

한국어 웹 사전 사용자의 행동에 관한 생각

정 철

카카오 검색 기획 분야

1. 사용자의 행동은 읽기 어렵다

나는 웹 사전을 기획하는 사람이라 언중들이 사전을 어떻게 이용하고 있는가 알고 싶다는 욕망은 이전부터 계속 가져 왔다. 그래서 설문 조사나 심층 면접(FGI, Focus Group Interview) 등을 종종 해 보기도 했지만 거기서 얼마나 적절한 결론을 이끌어 냈는지는 의문이 있다. 사람들은 자신의 행동을 개별적으로 인식하고 있지도 않을 뿐 아니라 그것을 있는 그대로 얘기하는 것도 어려워하기 때문이다. 이상한 공간에서 누가 관찰한다고 했을 때 사용자가 평소대로 행동할 것이라 100% 믿기는 어렵다.

그래서 사전 기능을 추가할 때는 아무래도 주관적인 발상을 통해 먼저 기획을 진행한 뒤 주변에 물어봐서 반응을 보고 결정하는 경우가 많았다. 물론 그것으로 인해 얼마만큼의 사용자 만족도가 증가했는가 등도 측정하기 어려웠다. 대략 주변인들의 평가 등을 통해 정성적으로 예상하는 것이 보통이다. 이상하게 만들면 나쁜 피드백이 돌아오므로 별 피드백이 없고 사용률이 떨어지지 않으면 ‘아, 괜찮은 기획이었나.’ 하고 생각하곤 했다.

사용자 만족도를 측정할 수 있는 극단적인 기회가 하나 있는데 그건 유지하던 기능을 없애는 것이다. 기능을 없앴다 살리는 방식으로는 기능

유지 여부를 결정하진 않지만 가끔 특정 기능을 없애게 되면 그 기능을 쓰던 사용자들의 문의가 다수 들어오기 때문에 ‘아, 이것을 사용하는 사람들도 꽤 있구나.’ 하는 피드백을 얻을 수 있다.

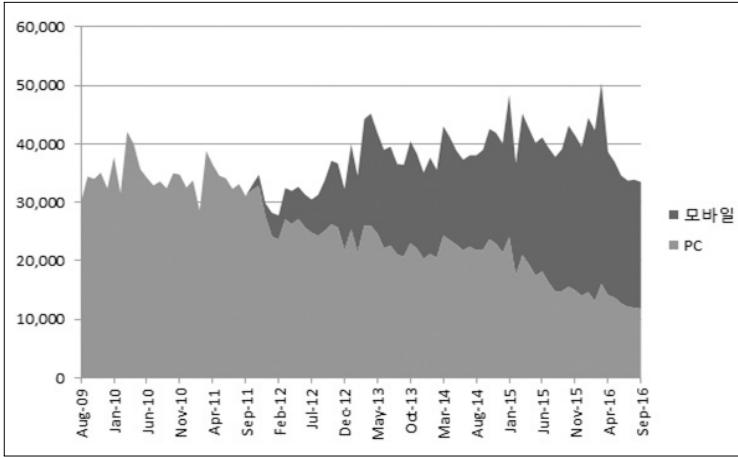
하지만 그것도 잠시뿐이고 다수의 사용자들은 침묵하거나 금방 적응해 버린다. 심지어 질적으로 더 좋은 사전을 제공해 주어도 기존의 익숙한 사전을 도로 내놓으라 하는 경우까지 있기 때문에 사용자의 요구와 행동 양상을 제대로 파악하는 것이란 보통 힘든 것이 아니다. 그래서 사전의 기능이나 내용은 점차 비대해질 뿐 내적으로 충실해지거나 오류를 지워 나가는 일들은 잘 일어나기 어렵게 된다. 다시 말해 사전 콘텐츠를 효과적으로 정리하고 사전 기능을 꼭 필요한 기능 위주로 최적화하는 작업은 아직 걸음마 단계라고 볼 수 있다.

그래도 포털 사이트 내에서 서비스를 만드는 입장에서 기존 국어학 전공자들이 보는 시각과는 다른 관점과 정보를 일부 가지고 있어 그것을 공유하는 마음으로 이하 내용을 기술하려 한다. 이용자 행태를 좀 더 본격적으로 추적하기 위해서는 목적을 먼저 정하고 그 목적에 맞는 자료를 지속적으로 수집하는 것이 먼저임을 통감하는 계기가 되었다.

2. 웹 사전 이용자의 숫자

최근 10년간 개인용 컴퓨터(PC) 웹 사전 이용자의 숫자는 아래와 같다. 2008년, 2009년에 하루 이용자 1,600만 명으로 정점을 찍고 점차 감소하고 있다. 이 감소분은 다수가 스마트폰 환경으로 이동한 것이다. 웹 사전 이용자의 숫자는 줄어들지 않았다.

표 3 개인용 컴퓨터+모바일 웹 사전 월 이용량 추이



(상대 수치, 출처: 다음 내부 자료)

거칠게 해석하자면 스마트폰 보급 이후 웹 사전을 이용하는 경로는 늘어났고 사전 이용량도 늘었지만 크게 보아 최근 10여 년간 웹 사전 사용량 자체는 비교적 일정했다. 한국은 개인용 컴퓨터 환경일 때부터 인터넷 보급률이 높았고 또 인터넷 초기부터 사전 서비스가 제공되어 왔으므로 웹 사전을 이용하는 사람들의 숫자는 어느 정도 포화된 느낌이 있었다.

그리고 [표 1]에 보이는 1,600만 명 등의 수치는 정확하다고 말하긴 어렵다. 왜냐하면 사전 검색 결과는 포털 사이트의 통합 검색 결과에 노출되는데 다수의 사람들이 통합 검색 결과만 보고 이탈해 버리는 경우가 많기 때문이다. 사전 상세 페이지까지 와서 내용을 읽는 사용자는 드물다. 다시 말해 검색 결과 자체가 사전적인 효과를 가지고 있으므로 꼭 사전 내용이 노출되지 않더라도 사람들의 의문은 검색 행위 자체만으로 해소된다. 그럼에도 불구하고 '1,600만'이라는 숫자는 언어 학습 욕구가 있는 대부분의 사람들이 필요할 때 사전을 찾아볼 수 있는 환경 속에 있다는 것을 의미한다.

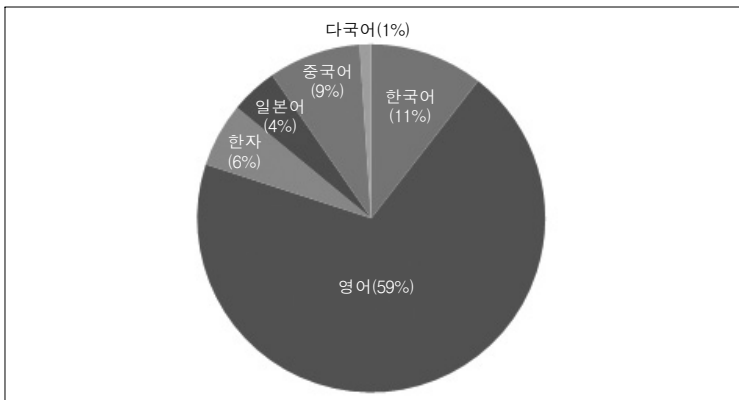
3. 사전 이용 비율

다음에서 어떤 사전을 많이 사용하는가를 보면 압도적으로 영어 사전이 많다. 영어 사전이 69%, 한국어 사전이 11%, 중국어 사전이 9%, 한자 사전이 6%, 일본어 사전이 4%이다. 다국어 사전은 모두 합해도 1% 정도의 비율을 차지한다.(표 4)

한국에서 프랑스어, 스페인어, 러시아어 등 주요 유럽 언어를 직접 번역할 사람이 줄어들까 봐 걱정하는 것도 당연하다. 중국어와 일본어가 예전엔 비교적 엇비슷한 비율을 차지했었지만 지금은 일본어가 중국어의 절반 정도에 불과하다. 다음 일본어 사전은 고지엔(廣辭苑)이라는 훌륭한 일한사전을 서비스 중인데도 그러하다.

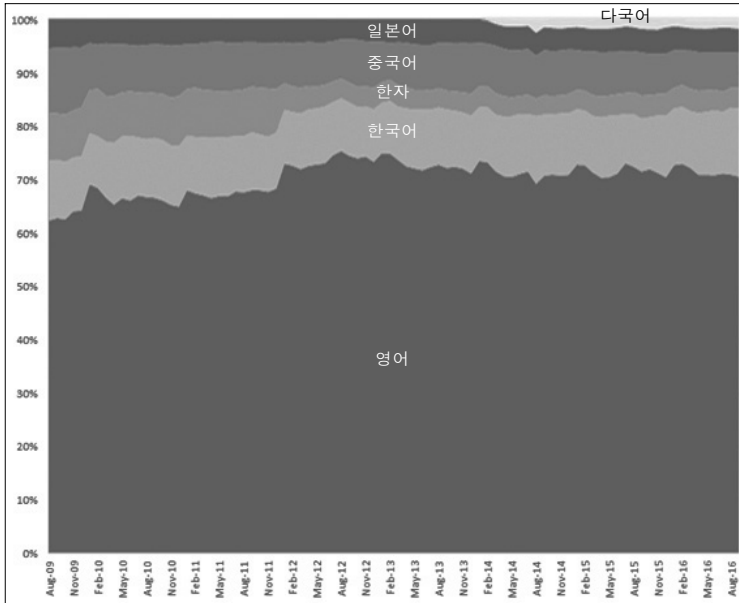
반면에 한자 사전의 6%라는 숫자는 꽤 고무적이다. 다음 한자 사전은 웹 한자 사전 중에서는 품질이 좋은 편이 아니므로 앞으로 충실한 사전으로 만들어 낸다면 숫자는 더 올라갈 것이다. 사람들의 언어생활 속에서 한자가 차지하는 비중은 의외로 엄청나게 높다. 한국어 사전을 찾는 비율이 한자 사전의 두 배밖에 되지 않는다는 것은 특기할 만한 사실이다.

표 4 2009년~2016년 월 평균 페이지 뷰의 사전별 이용 상대 비율



이러한 추이는 시계열적으로 살펴봐도 큰 차이가 없다. 영어 사전 이용은 60% 중반에서 후반으로 오히려 점차 상승했고 일본어와 중국어의 이용은 감소했다. 한국어 사전의 이용 비율은 비교적 일정하다.(표 5)

표 5 2009년~2016년 월 평균 사전별 이용 추이



4. 무엇을 알고 싶어 하는가: 맞춤법, 철자, 띄어쓰기, 외래어

나는 사전 검색어 로그를 분석하여 그 결과를 정리한 것으로 석사 논문을 썼다. 그중 일부의 결과를 여기에 소개해 본다. 이것을 보면 사람들이 사전에서 무엇을 알고 싶어 하는지 대략적으로라도 살펴볼 수 있기 때문이다.

이 자료는 2009년의 검색어 일부를 전수 검수 해서 정리한 것이다. 도저히 다시 정리할 엄두가 안 나는 작업이므로 시간이 좀 지났지만 다시 한 번

사용하기로 한다. 괄호 안의 숫자는 2009년의 한국어 사전 검색어 로그 122만 건 중에서 출현한 빈도수를 의미한다.

(1) 띄어쓰기

와 달다(1,346)	와달다(786), 와닫다(222), 와땀다(61), 와땀다(57), 와다왔다(50), 와닿는(48), 와땀다(47), 와닿는다(41), 와닫는(23), 와땀다(11)
뿐만(622)	뿐 만(622)
더 할 나위(259)	더 할 나위(259)
하 수상(195)	하수상(195)
본의 아니게(166)	본의아니게(166)
것이에요(925), 거예요(500)	거예요(290), 예요(83), 것 이에요(27), 것이에요(17), 거예요.(8)
내 것(149)	내꺼(149)
피치 못할(139)	피치못할(139)
허구한 날(788)	허구헌날(323), 허구한날(317), 허구헌(64), 허구헛날(47), 허구언날(37)
하는 대로(197)	하는대로(197)

가장 왼쪽의 것이 옳은 형태이고 오른쪽의 것들은 실제로 출현했던 오류 형태이다. 한국어 띄어쓰기 규정은 한국인들에게도 극히 어려우며 그것을 제대로 사용할 수 있는 사람들은 언론 출판계 종사자들 정도일 것이다. 그들조차도 자신만의 관행을 유지할 뿐 과연 띄어쓰기를 얼마나 엄밀하게

지키고 있는가 의문이 든다. 그럴 경우 사람들은 자신이 제대로 띄어쓰기를 지키고 있는지를 확인하기 위해 사전을 검색해 본다. 사람들이 찾는 것들 중 다수는 구이거나 그냥 연어일 뿐이어서 당연히 사전에 등재되어 있지 않다.

물론 이 검색어들은 어휘가 아니므로 사전에 등재되지 않는 것이 맞는다. 하지만 고빈도로 출현 중이고 또 띄어쓰기를 안 하는 경향이 있다는 것은 한 단어로 점차 인식되어 가는 도중에 있다고도 할 수 있다. 따라서 완전한 어휘가 아니지만 교육을 위해서 어휘 취급을 해줄 필요가 있다. 사전에 등재해 놓고 문법적으로 어떠한 것인지 설명을 추가적으로 해 주는 것이 바람직하다.

그러나 띄어쓰기와 맞춤법이 동시에 틀린 경우는 자동 띄어쓰기 모듈을 개발하여 사용한다 하더라도 자동으로 고쳐 주기 어렵다. 따라서 고빈도로 출현하는 연어들의 사전 등재에 대해 고민할 필요가 있다. 인터넷 시대의 웹 사전은 ‘단어인가, 아닌가.’보다도 검색 가능성을 중요하게 바라봐야 하기 때문이다.

(2) 맞춤법, 철자

멋쩍다	머쩍다(59), 멋쩍다(26), 멋적다(25), 멧쩍다(21), 머쩍은, 멋쩍은
쑥스럽다	쑥쓰럽다(233), 쑥수럽다(36), 쑥쑤럽다
주꾸미	쭈꾸미(391), 쭈꾸미(36), 주꾸미
착잡하다	착잡, 착잡하다, 착잡한(513), 잡잡하다
찌뿌듯하다, 찌뿌둥하다	찌푸둥, 찌뿌둥, 찌뿌둥하다, 찌부둥, 찌뿌드, 찌부듯하다, 찌뿌드하다, 찌부드드하다, 찌푸둥하다, 찌부드드(254)
늦깎이	늦깎이(457), 늦깎기(96), 늦각이

－ㄱ	닥다(154), 복음밥(33)
－ㄱ	바닷(473), 낚지(149))
언짚다	언짚다(343), 언찰다(111), 엇짚다, 언짚다, 언잔다, 엇짚다, 언찬다
점짚다	점잔다(31), 젊잔다(18)
같짚다	가짚다(29)
안－	않되다(1,225), 앓하다(585), 앓좋다(292)
외궂	외골수(317)
－를	용적울(577), 황금울(45), 달성울(148), 가동울(129), 유병울(118)
도롱농	도롱농(92), 도롱용(18), 도롱농(12), 도롱롱(12)
갑논을박	갑논을박(34), 갑을박론(24)

맞춤법 오류는 끝없이 발굴 가능하다. 사람들이 얼마나 희한한 철자를 입력하는지 보고 있으면 경이로울 정도이다. 사전 이용자들이 오류형을 입력했을 때도 오류를 수정한 정답을 잘 제시하는 수밖에는 다른 대안이 없다. 룰/울, 룡/농 등을 구분해서 일반인들에게 사용하라는 요구는 아무래도 좀 무리하게 느껴진다. 한 어휘 내에서의 맞춤법 오류는 현재의 자동 교정 기능으로 충분히 해결 가능하다.

(3) 외래어

컨펌(confirm[kənʔɜ : m])	کمکم(106), 컨폼, کمکم, کمکم
플래카드(placard[plæka : d])	플랜카드, 플랭카드, 프랭카드, 프랜카드, 프래카드, 플랑카드, 플랫카드, 플렌카드, 프렌카드

레인지(range[réindʒ])	전자렌지(120), 가스렌지(156)
어드벤처지 (advantage[ədʒvæntɪdʒ])	어드벤처지(28), 어드벤처지(16)
스테인리스 (stainless[stéinlɪs])	스테인레스(242, 스테인레스강 포함), 스텐레스(192, 스텐레스, 스텐리스 포함)
스테이플러(stapler[stéiplə])	스테플러(215, 스테플, 스테플러 포함)
라이선스(license[láisəns])	라이선스(543)
로브스터(lobster[lɒbstə])	랍스타(66), 랍스터(관행에 따라 표준 표기로 인정, 2016)
컨셉트(concept[kɒnsɛpt])	컨셉(1838), 컨셉트(361), 콘셉, 컨셉, 콘셉트, 콘세트
리플릿(leaflet[líflɪt])	리플렛(878), 리후렛(191), 리홀렛, 리플렛, 리프렛

외래어 표기 역시 맞춤법만큼이나 어렵기 때문에 사람들이 헛갈려서 검색해 보는 것은 당연한 일이지만 이 중 어떤 것은 표준으로 정해진 표기가 관습적인 표기와 차이가 커서 발생하는 문제이기도 하다. 이걸 인명/지명의 고유 명사를 표기할 때 더욱 큰 문제가 되곤 한다. ‘레오나르도 디캐프리오’의 바른 표기 ‘리어나도 디캐프리오’는 정말 정착할 수 있을지 심히 의문일 뿐 아니라 텔레비전 방송에서도 자막에선 ‘리어나도’, 출연자들의 입말에선 ‘레오나르도’가 등장하는 웃지 못할 경우까지 있었다.

5. 무엇을 고치고 싶어 하는가

천상천하 유아독존: 뜻풀이에 철학적 오류가 있음을 지적.

가족: 혈연 중심으로 설명되어 있어 입양 가족이 배제되어 있음.

독감: 독감과 감기는 다른 것임.

대축일: 종교에 대한 설명의 본문에서 ‘따위’라는 표현은 피하는 게 좋겠다고 지적.

폐백: 뜻풀이에 오류가 있음.

닭싸움: 깨금발 싸움과 닭싸움은 다름.

적시다: 뜻풀이에 정조를 빼앗김이 들어 있는 것은 유교적임.

사레: 재채기와 기침은 다름.

간호조무사: 법정 자격과는 무관함.

입양: 가족이 되는 새로운 방법임.

입갈굴정: 왜 같이천정은 없는가.

제주일보: 과거의 제주일보가 서술되어 있음.

바보: 비하하는 표현이 있음.

도급: 청부인과 도급인은 다른 것임.

제법: 품사 설명이 《표준국어대사전》과 다름.

다음 《한국어사전》에 대한 수정 요구는 지속적으로 들어온다. 이것들을 그대로 여기 옮길 수는 없으므로 유형별로 내용을 요약해 보았다. 이것들은 내용상의 오류, 이익 단체나 직업군의 입장 차이, 문법적인 질문(표준어와의 차이), 시대에 맞지 않는 뜻풀이 등 몇 가지 유형으로 나눌 수 있다. 아무래도 불편하게 느꼈던 것들에 대한 것을 굳이 고객 센터에까지 연락해 가며 고치려고 하는 것들이기 때문에 일리 있는 내용들이 많으며 이것들은 다음 《한국어사전》의 콘텐츠 공급자인 고려대 민족 문화 연구원에 전달되

어 개별적으로 수정, 반영되고 있다.

표준어와의 차이라는 부분은 사람들이 얼마나 표준어, 어법에 민감하게 반응하는가를 보여주는 지점이다. 사람들은 《표준국어대사전》에 있는 것을 표준어라고 생각하는 경향이 강하며 고려대 《한국어 대사전》과 《표준국어대사전》에 차이가 있으면 무척 혼란스러워한다. 두 사전의 표제어가 왜 일치하지 않느냐는 의견을 많이 주고 있으며 고려대 《한국어 대사전》에 수록된 어휘가 《표준국어대사전》에 없을 경우 《표준국어대사전》에 없는 단어가 왜 나오느냐 하고 제보를 하기도 한다. 그만큼 《표준국어대사전》의 명칭에 담긴 ‘표준’이라는 단어가 사람들을 강력하게 장악하고 있다고 말할 수 있을 것이다.

또 자신의 언어 감각과 다를 때 이런 말이 한국어에 있느냐고 따지는 경우들이 꽤 있다. ‘충진하다’처럼 예전에 쓰였으나 최근에 그 쓰임이 적은 것들에 대해 들어오는 제보들로, 이것은 한국어 모국어 화자라는 자신감에 기반을 둔 제안으로 보인다. 언중들은 사전을 언어의 역사적 기록이라는 생각까지는 크게 하지 않는 것인지도 모르겠다.

시대에 맞지 않는 뜻풀이의 경우는 페미니즘이나 가족 관련 어휘들에 대한 의견들이 많았는데 이것은 우리 사회의 여성관이 점차 개선되면서 벌어지는 현상이다. 이런 제안들도 이전의 뜻풀이가 사전에 담겨 있는 것이라는 생각보다는 지금과 다르기 때문에 틀린 뜻풀이로 인식하여 발생하는 제보들이다. 이런 뜻풀이들에 대해서는 해당 뜻풀이가 어느 시대에 한정적으로 쓰였다는 정보를 보강해서 해결할 문제들이다.

이익 집단의 제보는 자신이 속한 집단과 관련된 내용에 오류가 있을 때 들어오는 것들로 비교적 당연한 것들이 많다. 하지만 특정 직업의 자격증은 국가 자격이 아니므로 자격이라는 내용을 지워야 한다거나, 사회적으로 통용되는 표현임에도 불구하고 법적으로 적절하지 않은 표현이라 해서 수정을 요구하는 경우가 꽤 있었다. 주로 경쟁 관계에 있는 업계에서 자신들의

이익이 침해된다는 이유로 요구하는 편이지만 그 제보의 내용을 살펴보면 실제로 그들의 이익이 얼마나 침해당했는지는 의문이며 경쟁자들의 지위를 약화하는 것 자체가 목적으로 보인다. 이것은 포털 사이트뿐 아니라 국립국어원이나 고려 대학교 등 사전을 만드는 모든 곳에서 받는 제보 유형이다.

6. 결론

언중들이 다음 《한국어사전》에서 언급자 하는 것은 ‘확실성’이다. 그 ‘확실성’의 상당 부분을 차지하는 것은 어문 규범이다. 내가 올바른 한국어를 사용하고 있는가에 대해 상당히 예민하게 반응하고 있으며 그래서 다른 사전을 보더라도 《표준국어대사전》을 함께 검색해 보는 경우도 흔하다. 맞춤법이라는 말에 ‘法’이라는 한자가 들어 있기 때문인지는 모르겠으나 규범을 지키지 못하면 법률을 위반한 양 불편해하는 사람들이 많다. 그런가 하면 맞춤법을 잘 모르면서도 열심히 사전을 검색하는 사람들 또한 많다. 그들은 의미를 알고 싶어 하기도 하지만 좀 더 올바른 규범을 알고 싶어 하기도 한다.

‘확실성’의 또 한 가지는 내용상의 확실성이다. 내용이 궁금해서 질문하기도 하고 또 내용상의 오류를 발견하여 수정을 요구하기도 한다. 해당 내용에 대해서 지적하는 사람들은 상당수가 전문직 종사자여서 자신의 지식에 근거해 제보하곤 하지만 어떤 이들은 자기 직업군의 이익에 충실하게 작성해서 수정할 것을 요구하기도 한다. 종종 법원의 판결문까지 첨부하는 공문을 발송하는 사례까지 있었다. 이 경우 사전 편찬자는 이익 당사자들의 입장을 최대한 고려하되 중립적으로 서술하는 수밖에 없다. 사람들은 여전히 국어 사전에 공신력이 있다고 생각하는 것이다. 위키 백과 이후 ‘사전은 옳다.’라는 신화가 급격히 깨져 나가고 있는 시대이지만 그럼에도 불구하고 사람들

이 원하는 것은 ‘옳은 사전’인 것이다.

다음 《한국어사전》 이용자들 다수가 한국어를 모국어로 삼는 사람들이기 때문에 학습을 위해 이용하는 경우는 드물다. 그렇기 때문에 학습자가 많은 영어 사전만큼은 이용하지 않지만 일상 속의 궁금함을 확인하거나 올바른 우리말을 쓰고 있는지 검증하기 위해 한정적으로 사전을 사용함에도 불구하고 전체 사용량의 15% 이상을 《한국어사전》(과 한자 사전)이 차지하고 있다. 일본이나 다른 나라들의 사전 이용 비율이 궁금해지는 지점이다.

종이 사전의 사용량은 확실히 줄어들었으며 이것은 2005년, 2010년의 국민의 언어 의식 조사 보고를 봐도 명확하다. 시디롬(CD-ROM) 사전이나 전자사전의 영향도 서서히 받았지만 개인용 컴퓨터 웹 사전에 이어 스마트폰 모바일 사전이 등장하면서 사전 이용 매체가 결정적으로 변화하고 있다.

참고 문헌

- 김창영(2015) 외, 2015년 국민의 언어 의식 조사 보고서, 국립국어원.
양명희(2005) 외, 2005년 국민의 언어 의식 조사 보고서, 국립국어원.
양정환(2010) 외, 2010년 국민의 언어 의식 조사 보고서, 국립국어원.
정철(2011), 웹 사전 검색 실패어의 유형 연구: 사전학적 활용을 중심으로,
연세대학교 석사학위 논문.
특별취재팀(2014), '死典' 된 국어사전 …… 죽어 가는 민족 지혜의 심장, 2014년
6월 30일 자, 세계일보.