



# 남북한 언어정보처리 표준화에 대하여

최기호 / 상명대학교 사범대학장 · 한국어정보학회 회장

## 1. 머리말

1994년 7월 23일 남북한의 학자와 중국, 미국 등의 학자들이 중국 옌지에 모여서 ‘코리안 컴퓨터처리 국제학술대회(ICCKL)’를 열었다. 우리 한국어정보학회는 남북한이 분단된 이후 처음으로 북한의 정보기술(IT) 관련 학자들과 자리를 같이 하여 학술교류를 시작한 것이다. 그 이후 여러 가지 정보기술 관련 주제를 가지고 남북의 학자들이 다섯 차례의 학술회의를 지속적으로 가지면서 언어의 정보처리와 표준화 문제 등에 대하여 공동연구하고 토론을 하였다.

이들 학술대회에서는 다양한 주제가 다루어졌다. ‘정보통신 용어’를 비롯하여 국제표준규격(ISO)에서 ‘언어의 명칭’과 ‘자모 순서’, 그리고 ‘컴퓨터 자판’, ‘기계번역’, ‘음성인식’, ‘한손자판’, ‘말뭉치’, ‘한글의 세계화’, ‘글꼴’, ‘자국어 인터넷 찾기’ 각종 ‘표준화 문제’ 등 여러 가지 주제를 가지고 공동 연구하고 토론하였다.

여기에서 제일 먼저 대두되는 것이 말과 글, 즉 언어 문제였다. 인류의 문화는 크게 네 차례의 변혁이 있었는데, 거기에는 말과 글과 인쇄술과 뉴미디어가 핵으로 작용하였다는 것이다.

첫째는 사람이 말을 함으로써 수렵사회가 형성되어 사람보다 힘이 센 다른 동물을 제압할 수 있었다. 둘째는 글자를 만들어서 여러 가지 지식과 기술을 기록하고 축적하여 농경사회를 형성하였다, 셋째는 인쇄술의 발달로 기술정보를 다량으로 공유하여 산업사회를 만들었다. 그리고 넷째는 우리가 살고 있는 지식정보사회로서 컴퓨터를 중심으로 하는 뉴미디어가 변혁의 주인공이 되어 인류문화를 혁신시키고 있는 것이다.

우리 민족은 일찍이 한국어를 가지고 있었으며, 금속활자를 세계 최초로 개발하였고 음소문자인 훈민정음을 창제한 우수한 민족이다. 이러한 문화와 지식기반의 역사 속에서 한국의 근대산업화는 조금 뒤졌지만, 다행히 우리는 현대 지식기반 정보사회에서 뉴미디어의 연구개발로 세계에 앞장설 수 있는 위치에 있는 것이다.

그러나 안타까운 것은 이렇게 발달된 정보문화를 우리 민족이 모두 누리지 못한다는 것이다. 우리나라가 남북으로 갈라져 있으며, 특히 북한은 정보기술 분야가 매우 뒤져있고 정보기술 설비가 열악한 편이다.

만일 이런 상태에서 통일이 된다면, 남한의 정보사회와 북한의 미정보사회 간의 갈등이 크게 증폭될 것이고 이것은 통일을 지향하는 우리 민족의 장애에 큰 장애가 될 것이다. 조국의 통일을 위하여 남북의 6.15 공동선언을 바탕으로 정보기술을 서로 교류하고 표준화해야 하는 것이 무엇보다도 중요한 과제라고 생각한다.



## 2. 한글의 우수성과 언어정보처리 실제

미국의 과학자 제어드 다이아먼드는 권위있는 과학잡지 <디스커버>에 장문의 논문을 통하여 한글이 이 세계에서 가장 훌륭한 문자라고 극찬한 바 있다. 가령 영어는 '에이(A)' 라는 글자가 [ei, ə, a, e, ɪ, ai] 등 여러 가지로 발음되어 어떻게 읽어야 할지 모르기 때문에 발음기호를 꼭 써야 한다. 그러나 한글은 일자일음(一字一音), 일음일자(一音一字) 원칙이 지켜지는 진정한 표음문자로서 음성인식에서 뛰어나다는 것이다. 사실 한글 '아' 는 어떤 조건에서도 [a]로 발음되고, '야' 는 항상 같은 글꼴 '아' 로 표기한다.

이처럼 세종대왕이 창제한 한글은 이 지구상에서 가장 정보사회에 알맞은 문자로서 뉴미디어와 절묘한 조화를 이루며 정보사회를 이끌어가고 있다. 한글은 가장 발달한 음소문자로서 컴퓨터 구조에 잘 어울리며, 자판 배열에서 '좌자우모(左子右母)' 즉 왼쪽에는 자음을 배열하고 오른쪽에는 모음을 배열하여 자유롭게 문자를 조합하고 생성하는 아주 조직적이고, 체계적인 문자이다.

또한 손전화(hand phone)의 자판은 12개의 타건으로 짜여져야 하는데 중국의 수많은 한자나 100글자가 넘는 일본의 가다가나 문자로는 아주 불편하여 정보화에 큰 장애가 되고 있다. 그러나 한글은 모음을 천지인(· 一 1) 세 글자의 조합으로 모든 모음을 처리하고 나머지 자음을 배열하기 때문에, 삼성과 LG 등의 손전화는 세계 일류 상품으로 우리나라 경제의 주역이 된 것이다.

그리고 우리나라 청소년들이 문자메시지를 세계에서 가장 자유롭게 사용하고 있는데 이것은 한글이 우수하기 때문이다. 일본이 애니메이션 같은 분야에서는 뛰어나지만 문자가 들어가는 정보 분야에서는 우리나라에 뒤지는 것은 문자의 우열 때문으로 보고 있다.

한편 한글 입력 글자판의 구조는 영문자를 중심으로 개발된 것이라 한글을 쓰는 우리로서는 불편함을 느낄 때가 많다. 한글을 정보화하려면 자모를 자판에 입력하는 것이 중요하다. 한글은 초성, 중성, 종성의 3성음으로 구성되어 모아쓰기를 해야 하므로 로마자와는 달리 타건 입력 방식에는 여러 가지가 있다.

현행 한국 한국자판의 비효율성은 이미 밝혀져 있다. 이와 같은 비효율성을 지닌 현행 표준 규격을 그대로 유지할 것인가, 아니면 새로운 대안을 개발하여 보급할 것인가가 문제이다. 이미 미국, 독일과 같은 정보사회의 선진국에서는 과거의 표준규격을 그대로 살리되, 새 표준 개발에 적극적으로 대응하여 이미 복수 표준을 채택하고 있다. 그 이유는 새로운 기술을 창조하기 위해서 표준 기술에 대한 적극적 정책이 필요하기 때문이다.

한편 우리는 남북한이 공통의 문자체계인 한글을 사용하면서도 서로 다른 규격기술을 사용하여 왔다. 가령 한글의 가나다 순서가 남북한이 다르므로 표준화와 정보의 교류에 문제가 된다. 컴퓨터의 자판의 경우를 말하면 다행히 1994년부터 남북한 학자들의 협력, 협동 연구를 통하여 남북공동안 개발에 힘썼고 그 결과, 남북공용표준시안을 만들었고, 보완연구를 수행해 왔다.

최종안은 남북학술회의를 통하여 종합적 평가 검토가 이루어졌고, 남북 양 정부에 대한 최종 권고안을 마련하였다. 그리고 국제표준기구(ISO)에 남북공용 한글표준자판으로 등록하는 노력을 하였다.

국제정음기호(OPA) 자판은 IPA 체계보다는 훨씬 간편하고 익히기 쉬워서 우리나라 뿐 아니라 다른 나라 사람들도 편리하게 외국음을 표기할 수 있으며, 기호의 수를 52개 이내로 하여 현행 자판상에서 직접 타자할 수 있다

# 표준화 논단

록 하였다는 장점이 있다. 아직은 표준안이 없으므로 다음 2가지 방안을 생각할 수 있다.

그리고 한글 입력 전자기 타건은 현재 표준안이 상정되어 있지만, 새로운 표준안은 다음과 같이 고려할 수 있다. 현재의 방식과 두드러진 차이는 기호 이외에도 그 숫자 방식으로 입력시킨다는 점이다. 배치구도는 훈민정음 창제 원리 순으로 나열한다는 것이다.

현재 한글 부호는 2 바이트를 사용하고 있으며 로마자의 ASCII는 7 비트로 구성되어 있다. 그러나 한글 7 비트는 KS C 5701 부속서 4에 규정하고 있으나 일대일 대응이 되지 못하여 한글의 코드 변환이 어렵게 되어있다. 일본이나 한자를 본뜬 듯한 구조이다. 완성형이든 조합형이든 국제 규격을 만족하지 못하고 있다. 그 외 한글의 내부 처리코드가 회사마다 또는 기종마다 틀려서 한글을 다른 프로그램을 통해 읽고 수정하는 작업이 쉽지 않다. 따라서 통일된 내부처리코드가 필요하다. 한글 시스템에서 코드 구성의 주목표는 입력변환, 내부정렬, 정보교환, 표시재현시 일관작업이 가능할 수 있도록 국제정음기호 코드를 구성하는 것이며, 또한 그 국제정음기호 코드가 인터넷에 두루 통용될 수 있는 7비트 구조를 우선 고려해야 한다.

ISO 10646/유니코드에는, 한글과 비슷하게 모아 쓰기를 하는 문자가 14개 들어 있는데, 한글만 유일하게 조합형과 완성형 두 가지 모두 들어 있을 뿐, 나머지 13 문자는 모두 조합형으로만 들어가 있다. 아랍 글자 극히 일부가 완성형으로 되어 있지만, 보조 부호계이다. 그렇다면, KS C 5601 완성형(2,350 소리마디)과 상용 조합형은 어떻게 될 것인가? 얼마 동안은 완성형과 상용 조합형을 계속 많이 쓰겠지만, 시간이 지남에 따라서 ISO 10646/유니코드를 점점 많이 쓰게 되다가, 결국에는 ISO 10646/유니코드를 주로 쓰게 될 것이다.

### 3. 남북한 언어정보처리 표준화의 중요성

국제 표준 규격에서 남북한의 협조가 매우 필요하다. 1994년 여름에 남북 학자들이 중국 옌지에서 만난 뒤로 국제 표준 분야에서 한글과 관련된 일로는, 1994-95년에 걸쳐서 진행중인 국제 표준 문서 ISO 10646-1을 고치는 작업을 하였다. 여기서 두 가지 점이 문제가 된다.

첫째, ISO 10646에서 결국에는 6,656 소리마디를 빼고, Oxac-d7a3에 11,172 소리마디를 넣는 작업을 하면서, 현재 모두 남한의 가나다순을 따르고 있다. 남북한이 11,172 소리마디의 순서에 대해서 협의하여 정하기가 현실적으로 아주 어렵다는 점은 알고 있지만, 남북한 조정을 할 수 없다는 것이다.

둘째는 새로 들어갈 11,172 소리마디의 이름을 로마자로 나타낼 때 현재 쓴 원칙은, 기술보고서(Technical Report)에 나온 안에 따르고 있다. 그런데, 일단 남한 안을 따르고 있다.

1992년에 국제 표준문서 ISO 10646-1을 만들 때 다음과 같은 문제점이 있었다. ISO 10646-KSC5601, -KSC 5657, -KSC 0000 등에 6,656 소리마디를 넣을 때, 모두 남한의 가나다순을 따랐다. 그 때 현실적으로 남북한이 가나다순에 대해서 협의할 수 있는 가능성이 거의 없었다.

조합형 238 글자의 이름, 6,656 소리마디의 이름, 그밖에 한글이 나오는 여러 부호값에 대한 이름을 로마자로 나타낼 때 쓴 원칙을 1992년 6월 16-17일에 파리에서 남북한이 합의한 안에 따른 것이다. 그 때 로마자로 적는 안



은 아직 CD상태도 아니었기 때문에 그것을 따르는 것이 바람직하지 않다는 의견이 있었다. 그 뒤로 그 합의한은 결국 ISO 표준에까지 이르지 못하고, 현재는 기술 보고서에 정리되어 있다.

따라서, 10646 안에서조차 한글을 로마자로 적는 법이 두 가지가 섞여 있어서 통일성이 없다. 10646의 한글로마자 표기법은 현재 기술 보고서를 기준으로 하면, 1992년에는 북한의 안을 따랐고, 1995년 현재 작업 중에는 남한 안을 따르고 있다. 물론 홀소리 글자에 대한 표기법은 남북한이 같으므로, 실제 문제가 되는 것은 닿소리 글자이다.

한편 국제 등록부(International Reigster)에 등록된 KSC 5601에 나오는 한글 2,350 소리마디는 남한의 가나다순에 따르고 있다. 일반적으로 국제 등록부에는 같은 문자(한글)에 대해서도 여러 부호계를 등록할 수 있기는 하지만, 가나다순에 대해서 남북한 조정이 된다면 아주 바람직하다고 본다. 앞으로는, 2022를 따르면서 한글을 제대로 지원하는 한글 부호계를 남북이 합의하여 국제 등록부에 등록할 수 있다면 아주 바람직하다고 본다. 글자 이름은 안 들어갈 것이므로, 로마자 표기법은 문제되지 않으며, 한글 가나다순이 문제가 된다고 본다.

세계 여러 나라의 모아쓰기 글자를 지원하는 전반적인 추세가 조합형이라는 사실을 감안할 때, 바람직한 한글 부호계 운용 방안은 다음과 같이 요약할 수 있다. 10646 한글 부호계를 아무런 변형 없이 현재 표준 그대로 쓸 경우만을 생각해서 나온 것이다. 보기를 들어, 각정 UTF(UCS transformationformat) 방식이나 EUC 방식은 고려하지 않았다.

10646을 처음 수용할 때 “잠깐 동안” ISO 10646-KSC 5601 완성형 2,350 소리마디만 쓰는 방안을 쓸 수도 있지만, 되도록 처음부터 조합형과 ISO 10646-KSC 5601을 같이 지원하는 방안을 채택하는 것이 가장 바람직하다고 본다.

ISC 10646-KSC 5601의 완성형 2,350 소리마디도 점차 버리면서, 조합형만 쓰는 방안으로 나아가야 할 것이다. ISO 10646-KSC 5657 완성형 1,930 소리마디나, ISO 10646-KSC 0000 2,376 소리마디는 처음부터 아예 쓰지 않는 것이 바람직하다고 본다.

#### 4. 남북한 언어정보처리 표준화의 방향

우리 한국어정보학회에서는 1994년에 남북의 학자들이 함께 모여서 국어정보 교류에 관한 학술대회를 했고 실질적인 협력방안을 모색해 왔다. 그 후 지속적으로 남북한 학술회의와 정보교류를 통하여 컴퓨터 자판과 코드 문제, 음성인식 기술문제, 국제표준 규격문제, 정보기술 용어통일 문제, 글꼴 문제 등 여러 분야의 연구와 협력문제를 협의하여 왔다.

그리하여 그 결실의 하나로 ISO2382기준 <정보기술 표준용어사전>을 우리 한국어정보학회와 중국조선어신식학회와 북한의 조선교육성 프로그램교육센터 공동으로 한, 영, 조, 중, 일판(975쪽)과 조, 영, 한, 중, 일판(973쪽) 두 권의 <정보기술용어사전>을 출간하였다.

남북한 화해무드와 남북 경제협력이 강조되는 상황에서 남북한이 공존하며 발전할 수 있는 전략 마련의 일환으

# 표준화 논단

로 남북한 정보통신 협력의 현황을 진단할 필요성이 있다. 이를 위해 북한의 과학기술정책, 북한의 정보통신정책, 북한의 신사고론, 북한의 정보통신 현황 등을 잘 알아야 할 것이다.

구체적인 사업별 과제로는 통신교류의 협력, 정보통신 인력양상 남북한 협력 방안, 남북한 소프트웨어 협력, 남북한 정보통신합작회사의 설립, 음성인식기술 공동개발, 남북한 정보통신용어의 통일, 남북한 통일 도메인 구성 등을 논의해야 할 것이다.

북한의 주요 정보화기관으로는 평양 프로그램센터, 조선컴퓨터센터, 김일성 종합대학, 북한 과학아카데미, 북한 국제통신센터, 중앙 과학기술 정보연구소, 평성 이과대학, 김책 공업종합대학, 평양 전자계산기 단과대학, 평양 정보과학 기술대학 등이 있다.

21세기 정보화 시대에 무엇보다 중요한 것은 남북한 정보화의 균형적 발전이다. 통일의 시대를 맞이하여, 우리의 정보화는 남북이 함께 하는 정보화, 민족의 정보화가 되어야 하며 균형적 발전을 해야 한다. 그러기에 정보기술의 표준화는 무엇보다 선행되어야 할 것이다. 



## < 약 력 >

- 1963 ~ 1971 연세대학교 국문학과 졸업(문학사)
- 1979 ~ 1985 연세대학교 대학원 졸업(문학박사)
- 1980 ~ 2002 상명대학교 조교수, 부교수, 교수(현)
- 1990 ~ 2002 세종대왕기념사업회 상무이사(현)
- 1989 ~ 1992 방송심의위원(역임)
- 1988 ~ 1999 교육부 국어교과서 편찬심의위원
- 1992 ~ 2000 한국몽골학회회장
- 1993 ~ 2002 KBS방송언어자문위원(현)
- 1998 ~ 2002 한글학회감사(현)