

《우리말샘》의 정착과 발전을 위하여

최경봉

원광대학교 국어국문학과 교수

말뭉치와 컴퓨터를 활용한 사전 편찬 방법론이 대두하면서 새로운 형식과 내용의 사전이 등장할 수 있는 기반이 조성되었지만, 매체 환경의 변화로 종이 사전 시장이 위축되면서 사전 편찬의 동기는 급격히 떨어졌다. 기대와 우려가 뒤섞인 상황에서 사전의 존재 이유와 존재 방식을 묻는 근본적인 질문이 제기되었다.

“우리에게 사전은 무엇인가?”

“현재의 사전이 미래에도 존속할 수 있을까?”

질문의 답을 찾는 외중에 국립국어원이 주도하여 구축한 웹 사전, 《우리말샘》이 출현하였다. 《우리말샘》의 출현으로 사전의 존재 이유와 존재 방식을 묻는 질문의 답은 좀 더 명확해졌을까? 그런데 사전의 혁신을 통해 사전의 존재 이유를 찾고자 했던 사전 편찬자들에게조차 《우리말샘》의 출현은 당혹스러운 현상일 수 있다. 《우리말샘》의 편찬과 운영 방식 때문이다.

《우리말샘》은 ‘대중이 편찬에 참여하여 집필, 수정, 보완을 거듭할 수

있는 사전'이고, '언어 자료를 지속적으로 수집, 분석하여 곧바로 등재하는 시스템을 갖춘 사전'이라는 점에서 혁신적이다. 그러나 규범성과 체계성을 사전의 정체성으로 보는 사전 편찬자라면 그가 아무리 사전의 혁신을 강조했더라도 이러한 편찬과 운영 방식을 선뜻 받아들이기는 어려울 것이다. 이처럼 사전의 정체성 문제가 제기되는 상황이라면, 일단 "사전의 본질이 무엇인가?"라는 질문을 던져 보며 《우리말샘》의 성과를 검토할 필요가 있을 것이다. 이는 《우리말샘》의 정착과 발전 방안을 모색하는 일이기도 하다.

사전의 본질이 무엇이라는 질문에 대한 답은 다양하겠지만, 그 답들의 공통분모는 '사전은 어휘화된 지식을 구조화해 체계적으로 저장해 놓은 것'이라는 정의로 요약할 수 있을 것이다. 이에 비춰 보면 《우리말샘》의 특징이자 성과로 강조되었던 두 가지 사항, 즉 '규범으로 인정되지 않은 신어를 다량 등재하고, 사전 편찬 전문가가 아닌 일반 대중이 사전 편찬에 참여한다는 것'이 사전의 정체성을 훼손할 수 있다는 우려는 본질에서 벗어난 것임을 알 수 있다. 사실 언어의 사용 양상을 수록하는 기술 사전과 언어적 규범을 제시하는 규범 사전은 역사적으로 공존해 왔고 앞으로도 그럴 것이다. 또한 사전 편찬에 참여하는 인력의 운용 방식이 사전의 정체성을 결정하는 요소로 거론된 적도 없었다.

이를 감안하면 《우리말샘》의 성과는 분명하다. 《우리말샘》은 정보통신 기술을 활용해 이상적인 기술 사전에 한 걸음 다가갔으며, 사전 편찬자의 폭을 획기적으로 넓힘으로써 어휘 의미의 역동성을 사전에 반영할 수 있는 가능성을 높였다. 게다가 《우리말샘》은 《표준국어대사전》의 내용을 받아들임으로써 어휘화된 지식을 구조화하고 저장하는 체계를 안정적으로 구축하는 한편, 다양한 정보를 추가하고 이를 검색하는 방식을 다양화하였다. 따라서 《우리말샘》의 정착과 발전 방안을 모색하는 일을 곧 사전의 존재 이유와 존재 방식에 대한 고민을 확장하는 출발점으로 삼는 것은 의미

가 있다. 그렇다면 《우리말샘》의 정착과 발전 방안을 모색하는 데에서 반드시 검토되어야 할 문제는 무엇인가?

첫째, ‘어휘화된 지식’을 선정하는 데에서 적용되는 가치 중립적인 편찬 원칙에 주목할 필요가 있다. 《우리말샘》의 가치 중립적인 편찬 원칙은 “인터넷상에서 언어 자료를 지속적으로 수집, 분석한 후 신어를 추출, 반영하고 그에 따른 용례도 정비, 반영할 수 있는 실시간 어휘 수집·분석 시스템을 갖추었다.”라는 설명에 잘 나타나 있다. 그런데 어휘화된 지식을 최대한 망라할 수 있는 체계를 갖춘 것은 성과이지만, 사전으로서의 문제도 여기에서 비롯된다는 데 주의할 필요가 있다. 우리말로 된 어휘 지식을 집대성한다는 것을 한글로 표기되어 쓰인 모든 어휘를 등재한다는 것으로 생각한다면 우리말로 어휘화된 지식의 범위에 대한 혼란이 생길 수 있기 때문이다. 이러한 문제 제기는 규범적인 관점에서 유행어, 임시어 등의 가치를 판단하여 등재해야 함을 강조하는 것과는 차원이 다른 것이다. 유행어나 임시어 등을 망라해 등재하는 것은 우리말의 다양성과 역동성을 보인다는 의미가 있다. 그런데 지식의 국제적 교류 과정에서 접하게 되는 외국어 어휘를 수록할 때는 우리말 어휘 지식의 경계 문제가 제기될 수밖에 없다.

한 예로 현재 《우리말샘》에서 영어 ‘언더(under)’가 포함된 단어를 검색하면 120여 개의 표제어가 검색된다. 이는 현재 《표준국어대사전》에 ‘언더(under)’가 포함된 표제어가 20여 개 등재된 것과 대조된다. 《우리말샘》에서 새로 등재한 표제어 중 일부를 제시하면 다음과 같다.

“언더그라운드 레일로드, 언더그라운드 록, 언더그라운드 셔블, 언더플레이즈, 언더드레스, 언더라이팅, 언더라이닝 그래프, 언더레이, 언더리치, 언더마이닝, 언더 메이크업, 언더베스트, 언더벨리, 언더사이즈드 플레이어, 언더슈트, 언더슛, 언더스로 캐스팅, 언더스코어링, 언더스터디……”

문제는 《우리말샘》에 새로 등재한 위의 표제어들이 특정 분야의 사람들이 제한적으로 사용하는 외래 용어라는 데 있다. 더구나 이 중 ‘언더그라운드 서블, 언더벨리, 언더스코어링’ 등은 위의 표제어 의미로는 거의 검색되지 않는 단어이다. 그렇다면 이들이 특정 전문어 사전에 수록된 것이라 하더라도 우리말로 어휘화된 지식의 범위에 포함할 수 있느냐는 의문이 생길 수밖에 없다. 《우리말샘》이 ‘지식 관리 체계’가 아니라 ‘언어 지식 관리 체계’를 지향한다면 외국계 전문어를 우리말로 된 어휘 지식의 범위에 포함하는 기준을 마련할 필요가 있다.

둘째, 우리말로 어휘화된 지식을 남김없이 수용한다는 것은 생동하는 언어의 용법을 기록한다는 의미와 더불어 언어 사용의 경향을 역사적 자료로 남긴다는 의미도 있다. 그런 측면에서 보면 신어를 수집하는 것뿐만 아니라 과거의 어휘를 수집함으로써 공시적, 통시적 차원에서 우리말 어휘를 망라하는 것은 《우리말샘》의 취지에 부합하는 것이다. 《우리말샘》에서는 통시적인 용법 변화 과정을 보여주기도 하지만 이는 일부 표제어에 국한될 뿐이고, 실질적으로는 과거의 어휘 지식 수용은 《표준국어대사전》의 성과를 수용하는 것으로 같음하고 있다. 그러나 근현대 신문, 잡지, 학술서뿐만 아니라 현재 출판 유통되는 근현대 문학 작품의 어휘 중에도 현재 국어사전에서 확인할 수 없는 어휘(다양한 이형태들 포함)는 많다. 이런 점에서 신어 수집 작업과 더불어 과거의 어휘를 수집하는 작업이 꾸준히 이루어질 필요가 있다. 이처럼 공시적, 통시적 차원에서 어휘 지식을 집적하게 되면 새로 생성되는 단어와 사멸한 단어를 판단하는 방안을 마련할 필요가 있다.

셋째, 어휘화된 지식을 구조화해 체계적으로 관리하는 문제를 살펴볼 필요가 있다. 이 문제는 사전의 기술 방식을 결정하고 검색 체계를 구축하는 것과 관련되는 것인데, 여기에서 주목하는 것은 어휘 지식을 제시하는 방식이다. 현재 표제어의 기술 방식은 《표준국어대사전》의 기술 방식을 준용

하고 있다. 그런데 신어의 대부분을 차지하는 전문어의 어휘 정보를 기술하는 데에서 언어 규범 사전에서의 전문어 기술 원칙을 적용하기는 쉽지 않다. 실질적으로 언어적 지식과 백과사전적 지식의 구분이 모호한 상황에서 전문어 사전의 콘텐츠를 상당 부분 반영할 수밖에 없다면, 《우리말샘》과 같은 개방형 사전에서 언어 규범 사전과 같은 전문어 기술 원칙을 세우는 것은 불필요한 일일 수 있다. 그간 국어사전 편찬 시 전문어 기술이 문제가 된 것은 언어적 지식과 백과사전적 지식의 구분이 모호한 데다가 국어사전에서 제시하는 전문어 풀이가 권위성을 지니기도 어렵다는 점 때문이었다. 그렇다면 개방형 웹 사전인 《우리말샘》의 특성과 ‘국가 언어 지식 관리 체계’의 구축이라는 《우리말샘》의 취지를 살리기 위해서는 웹 사전의 형태로 제공되는 각종 백과사전과 《우리말샘》을 연계하는 체계를 갖출 필요가 있다.