

용치 언어학 : 사전 편찬의 필수적 개념

이 상 섭

연세대학교 부설 한국어 사전 편찬실

Corpus Linguistics as Necessary Concept for Korean Lexicography

Lee, Sangsup

Lexicographical Center, Yonsei University

기존 한국어 사전들은 자연 언어로서의 한국어에 대한 실질적 조사 연구에 근거하고 있지 않다는 점에서 치명적 결함을 안고 있다. 최근 유럽에서 개발·용용되고 있는 「용치 언어학」(corpus linguistics)은 컴퓨터의 급격한 발전에 힘입어 대규모 용량의 자연 언어 자료를 다각적으로 처리할 수 있는 방법을 고안할 수 있게 해주고 있다. 예컨대 영국 버밍엄 대학의 COBUILD 계획은 전혀 새로운 개념의 영어 사전을 편찬하는 데에 성공했다. 한국어 사전의 편찬도 용치 언어학적 방법의 도입으로 가능할 것으로 믿어, 필자가 작성한 작은 「용치」로부터의 실패를 제시한다.

우리말에 관한 최초의 사전들인 「한불사전」과 「한영사전」은 19세기말에 서양인들이 기독교의 전파를 위하여 선교사들의 한국어 안내서로 만든 것이었다. 20세기에 들어서 일본인들은 한국을 지배하기 위하여 조선어-일본어 사전을 만들었다. 우리는 불행히도 외국인들이 특수한 목적으로 만든 이들 사전의 한국어 어휘 자료를 우리말 사전의 기본 자료로 사용하지 않을 수 없었다. 답은 태고 보편 것, 특히 보편 것이 많고, 그 뜻매김과 뜻합이는 우리가 했다.

그런데 외국인들이 불충분하게 모아 놓은 어휘 자료에 우리가 보편 것은 무엇인가? 외국인들은 주로 우리말의 기본 어휘, 즉 사용 빈도가 높은 낱말들을 조사하여 모아 놓았는데, 우리가 거기에 보편 것은 주로 그들이 실수로 빠뜨린 기본 낱말들과, 그들에게는 그리 중요하지 않은 이른바 「어려운」 낱말들, 즉 빈도가 낮은 한자어, 외래어 등이었다. 그리고 수록된 낱말의 수가 많아야 한다는 압박 관념 비슷한 것으로 말미암아, 넣을 만하다고 생각되는 것은 모두 넣는 습관이 붙어서, 온갖 백과사전적 항목들을 찾아다 넣게 되었다. 이런 일은 각종 참고서에서 베껴 넣으면 되지만 우리말의 어휘는 사전 편찬 관련자가 책을 읽다가 우연히 발견한 것이 아니면, 중국어나 일어 사전에 올라 있는 한자어 중 우리말에 쓸 수도 있겠다 싶은 낱말을 베껴 넣은 것이다. 심지어 영어 사전까지도 한국어 사전의 밑천이 되고 있다.

바꾸어 말하면 기존 한국어 사전들은 외국인들이 특수

한 목적으로 제작한 사전들과 국내외의 정보 소용용 참고서에 주로 의존하고 있을 뿐, 실제로 광범위하게 한국어 문헌을 훑으면서 낱말을 수집한 것이 아니라는 치명적 결함을 안고 있다.

이 근본적 결함은 한국어의 실상을 광범위하고도 치밀하게 조사 연구하지 않고는 극복할 수 없다. 한때에는 국어학 이론이 발전되면 좀 더 좋은 국어 사전이 편찬될 수 있을 것이라는 기대도 없지 않았으나, 그것은 근거가 빈약한 기대에 지나지 않음이 드러나고 있다.

방대한 분량의 한국어의 실태를 면밀히 조사하는 엄청난 노력을 의미할 수는 없다. 다른 나라에서 50년, 100년 걸려서 한 일을 우리만은 면제 받을 수가 없다.

천만 다행으로 최근의 정보 처리를 위한 기계와 기술의 급격한 발전에 힘입어 동서양을 막론하고 사전 편찬 작업은 대대적인 전기를 맞고 있다. 뿐만 아니라 언어 정보 처리가 정보 공학이라는 첨단 과학 기술의 가장 중요한 과제로 부상하면서 사전 편찬은 인문과학적 의미에 더하여 전혀 새로운 의미를 가지게 되었다.

실제의 자연 언어에서 어휘와 문법을 추출하여 사전과 문법을 편찬하는 오랜 전통을 가진 유럽 선진국들은 기존 사전의 정보 과학적 전용을 대체로 쉽게 할 수 있지만, 그렇지 못한 우리 사전들은 언어 정보적 가치가 매우 낮다고 할 수 밖에 없다.

따라서 한국어의 실상을 정밀하게 추정할 수 있는 새로운 방법을 모색해야 하는바, 요즘 유럽에서 개발한 실

제 언어 연구 방법의 도입을 심각하게 고려하여야 한다.

연세대학교 한국어 사전 편찬실에서는 영국의 버밍엄 대학의 COBUILD Project<sup>1)</sup>에 꽤 오래전부터 관심을 가지고 연구하고 있다. 그것의 여러 양상중에서 필자는 특히 「용치 언어학」(corpus linguistics)<sup>2)</sup>의 개념에 유의하고자 한다. 이른 언어학과는 달리 용치 언어학은 일정한 원칙에 따라 선정된 대량의 말뭉치(문헌)를 전산기로 처리하여 각종 언어 정보를 추출하는 방법이다. 어휘의 빈도 등 각종 빈도의 측정은 용치 언어학의 가장 기본적인 실행이다.

처리될 말뭉치가 커야 한다는 것은 기본적인 전제가 된다. 100만 마디(tokens)의 용치는 실질적 조어 정보를 얻기 위한 최소한의 분량으로 여겨지고 있다. 「브라운 말뭉치」(Brown Corpus)는 100만 마디 분량의 말뭉치로서, 1967년에 그것을 자료로하여 작성한 각종 빈도표는 영어의 실상에 대한 많은 새로운 정보를 제공해 주었다. 오늘날의 시점에서, 그것은 작은 용치에 불과하며, 사전 편찬의 자료로서는 매우 부족한 것으로 여겨진다.

버밍엄 대학의 언어학자들은 대형 출판사인 콜린즈(collins) 대학의 후원을 받아 1980년에 COBUILD란 이름의 계획을 세우고, 현대 영어 문헌을 수집하여 전산언어학적으로 처리하였다. 이 계획의 주요 목표는 대량 자연 언어 자료의 전산 처리 결과물 기초로한 현대 영어 사전을 편찬하는 것이었다.

그들은 사심한 검토 끝에 선정된 현대 영어의 문헌 중에서 600만 마디를 선정하고 거기에 녹음 자료에서 뽑은 130만 마디를 합쳐, 도합 730만 마디의 용치를 만들기로 이룰 기본 자료로 하여 사전 편찬의 계획을 세웠다. 이 기본 자료를 뒷받침하기 위하여 1,200만 마디의 「배경 용치」를 또한 작성했다. 그러나 약 2,000만 마디 분량(20m running words), 즉 보통 책 5만쪽쯤 되는, 또는 500쪽짜리 책 100권쯤 되는 분량의 말뭉치가 사전 편찬의 자료가 된 것이다.

왜 이처럼 큰 용치가 필요한가? 문법 이론 학자들은 글방에 혼자 앉아서도 국어에 대한 온갖 진실을 다 말할 수 있는데, 이른바 「용치언어학」들은 왜 그리 큰 말뭉치를 가지고서야 언어에 대한 설명을 할 수 있다고 하는가? 이것은 자연 언어에 대한 중대한 개념 차이를 나타낸다. 일반 언어학자들은 주로 자신의, 또는 소수의 언어 사용자의 습관과 직관에 의존하여 언어에 대한 판단을 내린다. 그들에게는 언어 사용자의 습관과 직관이 사실상 유일한 판단 기준이 된다. 그러나 개인 또는 소수의 습관과 직관은 틀림없이 적당하다는 보장은 없다. 그것들은 확률적일 뿐이다. 문법 이론가의 국어에 대한 판단은 절대적이 아니라 확률적일 뿐이다.

확률은 많은 실제 사실들로부터의 귀납의 결과이다. 언어에 대한 개인적 직관이란 결국 개인의 언어 경험에서

귀납된 내용이지, 선험적인 것이 아니다. 따라서 한 언어의 실례의 대량 수집은 그 언어에 대한 판단의 타당성의 확률을 높여 준다.

우리말에서 남자란 말이 더 많이 쓰일지 또는 여자란 말이 더 많이 쓰일지는 아무리 우수한 언어 이론가라도 직관적으로 알아낼 도리가 없지만 적절한 문헌을 대량으로 수집하여 실제로 통계를 내 보면 타당도가 매우 높은 답도 얻을 수가 있다. 한국어의 기본 어휘 500 낱말은 개인적 직관이 알아낼 수 있는 사실이 아니다. 오히려 우수한 처리 능력을 가진 전산기의 도움으로 용치 언어학이 타당성의 확률이 높은 목록을 제공할 수 있다.

문법학자에 의하여 「비문법적」이라는 판정을 받은 문장이 광범위한 실제 언어 자료에서 여러번 효과적으로 쓰인 실례가 발견될 수도 있다. 이 경우 그 문법학자의 판정은 결코 성공했다고 볼 수 없다.

이제는 언어의 확률적 사실에 대한 판단은 개인의 「직관」에 의해서가 아니라 대량 언어 자료의 처리를 통해서, 즉 「용치 언어학적」 방법에 의해서 내려지는 시대가 된 듯하다.

그런데 대량 언어 자료를 무슨 원칙에 의하여 수집하는지는 가장 중요한 문제가 된다. Brown Corpus는 1961년도에 출간된 신문, 잡지, 소설, 오락물 등 20여종의 문헌에서 한 종류당 50,000마디(tokens)씩 선정하여 도합 100만 마디의 용치로 만든 것이었다. 즉 사회에 유포되는 중요 문헌을 20종으로 구분하고, 그 각각에 동일한 중요성을 배분했던 것이다.<sup>3)</sup>

COBUILD Project는 1970년 이후 영어 문헌 가운데에서 문헌의 주제를 크게 중요시하지 않고, 다만 독자를 많이 가졌던 것으로 알려진 문헌을 기본 자료로 삼았다. 즉 주제보다도 보급률이 기본 원칙이 되었다.

연세대학교 한국어 사전 편찬실에서는 정찬실 등이 1988년도에 실시한 사회 조사의 결과를 따라 한국어 문헌을 22종으로 나누고, 그 각각에다 상대적 중요도를 부여하였다.<sup>4)</sup> 예컨대 신문의 정치면은 10%의 중요도가 있고 소설은 8%의 중요도가 있는 것으로 한국인의 독서 경향에 대한 사회 조사는 말해 주고 있다. 이 상대적 중요도의 비율에 따라 우리는 1980년대 이후의 문헌을 종류별로 수집하여 총 3,000만 마디의 용치를 작성, 그것을 기초로 하여 현대 한국어 사전을 제작할 계획인 것이다. 이 원칙에 의하여 우리는 100만 마디 분량의 문헌을 수집하였고, 그 중 10만 마디를 전산적으로 실험하고 있는 중이다.<sup>5)</sup>

용치언어학적 방법이 오늘날 사전 편찬의 가장 타당성이 높은 방법이라는 사실은 COBUILD Project가 작업 개시 7년여 만에 Collins COBUILD English Language Dictionary(1987)를 출간함으로써 확실히 입증되었다.

그들은 우선 수집된 용치의 낱말 및 색인(concordance)

활용형태	동사적 사용	명사적 사용	형용사적 사용	합 계
decline	14번(12%)	108번(88%)	0번(0%)	122번(100%)
declined	76번(100%)	0번(0%)	0번(0%)	76번(100%)
declining	12번(33%)	0번(0%)	26번(67%)	38번(100%)
declines	8번(89%)	1번(11%)	0번(0%)	9번(100%)
합 계	110번(46%)	109번(44%)	26번(10%)	245번(100%)

(표 1) COBUILD 어휘 색인에 나타난 decline의 활용 형태

을 작성하였다. 이는 오늘의 전산기로서 쉽게 해낼 수 있는 일이다. 낱말표(word form, token)를 중앙에 위치케 하고 그 좌우로 그것의 문맥을 제시하는 이른바 문맥속의 중심어(Keyword In Context = KWIC) 색인 방식을 개발했다. 예컨대 영어의 decline이란 낱말의 여러 짧은 730어만 마디 통치에 246번 나타나는데 그 내역은 (표 1)과 같다.<sup>6)</sup>

(표 1)에서 보던 decline은 주로 동사나 명사로 쓰이고, 동사로 쓰이는 경우에는 주로 과거 시제(declined)로 쓰이며, 진행형인 declining은 통사적이라기 보다는 형용사적으로 쓰인다는 사실이 드러난다. 이러한 중요한 사실들은 통치 언어학적 방법에 의하여서만 밝혀질 수 있는 것이지 누구 한 사람의 직관으로 발견될 성질의 것이 아니다.

1 d the Governor in respect to Hearst.' When Hearst  
2. nndate for East Surrey But on each occasion, he  
3. ng (P 208) was to say. "Have some eland", which I  
4. MSELF STRONGLY IN FAVOUR OF FREE SPEECH. BUT HE  
5. ived at the television studio perfectly sober and  
6. ket of Bastos, offered one to the Englishman, who  
7. article on "Male chauvinism, British style." She  
8. onced before telling him about it, he gracefully  
9. ipal Ownership League on 4 October 1905, modestly  
10. ph was in his own performance. The battles he had  
11. g it all at Sujhir's disposal. (P 214) but Sudhir  
12. n we asked Gopal if he would not dance for us, he  
13. last Friday "the darkest [night] of my life"—and  
14. chool in the pony trap for the Easter Egg Hunt. I  
15. us: he had asked an enormous fee, we regretfully  
16. ython era was invited to appear in this film but  
17. lled down his throat. He offered me the bottle. I  
18. dy wished to see me urgently. I made excuses and  
19. trip on this railway from Greenwich to London. He  
20. t a special part was being written in for him. He  
21. n think of no way of stopping him, though he has  
22. cement and to avoid the fellow-traveller's tag. He  
23. ng them. (He was very impatient with students who  
24. nd grew their hair but a few of the more ruthless  
25. ned the invasion in the most forthright terms but  
26. e allowed to address the Court, but the Governors  
27. at at short notice." The Vice-Chancellor's Office  
28. ake it worse! [w]-Ed. Je frequently had requests, always  
29. the cause of death was not made public, Mr Santos  
30. ; "except one, ex-President Grover Cleveland, who  
31. ave other qualities," said Oliver Barrett III but  
32. was now to enter—the lad gravely and resolutely  
33. ion. I think it very significant that the UDA had  
34. isitor. What if I had turned the handle, the door  
35. that the four-man crew included one Briton, but  
36. himself strongly in favour of free speech. But he

기존의 영어 사전들을 보면 decline과 관계가 있는 declension, declinate, declination, declinometer, declinable, decliner, declensional, declensionally, declinational 등과 같은 '어려운, 낱말들이 수록되어 있는데 COBUILD의 통치에는 이것들은 한 번도 안 나타난다. 즉 이것들은 실제 영어에서는 사용될 확률이 매우 적은, 다른 말로 하자면 일반 정보 가치가 0에 가까운 특수 전문 용어들이다. 그럼에도 불구하고 이것들이 일반용 탁상 사전에 버젓히 자리잡고 있다.

앞서 언급한 KWIC 방식은 예컨대 730만 마디 통치에서 발견되는 245개의 decline의 여러 활용 형태들의 행태를 조사하기 위하여 그것들이 나타나는 문맥을 전산기 모니터에 나타나게 하는 방법이다. (표 2)는 245개의 용례중 일부의 예를 보인 것이다.

declined. Al Smith continued with his meeting. den  
declined. He was a zealous Coblenite free trader wh  
declined. It was a dish of minced eland, the big b  
DECLINED TO RE-PUDIATE CONTROVERSIAL TACTICS OF  
declined all refreshments thereafter, but absurd,  
declined, and lit one for himself. "That is not so  
declined, and they didn't ask again. She found she  
declined and shortly afterwards Sir Robert Bignold  
declined, and then accepted five days later as a p  
else-where. He won in the face below his  
declined everything, even to sit down, so they par  
firmly but said his Gura would perform in  
declined further comment. The last word belonged t  
declined her offer. She urged that I should support  
declined his services. There was an awkward interv  
declined it—because I think Brooks has actually  
declined it, as he knew I would. I had rather kiss  
declined politely but two minutes later my bedroom  
declined saying it was a "needless risk to run". A  
declined the offer and returned to his garret. I  
declined the invitation to address Boon as "Charle  
declined to speak at the Labour Party rally in Hyd  
declined to copy.) He had trouble with upright sha  
declined to do so, they had clearly not addressed  
declined to take any action whatever, however dipl  
declined to give an immediate answer to the studen  
declined to comment.—Rummidge Morning Post RIOT  
declined to keep things out of the paper. Several  
declined to comment. Brody looked up from the pap  
declined to allow his sorrow for those who died to  
declined to elaborate. (I doubt if he could have.)  
declined to carry the thing a step further. And at  
declined to attend this mass rally in the past, a  
declined to open and someone had observed my actio  
declined to identify him. The others were said by  
declined to repudiate controversial tactics of pro

(표 2) KWIC 방식으로 나타난 decline의 용례중 일부

이러한 KWIC 방식에 의하여 각 낱말의 활용 형태의 많은 것 뿐만 아니라 활용 형태별 통사 및 의미 정보도 통계적으로 얻을 수 있다.

Collins English Dictionary(CED)라는 명망 높은 영어 사전에서 동사 decline의 의미를 '1) 거절하다, 2) 줄어들다, 3) 기울다, 4) (건강이) 악화되다, 5) (문법용어로서) 체언의 굴절 형태를 나열하다.'의 다섯 가지로 풀이하고 있는데, COBUILD의 110개 용례중 1)의 뜻(거절하다)으로 쓰인 용례는 48개 밖에 안되는데, 그것도 과거 시제(declined)의 경우에만 76개중에서 38개가 '거절하다'의 뜻으로 쓰인 것이 드러난다. 245개의 용례를 검토해 보면 오히려 '줄어들다, 악화되다'란 의미로 쓰인 예가 가장 많고 다음으로는 '줄어들음, 악화'라는 명사적 용례가 많으며 '거절하다'란 뜻으로 쓰인 예는 그중 가장 적으며, 그것도 과거형 또는 과거분사형으로 쓰이며, 문법 용어로서의 '굴절하다'라는 뜻으로는 일반 문헌에서는 쓰이지 않음이 밝혀진다. 그러니까 decline을 '거절하다'라는 뜻을 가진 동사로만 앞에 내세운 CED 사전은 그만큼 틀린 정보를 제공하고 있다. COBUILD 사전에서는 그

러한 기존 사전의 오류를 바로 잡아 '1) 줄어들다, 악화되다, 2) 줄어들음, 악화, 3) 거절하다'의 순서로 decline의 의미를 서술하고 있다.

오랜 수정·증보의 과정을 거쳐서 형성된 이른바 영어 사전도 통치 언어학적 방법에 의하여 그 어구적 성격이 발견되는 형편이니, 그러한 전풍이 없이 우연히 용쳐진 우리의 기존 사전들은 아마 통치 언어학적 방법의 적용을 받는 즉시 거의 모두 와해할 것이다.

필자는 오래전에 사전 편찬과는 전혀 관계가 없이 한국 문학의 고전인 한용운의 사집 『님의 침묵』의 KWIC 색인을 만들어 본바 있다.<sup>7)</sup> 겨우 8,000어 마디(tokens)으로 된 작은 통치이지만 사전 편찬학적 관점에서는 흥미로운 점이 적잖이 발견된다. 그중 명사 '물'이 나타나는 문맥을 다음에 인용한다.

- 1) 달도 아니었던 산 넘고 물 넘어 나의 마음을 바칩니다.
- 2) 오지 않습니다. 꽃나무를 심고 물 주고 복돋우던 일도 아니합니다.

- 3) 가슴을 당신이 만질 때에는 물결이 보드랍지만은 당신의 위엄을 위하어는
- 4) 그 길에는 고개도 많고 물도 맑습니다. 갈 수가 없습니다.
- 5) 고요하고 밤은 고요하고 밤은 물로 씻은 듯합니다.
- 6) 당신을 기다리고 있습니다. 당신은 물만 건느면 나를 돌아 보지도 않고
- 7) 청춘을 잠재우는 남강의 흐르는 물아, 모란봉의 키스를 받고 계절향의
- 8) 비에다 무슨 글을 써서 물에 띄우고 입김으로 불어 서 서로
- 9) 없지마는 산에는 사다리가 없고 물에는 당신이 없 어요. 뉘라서 사다리틀
- 10) 다시 뉘우쳐 합니다. 그들은 물에다 발을 잠그고 반비속이 누워서
- 11) 짓밟습니다. 나는 당신을 안고 물을 건너갑니다. 나는 당신을 안으면
- 12) 아침에 일어나서 세수하려고 물을 떠다 놓으면 당신은 다아
- 13) 략차서 보기도 어려운 맑은 물을 만듭니다. 나의 노래는 님의
- 14) 행복해죽 웃는 온하수의 강안에서 물을 한 줍씩 쥐 어서 서로

(표 3) 한용운의 시집 『님의 침묵』에서  
추출된 「물」의 용례

8,000 마디 동치에 「물」은 14번 나타나는데, (표 3)은 KWIC 방법에 따라 「물」을 중앙에 위치케 하고 그 앞과 뒤에 각각 4개의 낱말을 계속 인용하여 그 문맥을 보이려 한 것이다. 여기서는 편의상 4개의 앞뒤의 낱말(띄어쓰기를 한 문자 기호의 단위)을 인용하는 것에 그쳤으나 전산기 화면이 허락하는 만큼 5개, 6개, 그 이상도 인용할 수 있다.(의문이 생기는 경우에는 그 부분만 길게 불러올 수도 있다.)

(표 3)에서 보면 명사 「물」에는 토씨(조사)가 생략된 예가 2개이고, 「같이」, 「도」, 「로」, 「만」, 「아」, 「에」, 「에는」, 「에다」는 각각 1번씩, 「을」이 4번 쓰이고 있다. 보통 명사로서의 「물」은 아마도 사람의 어떤 행위의 목적어로 가장 자주 쓰인 듯하다. 이처럼 특정 명사에 따른 조사의 사용 빈도도 동치 언어학적으로 세밀하게 추정할 수 있다.

통사론적으로 흥미있는 사실도 보인다. 1)의 「산 넘어 물 넘어」는 6)의 「물만 건느면」과 비교할 때, 수어적임을 알 수 있다. 물은 건너는 것이지 넘는 것이 아니지만, 「산 넘어 물 넘어」와 같은 수어적 표현에서는 「물 넘어」가 쓰인다. 이 관용구는 사전에서 설명되어야 한다. 겨우 8,000 마디 동치에서도 이런 중요한 사실이 발견된다.

2)의 「물 주고」는 하나의 합성동사 [물주다]로 다룰 수도 있겠다. 꽃이나 밭에는 「물을 주」기로 되어 있다. 역시 중요한 연결구조(collocation)이다. 「주다」의 뜻을 가진 「제공하다, 수여하다, 등은 꽃이나 밭과 연결될 수 없다. (\*「꽃에 물을 수여하다」)

5)의 「물로 씻은 듯하다」는 표현도 관용적이다. 깨끗 하기도 신선한 느낌이 들도록 잘 정소한 상태를 「물로 씻 은 듯하다」고 하지 「비누로, 하이하이로, 합성세제로, 휘발유로, 씻은 듯하다고 하지 않는다.

12)의 「물을 떠다 놓으면」에서 「물을 뜨다」도 관용적이다. 「뜨다」라는 말은 그릇에다 물 또는 액체를 한번 사

용할 분량만큼 퍼 담는 것을 가리키는 표현이다. 「살을 뜬다」고 하지 않으며 「큰 물탱크에 물을 뜬다」고 하지도 않는다.

의미론적으로도 「물」은 여러 다른 뜻을 가지고 있음이 위의 14가지 예문에서 드러난다.

1), 4), 6), 7), 8), 10), 11)의 「물」은 주로 하천을 뜻한다. 즉 지형 지물의 한 가지이다.(이 문학 작품의 중요한 주제인 「여행」과 관계가 있다)

2), 5), 12), 14)는 물질 명사로서의 「물」을 가리킨다. 우리의 생활에서 쉽게 다룰 수 있는 물질이다.

3), 13)의 「물」은 물의 추상적 성질, 즉 물의 부드러운 가 맑은 것을 나타낸다. 물다 비유적으로 쓰였다.

이 동치에서 「물」은 적어도 3가지 다른 뜻으로 쓰이고 있다. 겨우 8,000 마디 동치에서도 그런 사실이 드러나니, 그 4,000배인 3,000여만 마디의 동치에서는 얼마나 많은 뜻으로 쓰일지 실제로 조사해 보기 전에는 짐작할 수가 없다. ㅁ출판사의 국어 사전을 보니까 「물」은 「수소 2와 산소 1의 화합물 온온」하고 풀이했는데, 그것은 화학 용어 사전의 뜻대집이지 자연 언어로서의 한국어의 뜻대집이 절대로 되지 못한다.

이처럼 시험적인 작은 동치만 가지고도 사전의 편찬에 관련된 여러 문제를 확인할 수 있다.

동치 언어학적 방법은 오늘의 사전 편찬에 있어서 반드시 고려되어야 하는 방법임이 확실하다. 특히 우리말의 사전적 처리에 꼭 필요한 방법이다.

[1] COBUILD는 Collins Birmingham University International Language Database의 약자이다. COBUILD Project에 대해서는 *The Collins COBUILD English Language Dictionary*(London and Glasgow: Collins, 1987)의 서설과 J.M. Sinclair, ed., *Looking Up*(London and Glasgow: Collins ELT, 1987)에 소상하게 적혀 있다.

[2] Corpus linguistics는 COBUILD 계획과 더불어 널리 유포된 새로운 개념의 분야이다. 필자는 1988년에 이를 국내 처음 소개하면서 「동치 언어학」이라고 옮겼다. 필자의 논문 「동치 언어학적으로 본 사전 편찬의 문제」, 『사전 편찬학 연구』 제2집 (서울: 탑출판사, 1988) 162-193 참조.

[3] Brown Corpus에 대해서는 정한섭의 「우리말 낱말 빈도조사 표본 선정 기준」, 『인문과학』 61집 (연세대 인문과학 연구소, 1989) 5쪽에 언급되어 있다.

[4] 위의 논문 참조.

[5] 이상섭, 「낱말 빈도를 측정하기 위한 말뭉치 자료 수집의 실제」, 『인문과학』 61집 (연세대 인문과학 연구소, 1989), 61-67 참조.

[6] 이 포와 그에 관련된 내용은 모두 J.M. Sinclair, "Lexicographical Evidence," *Dictionaries, Lexicography and Language Learning*, ed. Robert Ilson(Oxford: Pergamon Press, 1985), 81-94에서 따른 것이다.

[7] 이상섭, 『님의 침묵의 어휘와 그 활용 구조: 용례 색인』 (서울: 탐구당, 1984).