
21세기 세종 계획 사업 성과 및 과제

홍윤표 · 전 연세대학교 교수

1. 서론

21세기 세종 계획은 1997년에 그 계획이 수립되어 1998년부터 2007년까지 10년 동안 시행된, 한국의 국어 정보화 중장기 발전 계획이다. 21세기 세종 계획(이하 세종 계획이라 약칭함)이 실행되기 시작했던 1998년과 이 계획의 실행이 끝난 오늘날의 국어 정보화 양상을 비교하면 그 차이는 엄청나다고 할 수 있다. 물론 국어 정보화가 아닌 일반 정보화의 발전에 따른 부가적 효과도 있었겠지만, 세종 계획에 의해 이루어진 업적은 그 이전에 있었던 다른 학문적 업적보다도 크다고 할 수 있다. 앞으로 국어국문학의 연구사를 기술하게 될 때에, 세종 계획은 시대를 구분하는 중요한 기준이 될 것이라고 생각한다. 세종 계획이 실행된 결과로 우리나라의 국어 정보화는 여러 분야에서 비약적인 발전을 하여 왔기 때문이다.

그러나 세종 계획의 실현이 끝났다고 국어 정보화가 완결된 단계라고는 할 수 없다. 왜냐하면 21세기의 정보화 사회는 더 많고 다양한 국어 정보화 작업을 요구하고 있고, 또 국어 정보화 작업이 지속적으로 실행되어야 미래의 정보 사회에 대응할 수 있기 때문이다.

세종 계획의 10개년 계획의 실행이 마무리된 이 시점은 지난 10년 동안 시행되어 온 세종 계획을 되돌아보면서 수정·보완하고 시정하여야 할 점들을 짚어 보고, 이를 통해 미래의 국어 정보화가 나아갈 길을 다시 설정하여야 할 때라고 생각한다.

2. 세종 계획의 수립 과정

세종 계획은 1997년 그 당시의 문화체육부(지금의 문화체육관광부)의 국어정책과(지금의 국어민족문화과)에서 제안하여 1997년 3월 1일부터 8월 30일까지 6개월 동안 6명의 학자들이 중심이 되어 각계의 의견을 수렴해서 수립한 국어 정보화 10개년 계획이다.

세종 계획은 국어국문학 전공자 3인과 전산학과 국어 정보 처리 전문가 3인, 모두 6명과 국어정책과 담당자 2인 등의 8인 및 연구보조원 4명 등, 모두 10여 명이 8번의 합숙 과정과 문화체육부, 정보통신부, 과학기술처의 부서 책임자들과의 토의 과정을 거쳐 입안한 내용을 공청회를 거쳐서 최종안이 마련되었다.

국어국문학 전공자로서는 홍윤표(국어학, 연구책임자), 김흥규(고전문학), 김병선(현대문학) 교수가 참여하였고, 정보학 관계자로서는 박동인(정보학), 박세영(자연언어처리), 이균하(컴퓨터 공학) 교수가 참여하였다. 이렇게 만들어진 세종 계획안은 국어심의회 국어 정보화 분과의 심의를 거쳐 확정되었다. 그 당시 문화체육부 국어정책과의 세종 계획 담당자는 문영호 사무관과 이종인 서기였는데, 모든 모임에 연구진과 국어정책과 담당자는 모든 모임에 한 번도 빠짐없이 참석할 정도로 적극적이었다.¹⁾

1) 특히 국어정책과의 문영호 사무관의 의지와 열성은 세종 계획의 입안과 실행 및 예산 책정에 결정적인 역할을 하였다. 비록 지면상이나마 그 노고와 열정에 크게 감사한다.

3. 세종 계획의 조감

(1) 세종 계획 명칭의 의미

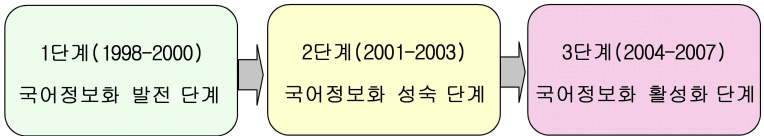
국어 정보화 계획에 ‘세종’의 명칭을 붙인 것에는 특별한 이유가 있었다. 15세기에 세종이 훈민정음을 창제한 목적이 백성들의 문자 문맹률 0%에 있었다고 한다면, 21세기 세종 계획의 목적은 국어 정보를 통한 정보 문맹률 0%에 두었기 때문이다. 결국은 15세기에 세워진 세종의 큰 뜻을 21세기에 와서 정보화로 승화시키자는 뜻이었다.

(2) 세종 계획의 사업 목적

세종 계획의 목적은 바로 ‘국어 정보화’였다. 그리고 국어 정보화의 최종 목표는 ‘우리말과 우리글을 바탕으로 하는 정보 사회 건설’에 있었다. 이를 위해 우리말 정보화, 우리말 체계화, 우리말 세계화라는 표어를 내 세우고 그 발전 단계를 3단계로 나누어 발전 단계(3년), 성숙 단계(3년), 활성화 단계(4년)로 구분하여 실행하기로 하였다. 그것을 표로 보면 다음과 같다.

우리말과 우리글을 바탕으로 하는 정보 사회 건설

- ☞ 첫째, 세계수준의 국어 기초 언어 자료 베이스 구축을 통한 우리말 정보화
- ☞ 둘째, 표준화된 전자사전 구축을 통한 우리말 체계화
- ☞ 셋째, 한민족 언어 정보화를 통한 우리말 세계화



이러한 목적을 구현하기 위해서는 정부의 각 부처가 효율적으로 보완하면서 이 과제를 수행하기로 각 부처 담당자들과 합의를 하였다. 부처별 역할은 다음과 같은 것이었다.

부처	역할
문화체육부	컴퓨터 활용 및 언어학적인 입장에서 공통적으로 필요한 언어 정보 기반을 구축
과학기술처	구축된 언어 정보 기초를 바탕으로 언어 처리 도구 및 응용시스템을 위한 핵심기술 개발
정보통신부	언어 처리 응용시스템(기계 번역, 정보 검색, 지능형 에이전트) 개발

세종 계획의 목표는 사업을 수행해 가는 과정에서 변화하게 되었다. 즉 ‘우리말과 우리글을 바탕으로 하는 정보사회 건설’에서 ‘언어학적인 관점에서 공통적으로 필요한 언어 정보 기반 구축’이라는 목표로 변화한 것이다. 이 목표는 그 당시의 문화체육부가 다른 부서와 분담한 임무 내용을 기준으로 하여 본다면 그 역할을 충분히 수행한 것이 될 것이다. 그

러나 과학기술처와 정보통신부에서 담당한 과제들이 함께 이루어지지 않는다면 넓은 의미의 국어 정보화는 절름발이 상태의 실행이라고 할 수밖에 없다. 세종 계획의 목표가 이렇게 축소된 배면에는 다음과 같은 이유가 있었다.

첫째는 국어 정보화의 효율적인 시행을 위해서는 정부 부처 간의 원활한 협조가 필수적인 것인데, 문화부의 과제에 대해 정보통신부와 과학기술처와의 협조 관계가 거의 이루어지지 않았던 점을 들 수 있다.

둘째는 국어 정보화의 개념에 대한 차이에 기인한다. 넓은 의미의 국어 정보화는 말뭉치 국어학과 전산언어학과 국어 정보 처리를 모두 포괄하는 개념으로서 그 최종 목표는 컴퓨터에 국어 능력을 인식하도록 하는 것에 있다. 그러나 좁은 의미의 국어 정보화는 소위 말뭉치 국어학에 국한된다. 넓은 의미의 국어 정보화는 국민 전체를 대상으로 한 것인데 비해, 좁은 의미의 국어 정보화는 일부 국어국문학에 관여된 전문가만을 대상으로 한 것이라고 할 수 있다. 세종 계획은 국민들의 국어 정보화 능력을 높이기 위한 것이지, 국어학자들의 연구를 돕기 위한 것이 아니어서 세종 계획에서 말하는 국어 정보화는 넓은 의미의 개념으로 설정된 것이다. 그래서 실행 단계를 국어 정보화의 기반 구축 사업과 국어 정보화 여건 조성으로 구분하였던 것이고, 특히 여건 조성을 위해 문자코드의 표준화나, 글꼴 개발 보급 지원이나 국어 정보화 인력 양성 등이 포함된 것이었다. 세종 계획의 국어 정보화는 그 개념이 축소되어 주로 말뭉치 국어학에 치중하게 되고, 그 결과로 국어국문학자, 특히 그 중에서도 국어학자들과 같은 전문가들만을 위한 과제로 변질되어 갔다고 할 수 있다.

셋째로 세종 계획을 행정적으로 담당한 기관의 국어 정보화에 대한 인식 부족에 기인한다고 할 수 있다. 국어 정보화에 대한 정확한 개념을 인식하지 못한 채, 세종 계획의 실행에 깊이 관여하면서 그 방향을 잘못 유도해 간 것이다. 심지어 '21세기 세종 계획'의 계획서조차도 제대로 읽어 보지 않은 것으로 보인다. 그 결과로 세종 계획의 참여자는 원래 국어국

문학을 비롯한 인문학자 70%, 그리고 정보 처리나 자연언어 처리 전문가 30%가 참여하는 것으로 계획되었으나, 연구진의 배정이 잘못되어 거의 대부분이 국어국문학을 비롯한 어문학자들만 참여함으로써 폭넓은 국어 정보화의 길을 가지 못한 것이다.

넷째로 세종 계획의 과제 참여자의 변동에 기인한다고 할 수 있다. 세종 계획을 실현하면서 처음에는 1차년도 과정에서 보는 바와 같이, 정보화의 저변 인구를 확대하기 위해 전국의 각 대학 소속 교수들이 연구자들이었다. 그러나 2차 년도부터 각 중요 과제들의 연구책임자 소속 대학의 연구자들이 주로 참여하면서, 국어 정보화 사업이 일부 한정된 대학(그것도 서울의 일부 대학)의 연구비 지원처럼 변질되었다고 할 수 있다. 10년 동안 연구자들이 전국적인 분포를 보이면서 과제를 시행해 간 곳은 ‘한민족 언어 정보화’ 과제뿐이었다.

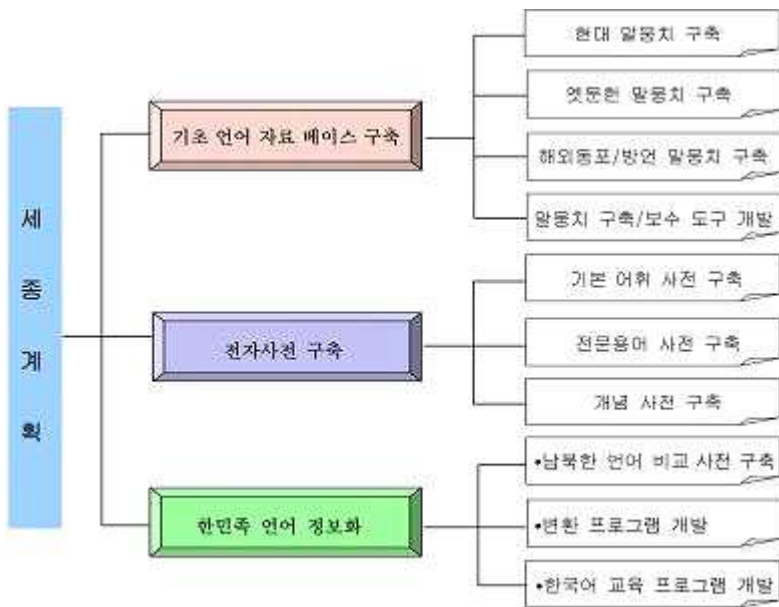
다섯째로 세종 계획의 수행 체계에 대한 문제점에 기인한다. 운영상의 문제는 주로 운영위원회에서 논의·수정·결정되는데, 운영위원들의 대부분이 과제 연구 책임자들이어서, 문제점을 수정하고 보완해 나가는데 어려움이 많았다. 뿐만 아니라 각 과제 연구자들 간의 토의와 논의를 통해 세종 계획을 더 발전시켜야 함에도 불구하고, 그러한 모임이 거의 없었다는 사실도 이해하기 어려운 점이다.

이러한 국어 정보화의 방향이 변화를 했어도, 국어 정보화는 말뭉치 언어학이라는 좁은 범위의 영역 안에서나마 계속 시행되었다.

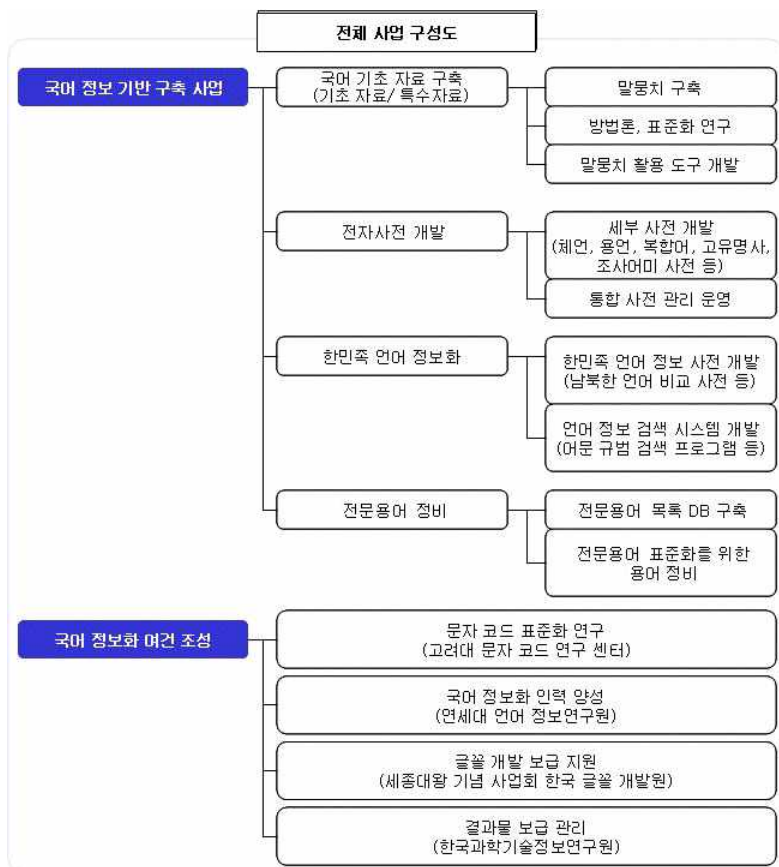
(3) 세종 계획의 사업 구조

세종 계획의 사업 구조는 세종 계획의 목적에 따라 크게 세 가지로 구분하였다. 즉 우리말 정보화를 위한 ‘기초 언어 자료 베이스 구축’(후에 ‘기초 자료 구축’으로 변경), 우리말 체계화를 위한 ‘표준화된 전자사전 구축’, 그리고 우리말 세계화를 위한 ‘한민족 언어 정보화’의 세 분야로

나누고, 각 분야에서 이를 위하여 세부적인 목표를 설정하여 과제를 이행해 나아가기로 하였다. 그 결과 기초 자료 구축을 위해서는 현대 말뭉치, 옛문헌 말뭉치, 해외동포/방언 말뭉치를 구축하기로 하고 그 말뭉치를 구축하고 보수하는 도구까지도 개발하기로 하였다. 전자사전을 구축하기 위해서는 기본 어휘 사전, 전문용어 사전, 개념 사전을 구축하기로 하였으며, 한민족 언어 정보화 분야에서는 '남북한 언어 비교 사전을 구축하고, 변환 프로그램을 개발하고 한국어 교육프로그램을 개발하기로 하였다. 이를 표로 보이면 다음과 같다.

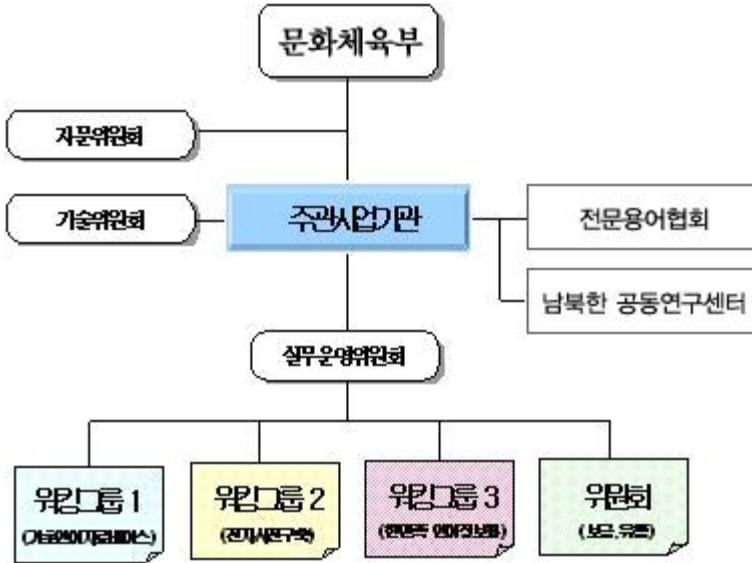


세종 계획은 위에서 보는 바와 같은 사업 구조로 과제를 실현하기로 되어 있었다. 첫해부터 몇 가지 변동 상황이 일어났다. 그것은 '한민족 언어 정보화' 분야의 각종 사업이 국어정책과에서 새로 입안하여 시행하게 된 '한국어 세계화'라는 과제로 흡수된 점이다. 그래서 '한민족 언어 정보화'는 그 과제가 축소되었을 뿐만 아니라, 모든 계획들이 전면적으로 수정되었다. 변경된 세종 계획의 사업 구조를 보면 다음과 같다.



(4) 세종 계획의 수행 체계

세종 계획의 수행 체계의 목표는 다음과 같았다.



이 수행 체계는 그대로 시행되었으나, 단지 자문위원회와 기술위원회가 설치되지 않았다. 자문위원회는 국어심의회 국어 정보화 분과가 그 일을 대신할 수 있었기 때문이었고, 기술위원회의 위원들은 실무운영위원회 속에 들어 있었기 때문이다.

(5) 세종 계획의 단계별 세부 내용

세종 계획의 단계별 세부 내용은 다음과 같았다.

① 기초 언어 자료 베이스 구축

(단위는 어절수)

구분	1단계 (1998~2000)	2단계 (2001~ 2003)	3단계 2004~ 2007)	합계
현대국어 말뭉치	1억(+7,000만)	1억 3,000만	2억	5억 어절
현대국어 분석 말뭉치	1,100만(+200만)	2,200만	4,400만	7,900만 어절
북한/해외 한국어 말뭉치	1,500만	1,500만	1,000만	4,000만 어절
옛 문헌/방언/구비문화 말뭉치	900만	850만	850만	2,600만 어절
대역(한국어-외국어) 말뭉치	200만	400만	400만	1,000만 어절

② 전자사전 구축

구분	제1단계	제2단계	제3단계
기본 사전	10만 단어 구축	20만 단어 구축	50만 단어 구축
전문용어 사전	10개 분야 5만 단어 구축	20개 분야 10만 단어 구축	50개 분야 20만 단어 구축
언어 정보 사전	20만 언어 정보 구축	50만 언어 정보 구축	100만 언어 정보 구축
대역 사전	한영 2만 단어 구축 한일 2만 단어 구축	한영 5만 단어 구축 한일 2만 단어 구축	한영 10만 단어 구축 한일 10만 단어 구축
개념 사전	5만 개념 구축	10만 개념 구축	15만 개념 구축

③ 한민족 언어 정보화

구분	1단계 (1998-2000)	2단계 (2001-2003)	3단계 (2004-2007)
신조어·전문용어 순화	전산·전자 분야	문화·산업·경제 분야	전 분야
남북한 언어 비교 사전	기초 연구, DB 구축	비교 사전 편찬	DB 및 사전 확충
한국어 교육 지원	기초 자료 지원	조직화된 DB 지원	지식 DB 및 온라인 평가
언어정보 통합 네트	기반 구조 형성	서비스 체제 확충	워드 네트 구현
남북한 및 해외동포의 정보화	개인 컴퓨터용 소프트웨어 개발	인터넷 등 통신용 및 방송망용 프로 그램 개발	양적, 질적 확장 단계
남북한 간 이질성 극복 사업	○ 해외 한국어학 지원 ○ (연변) 남북 교류센터 지원	○ 남북 자판, 부호 표준 연구 활동 지원 ○ 남북한 공동 사전	○ 남북한 언어 (어휘) 비교 사전

④ 국어 정보 처리 표준화 분야

구분	제1단계	제2단계	제3단계
기준 부호계	기준 부호계 틀 마련	사용자 인터페이스 마련	통신망(인터넷 등) 을 통한 원격 접근 시스템 구축
비표준문자 등록 및 표준화	비표준문자 수집	비표준문자 관련 정 보 연구(음, 훈 등)	비표준문자 등록 및 수집상황 보급 상설 운영체제 구축
OCR 전용 글꼴	국내 보급 OCR시스 템들의 성능 및 오류 특성 분석	OCR 전용 글꼴개발 (시험안 마련)	OCR 전용 글꼴보완 (확정안 마련)
한글 자형 표준화	한글 한자의 국내 자체의 정리 표준화 대상 목록 작성	한자 자형 사전의 편찬 옛 한글 자형 사전 의 편찬	기초 상용한자의 대표 자형 지정 표준 자형과 약자체의 결정

⑤ 핵심 소프트웨어 개발 분야

구분	1단계(1998-2000)	2단계(2001-2003)	3단계(2004-2007)
국어 처리 소프트웨어	빈도조사 프로그램 등	문체 분석기	지능형 문서작성기
말뭉치 구축 소프트웨어	형태소 해석기	구문 분석기	국어 생성기
말뭉치 처리 소프트웨어	언어(連語)정보 추출기	어휘 구문 통계 추출기	문장분석 퇴고시스템
언어지식처리 소프트웨어	개념분류 체계	분야별 시소러스 조사 결정	지식 획득 소프트웨어

⑥ 예산

원래 세종 계획을 실행하기 위해 소요되어야 할 예산은 다음과 같다.

(단위 백만원)

사업명	총투자	'96예산	'97예산	1단계('98-2000)			2단계 (2001-2003)	3단계 (2004-2007)
				'98	'99	'2000		
기초언어 자료 베이스	5,704	514	250	480	704	774	1,471	1,511
국어정보 처리 표준화	1,640			95	154	169	603	619
전자사전 구축	3,060		90	300	462	508	840	860
핵심 기술 개발	4,983		20	100	230	277	2,163	2,193
환경 조성 및 보급	980			25	95	115	357	388
계	16,367	514	360	1,000	1,645	1,843	5,434	5,571

4. 세종 계획의 결과

이러한 세부 내용에 따라 10개년 동안 이루어진 결과는 다음과 같다.

(1) 기초 자료 구축

구분	계획	결과	
현대국어 말뭉치	5억 어절 (+7,000만)	9316만 어절(구어 466만 어절, 전문용어 200만 어절 포함) 9 + 7000만 어절	18.6%
현대국어 분석 말뭉치	7,900만 어절(+200만)	1,230만 어절	15.6%
북한/해외한국어 말뭉치	4,000만 어절	1,083만 어절	27%
옛문헌/방언/구비문학 말뭉치	2,600만 어절	637만 어절	24.5%
대역(한국어-외국어) 말뭉치	1,000만 어절	766만 어절	76.6%
계	6억 5500만 어절	2억 32만 어절	30%

기초 자료 구축은 곧 말뭉치 구축을 의미한다. 10년 동안 약 2억 어절을 구축한 셈이다. 이 2억 어절 속에는 이전에 구축하여 온 각종 말뭉치, 즉 국립국어원에서 표준국어대사전 편찬을 위하여 만들어 놓은 말뭉치와 정보통신부의 ‘STEP 2000’이란 국가 과제로 입력해 놓은 말뭉치가 포함되어 있다. 그러나 이들을 하나로 묶어 국가 말뭉치(명칭을 ‘세종 말뭉치’라고 하였다)로 정비한 것은 매우 효율적인 것이었다고 할 수 있다.

뿐만 아니라 다양한 말뭉치를 구축한 셈이다. 현대국어 말뭉치를 비롯해서 역사 말뭉치 및 방언 및 구비문학 말뭉치, 그리고 각종의 주석 말뭉치와 대역 말뭉치를 만들어서 각종 분야에서 활용할 수 있도록 한 셈이다. 특히 북한 말뭉치와 해외 말뭉치(주로 중국 및 중앙아시아의 동포사회에서 사용하는 한국어 말뭉치)를 만들어 남한 내에서만 이루어지는 국어 정보화가 아니라 전 세계에 흩어져 사용하는 우리말의 정보화를 겨냥하였다.

말뭉치 구축은 원래의 목적에 비해 약 30%만 시행된 셈인데, 이러한 결과를 가져 온 이유는 다음과 같이 해석할 수 있다.

- ① 갑작스런 국제금융(IMF)으로 인한 예산의 대폭 축소
- ② 물가 상승으로 인한 인건비 및 연구비의 부족
- ③ 과도한 계획

(2) 전자사전 구축

구분	목표	결과
기본 사전	80만 단어	60만 항목
전문용어 사전	35만 단어	
연어 정보 사전	170만 연어	
대역 사전	31만 단어	
개념 사전	30만 개념	

전자사전 구축은 원래의 계획에 비해 매우 축소된 셈이다. 그 이유는 일반 국민들이나 국어 정보 처리 연구자들의 요구하는 대용량의 전자사전이 아니라, 언어학적으로 매우 정밀한 전자사전을 구축하려고 한 데에 기인한다. 따라서 전자사전 구축은 언어학적 연구로서는 성공한 셈이지만, 다양한 방면으로 활용하려는 용도로서는 그리 효율적이지 못한 편이다. 왜냐하면 일반 용도로 사용되는 전자사전은 그 용량이 커야 하기 때문이다. 특히 그러한 용량의 전자사전을 구축하지 못한 데에는 예산 및 인력의 태부족에 기인하는 것으로 여겨진다. 그래서 중도에 다른 분야에 비해 전자사전의 예산이 증액된 편이지만, 그래도 그 정도의 증액만으로는 소기의 목적을 이루는 데에는 도달하지 못하였다.

(3) 한민족 언어 정보화

구분	목표	결과
신조어·전문용어 순화	전산·전자 분야 문화·산업·경제 분야 전 분야	신조어 순화 : 없어짐 전문용어 : 전문 용어 센터로 이관
남북한 언어 비교 사전	기초 연구, DB 구축 비교 사전 편찬 DB 및 사전 확충	13,000 항목 구축
한국어 교육 지원	기초 자료 지원 조직화된 DB 지원 지식 DB 및 온라인 평가	한국어 세계화 과 제로 이관
언어정보 통합 네트	기반 구조 형성 서비스 체계 확충 워드 네트 구현	없어짐
남북한 및 해외동포 의 정보화	개인 컴퓨터용 소프트웨어 개발 인터넷 등 통신용 및 방송망용 프로그램 개발 양적, 질적 확장 단계	없어짐
남북한 간 이질성 극복 사업	해외 한국어학 지원 (연변) 남북 교류센터 지원 남북 자관, 부호 표준 연구 활동 지원 남북한 공동 사전 남북한 언어(어휘) 비교 사전	없어짐
신규 사업 계획	국어 어문규정 검색 프로그램	54,245 항목
	한국 방언 검색 시스템	38,011 어휘
	국어 어휘의 역사 검색 프로그램	5,236 단어
	한민족 언어 정서법 변환 프로그램	
	문학 작품에 사용된 방언 검색 프로그램	2,186 어휘
	한국 전통문화 어휘 검색 프로그램	322 어휘

한민족 언어 정보화의 성과물의 특징은 다른 분야의 성과와는 차별점이 있다고 할 수 있다. 다른 분야의 사업이 주로 국어국문학 전공자들을 위해 이루어진 성과라고 한다면, 한민족 언어 정보화의 성과물들은 일반 국민들이 요구하는 내용이라고 할 수 있다. 그래서 중도에 한민족 언어 정보화는 많은 시련을 겪게 되었다. 즉 세종 계획의 행정 담당자들은 모두 국어학 전공자들이었는데 이들이 인식하고 있는 국어 정보화의 목적

은 말뭉치 언어학에 국한되었기 때문에 한민족 언어 정보화는 원래의 계획에서 크게 후퇴하였다. 뿐만 아니라 예산의 대폭 삭감과 잦은 사업 변경으로 인하여 중도에 연구책임자 및 연구자 모두가 사퇴하는 지경에까지 이르렀으나, 국립국어원에서 이를 수정하겠다는 약속을 받고 철회한 적도 있다. 그러나 그 약속은 지켜지지 않았다. 그런데 흥미로운 사실은 오늘날 일반인들이 가장 많은 관심을 갖고 그 결과물을 이용하려는 것은 한민족 언어 정보화 분야의 결과물이라는 점이다.

한민족 언어 정보화의 사업이 이렇게 변질된 이유는 다음과 같다.

① 신조어·전문용어 순화: 신조어 순화는 국립국어원의 세종 계획 담당자의 개인 연구 과제와 중복되어 없앤 것으로 추정된다. 그러나 그 개인 연구과제와 정보화 근거의 신조어 순화는 전혀 다른 개념인데, 그것을 이해하지 못한 것이다. 전문용어 순화는 전문용어 센터로 이관하게 되었는데, 이것은 세종 계획의 연구자 중 30%는 정보 처리 전문가들에게 위임한다는 계획과 이전의 과학기술처에서 시행하여 왔던 STEP 2000 연구진들을 세종 계획에서 포괄하려는 계획에 의해 이루어진 결과로 보인다.

② 한국어 교육 지원: 한국어 세계화란 새 과제의 실현으로 인하여 소멸되었다.

③ 언어 정보 통합 네트: 21세기 세종 계획 홈페이지로 변환함으로써 소멸되었다.

④ 남북한 및 해외동포의 정보화와 남북한간 이질성 극복 사업: 남북한 관련 과제가 소멸된 이유는 알 수 없다. 단지 그 당시에 한국어 정보 학회가 이 과제로 남북 접촉을 하고 있었던 것에 기인한 것으로 추정될 뿐이다.

⑤ 신규 사업 계획: 위의 여러 가지 사업이 소멸되어서 새로 개발해 낸 과제들이었다. 그런데 흥미롭게도 이 분야가 한민족 언어 정보화의 핵심 사업으로 변화하였다. 원래의 계획과 너무 달라진 셈이다.

(4) 국어 정보 처리 표준화 분야

구분	목표	결과
기준 부호계	기준 부호계 틀 마련 사용자 인터페이스 마련 통신망(인터넷 등)을 통한 원격 접근 시스템 구축	없어짐
비표준문자 등록 및 표준화	비표준문자 수집 비표준문자 관련 정보연구(음, 훈 등) 비표준문자 등록 및 수집상황 보급 상설 운영체제 구축	비표준문자등록센터에서 수행
OCR 전용 글꼴	국내 보급 OCR시스템들의 성능 및 오류 특성 분석 OCR 전용 글꼴개발(시험안 마련) OCR 전용 글꼴보완(확정안 마련)	글꼴 개발센터 과제 지원 중단으로 없어짐
한글 자형 표준화	한글 한자의 국내 자체의 정리. 표준화 대상 목록 작성 한자 자형 사전의 편찬. 옛 한글 자형 사전의 편찬. 기초 상용한자의 대표 자형 지정. 표준 자형과 약자체의 결정.	없어짐

이 분야는 특히 비표준문자 등록 센터의 활동이 가장 인상적이었다. 옛 한글 및 한자코드를 정비한 것은 국민들의 국어 정보화를 원활하게 하려는 의도에 따른 것이기 때문이다. 그러나 나머지의 과제들은 소멸되었는데, 그 이유는 다음과 같은 것으로 해석된다.

① 기준 부호계: 기술표준원의 일과 중복

② OCR 전용 글꼴: 국어 정보화에 대한 개념이 축소되면서, 그 필요성에 대한 인식의 부족으로 인하여 소멸되었다.

③ 한글 자형 표준화: 국어 정보화에 대한 인식의 부족으로 인하여 과제에서 제외되었다.

(5) 핵심 소프트웨어 개발 분야

구분	계획	결과
국어 처리 소프트웨어	빈도 조사 프로그램 등 문체 분석기 지능형 문서 작성기	용례 추출기(글잡이) 문장 분리기
말뭉치 구축 소프트웨어	형태소 해석기 구문 분석기 국어 생성기	지능형 형태소 분석기 구문 태그 부착기 지능형 구문 분석 도구
말뭉치 처리 소프트웨어	연어(連語) 정보 추출기 어휘 구문 통계 추출기 문장 분석 퇴고 시스템	한마루 기타 H2B.EXE H2K.EXE WDcount.EXE WDC_TEL.EXE CWORD.EXE Jamo.EXE Hsort__H.exe Correct.EXE Chk_TEL.exe
언어 지식 처리 소프트웨어	개념 분류 체계 분야별 시소러스 조사 결정 지식 획득 소프트웨어	개발 못함

① 국어 처리 소프트웨어: 문체 분석기를 제외하고는 모두 개발되었다. 문체 분석기는 아직은 국어의 문체에 대한 연구 부족으로 정보 처리에서 제외되었다.

② 말뭉치 처리 소프트웨어: 문장 분석 퇴고 시스템은 연구가 부진하여 실행이 가능하지 않은 상태로 이해된다.

③ 언어 지식 처리 소프트웨어: 아직은 기술력이 부족한 것으로 생각된다.

그러나 나머지 개발된 프로그램들은 국어 정보화 관련 전문가나 국어학 전공의 대학원 학생들이 크게 이용하는 것으로 되었다.

(6) 예산

(단위: 백만 원)

사업명	총투자	결과
국어 기초 자료 구축	5,704	5,174
국어 정보 처리 표준화	1,640	2,905 (전문용어 799, 기타 2,106 포함)
전자사전 개발 구축	3,060	3,748
핵심 기술 개발	4,983	
환경 조성 및 보급	980	1,583 (문자 코드 연구 센터, 인력 양성, 글자체 지원 등)
한민족 언어 정보화		1,656
계	16,367	15,066

이 예산은 천신만고 끝에 확보된 것이다. 세종 계획의 입안이 끝날 무렵 갑자기 닥친 외환위기로 모든 신규 예산이 채택될 가능성이 전혀 없는 상태였었는데, 신규 예산으로 배정된 것은 여러 사람들의 노력 덕분이였다. 한민족 언어 정보화 연구 계획을 세운 연구진들은 직접 경제기획원 담당자들을 만나 국어 정보화의 필요성을 설득하였으며, 특히 김홍규 교수는 예산 담당자들을 직접 만나 그 필요성을 역설하여, 새로운 국가 과제로 선정되는데 결정적인 역할을 하였다. 물론 국어정책과의 노력도 중요한 계기가 되었다.

그러나 이 예산은 실행과정에서 변화가 있었다. 즉 2003년까지 증가하다가 2004년부터 감소하였다. 그 이유는 알 수 없다. 연도별 투자금액을 보면 다음과 같다.

(단위: 백만 원)

연도	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
투자액	850	1,250	1,482	1,730	1,738	1,969	1,871	1,684	1,471	1,021

5. 세종 계획의 성과

세종 계획은 여러 가지 어려운 여건 속에서도 중요한 성과를 이루어 냈다.

즉 짧은 기간과 적은 예산에 비하면, 많은 국어 관련 정보를 구축한 편이라고 할 수 있다. 그 중 뛰어난 업적이라고 볼 수 있는 내용은 다음과 같다.

① 말뭉치의 구축: 약 2억 어절을 구축하였는데, 정보화 선진국에 비해 절대 뒤지지 않는 양이라고 할 수 있다.

② 전자사전의 구축: 약 60만 항목을 구축하여 그 항목 수에서는 부족함이 있다고 할 수 있으나, 구축 정밀성에서는 선진국보다 나은 편으로 평가하고 있다.

③ 어문 생활과 연관된 국어 정보화: 어문규정 검색 시스템, 남북한 언어 비교 사전 검색 시스템, 한국 방언 검색 시스템, 어휘 역사 검색 시스템, 문학 작품에 사용된 어휘 검색 시스템, 전통 문화 어휘 검색 시스템 등을 개발하였는데, 이러한 프로그램들은 어느 개인 업체나 기관, 또는 연구소 등에서도 전혀 개발할 수 없는 시스템을 개발한 것이라고 할 수 있다. 왜냐하면 이 프로그램들은 상업성을 지니는 것이 아니어서 국가에서 개발하여 국민들에게 보급하여야 하는 것들이라고 할 수 있다. 국민들의 국어에 대한 주된 관심사들을 정보화하여 보급한 것이어서 국어학 전문가들뿐만 아니라 국민들 전체를 대상으로 하여 만든 것이라고 할 수 있다.

④ 문자 코드 표준화: 옛 한글, 한자 등에 대한 코드를 표준화하였는데, 이 사업은 국어 정보화의 여건을 조성하는데 필수적인 일이다. 특히 국가 코드 표준 등을 결정하는 중대한 역할을 수행하였다고 할 수 있다.

⑤ 전문용어 DB 구축: 전문용어 DB 구축 및 표준화를 시행한 것인데,

앞으로 계속될 전문용어의 중요성을 인식하고 수행하여 온 것이라고 할 수 있다. 이 작업은 이제부터 시작이라고 할 수 있다.

⑥ 한글 서체의 정리: 한글 글자체에 대한 조사와 새로운 서체를 발굴 하였는데, 이 일은 한글 글자체에 대한 중요성을 인식하게 하였을 뿐만 아니라, 국어국문학자 이외의 여러 분야 전문가들에게 한글에 대한 새로운 인식을 심어 주는데 매우 중요한 역할을 하였다고 생각한다.

⑦ 각종 프로그램 개발: 각종 프로그램을 개발하여 공개하였는데, 이 프로그램들은 현재 국어국문학 전공자들과 일반인들에게는 없어서는 안 될 국어 연구의 중요한 도구로 자리 잡게 되었다.

⑧ 전문 인력 양성 및 보급: 이것을 통하여 각 대학에 국어 정보화 강의가 설정되었고, 이에 따라 국어 정보화 전문 교수 확보를 위한 대학의 노력이 보이기 시작하였다. 그리고 이를 계기로 하여 각 학회에서 국어 정보화에 대한 집중 학술대회가 성행되었다. 뿐만 아니라 21세기 세종 계획 성과물을 이용하여 연구한 논문이 상당수에 이르게 되어, 앞으로 국어 정보화가 더 발전할 수 있는 토양을 마련한 셈이다. 또한 많은 인력이 동원되어서, 국어 정보화에 대한 인식을 넓히고 저변 인구를 확대하는데 중요한 공헌을 하였다고 할 수 있다.

6. 세종 계획의 문제점

위에 제시한 중요한 성과에도 불구하고 수행 과정에서 여러 가지 문제점도 드러났다. 그중에서 반드시 수정·보완되어야 할 문제점을 제시하면 다음과 같다.

① ‘한민족의 정보화’에서 ‘남한 내 정보화’로 축소된 점

이처럼 내용이 축소된 것은 ‘한국어 세계화’의 별도 계획 수립으로 인하여 세종 계획의 목표에 변동이 일어났기 때문이다. 그 결과로 정보화의

남북문제 및 세계 우리 민족의 국어 정보화가 전혀 실현되지 못하였다. 앞으로 반드시 보완해야 할 과제의 하나이다. 그렇지 않으면 한국어와 한글은 남한 내의 언어와 문자로 축소되고 한국어의 세계화나 소위 한류열풍이란 말은 한갓 구호에 지나지 않게 될 것이다. 한국어의 세계화나 한류열풍의 확산과 지속은 한국어 정보화가 밑받침되어야 하기 때문이다.

② 국어 정보화 사업이 연구 중심으로 치우친 점

세종 계획이 ‘정보화의 생산, 유통, 소비’ 중심이어야 함에도 불구하고, 모든 것을 ‘연구 중심’으로 진행시켜서, 국어 정보화의 저변 인구 확대에 지장을 초래하였다고 할 수 있다. 이 문제는 연구자들이 주로 교수 중심이었다는 데에 기인한 것으로 보인다. 기업체 등도 참여하였어야 하겠지만, 예산의 부족으로 참여시키지 못한 것으로 생각된다.

③ 세종 계획의 홍보 및 후속 단계에 대한 관심 소홀

세종 계획에 대한 대국민 홍보 활동이 부족하였다고 할 수 있다. 국어 국문학계나 정보 처리 전문가들에게 세종 계획은 널리 알려져 있었지만, 정작 잘 알아야 할 국민들에게는 전혀 알려지지 않았을 뿐만 아니라, 일반 국민들은 국어 정보화에 대한 어떠한 정보도 제공받지 못했다고 할 수 있다. 그것은 자료의 배포 및 보급이 전문학자 그것도 일부 국어학자와 전산학자에 한정되고, 원래 계획대로 국민들에게 배포 보급하는 것에 등한시한 결과이다. 원래의 계획에는 세종 계획의 결과물들은 각급 학교 등에 배포하고 각종 중요한 프로그램(예컨대 ‘훈글’이나 ‘워드’) 속에 이 자료들이 같이 탑재되도록 하였던 것인데, 그렇게 되지 못한 것이다.

④ 세종 계획 이후의 속(續) 국어 정보화 사업이 진행되지 못한 점

세종 계획은 10개년으로 모든 국어 정보화가 완성된 것은 아니다. 이 국어 정보화 작업은 계속되어야 할 일이며, 이 작업은 어느 개인이나 학

교 또는 연구소에서 감당할 수 있는 일도 아니다. 따라서 세종 계획 이후의 지속적인 과제를 위하여 노력하여야 했음에도 불구하고 그러한 노력이 미진하여, 결국은 10년 계획으로 마무리 된 것이라고 할 수 있다. 우리의 노력과 관심과 의지의 부족으로 인식되어 부끄러울 뿐이다. 지금이라도 세종 계획의 다음 단계 계획을 별도의 과제로 하여 전문가들에게 연구시켜야 하며, 동시에 예산을 확보할 수 있는 방안도 제시하여야 한다.

7. 앞으로의 과제

세종 계획의 10년 동안의 수행결과와 성공 여부는 그 결과들을 어떻게 활용하는가에 따라 결정될 것이다. 그래서 지난 10년이 중요한 것이 아니라 앞으로의 10년이 더 중요하다고 할 수 있다. 앞으로 우리가 국어 정보화를 위해 하여야 할 일들을 제시하면 다음과 같다.

(1) 세종 계획의 후속 계획 마련

전술한 바와 같이 세종 계획의 후속 계획을 시급히 마련하여 예산을 확보하고 이를 시행시키는 일이 앞으로 남은 가장 큰 과제의 하나일 것이다. 현재 국립국어원에서 제시한 국어 정보화 10개년 계획이 있지만, 이는 국립국어원 자체를 정보화시키는 계획일 뿐이지, 국민들을 대상으로 한 국어 정보화 계획이라고 하기 어렵다. 따라서 21세기 세종 계획의 후속 계획을 수립하기 위한 대책을 시급히 마련하여야 한다. 이 후속 계획이 마련된다면, 그 계획 속에는 세종 계획에서 여러 가지 사정으로 이루어 놓지 못한 과제들을 수정·보완하여야 한다.

(2) 세종 계획의 정리

세종 계획의 후속 조치가 없어도 당장 계속하여야 할 일들도 많다. 그것을 제시하면 다음과 같다.

① 세종 계획 결과물 정리 작업

세종 계획에서 이루어 낸 결과물들을 정리하여 일목요연하게 알아볼 수 있도록 해 놓아야 한다. 그러기 위해서는 말뭉치의 분류도 동시에 이루어져야 할 것이다.

② 세종 계획 결과물의 관리 체계 수립

세종 계획을 통해 나온 결과물들을 홈페이지에 공개만 해 놓으면 마치 국어 정보화가 다 이루어진 것처럼 인식하는 일이 없어야 한다. 그 일은 국어 정보화 작업의 일부분일 뿐이다. 가장 중요한 일은 이러한 결과물들을 관리하고 이를 활용할 수 있는 제도적 장치를 마련하는 일이다. 한때 KISTI에 일임하였던 ‘국어 정보 관리 보급 센터’는 국어 전문가들에게 위임하여야 한다. 전산학자 중심의 전문가보다는 인문계 전문가를 중심으로 하고 전산학 전문가가 조언해 줄 수 있는 체계로 전환되어야 정보의 유통과 소비가 원활하게 이루어질 것이다.

(3) 한국어 말뭉치 관리 센터의 설치

말뭉치는 구축만 해 놓았다고 해서 그 사명이 끝나는 것이 아니라, 계속해서 그것을 관리·운용·활용하여야 한다. 이러한 일을 담당할 기구는 개인이 할 수 있는 일이 아니다. 그렇다고 국립국어원에서 모두 관리하는 것에도 문제가 있다. 말뭉치는 그것을 필요로 하는 사람들이 관리하면서, 이를 국민들에게 보급, 활용할 수 있는 길을 계속 모색하고 또 이를

추진해 나가야 한다.

이 말뭉치 관리는 그 영역을 확대시켜야 한다. 즉 남한의 말뭉치뿐만 아니라 북한 및 해외 한국어 말뭉치까지도 수집, 정리하여야 한다. 또한 한국어 말뭉치나 대역 말뭉치만을 구축 관리하는 것이 아니라, 전 세계의 모든 언어와 문자에 대한 말뭉치 구축과 관리도 동시에 행하여야 할 것이다. 그것이 한국어 세계화 내지 한국의 세계화에 기여할 것이기 때문이다.

말뭉치들을 관리하고 활용하기 위해서는 가칭 ‘한국어 말뭉치 관리 센터’를 설치하고 이를 운영할 필요가 있다. 국립국어원 등의 국가 기관에 위임하면 지금까지 구축된 말뭉치들이 사장될 위험이 크다. 왜냐하면 세 종 계획의 수행과정에서 확인된 사실이지만, 그곳에서는 이들을 적극적으로 활용할 수 있는 인력이 부족하고 또 전담 부서가 없기 때문이다. 행정 중심의 체제에서는 능동적으로 일을 진행시킬 수가 없는 것이 우리의 현실임이 안타깝다.

이 기구에서는 ‘디지털 한글 박물관’, ‘한글 100대 문화유산’, ‘국립국어원 자료’ 등의 모든 자료를 관리하여야 할 것이며, 텍스트 자료, 음성자료, 이미지 자료 등을 모두 포괄하여야 할 것이다. 뿐만 아니라 정보통신부의 ‘한국 역사 정보 통합시스템’, 문화재청의 ‘국가기록유산’ 등을 각 도서관과 연계시켜 효율적으로 이용할 수 있는 방안이 마련되어야 할 것이다. 특히 ‘한국 역사 정보 통합시스템’과의 연계 구조를 염두에 두는 것이 좋을 것이다.

(4) 각종 말뭉치 활용 방안에 대한 연구

지금까지 구축된 각종 말뭉치를 어떻게 활용할 것인가를 능동적으로 연구하여 이에 적극적으로 대처하도록 하여야 한다. 그러기 위해서는 국립국어원에서 공모를 통해 그 연구가 이루어지도록 하고, 그 연구에 따라

말뭉치를 활용할 수 있는 방안을 마련할 수 있도록 하여야 할 것이다. 우선 국어 연구 및 국어사전 편찬에 이용하기 위한 작업이 이루어지도록 하여야 할 것이다. 예를 들면 다음과 같은 주제들이 있을 수 있다.

- ① 세종 말뭉치를 활용한 한국어 의성 의태어(상징어) 사전 편찬
- ② 세종 말뭉치를 활용한 한자어 사전 편찬
- ③ 세종 말뭉치를 활용한 외래어 사전 편찬
- ④ 세종 말뭉치를 활용한 세기별 국어사전 편찬
- ⑤ 세종 말뭉치를 활용한 어휘 역사 사전 편찬
- ⑥ 세종 말뭉치를 활용한 방언 사전 편찬
- ⑦ 세종 말뭉치를 활용한 문학 작품에 나타난 방언 어휘사전 편찬
- ⑧ 세종 말뭉치를 활용한 접사 사전 편찬
- ⑨ 세종 말뭉치를 활용한 신문어 사전 편찬
- ⑩ 세종 말뭉치를 활용한 구어 사전 편찬
- ⑪ 세종 말뭉치를 활용한 고소설 어휘 사전 편찬
- ⑫ 세종 말뭉치를 활용한 신소설 어휘 사전 편찬
- ⑬ 세종 말뭉치를 활용한 전문용어 사전 편찬
- ⑭ 세종 말뭉치를 활용한 신조어 사전 편찬
- ⑮ 세종 말뭉치를 활용한 종교어(불교, 기독교, 유교 등) 사전 편찬

말뭉치를 홈페이지에 올려놓으면 국어 정보화가 다 이루어진 것이 아니다. 국민들이 이것을 활용할 수 있도록 적극적으로 홍보하고 대응하여야 한다. 예컨대 국어 정보화 활용지침을 만들어 각급 학교나 직장에 배포하고 특강 등을 통해 알리는 일도 매우 중요할 것이다. 전술한 바와 같이 세종 계획 입안 당시에는 모든 자료(저작권 해결된 것)와 프로그램은 각급 학교에 배포하도록 되어 있었으나, 시행 과정에서 예산 문제로 시행되지 못하였다. 오히려 예산은 이러한 용도로 사용하는 것이 더 효율적이다.

(5) 새로운 말뭉치의 지속적인 구축

시대가 흐르면서 구축해야 할 말뭉치의 내용이 달라지므로, 새로운 말뭉치를 지속적으로 구축해야 할 것이다. 즉 세종 계획 이후의 말뭉치들을 계속 입력하여야 할 것이다.

말뭉치 구축은 일정한 기간 동안만 이루어지는 것이 아니라, 지속적으로 이루어져야 하기 때문에, 연차적인 계획을 세워서 말뭉치를 구축하여야 한다. 예컨대 2010년도에는 2008년도에 이루어진 내용들을 중심으로 말뭉치를 구축하는 방식 등이 그것이다.

이 새로운 말뭉치를 구축하기 위해서는 다음과 같은 점에 유의하여야 할 것이다.

① 지금까지 구축된 자료를 분석하여 부족한 말뭉치와 수정·보완하여야 할 말뭉치를 구축하여야 한다.

② 주식 말뭉치가 태부족이어서 이를 확대해서 구축하여야 한다.

③ 음성 말뭉치와 문자 말뭉치 및 동화상 말뭉치를 구축·수집·정리하여야 한다. 특히 국립국어원의 지역어 조사 관련 자료들도 음성 말뭉치와 텍스트 말뭉치로 활용할 수 있을 것이다.

④ 전 세계의 한국어 말뭉치 관련 기관과 연계하여야 한다. 예컨대 중국의 연변대학교에서는 2008년도부터 중국 조선어 말뭉치 구축 계획을 세워 상당한 양의 말뭉치를 구축하였다. 앞으로 4년간 국가 예산으로 더 많은 말뭉치를 구축할 예정인데, 이와 연계하여 말뭉치를 교환하는 방안도 마련하는 것이 좋을 것이다.

⑤ 북한의 말뭉치 구축 교류 방법도 고안해 내어야 할 것이다.

(6) 국어 정보 처리 프로그램에 대한 지속적인 개선

국어 정보를 처리하는 프로그램에 대한 지속적인 개선이 필요하다. 세

중 계획에서 개발된 프로그램들은 그것이 개발되던 시점의 한글 코드를 사용하였기 때문에, 프로그램에서 이용하는 한글 및 한자의 코드가 달라서 이용에 불편한 점이 많다. 예컨대 지능형 형태소 분석기는 ‘시스템 기본값’으로, 그리고 어느 것은 ‘유니코드’로 또 어느 프로그램은 2바이트 코드로 이용하게 되었거나, 어느 것은 DOS용 프로그램으로 만들어져 있어서 이용자들의 혼란을 야기하고 있다. 이를 유니코드 체제로 모두 변환시켜 활용할 수 있도록 지속적인 개선이 이루어져야 할 것이다.

한자 문제는 문자 코드 표준화 작업을 하면서 IRG의 CJK에서 작업을 하였지만 이에 대한 표준 석음이 달려 있지 않아서 국민들의 어문 생활에 큰 불편을 겪고 있으므로 이에 대한 개선도 필요한 일이다.

8. 맺음말

21세기 세종 계획은 이제 끝이 났지만, 국어 정보화는 이제부터 새로운 시작이라고 할 수 있다. 국어 정보화의 기틀은 마련되었지만, 이제부터 국어 정보화의 성숙 단계로 들어가고 있으며, 이것을 완성 단계로 끌어 올리는 일이 우리에게 남겨진 사명이다.

그러기 위해서는 세종 계획의 결과를 하나하나 검토하고 그 과정에서 미진한 부분들을 검토 보완하고 앞으로 국어 정보화를 위해 우리가 해야 할 일들을 제시하여 이를 적극적으로 시행하여야 한다. 그러나 그러한 과제는 개인이 하기 어려우므로 국가 기관에서 주도적으로 유도하여야 할 것이다. 특히 지금까지 세종 계획을 담당해 온 국립국어원은 세종 계획이 끝났다고 팔짱만 끼고 있어서는 안 되고, 앞에서 제시한 여러 가지 과제들을 실현하기 위해 분발하여야 할 것이다.

그러나 가장 중요한 일은 21세기 세종 계획을 전면적으로 검토하는 자리를 마련하고 여러 사람들의 의견을 들었음에도 불구하고, 세종 계획을 검토했다는 사실만으로 만족해 버리지 않아야 한다는 점이다. 과거에도

그렇듯이, 국어 정보화 중장기 발전 계획인 21세기 세종 계획에 대한 전면적인 검토가 한갓 통과의로 끝나고 후속 조치에 대한 노력이 없게 된다면, 그 자체가 또 다른 문제로 남게 될 것임을 명심해야 한다.