

21세기 세종 말뭉치 제대로 살펴보기 — 언어정보나눔터 활용하기

황용주·최정도
국립국어원

1. 서론

잘 알려져 있다시피 ‘21세기 세종계획’은 1997년에 그 계획이 수립되었고 이듬해인 1998년부터 2007년까지 10년 동안 시행된 한국의 국어 정보화 중장기 발전 계획이다(홍윤표, 2009). 이 사업을 통해 한국어 연구를 목적으로 하는 기초 자료에 대한 갈증을 상당히 풀게 된 것이 사실이다. 최초 이 글은 ‘21세기 세종계획’이 끝난 지 약 10년이 되는 지금 시점에 다시금 ‘21세기 세종계획’과 그 성과를 소개하고 세종 말뭉치에 대해 전망하는 것이 목적이었다. 그러나 2009년 《새국어생활》 봄 호에 이와 관련된 특집이 실려 상세히 전달되고 있는 바 ‘21세기 세종계획’의 성과와 전망은 모두 홍윤표(2009), 서상규(2009), 홍종선·남경완(2009)에 기대기로 하고, 여기서는 ‘21세기 세종계획’의 결과물을 간단히 알아보고 실용적인 차원에서 세종 말뭉치의 활용과 관련된 몇 가지를 소개하고자 한다. 그리고 ‘21세기 세종계획’의 결과물과 관련된 사항 중 그간 알려지지 않았던 몇 가지 도움말을 전하고자 한다.

2. ‘21세기 세종계획’ 결과물의 구성

‘21세기 세종계획’의 성과물은 최초 공개 성과물을 조금씩 수정해 가면서 총 네 가지 판으로 배포되었다. 이들을 간단히 소개하자면 아래와 같다.

- ‘21세기 세종계획’의 성과물

- 첫 번째 배포판: 2007년, DVD 4장
 - DVD 1: 세종 말뭉치
 - DVD 2: 전자 사전
 - DVD 3: 한민족 언어 정보화
 - DVD 4: 전문 용어/문자코드 표준화/글꼴 개발/정보화 인력 양성
- 두 번째 배포판: 2009년, DVD 2장(2007년의 수정판)
 - DVD 1: 세종 말뭉치
 - DVD 2: 전자 사전/한민족 언어 정보화/전문 용어/문자코드 표준화/글꼴 개발/정보화 인력 양성
- 세 번째 배포판: 2010년, DVD 2장(2009년의 수정판)
 - DVD 1: 세종 말뭉치
 - DVD 2: 전자 사전/한민족 언어 정보화/전문 용어/문자코드 표준화/글꼴 개발/정보화 인력 양성
- 네 번째 배포판¹⁾: 2011년, DVD 2장(2009년의 수정판, ‘한마루2.0’ 포함)
 - DVD 1: 세종 말뭉치
 - DVD 2: 전자 사전/한민족 언어 정보화/전문 용어/문자코드 표준화/글꼴 개발/정보화 인력 양성

1) 네 번째 배포판이 처음 공개되었을 때에는 DVD 케이스가 노란색이었으나, 수량 부족으로 좀 더 제작하면서 다시 공개된 것은 연두색으로 배포되었다. 그 내용은 서로 같다.



다양한 판이 존재하기에 사용자가 혼란을 겪을 수 있겠으나, 처음 배포된 것보다는 최신 배포판을 사용하는 것이 무난할 것으로 보인다. 현재는 DVD 형태로는 더 이상 배포하지 않고, 2013년 개통한 ‘언어정보나눔터(<http://ithub.korean.go.kr/>)’를 통해 가장 마지막에 배포된 DVD의 내용을 모두 공개하고 있다.²⁾

2) ‘언어정보나눔터’는 사용자가 직접 가입을 해야 이용할 수 있다. 기존 DVD에 들어 있는 내용물은 모두 ‘말뭉치’ 또는 ‘통합자료실’ 탭의 ‘기타 참고자료’에 유형별로 압축되어 등록되어 있다. 따라서 필요한 것을 직접 내려받아 사용할 수 있다.

2.1. 세종 말뭉치³⁾

‘세종 말뭉치’는 크게 ‘문어, 구어, 병렬(한영, 한일), 역사’ 말뭉치로 구성되어 있다.⁴⁾ 사용자의 목적에 맞는 말뭉치를 선택하여 연구할 수 있다.

- 세종 말뭉치의 구성

- 문어 말뭉치
- 구어 말뭉치
- 병렬 말뭉치(한영, 한일)⁵⁾
- 역사 말뭉치

2.2. 전자 사전

세종 전자 사전은 해당 어휘가 가질 수 있는 언어적 정보를 모두 ‘코드화’하여 기술하고 있는 사전이다. 여기에는 기본적으로 해당 어휘의 의미와 의미 분류, 예문이 포함됨과 동시에 종이 사전에서는 제시하기 힘들었던 결합 정보, 격률 등등을 갈래뜻(sense) 단위로 부가한 것이기에 사용자에게 큰 도움이 될 것이다.

- 전자 사전(전체 약 5만 항목)

- 체언 25,458항목
- 동사 15,180항목
- 형용사 4,398항목
- 부사 4,320항목

3) 각 말뭉치에 대한 자세한 설명은 3장에서 기술한다.

4) ‘복한 및 해의 말뭉치’는 저작권 등의 문제로 배포되지 못하고 있다.

5) ‘21세기 세종계획’에서는 ‘한영 병렬 말뭉치’와 ‘한일 병렬 말뭉치’가 주로 구축되었다. 이 외에 ‘한러, 한불, 한중’ 말뭉치가 시범적으로 구축되었으나 양이 적은 편이기에 연구 결과의 신뢰성을 담보하기는 어려울 듯하다. DVD를 통하여 배포되지는 않았으나 적은 양이나마 사용하고자 하는 연구자는 ‘언어정보나눔터’를 통하여 담당자에게 연락하면 자료를 접할 수 있을 것이다.

사용자는 낱낱의 파일을 각종 텍스트에디터에서 열어 확인할 수 있으며, 함께 제공되는 검색 도구인 ‘~단순 검색기/상세 검색기’를 통하여 연구자가 원하는 어휘 정보를 추출할 수 있다. 참고로 ‘단순 검색기’보다 ‘상세 검색기’로 검색하면 검색하는 자료 자체가 상세 사전이기 때문에 더욱 풍부한 내용을 확인할 수 있다.

2.3. 한민족 언어 정보화

‘한민족 언어 정보화’의 결과물은 다음과 같이 구성되어 있다.

- 한민족 언어 정보화⁶⁾

- 국어 어문 규정(검색 프로그램)
- 국어 어휘의 역사(검색 프로그램)
- 한국 방언(검색 프로그램)
 - 남한 방언(검색 프로그램)
 - 북한 방언(검색 프로그램)
 - 중국 및 기타 지역(검색 프로그램)
- 문학 작품에 사용된 방언(검색 프로그램)
- 남북한 언어 비교 사전(검색 프로그램)

위 자료들은 각각이 ‘언어정보나눔터’에 파일의 형태로 등록되어 있는데, 사용자는 ‘통합자료실 > 프로그램’에서 ‘한민족 언어 정보화 통합 검색 프로그램’을 내려받아서 설치한 다음, 단어 단위의 검색어를 입력함으로써 자신이 원하는 정보를 얻을 수 있다. 예를 들어 검색창에 ‘부추’를 입력하면

6) 여기서 ‘국어 어휘의 역사’는 국립국어원 누리집의 ‘자료 찾기 > 국어 어휘 역사’에서도 확인할 수 있고, ‘한국 방언’과 ‘문학 작품에 사용된 방언’은 국립국어원 누리집의 ‘자료 찾기 > 지역어 자료 > 지역어 찾기’에서도 확인할 수 있다.

‘부추’와 관련된 어문 규정 사항을 알 수 있고, 국어 어휘의 역사에서는 ‘부추’가 어느 시기부터 사용되었는지, 한국 방언에서는 ‘부추’가 어느 지역에서 사용되고 있는지 등을 확인할 수 있어 사용자에게 유용한 정보를 제공한다.

2.4. 정보화 인력 양성⁷⁾

국어 정보화 인력 양성의 일환으로 해마다 ‘국어정보화아카데미’가 개최되었는데, 국어 정보화에 몸담고 있는 연구자들의 강의를 통해서 국어 정보화의 실태를 직접 체득할 수 있는 기회가 되었다. 여기 ‘정보화 인력 양성’에는 그때 사용되었던 강의 자료가 담겨 있어 참고할 만하다. 현재는 ‘언어정보나눔터’의 ‘통합자료실 > 기타 참고자료’에서 당시 ‘국어정보화아카데미’에서 열렸던 강의의 자료를 내려받아, 사용자의 연구에 참고할 수 있다.

3. 세종 말뭉치

아무래도 일반인(연구자)가 가장 많이 사용하는 ‘21세기 세종계획’의 결과물은 ‘말뭉치’가 아닌가 싶다. 말뭉치는 머릿속으로 쉽게 만들어질 것이라 예측되지만 실제로는 구축에 시간과 노력이 많이 투입되기에 만들기가 쉽지 않다. 그리고 저작권 등의 문제로 말미암아 쉽게 구해서 쓰기도 힘들다. 이러한 현실에서 ‘21세기 세종계획’의 결과로 배포되고 있는 ‘세종 말뭉치’는 한국어 연구에 지대한 역할을 하고 있는 것으로 보인다. 하지만 아직까지도 말뭉치가 무엇인지 모른다거나 어디서 어떻게 구할 수 있는지를 몰라 활용하지 못하는 사람들이 존재하는 바, 이 자리를 통해 몇 가지 유용한 사항을 전달하고자 한다.

7) ‘전문 용어/문자코드 표준화/글꼴 개발/정보화 인력 양성’ 중 일반인이 사용할 수 있는 것은 ‘정보화 인력 양성’이기 때문에 이에 대해서만 언급한다.

3.1. 세종 말뭉치의 내용

우선 세종 말뭉치는 2015년까지 배포되었던 DVD에 포함되어 있기 때문에, DVD를 가지고 있는 사용자는 네 가지 판 중 어떠한 DVD를 사용하여도 같은 내용의 말뭉치를 이용할 수 있다. 하지만 그 이후로는 DVD가 배포되지 않고 그 내용이 모두 ‘언어정보나눔터’를 통하여 배포되고 있기 때문에, 간단한 절차로 ‘언어정보나눔터’에 가입하고 로그인을 하면, DVD를 받았을 때와 같이 ‘21세기 세종계획’의 결과물을 모두 만나 볼 수 있다.⁸⁾

앞서 세종 말뭉치에는 ‘문어, 구어, 병렬, 역사’ 말뭉치가 있다고 했는데 이들은 말뭉치의 종류(유형)에 따라 다시 여러 가지로 구분된다. 말뭉치의 종류를 부가되는 정보의 수준에 따라 구분하면 아래와 같은 말뭉치들이 존재한다고 할 수 있다.

(1) 말뭉치의 종류

- 원시 말뭉치: 원문을 입력해 놓은 말뭉치
- 형태 분석 말뭉치: ‘원시 말뭉치’에 형태(어휘) 단위의 정보를 부가한 말뭉치
- 의미 분석 말뭉치⁹⁾:
 - ① ‘형태 분석 말뭉치’에 동형어 구분 표지를 부가한 말뭉치
 - ② ‘형태 분석 말뭉치’에 갈래뜻(sense) 구분 표지를 부가한 말뭉치

8) 꼭 로그인을 해야 하는 이유는 자료를 이용하는 데 필요한 ‘사용 각서’의 작성을 대신하기 위한 것이다. DVD를 신청하여 받을 때에도 이와 같은 절차를 거쳤는데, 사용자가 서면으로 작성하고 다시 우편으로 제출해야 하는 불편함을 줄일 수 있는 것으로 결국 사용자의 편의를 위한 것이다.

9) 세종 말뭉치의 ‘의미 분석 말뭉치’는 ①에 해당한다. 만약 사용자가 ②와 같은 효과를 얻고 싶다면 ①의 검색 결과에서 자신이 원하는 방식으로 갈래뜻(sense)을 구분해 주는 작업이 병행되어야 한다.

- 구문 분석 말뭉치: ‘의미 분석 말뭉치’에 통사 구조/기능 표지를 부기한 말뭉치

(2) 말뭉치의 층위

- 원시 말뭉치 ⇒ 형태 분석 말뭉치 ⇒ 의미 분석 말뭉치 ⇒ 구문 분석 말뭉치

원시 말뭉치에서 구문 분석 말뭉치로 갈수록 언어학적 정보가 더 부가 되기 때문에, 사용자는 자신이 원하는 언어 정보가 무엇인지를 고민한 다음 대상 말뭉치를 선택하면 된다. 사용자가 검색어로 입력하는 것(검색어)이 보통 음절이 길다거나 동형어가 없는 어휘(나 형태)의 경우는 ‘원시 말뭉치’나 ‘형태 분석 말뭉치’를 사용할 것을 권하고, 음절이 짧다거나 동형어가 존재하는 어휘(나 형태)의 경우는 ‘의미 분석 말뭉치’를 사용할 것을 권한다.

다음으로 현재 배포되고 있는 세종 말뭉치의 양에 대해서 알아보자.

(3) 세종 말뭉치의 양¹⁰⁾

- 문어 말뭉치

- 원시 말뭉치: 약 3,700만(36,879,143어절)
- 형태 분석 말뭉치: 약 1,000만(10,066,722어절)
- 의미 분석 말뭉치: 약 1,000만(9,071,054어절)
- 구문 분석 말뭉치: 약 45만(433,839어절, 43,828문장)

10) 여기서는 ‘21세기 세종계획’의 전체 보고서에 나와 있는 구축량이 아니라, 실제 DVD 혹은 언어정보나눔터를 통하여 배포되고 있는 말뭉치의 양을 제시한다. 연구자들이 사용할 수 있는 자료는 현재 배포되고 있는 자료가 전부이기 때문에 혼동을 주지 않고자 함이다. 세종 보고서와 결과물의 양이 다른 이유는 저작권과 같은 문제 때문인 것으로 볼 수 있다.

- 구어 말뭉치

- 원시 말뭉치: 약 80만(805,646어절)
- 형태 분석 말뭉치: 약 80만(805,646어절)

(2)에서 확인한 바와 같이 언어 정보가 많이 부가될수록, 말뭉치의 구축이 어렵다. 따라서 ‘원시 말뭉치’에서 ‘구문 분석 말뭉치’로 갈수록 양이 적어지는 것을 확인할 수 있다. 일반적으로 문어보다는 구어 말뭉치의 구축이 상당히 힘들기 때문에 (3)에서 보듯 그 양에서도 차이가 있으며, 문어에서 구축된 의미 분석 말뭉치가 구어에서는 구축되지 못하여 아쉬운 점이 있다.

한편 2000년대 초반에 세종 말뭉치가 한 번 배포된 적이 있다는 것을 소개한다. 이는 ‘21세기 세종계획 균형말뭉치’라는 이름으로 배포된 것인데 그 내용은 아래와 같다.

- 21세기 세종계획 균형말뭉치

- 1,000만 말뭉치
- 문어 대 구어 비율: 90%:10%(BNC 참조)

‘21세기 세종계획’이 진행되고 있을 때 배포된 말뭉치로 영국의 국가 말뭉치인 BNC(1억 어절)의 구성을 참조하여 전체 1,000만 어절로 만들어진 말뭉치이다. 이 말뭉치는 장르 구성이 이미 짜여져 있어 사용자가 고민하지 않아도 되는 균형 말뭉치이며, 자료가 모두 글잡이II(색인)용으로 인덱싱이 되어 있어 바로 형태 중심의 검색과 빈도 산출이 가능하다.¹¹⁾ 이 자료는 언어정보나눔터(‘말뭉치 > 기타 참고자료’)에서 내려받을 수 있다.

11) 원시 말뭉치를 사용하는 글잡이II(직접)에서는 ‘인덱싱’ 과정이 필요 없지만, 형태 분석 말뭉치를 사용하는 글잡이II(색인)에서는 ‘인덱싱’이라는 과정이 꼭 필요하다.

3.2. 세종 말뭉치의 활용

여기서는 세종 말뭉치를 어떻게 활용할 수 있는지를 살펴보자. 즉 어떠한 연구에 활용할 수 있는지를 소개하는 것이 아니라, 어떠한 말뭉치를 사용할 경우 어떠한 도구를 사용해야 한다는 것 등과 같은 실용적인 방법을 소개한다. 아래는 배포되고 있는 각종 도구이다.

- 말뭉치 구축 및 검색 도구
 - 지능형형태소분석기
 - 글잡이II(직접, 색인)
 - 한마루, 한마루2.0
 - 한영 병렬 말뭉치 용례 검색기(hepman)

‘지능형형태소분석기’는 사용자가 자신의 자료를 만들고자 할 때 이용할 수 있는 형태 분석기이다. 원본이 입력된 자료(원시 말뭉치)만 있으면 언제든지 쉽게 분석이 가능하다. 이 도구를 이용한 분석 결과는 바로 ‘글잡이II(직접, 색인)’에서 검색과 빈도 산출이 가능하다¹²⁾.

‘글잡이II(직접, 색인)’는 검색기이자 빈도 산출 도구인데, 사용자가 직접 구축한 말뭉치를 이용하고자 할 때 사용하기에 적합한 도구이다. 다만 아쉬운 점으로는 2000년대에 개발되었기 때문에 유니코드(UTF)를 지원하지 못하는 것과 닫힌 프로그램이기에 앞으로 수정이 될 가능성이 낮은 것을 꼽을 수 있다. 따라서 사용자는 ‘글잡이II’가 지원하는 형식의 자료로 변환해서 사용하여야 한다. 즉 이 말은 현재 배포되고 있는 세종 말뭉치는 전체가 다 인코딩이 유니코드(UTF-16)로 구축되어 배포되고 있기 때문에 ‘글잡이

12) ‘지능형형태소분석기’는 기본적으로 유니코드를 분석하지 못한다(UTF-8의 자료를 분석해 주기는 하지만, 결과물에서 코드가 깨어질 수 있다.). 그래서 세종 말뭉치를 모두 일반 텍스트 코드로 변환하여 분석해야 한다.

Ⅱ'에서는 사용할 수 없다는 말이다. 하지만 앞서 언급했듯이 유니코드(UTF-16)로 저장되어 있는 말뭉치를 '글잡이Ⅱ'의 입력 형식인 '일반 텍스트 코드(CP949, ANSI, ASCII, KS-5601, Euc-kr 등)'로 변환하면 세종 말뭉치를 '글잡이Ⅱ'에서 사용할 수 있다¹³⁾.

그렇다면 유니코드(UTF-16)로 되어 있다는 세종 말뭉치는 어떤 도구에서 사용할 수 있을까? 정답은 바로 '한마루'라는 프로그램이다. '한마루'는 세종 말뭉치에 최적화된 검색 도구이다. 다만 2011년 배포판 이전까지 공개되었던 '한마루'는 일정 수의 용례 추출과 구어 검색이 문제가 있었는데, 2011년판부터 공개된 '한마루2.0'은 그러한 문제를 모두 해소한 훌륭한 검색 도구이다. 따라서 앞으로 세종 말뭉치를 활용하고자 할 때에는 무조건 이 말뭉치에 최적화된 도구인 '한마루2.0'을 사용하자. 이 도구는 기본 검색은 빈도 산출(장르별 빈도 산출도 가능), 연어 검색 등등의 고급 기능을 제공하기 때문에 사용자에게 아주 유용한 프로그램이라 할 수 있다. 다만, 이 도구를 사용할 때에는 오직 세종 말뭉치만을 사용하거나, 사용자가 직접 자신의 말뭉치를 세종 말뭉치의 형식으로 구축해서 사용해야 한다는 것만 언급해 두기로 한다.

그리고 한영 병렬 말뭉치를 다룰 수 있는 도구로는 '한영 병렬 말뭉치 용례 검색기(hepman)'라는 프로그램이 있다.¹⁴⁾ 일반 사용자들은 프로그래밍을 하지 않는 이상 병렬 말뭉치를 그대로 이용하여 검색할 수 없다. 따라서 말뭉치를 이용한 한국어와 영어의 대조 연구를 진행하기 위해서는 이 도구를 이용해야 하고 또한 이 도구에 적합한 검색용 한영 병렬 말뭉치를 입력 말뭉치로 사용해야 한다. 한영 병렬 말뭉치 용례 검색기용 한영 병렬 말뭉치

13) 코드 문제는 도구마다 성격이 조금씩 달라 발생하는 것이지 그 자체가 문제점은 아니다. 한편 검색 대상이 되는 자료를 불러올 때(loading) 파일 단위가 아닌 폴더 단위로 불러와야 하는 점을 언급할 수 있겠다(즉, 한 개의 파일을 불러올 때에도 그 파일을 폴더에 넣어서 폴더 단위로 불러와야 한다).

14) 아쉽게도 다른 병렬 말뭉치를 검색할 수 있는 프로그램을 개발되지 않았다.

는 언어정보나눔터에서 배포하고 있다.

마지막으로 세종 말뭉치에는 (한국어) 역사 말뭉치가 있는데 이 말뭉치를 사용하고자 할 때에는 역사 말뭉치 전용 검색기를 사용해야 한다. 그러한 도구로는 ‘깜짝새’나 ‘유니콩크’라는 프로그램이 있는데 홍운표(2012)의 부록 CD에 첨부되어 있어 활용할 수 있다. 이때 주의할 것은 ‘깜짝새’라는 프로그램은 말뭉치의 입력 형식이 ‘.2b’라는 형식이어야 하고, ‘유니콩크’라는 프로그램은 말뭉치의 입력 형식이 유니코드(UTF-16) 형식이어야 한다는 것이다¹⁵⁾.

간혹 사용하고자 하는 말뭉치가 낱낱의 파일로 되어 있어 한테 묶어 사용할 필요성이 있다거나 ‘한글’ 워드프로세서로 입력되어 있는 여러 파일을 일괄적으로 텍스트 파일로 바꾸어 사용하고자 할 때에는 ‘디지털 한글 박물관’에서 배포하고 있는 ‘한글 자료 처리 프로그램’을 사용해 보자.¹⁶⁾

4. 결론

여기서는 ‘21세기 세종계획’의 성과와 전망에 대한 것은 특집으로 마련되었던 2009년 《새국어생활》 봄 호에 기대고, 구체적인 ‘21세기 세종계획’의 결과물을 간단히 소개함으로써, 세종 말뭉치의 활용 방법에 대한 도움말을 제시하였다. 아직도 세종 말뭉치의 존재를 모른다거나 혹은 세종 말뭉치를 손에 쥐고 있어도 제대로 활용하지 못하는 이들에게 작은 도움이나마 되고자 한 것이 이 글의 목적이다. 각각의 결과물들에 대한 사용법은 ‘21세기 세종계획’의 보고서나 각 도구의 사용 설명서로 밀어 두고, 주로 실제 말뭉치

15) 이 두 프로그램의 차이에 대해서는 이주현(2013)에 상세히 설명되어 있다.

16) ‘한글 자료 처리 프로그램’은 디지털 한글 박물관(<http://archives.hangeul.go.kr/>)에서 내려받을 수 있다.

나 도구를 사용하면서 경험적으로 터득할 수밖에 없는 도움말을 전달하고자 하였다. 그 외 자세한 모든 사항은 국립국어원 누리집의 ‘자료 찾기 > 연구 결과’에 탑재되어 있는 ‘21세기 세종계획’ 관련 보고서를 통해 확인할 수 있다. 각 보고서에는 사업이 어떻게 시작되고 진행되었는지, 그 결과물은 무엇인지에 대한 내용이 담겨 있고, 자료의 구축 방법과 도구 사용법 등이 소개되어 있어 유용하게 활용할 수 있다.

한편 현재 세종 말뭉치는 조금씩 나이를 먹어 가고 있다. ‘21세기 세종 계획’ 이후로 한국어의 전반적인 언어 사실을 담고 있는 자료가 구축되지 않아, 현재 배포되고 있는 세종 말뭉치가 조금씩 오래된 말뭉치가 되어 가고 있는 것이다. 일부 민간 단체에서 말뭉치가 구축되고 있기는 하지만 특정 영역의 말뭉치이거나 검색 결과만 확인할 수 있는 말뭉치가 대부분이다. 한시라도 바빠 한국 사람들이 시대별로, 영역별로 한국어를 어떻게 사용하고 있는지를 확인할 수 있는 자료가 구축되고, 그 자료에서 분석된 언어 정보가 말의 저장고인 사전에 오롯이 담겼으면 하는 바람이다.

참고 문헌

서상규·한영균(1998), 《국어 정보학 입문》, 태학사.

서상규(2009), 국어 특수 자료 구축의 성과와 전망, 《새국어생활》, 19-1 (봄), 국립국어원.

이주현(2013), 17세기 국어의 명사형 어미 연구, 연세대학교 석사학위 논문.

최정도(2011), 말뭉치를 이용한 사전 편찬에서의 몇 문제에 대하여, 《언어 사실과 관점》, 27, 언어정보연구원.

홍윤표(2002), 《한국어와 정보화》, 태학사.

_____ (2009), 21세기 세종계획 사업 성과 및 과제, 《새국어생활》, 19-1 (봄), 국립국어원.

_____ (2012), 《국어 정보학》, 태학사.

홍종선·남경완(2009), 국어 정보화 사업의 미래와 전망, 《새국어생활》, 19-1(봄), 국립국어원.