

## 남북 국어정보기술 표준안개발의 과제

특집  
02

### 목 차

1. 서론
2. 한글정보처리 기본기술의 표준화 과제
3. 한글정보처리 요소기술의 표준화 과제
4. 한글정보처리 응용기술의 표준화 과제
5. 한글 컴퓨터 처리기술의 남북 현황
6. 결론

정의성  
(선문대학교)

### 1. 서론

이 논문은 현안으로 떠오르고 있는 남북의 국어정보처리 표준기술에 대한 과제를 정리한 것이다. 같은 언어권임에도 불구하고 서로 다른 처리규격 및 기술을 유지하고 있는 남북이 앞으로 어떤 자세로 공통표준기술을 갖추어 나가야 할 것인가를 제안하고 있다.

훈민정음이 한자를 대신하여 국어를 적는 혁명적 발명이었음은 사실이다. 오늘날 디지털 혁명이라 일컬어지는 컴퓨터 기술에 힘입어 마땅히 한글의 컴퓨터 처리기술은 다른 언어권에 비해 혁명적이라 말할 수 있어야 한다. 그러나 상황은 그렇지 못하다.

80년대 초, 한글문자의 표시방법은 지극히 고급기술에 속하는 것이었다. 「우리는 세종 때문에 망한다.」라는 속설도 이때 통했던 유행어이다.

음성인식과 합성어, 기계번역 기술을 융합하여 컴퓨터 통역시스템을 개발한다는 고난도의 기술

연구가 진행중인 반면, 아직도 한글문자의 디지털화, 즉 한글코드를 어떻게 해야 하는가를 놓고 그것이 연구대상이 되어있다는 사실 하나만으로도 우리는 근본적으로 그 이유에 대하여 설명을 해야 할 때이다. 도대체 무엇이 문제인가.

본디 「규격화」란 어느 대상을 규격이라는 틀에 넣어 그 규격의 틀에 맞도록 대상을 변형시키는 것이 일반적인 이해였다.

규격이란 규제하기 위한 「틀」이므로 규격에서 정한 차례, 내용대로 쓰도록 강제 강요하는 일로만 알아 왔다.

전형적인 예가 「한글문자의 코드」 문제에서 일어났고 그 결과 코드를 위한 코드의 논쟁이 지속되고 새로운 규격의 제안이 있을 때마다 받아들인 결과, 한글 코드는 「누더기」화 되고 말았다.

이 문제는 마침 「남북한이 한글정보의 컴퓨터처리에 대한 표준화 규격 및 기술」의 단일화 방안을 놓고 1996년부터 본격적으로 남북의 과학기술자들이 발표, 토론한바 있는 남북한의 표준 「코드 규격」과 「자판 규격」 등에서도 불

수 있다.

「국제규격에 따른다」 또 「한글에 가장 알맞은 규격이다」 하고 주장되어 온 여러 규격(안)이 참으로 그러한가를 밝히는 연구가 필요한 시점이다. 이제라도 현행 기술 규격을 합리적으로 분석 검토하여 보다 나은 남북 통일규격으로서 「한글 처리 표준안」을 개발하기 위한 기틀을 마련해야 한다.

이 연구의 방법론은 남북한 규격 및 기술의 상대적 평가를 필요로 하며 국제 규격의 실체와 선진 기술의 실체에 견주어 남북한의 것들이 어떤 위치에 있는가를 판단하는 객관적 자료를 제공할 필요가 있다.

제 1절에서는 한글의 컴퓨터 처리를 위한 기본 기술, 제 2절에서는 요소기술, 제 3 절에서 응용 기술을 제시하고 현재 선진기술 연구동향과 견주어 어떤 표준화 연구가 시도되어야 하는가를 밝혀 앞으로 해야 할 남북한 공동연구 과제를 도출하는 자료로 삼고자 한다.

제 4절에서는 1996년 이후 남북의 학자간에 연구, 논의되었던 표준안에서 코드 문제와 컴퓨터 자판에 대해서 집중적으로 논하고 결론으로서 앞으로의 우리의 과제에 대해서 논한다.

## 2. 한글정보처리의 기본기술 표준화 과제

우리의 경우, ‘한글의 기계화’로 일컬어지는 기술 가운데에서 대표적인 것은 1950년대 영문타이프 라이터의 기계적 형태(메카니즘)를 그대로 한 한글문자용의 타이프 라이터의 발명이다.

3별식, 4별식, 5별식 등 소위 한글문자의 시각표현을 활자체에 가깝게 하는 글자포맷(character format)을 유지하기 위해서는 한글문자의 6가지 형태, 즉 가, 고, 기, 각, 곡, 꿩과 같은 글자포맷에서 초성을 몇 종류의 크기와 꼴로, 중성을 몇 종류의 크기와 꼴로, 종성은 몇 종류의 크기와 꼴을 들것인가로 기술기준으로 두고 여러 가지의 한글 타이프 라이터가 출현했다. 이 배경은 당시

의 기계적인 기구로서는 일차원적 한글 자모를 입력하여 한 음절에 해당하는 한글 한 글자를 활자와 같이 모양 좋게 인자하는 장치는 불가능했기 때문에 생겨난 기술이었다.

그 후 정부는 한글 타이프 라이터의 자모배열에 대한 표준 규격의 필요성에 따라 제정한 것이 4별식으로 국무총리 훈령 제 80호(1976)에 의해서 국가규격으로 채택되어 오늘날 남한 컴퓨터 자판의 밑바탕이 되었다.

1980년 이후 마이크로 프로세서의 발명과 함께 퍼스널 컴퓨터가 등장하여 1차원적 입력 자모를 2차원적으로 합성하여 문자를 생성하는 자동화 기술이 개발되었으나 당시로서는 한글문자의 시각표현에 맞는 출력기술이 한글처리를 위한 컴퓨터의 첨단기술의 하나였다.

한글을 컴퓨터로 처리한다 했을 때, 부수되는 기본기술은 한글 자모의 입력 장치, 한글문자의 출력장치, 그리고 1차원적 자모에서 2차원적 글자구조를 생성, 표시 또는 인쇄하는 소위 한글 오토마타(기술적 용어로서는 잘못임)가 발명되었으나 오늘날과 같은 적은 내부 상태 수로 한글을 합성, 생성하는 기술은 비교적 최근의 일이다. 지금은 한글 오토마타의 프로그램 소스가 컴퓨터 잡지에서 공개 개제되고 있는 것도 모두 그때 이후이다. 적은 오토마타가 누구에 의해서 제안되었는지 밝혀 볼일이다.

그 후 한글 워드프로세서의 등장은 단순히 한글 문자의 입력과 출력에만 신경을 써오던 문자 처리(character processing)기술이 워드처리(word processing)기술로 한 단계 올라갔다. 이것은 단순히 문자 수준의 입출력 기술이 단어 또는 문장의 수준에 이르는 편집 처리라는 고급기능의 기술에 이르렀다고 말할 수 있다.

컴퓨터에 의한 문자의 입출력, 즉 문자 처리기술에서 가장 필수적인 것은 문자의 입력방식, 문자의 컴퓨터 내부표현방식, 문자의 출력방식이다. 이 세가지 기능에 관련하는 기술이 한글의 컴

퓨터처리에 따른 기본기술이다.

한글문자의 입력방식에서 가장 보편적이며 안정된 방식으로 키보드 입력기술이 있고, 한글문자의 컴퓨터 내부 표현 방식은 한글문자의 코드 기술을 말하며, 한글문자의 출력방식이란 한글문자의 표시기술을 말한다.

지금까지 우리는 이와 같은 기술의 실현방법에만 몰두해 왔을 뿐 어떤 방법과 기술이 보다 한글을 위해 효율적이며 경제적인가 하는 소위 알고리즘적 접근법의 연구는 등한시 해왔다.

한글 입력 키보드의 한글 자모 배열이 과연 컴퓨터의 기술발전과 우리의 경제력과 기술력에 맞는 국민의 지적 생산도구인가라는 측면에서 기술 검토, 분석이 되어 본 적이 없다.

한글코드의 경우도 어떤 코드가 한글의 지적 처리를 위해 필요한가, 어떤 코드체계가 국제규격에 맞고 경제적이며 효율적인가 객관적 연구와 기술개발이 되지 못하고 학설, 기호에 따른 논쟁만 되풀이 되어 이것, 저것 모두 받아들이다 보니 세계에서 가장 코드체계가 복잡한 문자가 되고 말았다. 가장 합리적이며 과학적이라는 한글 문자체계가 컴퓨터의 기술을 추구하면서 세계에서 가장 처리하기 힘들고 돈이 많이 드는 문자체계로 전락하고 말았다.

ISO-10646의 규격으로 만든 한글 영역을 두고 코드제정에 참여했던 연구자가 전체 영역 중 1/4 가깝게 한글 한자를 위한 영역을 차지하게 되어 무척 다행스럽다고 언급하던 것은 코드체계와 실현에 얼마나 무지한가를 보인 단면이다.

도트프린터(dot printer)방식이 발명되어 한자 인쇄가 자연스럽게 해결되자 한글 문자의 인쇄 방식도 덩달아 해결을 보았다. 또 한글 글자꼴을 놓고 오랫동안 명조체니 고딕체니 하더니 누군가의 제안에 의하여 서체의 이름이 우리말 화되었다.

그러나 한글문자의 표시 기술은 한자와는 다른 표시 알고리즘으로 처리 가능한 기술개발이

기대되는 영역으로 남아 있기도 하다.

한글 문자는 알파벳, 한자와는 다른 데이터 구조를 가지고 있다. 보다 정보처리적 측면에서의 요구에 따른 한글문자를 위한 기본기술의 개발이 필요하다.

한글문자 데이터구조 또는 정보구조에 따른 입출력기술과 컴퓨터 외·내부의 효율적 표현방법에 대한 연구를 다시 원점부터 시작할 필요가 있다.

그렇게 한다면 가장 합리적이고 한글처리를 위한 기본기술의 발명의 가능성이 크며 그 평가는 반드시 정보과학적 측면에서 이루어져야 한다.

### 3. 한글정보처리의 요소기술 표준화 과제

한글을 컴퓨터 처리하는 기술 가운데에서 가장 보편화되고 있는 것이 워드프로세서(word processor)의 기술이다.

한글을 컴퓨터 처리를 위한 기본기술의 평가가 있고 없음에 관계없이 문자 그대로 단어를 처리단위로 하는 워드프로세서가 개발되어 쓰이고 있다.

「한글 워드프로세서」의 기능을 갖추는 소프트웨어의 구성은 다음과 같다.



입력한 문장을 교정하거나 서식(format)을 맞추도록 하는 소프트웨어가 문서편집 소프트웨어이다. 편집 기능에는 가장 기본적인 것으로 삽입, 삭제, 정정, 복사, 이동 등이 있으나 문서작성의 능력은 이와 같은 다양한 기능들을 손쉽게 배우고 쓰기 편해야 한다.

한글처리와 관련한 기술 사항으로서는 컴퓨터 내부처리에서 나란히 세우기(sorting), 찾기(searching)등의 문자열 처리에서 어떤 알고리즘이 가장 한글처리에서 알맞은 것인가를 가려내는 일이다.

알고리즘에 필요한 것은 처리하고자 하는 대상의 데이터 구조와 연산(operation)이다. 한글의 데이터 구조를 어떻게 지정하고 거기에 부수되는 연산은 무엇인가를 형식화하는 것이야말로 한글처리의 기본기술과 요소기술을 확립하는 지름길이다.

이런 저런 코드방식은 글자가 나오지 않고, 글자가 나오고 하는 투의 논쟁은 이제 삼가야 한다. 한글 컴퓨터 처리를 위한 새로운 제안은 반드시 처리 목적과 그 목적에 맞는 알고리즘, 그리고 그 알고리즘의 효율을 기존의 것과 비교하여 그 근거를 명시해야 한다.

예를 들면 한글 패턴매치 알고리즘은 그것에 따른 데이터구조와 연산 방법에서 영어의 패턴매치 또는 한자의 패턴매치 알고리즘과 크게 달리할 수 있다. 어떻게 하면 그와 같은 이론과 기술을 개발할 수 있는가를 고려하면 한글코드체계에서 어떤 정보를 코드 값으로 정해야 할 것인가를 자명한 일이다.

이와 같이 한글문자 정보처리에서 핵심이 되는 시스템으로 워드프로세서의 역할은 크다. 워드프로세서의 기술 비교에서 중요한 것은 워드프로세서의 각 기능에서 본 효율성이다.

알고리즘의 효율성을 평가하는 방법은 이미 확립되어 있어 제안하는 어떤 이론이나 기술도 객관적 평가가 가능하므로 지금부터의 한글 정

보처리 관련 규격 및 기술은 효율성과 경제성의 면에서 평가하는 연구가 주축을 이루어야 한다.

한글 정보처리의 요소기술로서 최적화 평가 대상이 되는 규격과 기술 관련 기능에는 다음과 같은 것들이 있다.

- (1) 입력기능 관련 규격 및 기술
- (2) 내부 기억기능 관련 규격 및 기술
- (3) 표시 기능 관련 규격 및 기술
- (4) 편집 교정 분류 추출 데이터처리 기능 관련 규격 및 기술
- (5) 인쇄 기능 관련 규격 및 기술
- (6) 통신 기능 관련 규격 및 기호

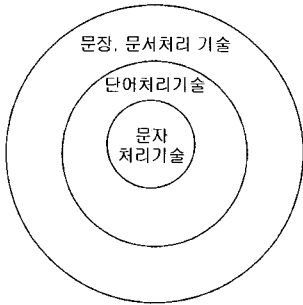
위에서 적은 기능의 최적화 평가는 물론 컴퓨터 처리상의 효율성도 중요한 문제이다. 휴먼 인터페이스(human interface)의 측면에서 쓰기 쉽고 배우기 쉬운 척도의 평가도 지금부터 중요한 과제이다.

#### 4. 한글정보처리의 응용기술 표준화 과제

한글 정보처리 영역을 단계적으로 분류한다면 문자단계, 단어단계, 문서문장단계로 나눌 수 있다. 이 절에서 논의하고자 하는 단계는 (그림 1)과 같이 수준별로 포함관계에 있다. 이와 같은 한글정보처리 기술을 이용하여 응용시스템을 개발할 수 있으나 이 분야는 기계번역, 음성인식 합성 시스템, 문자인식 시스템 등 한글의 언어정보를 응용한 보다 지적인 시스템 개발을 목표로 한다.

언어학과 컴퓨터 과학의 융합분야를 계산언어학(computational linguistics)이라 부른다. 이 분야에서는 언어와 관련하는 정보 및 지식을 컴퓨터로 처리하는 이론을 정립하는 것을 목적으로 하며, 이론이 정밀하면 할수록 컴퓨터 처리 효율은 좋아진다.

인간의 언어구조는 크게 음성 음운체계, 단어의 형태를 가리는 형태소론, 형태소가 모여 문장



(그림 1) 한글정보처리의 수준

을 이루는데 이 구조를 다루는 통사론, 문의 의미를 따지는 의미론, 또 문맥, 비유, 환유, 관용표현의 현상을 다루는 담화론으로 크게 나뉜다.

이와 같은 영역의 정보와 지식을 컴퓨터 처리하는 기술을 각각 음성인식 합성기술, 문외 형태소 해석기술, 문외 동사 해석기술, 의미해석기술, 담화해석기술이라 한다.

또 위의 기술을 이용하여 개발되는 시스템을 음성인식 합성 시스템, 문장해석시스템, 문장이해 시스템으로 크게 나뉘고 인간과 컴퓨터와의 대화를 보다 인간화 해가는 기술혁신이 제일 큰 목표로 되어 있다. 또 이것들을 복합기술로 하여 기계번역 시스템, 자동통역 시스템이 개발되고 있다.

국어의 컴퓨터 처리기술의 궁극적 목표가 이곳에 있고 여기에 부수되는 기술도 근본적으로는 기본기술과 요소기술 이외에 국어의 문법이론 체계도 크게 작용하게 된다.

예를 들어 남북은 국어의 문법체계가 다르다고 볼 수 있으나 공통분모적 부분의 정리는 어렵지 않다고 본다.

이와 같은 공통분모적 정보와 지식을 만들어 내는 작업 또한 중요한 과제이다.

## 5. 한글 컴퓨터 처리기술의 남북 현황

북한의 한글 정보처리 기술수준에 대한 현황을 설명하기란 쉬운 일이 아니다. 특히 남북간에

기술교류 및 인적 교류가 드문 상황인 만큼 더욱 그러하다. 그러나 94년과 95년 2차에 걸친 우리말 컴퓨터처리 국제학술대회에서 발표된 논문 가운데, 특히 94년도 발표 논문에서 북한의 컴퓨터 과학 기술인의 국어 정보처리 기술과 관련된 연구동향을 엿볼 수 있다. 또 문제 제기형의 논문에서 북한의 컴퓨터 과학기술이 해결하고자 하는 문제점과 해결방법의 제시가 남북한이 공통으로 알고 있는 문제점임을 확인하는 데는 그다지 어렵지 않다.

조선 컴퓨터센터 이수락 처장의 부록 논문 「전자계산기에 의한 조선어 처리기술에서 해결해야 할 과학기술적 문제와 그의 표준화에 대하여」 [1]에서는 조선글의 컴퓨터처리 기술의 문제점이 잘 정리되어 있다. 지적하고 있는 대로 남북한 그 어느 쪽도 만족할만한 규격과 기술을 확립시키지 못하고 있다.

가령 논문의 지적대로 한글의 컴퓨터처리를 위한 표준 규격, 기술이라 하면 그 목표는 한글을 컴퓨터에서 이용하여 처리하고자 하는 목적에 맞게 가장 합리적이어야 하며, 경제성, 생산성에서 최고의 것이 최선의 것이다라는 평가지표에 따라 국민적 합의하에서의 규격과 기술이어야 함은 물론이다.

도구를 발명한 사람들에 의해서 그것을 어떻게 쓸 것이냐 라는 대답도 나와야 한다. 그렇지 못하면 한글과 컴퓨터는 항상 표면적, 피상적 처리목표를 위해서만 존재할 뿐 그것을 이용한 응용 소프트웨어의 개발은 결코 풀지 못한다.

### 5.1 새로운 한글코드의 전망

문자를 이진 코드화(binary coding) 한다는 것은 문자가 가지고 있는 외적, 내적 표현을 컴퓨터 처리할 수 있도록 컴퓨터의 데이터 처리 단위인 비트열, 즉 바이트로 표현하는 일이다. 흔히 어떤 데이터가 가지는 외적, 내적 표현을 데이터 구조(data structure) 또는 정보구조(information

structure)라고도 한다. 따라서 코드화하고자 하는 문자셋트를 고려하여 그것의 구조에서 무엇을 중심으로 코드화 할 것인가가 중요하다.

알파벳트 로마자의 경우를 보자. 가장 작은 단위가 알파벳트, 즉 문자(character)이다. 이 문자가 모여 문자열(string)을 이룬다. 알파벳트의 컴퓨터처리에서 가장 작은 단위는 문자형이고 다음이 문자열의 형이다. 따라서 더 이상 나누어지지 않는 단위, 즉 문자(character)를 코드화 단위로 하여 알파벳트 26자를 각각 대문자, 소문자로 구별하여야 하는 문자구별 정보와 알파벳트의 순서라는 순서정보(ordered information)를 가지고 국제기준판, ASCII등의 코드계가 되어 있다. 한편 한자는 그것의 데이터 구조에서 보면 획과 변을 최소단위로 하여 구성되어 있다. 따라서 어떤 한자라도 데이터 구조에 따라 정의를 할 수 있으나 컴퓨터에서의 한자처리는 한자를 구성하는 최소단위인 획과 변보다는 표어성을 지닌 한글자의 형상과 음을 최소 처리단위로 보고 글자 한자 한자를 단위로 코드화되어 있다. 소위 한자권이라 하는 중국, 일본, 한국등이 한자코드계를 표준 규격으로 삼고 있으나 각각의 음에 따라 순서를 정해 배열한 한자에 대응한 코드값을 가지고 있다. 최근의 ISO-10646에서는 중국, 일본, 한국이 공통으로 쓰는 한자를 음가의 차이에서 오는 코드값의 다름을 없애기 위해 한자의 변은 중심으로 한 강희자전순으로 CJK 공통한자 20902 문자 표현이 가능하도록 새로운 한자 코드체계를 만들었다.

또, 한글문자의 경우를 보자.

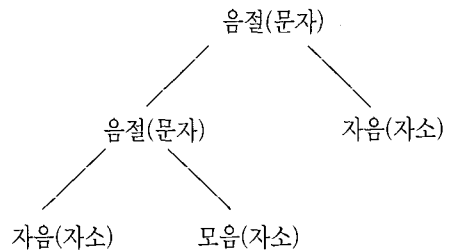
한글문자의 데이터 구조는 세계의 어떤 문자체계의 구조와는 사뭇 다르다. 언어학적으로 보면 한글문자는 표음문자(phonograph)이면서 한자와 같이 표어문자(logograph)이다. 표어문자란 한글자가 한 낱말의 의미를 갖는 문자를 말하며 한자(漢字)가 그 전형적인 예이다.

한글문자는 표음문자로서 유소문자와 음절문

자이다.

이와 같이 한글문자는 복합적 구조를 가진다. 따라서 한글문자의 최소 단위는 음소에 해당하는 자모이며 자모는 다시 음절에 해당하는 문자를 구성하고 문자를 다시 문자열을 이루는 계층적 구조를 가진다.

이와 같은 개념을 데이터 구조인 트리(tree)로 나타내면 아래 그림과 같다.



(그림 2) 한글문자의 데이터 구조 트리

한글 코드체계가 단순화되지 못하고 근본부터를 그르친 이유가 바로 이와 같은 데이터 구조에 따른 코드화를 하지 못하고 맵질식(ad-hoc) 해결법에 의존해 왔기 때문이다.

## 5.2 한글정보의 지적처리를 위한 만능코드에 대한 기대

한글정보처리 기술이란 한글로 표현하는 문서·문장·회화 구조에서 어떤 문제의 처리를 행하고자 할 때, 문제해결의 필요 충분조건을 만족하는 알고리즘이라 정의할 수 있다. 따라서 알고리즘은 보편적 정의에 따르면 처리대상의 데이터 구조와 그것에 필요한 제어연산자(control operater)로 구성되므로 처리하고자 하는 문제의 대상인 한글 데이터의 구조와 필요한 연산자를 설계하면 된다.

컴퓨터 처리가 효율적으로 가능한 대상은 수학적 구조를 가진 것이어야 한다. 수학적 구조란 대상을 수학의 개념으로 정의한 집합을 말하며

함수표현 또는 관계표현으로 하는 것이 일반적이며 추상수학의 범주이다.

수학적 구조에는 대상이라는 집합을 구성하는 요소와 요소 사이에 서로의 맺음 관계를 나타내는 대수구조(algebraic structure), 요소와 요소 사이의 차례를 나타내는 순서구조(ordered structure), 요소와 요소 사이에 서로 닮음의 관계를 나타내는 위상구조(topological structure), 또 구조끼리의 구조를 나타내는 메타구조(meta structure)가 있다.

컴퓨터로 처리하고자 하는 대상에서 해결하고자 하는 문제의 개념을 알고리즘화하는 것은 이와 같은 추상적인 데이터 구조를 어떻게 추출하느냐에 달려있다. 또 그와 같은 추상적 구조는 컴퓨터 내부에서 디지털화 표현되고 그 표현 위에서 필요한 연산 조작이 이루어지며 이 과정이 정보처리이다.

지적(intelligent)이란 정보처리의 대상에 대한 과제의 수준을 말하는 것으로 만일 한글정보에서 지적처리라 하면 음성인식과 합성, 또는 음성 이해, 기계번역 또는 기계통역, 문자인식, 정보검색의 한글화 등, 한글을 일상생활의 언어도구로 사용하는 사람들에게는 아무나 자연스런 언어지식이다. 이 언어지식을 컴퓨터 처리한다는 가정은 인간의 지식을 처리하는 것이 되므로 지적 정보처리라 하는 것이다.

결국 한글정보의 지적처리를 고려한다면 이것에 관련하는 모든 데이터는 디지털화되고 컴퓨터 처리단위로서 코드화된다. 또 단순코드는 데이터 구조에 따라 함께 구조화되므로 한글 내부의 구조에서 추상적인 수학적 구조를 정의할 수 있느냐 없느냐에 따라 마땅히 코드의 구조도 수학적 구조와 대응성을 이루느냐 어떠냐가 결정된다. 만약 결정적 수학적 구조에 따라 코드구조가 결정된다면 그 코드는 한글정보의 지적처리를 위한 만능코드(universal code)구조를 가지게 된다. 이런 측면에서 한글코드를 논한다면 단순

히 한글 입출력 처리만을 강조하는 코드체계는 한계를 보이게 된다.

### 5.3 새로운 한글코드의 조건

현재까지 쓰이고 있거나 제안되고 있는 한글 코드계의 대부분이 국제표준규격의 개정이나 신설 요구 또는 특정의 지지 그룹에서 제안된 것을 논의의 대상으로 삼아 구현되거나 실시되고 있는 것들이다.

위에서 설명한대로 문자코드도 문자가 가지는 본디의 구조원리(ontology)에 따라 무엇이 어떻게 코드화 함으로서 정보처리 및 교환에 어떤 효율성이 있는가, 또 국제 규격으로서의 품질은 보장되고 있는가 등에서 한글 코드체계가 개발되고 평가되어야 한다.

현행 코드체계인 ISO-10646 UCS 규격에 따른 KS X 5700에서만 보아도 그간에 쓰이고 있던 코드, 또는 제안된 코드에 대하여 모든 요구를 수용하려 하다 보니 256X256바이트 영역에서 근 1/4을 차지하는 역사적(?) 결과를 낳았다.

남북한에서 거의 같은 관점의 코드가 제기되고 있는데 조합형의 초성, 중성, 종성 자모별의 코드체계이다. 그러나 초성, 중성, 종성 자모의 정보를 추립으로 해서 한글정보처리 및 교환에서 어떤 효율성이 있는지 제시되지 않고 있다. 한글 입력방식에서 3별식을 지지하는 그룹의 입장에서 보면 입력방식과 한글의 내부 표현방법에서 대응성이 좋다는 주장에는 객관적 효율평가에서 어떤 우월성이 보증하고 있는 것인가 밝혀야 한다. 보다 지능적 기능을 가진 지적 정보시스템에서의 한글 문자 처리라면 추상적인 데이터로서의 한글문자이지 그것을 구성하는 구체적인 형태, 예를 들어 글자꼴이 문제라면 바른 글자꼴을 구성, 표시하는 일은 글자의 외부표시형식에 대한 지식을 시스템 자신이 가지고 처리할 일이지 사용자에게 일일이 그것의 표현방법, 표시방법을 의식시킬 필요가 없는 것이다.

새로운 한글코드 시스템은 다음과 같은 원칙 밑에서 개발되고 평가되어야 한다.

- (1) 훈민정음의 경제원리 : 스물 여덟자로서 굴러바뀌어 무궁하여 간단하고도 요긴하며 정교하고도 두루 통한다. 즉 작게 만들어 크게 쓴다라는 창제 정신을 구현하는 코드체계일 것.
- (2) 한글의 지적정보처리에서 필요로 하는 모든 알고리즘에서 효율성을 증명할 수 있는 코드체계일 것.
- (3) 국제규격의 틀을 지키고 현재의 어떤 한글코드체계와도 호환성을 보증하는 코드체계일 것.
- (4) 한글 문자의 내부 표현방식과 외부 표시방법을 분리한 중립적인 한글 데이터 표현을 기본코드로 하되 도형문자의 표현과 문자표시제어 코드로 구성하고 한글의 지적 정보처리에서 필요한 정보는 동적으로 생성가능한 코드체계일 것.
- (5) 문자셋트와 코드화 방식에서 국제규격의 흐름인 싱글바이트(7비트, 8비트), 멀티바이트(2바이트~4바이트) 체계와의 대응성, 연계성, 확장성을 가진 코드체계일 것.
- (6) JTC1/SC2에 의한 부호화 문자셋트에 관련하는 모든 규격은 문자개념에 대응하는 문자부호화 표현만을 규정하므로 한글표시에 관련하는 규격(간격, 크기, 모아쓰기, 풀어쓰기 등)은 독립적 또는 제어코드의 확장법에 의해 만들어진 코드체계일 것.
- (7) 한글 문자코드 셋트는 시각표현인 도형문자(graphic character)의 추상화(개념)에 대한 코드 표현이므로 음성, 음운표현인 음절문자, 즉 초성, 중성, 종성으로 이루는 소리마디(syllable)등 한글 문자표현의 내부적 구조를 사용자에게 보이는 코드체계가 아닐 것. 즉 3별식 내부코드를 대표코드로 하지말 것.
- (8) 한글 정보처리 뿐만 아니라 한글 문자정보의 교환 시스템에서 쓰이는 데이터 통신 규약(protocol)의 표준 모델에서도 투명도(transparency)를 유지하며 프리젠테이션 레이어(presentation layer)의 규격에 맞는 다목적 공용 코드체계일 것.

어(presentation layer)의 규격에 맞는 다목적 공용 코드체계일 것.

- (9) 창제 당시부터 오늘날에 이르기까지 진화과정을 거친 국어 표기현상을 거친 국어 표기현상을 표현·표시할 수 있고 국어학에서의 모든 요구에 대응할 수 있는 코드체계일 것.

이상과 같이 여러 조건을 만족하는 새로운 한글코드체계가 필요하다.

#### 5.4 남북한 한글코드의 통일안은 가능한가

95년 우리말 컴퓨터 국제학술대회에서 합의한 한글코드 공동사용단일안 마련을 위한 사항은 다음과 같다.

- (1) 현재 남한과 북한 및 중국에서 사용하고 있는 2바이트 완성형과 2바이트 조합형 부호계는 그대로 둔다.
- (2) ISO 2022를 따르면서 우리글을 제대로 지원할 수 있는 1바이트 조합형 부호계 작성의 필요성을 인식하고 공동안을 만들기 위한 연구를 진행한다.
- (3) ISO 10646-1의 우리글 부호계를 완성하기 위하여 검토, 연구한다.
- (4) 우리글을 좀더 폭넓게 지원할 수 있는 부호계를 공동으로 연구한다.
- (5) 남과 북, 중국의 서로 다른 부호계를 변환할 수 있는 프로그램을 공동으로 연구, 개발한다.

위의 합의 사항에는 첫째, 사실상의 표준코드로 정착되어 있는 완성형, 조합형의 남북, 중국안의 실용 상황을 인식하여 그대로 쓰되 새로운 코드체계의 연구에 대한 기대도 담겨져 있다. 둘째, ISO-2022에 따른 한글코드계는 2바이트 완성형이 되어 있으므로 7비트, 8비트계의 한글코드를 만들자는 제안이 합의점이나 초성, 중성, 종성의 조합형 코드는 기존의 것과 대동소이하며 다만 음절을 나타내는 문자 바이트의 크기를 2바이트로 고정시키지 않는다는 것이 다르다. 반드시 기술적 효율에서 우월을 가리는 싱글 바이트 코



드가 만들어져야 한다. 셋째, ISO-10646의 코드 체계에 넣을 수 있고 남북의 상이점을 흡수하는 한글코드가 연구되어야 한다. 넷째, 우리글의 폭 넓은 지원이란 현재 한글코드의 논점이 문자 생성 표현과 표시 수준에 머물고 있는만큼 고도의 문법처리등 요소기술, 응용기술에서 까지 효율 좋게 쓰일 한글코드계의 발상전환적 새로운 연구가 필요하다. 다섯째, 현행 코드체계간의 호환성을 보증하는 소프트웨어 개발을 통하여 상징성 있는 남북교류 공동연구의 결과를 내는 연구도 중요하다.

### 5.5 컴퓨터 한글자판의 남북통일안에 대한 기대

이철의 글은 북한 교육성 정보센터 소장 이수락 박사의 <컴퓨터 한글선택 표준자판 표준화에 관한 연구>에 대한 분석보고서[2]를 일부 그대로 옮긴 것이다.

현행 표준자판을 대신할 수 있는 새로운 표준자판은 적어도 다음과 같은 요구를 만족시켜야 한다고 인정한다.

- 1) 새로운 자판이 현실에 도입되려면 현행 표준자판과 본질적으로 구별되는 기술적인 우월성을 가져야 한다.
- 2) 새자판의 시안은 공정성을 기하기 위해 현행자판을 비롯하여 여러 가지 가능한 방안들과의 대비고찰에 기초하여 이론적으로나 실험적으로 그의 타당성이 충분히 론증되어야 한다.
- 3) 지난 시기 Korean 국제학술대회를 통해 통일자판이 합의되면서도 아무런 결실을 얻지 못하면서 상당한 교훈을 찾아 선택자판은 반드시 사회적인 인정을 받기 위한 절차를 밟아 현실도입을 위한 길을 열어놓아야 한다.
- 4) 표준이 지녀야 할 성질로부터 반드시 고쳐야 할 절실한 근거가 있는 부분은 제외하고는 기성표준을 될수록 변경시키는 일이 없어야 한다. 이러한 판단으로부터 이번 공동연구사업은 현행 국규자판을 당장 바꾸기 위한 새로운 자판의 개

발에 목적을 두는것이 아니라 현행표준자판을 당분간 그대로 두면서 건반작업을 많이 하는 작업자들에게 있어서 입력속도가 빠르면서 정신육체적인 피로를 적게 주는 비표준자판을 개발하는데 두기로 한다. 이러한 비표준자판의 사용자가 많아지고 그의 과학기술적타당성과 실용적 가치가 사회적으로 충분한 인정을 받을때에는 현행 표준자판을 대신하는 새로운 표준으로 채택될수도 있을 것이다.

#### 5.5.1 현행 우리글자판이 가지고 있는 부족점들

현행 표준자판을 대신할수 있는 새로운 자판의 설계를 위하여 기존 우리 글 자판들이 가지고 있는 부족점들을 종합적으로 분석한 결과는 요지 다음과 같다.

가. 현행 두벌식 자판이 가지고 있는 공통적인 부족점  
 현재 북과 남에서 리용하고 있는 국규 9256자판과 KS X 5002자판은 다같이 초성 자음과 종성 자음을 같은 건으로 입력하는 2벌식자판이면서 다음과 같은 공통적인 부족 점을 가지고 있다.

- 1) 자판배열이 최적화되지 못한것이다.  
 자판배열이 우리 글 자모의 출현빈도와 손가락들의 운지능력에 토대하여 합리적으로 배정되지 못하고 있다. 특히 KS X 5002 자판의 모음배열에서의 부족점이 매우 심하다. (그림 3.4)
- 2) 모음인 “ㅏ” 가 예외적으로 왼손 타건구역에 놓여 있는것이다.  
 왼손으로 자음, 오른손으로 모음을 타건하도록 훈련된 자를신경에 하나의 예외를 조성하는 이러한 타건법은 타건 리듬을 심히 손상시킨다.
- 3) 타건하기 어려운 Shift건의 리용률이 높은 반면에 운지능력이 높은 엄지손가락이 space bar의 리용에만 리용되고 있는 것이다.  
 손가락 이동거리가 큰 Shift나 Ctrl건의 사용은 blind touch입력에서 정신적인 피로를 주며 타건작업의 리듬감을 훼손시키는 작용을 한다.
- 4) 모음의 6%정도의 출현빈도를 차지하는 9개

ㅂ	ㅅ	ㅇ	ㄹ	ㅎ	ㅋ	ㅌ	ㅍ	ㅈ	ㅊ	ㅋ	ㅌ	ㅍ	ㅊ	[	]
10	7	6	4	9	7	6	5	8	9						
ㅈ	ㅊ	ㅇ	ㄹ	ㅅ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	ㅊ	;	,
8	3	1	2	5	4	1	2	3							
ㅋ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	/
14	13	12	11	18	13	16									

(그림 3) 국규 9256 자판의 건배렬과 출현빈도(수자는 출현빈도순위:부록1)

ㅂ	ㅅ	ㅇ	ㄹ	ㅅ	ㅌ	ㅍ	ㅈ	ㅊ	ㅋ	ㅌ	ㅍ	ㅊ	[	]	
10	7	6	3	5	13	7	14	8	9						
ㅅ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	ㅊ	;	,	
8	2	1	4	9	5	4	1	2							
ㅋ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	ㅊ	ㅌ	ㅍ	/
14	12	11	13	17	6	3									

(그림 4) KS X 5002 자판의 건배렬과 출현빈도(수자는 출현빈도순위:부록2)

의 겹모음을 두번의 건입력으로 입력하는 것으로 인한 타건속도의 저하가 존재하는것이다. 독자적인 건이 배정되어 있는 “ㅌ, ㅍ, ㅊ” 보다 출현빈도가 더 높은 “ㅅ, ㅌ, ㅍ” 를 비롯하여 9개의 겹모음을 런타로 합성입력하는것으로 인한 타건속도의 저하율은 3%정도로 평가된다.

5) 오른손에 비한 왼손 부담이 높은것이다.  
 현행 표준자판에서는 오른손/왼손의 리용비율이 4:6정도로써 오른손보다 운지능력이 낮은 왼손에 더 큰 부하부담을 주고 있다.

6) 《도깨비불》 현상에 의한 심리적 부담  
 2벌식입력에서는 다음 글자의 모음입력이 있기 전에는 앞글자를 확정할수가 없어 《도깨비불》 글자가 생기는것을 피할수 없으며 이로 인한 심리적부담이 존재한다.

실례 : “나라” 의 입력과정에서 생기는 중간글자인 “날”

5.5.2 남과 북의 일치되는 견해와 연구결과

2005년 7월에 넘겨 받은 연구보문 《남북통일 시대에 사용될 컴퓨터 한글선택 표준자판 표준화에 관한 연구》와 1983년부터 필자가 진행해 온 우리글 자판에 대한 연구 결과에서 일치되는 내용들은 다음과 같다.

가. 일치되는 견해

다음과 같은 문제들에 대해서는 우리도 같은 견해를 가지고 있다.

1) 현행 자판에 대한 객관적이며 과학적인 분석, 검토와 대안제시를 위한 연구사업이 필요하다.

현행 국규9256자판과 KS X 5002자판은 자판 배치에서 얼마간의 차이를 가질 뿐 2벌식자판으로서의 기본설계사상에는 별다른 차이가 없다.

또한 이 자판들은 자모배렬과 입력방법에 있어서 사용자들을 만족시키지 못한 본질적인 부족점들을 가지고 있다.

2) 현행 표준자판과 병존되는 보다 기술적으로 개량된 우리글 자판의 개발은 국가적인 표준화 사업에 저촉되지 않을뿐 아니라 표준화사업을 발전시키는데 긍정적인 작용을 논다.

현재 북에서 리용되고있는 IME에는 국규자판과 함께 《창덕》 자판과 《로마자입력》이 지원되어 있으며 남에서도 3벌식 자판을 비롯한 여러 가지 비표준자판들이 존재한다.

모든 표준은 고정불변한것이 아니라 현실발전에 따라 계속 개선되어 나가야 하며 우리글 자판을 개량하기 위한 연구 그 자체를 부정하는것은 과학기술발전을 부정하는 견해로 밖에 볼수 없다.

지난 시기 Korean 콤퓨터처리 국제학술대회를 통한 북남통일자판개발사업이 합의서 채택만으로 끝난것은 채택된 시안자체가 가지고 있는 부족점보다도 자판을 계속 개선하기 위한 연구사업 그 자체를 국가규격을 준수하는데 지장을 주는 해로운 일로 간주하는 낡은 관점에 기인되는 것으로 분석된다.

교육을 비롯한 일반사용에서는 국가적인 표준 자판을 준수하면서 건반작업을 많이 하는 사람들속에서 제한적으로 리용되는 보다 효율적이며 쓰기 편리한 자판의 개발은 그의 사용과정을 통하여 현실발전에 상응하게 국가표준을 보다 훌륭하게 개선해나가기 위한 전단계로써의 적극적인 역할을 늘수 있을 것이다.

3) 자판의 평가 범주 및 항목은 국제적인 표준에 맞추어 다음의 5가지 항목으로 하는데 대해서 우리는 견해를 같이 하고있다.

가) 손, 손가락의 작업평형척도

나) 입력속도

다) 오타률

라) 피로도

마) 학습상승도

그러면서 다음의 세 개 항목을 더 보충할것을 제안한다.

바) 장치적으로 특수하게 만들어진 건반을 쓰는 일이 없이 현존 건반상에서 소프트웨어적으로 가동하도록 해야 한다.

사) 입력모드를 전환함이 없이 기호와 수자들을 입력할수 있도록 우리글 자모입력을 위한 자판 구역은 될수록 영문자입력구역을 크게 벗어나지 말아야 한다.

아) 반드시 고쳐야 할 절실한 근거가 있는 부분을 제외하고는 기성표준을 될수록 변경시키는 일이 없어야 한다.

나. 일치되는 연구 결과

1) 우리 글 자모출현빈도의 조사결과

1980년대에 우리가 진행한 자모출현빈도 조사 결과는 제시된 조사결과와 기본적으로 일치된다. 북남간에 존재하는 철자법의 차이와 선택한 본문 Corpus의 차이로 인하여 수값적으로는 얼마간의 차이를 보여주고 있지만 자모별 출현 빈도의 순위에 있어서는 큰 차이가 없다

2) 손/손가락들의 운지능력순위에 대한 연구결과도 완전히 일치한다.

- 손 : 오른손 > 왼손

- 런타 : 좌우교차타건 > 오른손런타 > 왼손런타

- 입력단 : 중단 > 상단 > 하단

- 손가락 : 집게 > 엄지 > 가운데 > 약 > 새끼

3) 자모출현빈도와 운지능력의 견지에서 평가할 때 북남의 현행 2벌식자판들은 다 같이 완전한 최적화가 이루어져 있지 못하다.

그림에서 알수 있는것처럼 KS X 5002자판에서는 출현빈도가 비교적 큰 “ ㄱ, ㅡ ”를 타건하기 어려운 하단에 배치하는데 비해 국규9256자판에서는 이들을 타건하기 쉬운 기준렬의 중단과 상단에 배치한것을 비롯하여 얼마간 최적화수준이 높은 것으로 평가된다.

4) Shift건의 사용이 타건의 연속성을 깨뜨리고 사용자에게 스트레스를 주며 사고과정에 악영향을 주는 입력방법으로 된다는 평가.

대/소문자의 절환이 있는 영문입력에서는 Shift건의 리용이 불가피하지만 대/소문자 절환이 없는 우리글 입력에서는 손가락 이동거리가 큰 Shift건을 리용하지 않는 입력방법이 바람직하다.

## 6. 결 론

'95년 우리말 콤퓨터 처리 국제 학술대회에서 합의를 본 한글코드, 한글입력용자판, 콤퓨터 용어등 통일 규격에 대한 공동시안은 언젠가는 실용화 단계에 이를 것이다. 그러나 이와 같은 사안의 기술 및 규격은 한글정보의 콤퓨터처리 기술에서 보면 아주 기본적인 것임을 이미 지적했다.

한글 컴퓨터처리 기술의 세계화, 국제경쟁력의 확보를 위한 영역의 연구 및 국제 표준규격 마련을 위한 범주의 연구영역은 아직 많다. 남북한 과학 기술인끼리의 공동연구 과제로서 바람직한 대상을 살펴보면 다음과 같은 현안이 있다.

### 6.1 한글 폰트(font) 관련 기술

- 새로운 우리글 폰트의 공동개발 및 공용 기술 연구
- 한글 문서 기술언어 및 폰트 정보교환 규격의 공동연구(ISO/IEC JTC/SC18/WCT8의 ISO 9541, 10036)

### 6.2 한글 언어(language)프로세서의 기술 개발

- 한글문자 자동생성장치의 표준기술(하드웨어 및 소프트웨어) 공동개발
- 한글 워드 프로세서의 개념을 넘는 새로운 언어기술의 공동개발

### 6.3 한글 문법의 컴퓨터 처리 시스템 개발 기술

- 한글 형태소 처리를 위한 문법이론 단일화 연구와 처리 시스템 연구
- 한글 구문 처리를 위한 문법이론 단일화 연구와 처리 시스템 연구
- 한글 의미 처리를 위한 문법이론 공동연구 및 처리 시스템 연구

### 6.4 국어의 음성인식, 문자인식, 교정지원, 기계번역 시스템 개발을 위한 기반 요소기술의 공동 연구등이다.

21세기에는 인간과 컴퓨터 사이의 대화가 자연언어로 이루어질 것이라는 것이 지배적인 견해이다. 앞으로는 특별한 학습이 필요없이 인간이 태어나면서 익히는 말과 글로 컴퓨터를 부리는 시대가 온다는 말이다. 이와같은 기술추세에 따라 선진국에서는 자국어의 컴퓨터처리의 최적화는 지상과제가 되고 있다.

따라서 우리의 경우도 국어의 정보처리기술의 확립은 컴퓨터 기술에 있어서 21세기 국가 경쟁력을 가름하는 가장 큰 지표가 될 것임을 부인할 수 없는 사실이다.

우리나라는 남과 북이 나뉘어져 있다. 남과 북이 따로 따로 국어의 정보처리 기술을 연구, 개발하고 있다. 기술은 끊임없이 발전되므로 최적, 최선을 고정화 할 수는 없다.

국어 기술의 발전과 진보를 위해서는 남과 북이 연구하고 개발한 내용을 서로 공개하고 교환하고 또 선의의 경쟁과 협력체계를 갖추도록 해야 한다.

위와 같은 목표를 위해서는 국어정보의 컴퓨터처리 기술개발을 위한 남북의 정기적 모임을 갖는 일이 지름길이다.

### 참고문헌

- [1] 이수락 : 전자계산기에 의한 조선어처리기술에서 해결해야 할 과학기술적 문제와 그의 표준화에 대하여, '94 Korean 컴퓨터처리국제학술대회 논문집, 1~9쪽, 1994.
- [2] 이수락 : 컴퓨터 한글선택 표준자판 표준화에 관한 연구에 대한 분석보고서, 2006.
- [3] 김명규 : 우리글 부호계의 현실상태와 개선방안에 대한 몇가지 고찰, '95 Korean 컴퓨터 처리 국제학술대회 논문집, 중국엔지 1995.
- [4] 강영민 : 건반에서 조선글 자소 배열, '95 Korean 컴퓨터 처리 국제학술대회 논문집, 중국엔지, 1995.
- [5] 한국기계연구소 : 한글 타자기 자판 및 기구제도 개선에 관한 연구, 과거지, 1985.
- [6] 정희성 : 남북한 한글 입력 키보드의 생산성 비교, 548돌 한글날 기념 '94 우리말 정보화 가을철 학술대회 논문집, 국어정보학회,

1994.

- [7] 정희성 : 한글 입력 키보드의 생산성 평가를 위한 기술, '95 Korean 컴퓨터 처리 국제 학술대회 논문집, 중국 옌지, 1995.
- [8] 권오일 : 건반의 건배치에서 비교차 타건이 입구 속도에 미치는 영향에 대한 고찰, '95 Korean 컴퓨터처리 국제학술대회 논문집, 중국옌지, 1995.
- [9] 정희성 : 남북한 국어정보처리기술 비교평가 연구, 최종연구보고서, 한국과총, 1995.

### 저자약력



### 정희성

1987년 일본 동경대학 대학원 이학박사 (정보과학)  
 1987년 한국전자통신연구소 선임연구원  
 1991년 한국생산기술연구원 수석연구원  
 1992년 선문대학교 컴퓨터정보학부 교수  
 2002년 (주)네오펠드 대표이사  
 2006년 선문대학교 대학원장  
 관심분야 : 한글공학, 정보검색, 휴먼인터페이스