할 것이다.

이런 과제들을 충실히 이행하기 위해서는 충분한 인력과 예산이 필요하다. 하지만, 기존 사전들과 같이 편찬한 사전을 공개하고 나면 사전 편찬 사업이 끝난 것으로 인식하기 때문에 운영에 필요한 충분한 인력과 예산을 확보하기가 쉽지 않다. 《말샘》은 편찬이 계속되는 사전이라는 점과 그 의의를 이해시키기 위한 노력이 무엇보다 중요하다.

## 참고 문헌

- 국립국어원(2009), 《개방형 한국어 지식 대사전 구축》, 국립국어원.
- 남길임·송현주·최준(2012), 개방형 한국어 지식 대사전의 생활용어 구축, 2012 한국사전학회 제21차 전국학술대회 발표 자료집.
- 박보연(2012), 개방형 한국어 지식 대사전의 방언 구축, 2012 한국사전학회 제21차 전국학술대회 발표 자료집.
- 소강춘·이래호·주경미(2012), 개방형 한국어 지식 대사전의 신어 구축, 2012 한국사전학회 제21차 전국학술대회 발표 자료집.
- 송현주·남길임(2013), 《우리말샘》 관련어 유형 및 정비 구축의 실제, 《한국사전학》 제22호, 한국사전학회.
- 이승재(2012), 21세기형 사전 “개방형 한국어 지식 대사전”, 《한국사전학》 제20호, 한국사전학회.
- 조남호(2000), 《표준국어대사전》 편찬 경과, 《새국어생활》 제10권 제1호, 국립국어연구원.