

□ 특집 □

# 한글 부호계 제정에서 반복되는 시행착오의 고리를 끊으려면

변 정 용<sup>†</sup>

◆ 목 차 ◆

- |                              |                     |
|------------------------------|---------------------|
| 1. 서 론                       | 4. 시행 착오의 고리는 무엇인가? |
| 2. 부호계 제정지의 한글에 대한 근본 인식의 모순 | 5. 한글 부호계가 나아갈 방향   |
| 3. 바른 기술 보급을 위한 바른 용어 정의     | 6. 결 론              |

## 1. 서 론

언어를 인간 사고의 본체라고 할 때 컴퓨터가 가진 제한된 글자 표현 기능으로 인하여 우리가 사용하고 있는 말을 제대로 표현하지 못하고 가능한 다른 소리로 표현할 것을 강요받는다면 우리의 사고 체계는 큰 위기를 맞고 있다고 볼 수 있다. 인간의 지적 노동 생산성을 높이는 대표적인 도구인 컴퓨터는 연필과 원고지를 대신하여 그 활용 범위가 날로 넓어져 가고 있다. 하지만 컴퓨터를 사용하여 마치 연필로 글을 쓰듯이 사용함에는 부족함이 많다. 이러한 결과는 바로 한글 부호계의 비과학적 제정으로 비롯되고 있다. 오늘날 컴퓨터에서 현대 표준 한글 부호계를 따를 때 한글 맞춤법에서 정한 11172 음절의 20% 정도만 지원하며 옛 한글은 아예 엄두도 못 내고 있다. 최근 국제문자 부호계(UCS) KS C 5700-1995[1]를 수용하여 국내 표준으로 채택함으로써 그 안에 포함되어 있는 음절형(Hangul) 11172자와 자소형 240자, 그리고 자모형 96자 등

이 이제 우리가 원하는 현대 한글 및 옛 글자를 모두 지원하게 될 것[7]이란 말에는 허구가 있

다. 그 부호계는 훈민정음의 문자 구조원리에만 비추어 보아도 많은 문제점을 가지고 있는 것이 사실이다.

본 논문은 우리가 30여년간 한글 정보처리를 하면서 아직도 제대로 된 한글 부호계를 갖지 못하고 온갖 시행착오를 겪고 있는 이유를 밝히고 해결 하는데 있다. 컴퓨터를 사용하는 사람이 다양하고 그들에게 한글은 사용에 있어서 차별이 없어야 한다. 그래서 한글에 대한 관점을 1933년의 조선어학회 이전으로 올라가서 창제 당시의 훈민정음 해례에 담겨 있는 과학성을 규명함으로써 컴퓨터를 한국인의 지적 노동 생산성을 극대화 할 수 있는 가장 편리한 도구로 만드는데 두고자 한다. 다시 말해서 “마치 연필로 글을 쓰듯이” 컴퓨터를 사용할 수 있도록 한글 부호계를 만드는 방법을 찾아내고, 한글 문자의 과학성을 현대 과학의 산물인 컴퓨터를 통하여 입증하는 방향에서 현행 한글 부호계에 대한 개선 방안을 제안하고자 한다.

## 2. 부호계 제정자의 한글에 대한 근본 인식의 모순

우리는 문자에 관한 한 대단한 자부심을 가지고 있다. 언제 누가 창제하였는지에 대하여 너무나 잘 알고 있으며, 매년 10월 9일에 한글 반포 기념

<sup>†</sup> 정희원 : 동국대학교 경주캠퍼스 전자계산학과 부교수

식을 가지고 있다. 심지어 어느 미국 대학의 언어학 교수도 이 날을 기념하는 의식을 갖는다고 한다. 하지만 한글 문자에 대한 우리의 인식은 훈민정음 해례가 아니라 1933년 조선어학회가 제정한 한글 맞춤법에 두고 있다. 컴퓨터는 이제 우리 사회 전반에 사용이 보편화되면서 다양한 사용자를 가지게 되었고, 초기에 컴퓨터가 가진 계산 기능에 앞서 현재는 문서 작성이나 저장, 검색 및 출력 등에 더 큰 비중을 가지고 있다. 보편화는 결국 국어 학자나 관련 학생들의 경우 옛 글자를 문서에 표현하려는 요구가 커졌다. 한글 맞춤법은 첫소리 19자, 가운데 소리 21자, 끝소리 27+1자 범위 내에서 사용하도록 규정하여 옛 글자는 제외하였다. 이것이 오늘날 우리가 한글 문자 부호화에서 대상으로 삼는 11172자의 근거이다. 그런데 우리는 현재 옛 글자를 표현하려고 하면서 우리의 생각을 1930년대에 매어 두는 모순에 빠져 있다. 옛 글자를 표현하려면 당연히 1446년 반포 당시의 훈민정음 해례에 근거해야 마땅하다. 조선어학회의 결정은 훈민정음의 전체 집합에서 부분 집합을 선택한 것이며 이에 근거한 부호화나 구현은 결국 부분 집합의 구현으로 귀결되므로 옛 글자를 지원하는 것은 불가능하다.

### 3. 바른 기술 보급을 위한 바른 용어 정의

올 바른 의사 전달에 있어서 바른 어휘를 사용하는 것은 중요하다. 기술 보급에서 그것은 더욱 그렇다. 한글 문자 부호화 과정에서 용어를 살펴보면 국민적 자부심을 가지고 있는 문자임에도 불구하고 많은 잘 못된 용어를 사용하고 있다. 올 바른 의사 표현과 올 바른 기술 전수를 위하여 바른 용어의 사용을 제안한다.

#### 3.1 한글 부호계의 표준 규격 번호

현행 표준인 정보교환용 한글 부호계는 규격

번호가 KS C 5601-1987[10]이다. 이것은 개정되어 KS C 5601-1992가 된다. 규격 번호에는 제정 또는 개정 연도를 반드시 첨부해야 한다. 최초의 정보교환용 한글 부호계는 KS C 5601-1974이며, 그 다음 개정되어 KS C 5601-1982로 되었다가 1992년 다시 개정이 된 것이다. 최근 국제문자 부호계인 KS C 5700-1995[11]가 새로운 표준으로 제안되고 있으며, 1997년 8월 20일부터 기술품질원은 규격 번호를 시작 부분에 대하여 KS C에서 KS X 체제로 바꾸었다[12,13]. 본 논문에서는 아직 우리가 새로운 규격 번호에 익숙하지 않다는 점에서 기존 번호를 그대로 사용한다.

#### 3.2 부호계의 별칭을 바로 잡으면

현재 한글 부호계를 부를 때 완성형이니 조합형이니 하는 용어를 쓰고 있다. 한글 문자는 15세기 조선의 최고의 학자들이 만들어낸 걸작이다. 여기에는 당시의 음운학적 연구 결과를 바탕으로 해서 만들었으며 매우 수준 높은 과학을 적용하고 있다. 그렇기 때문에 20세기에 들어 우리가 붙이는 명칭이 단순한 감각에 바탕으로 한 것이거나 비전문가의 무지에서 비롯된 것이어서 안 된다. 그럼 현재 우리가 사용하고 있는 별칭의 사용 예를 보자.

정보교환용 부호계의 별칭으로 1974년 것은 날자형이라 하였고, 1982년에는 조합형, 1987년에는 완성형이라 하였다. 그리고 국제문자 부호계 KS C 5700-1995에서 한글 부호계는 한글(Hangul), 한글 자모(Hangul Jamo), 호환 한글 자모(Compatible Hangul Jamo)라고 이름을 붙이고 있다. 이들을 볼 때 조합형과 완성형이란 용어는 국어학에서 찾아보기 힘든 용어로서 컴퓨터 기술자들이 편의상 붙인 용어이다. 국어학에서 문자는 자음(닿소리)과 모음(홀소리)라 하고, 그리고 음절 구성에 대한 음소로서 초성, 중성, 종성으로 기본 글자를 구분하며, 이들을 다시 날 글자와 접 글자로 구성

한다. 소리를 내는 단위는 음절이며 자음+모음, 자음+모음+자음 또는 초성+중성, 초성+중성+중성으로 구성을 나타낼 수 있다. 부호계의 명칭은 부호화(coding)의 대상을 무엇으로 삼았느냐에 따라서 국어학에서 사용하는 용어를 바탕으로 붙여 주어야 한다. 따라서 그 대상이 자음과 모음일 때 ‘자음과 모음’의 줄인 말인 ‘자모’이므로 자모형이라 한다. 초성 중성 종성이 대상일 때는 이들을 음소라 하므로 음소형이라 한다. 여기서 음소는 훈민정음 당시는 소리와 글자의 명칭을 섞어서 사용하였기 때문에 문제가 없으나 음절 글자의 요소라는 뜻에서 음소(音素) 대신 자소(字素)라 함이 좋을 듯하다. 그래서 ‘자소형’이라 부르는 것이 적합할 것이다. 본 연구자가 1992년 훈민정음 해례에 기초하여 제정한 ‘정음형’ 부호계[5,6]는 대표적인 ‘자소형’이다. 그 다음으로 부호화 대상을 음절 글자로 한다면 당연히 ‘음절형’ 또는 ‘음절자형’이라 해야 한다. 그러면 기존의 부호계가 어떻게 고쳐질 수 있는 지 보자. 정보교환용 부호계 KS C 5601-1974는 ‘낱자형’에서 ‘자모형’ 또는 ‘겹자모형’이 되고, KS C 5601-1982는 ‘조합형’에서 ‘음절형’ 또는 ‘자소 있는 음절형’이 된다. KS C 5601-1987은 ‘완성형’에서 역시 ‘음절형’ 또는 ‘자소 없는 음절형’이 된다. 여기서 ‘자소’는 ‘음소’로 바꾸어 써도 좋다. KS C 5700-1995에 있는 ‘한글’은 ‘자소 없는 음절형’으로, 그리고 ‘한글 자모’는 ‘자소형’ 또는 ‘겹자소형’이 되며, ‘호환 한글 자모’는 ‘자모형’ 또는 ‘겹자모형’이 된다.

### 3.3 조합형과 완성형의 대결

우리가 인식하고 있는 한글은 주로 음절(글자)이다. 읽고 쓰기의 기본 단위이기 때문이다. 따라서 컴퓨터에서 사용하는 부호계의 별칭을 보면 부호화 대상에 관계없이 음절을 어떻게 구성하느냐에 따라서 이름을 붙이고 있다. 현재 통상적으로

로 사용하고 있는 용어를 보면 자모형이나 자소형 부호계라 하더라도 둘 다 조합형이라고 부르고 있다. KS C 5601-1982는 2 바이트를 아래쪽부터 5 비트씩 셋으로 잘라서 초성, 중성, 종성 글자들을 표현하였던 부호계로 이를 흔히 조합형이라 하고, KS C 5601-1987은 완성형이라 한다. 이런 결과로 볼 때 음절을 어떻게 구성하느냐에 관점을 두고 음절을 조합해서 구성하면 무조건 조합형이라 하고 미리 구성된 것이면 완성형인 것이다. 여기서 조합형이란 바로 ‘음절 조합 방식’을 뜻한다. 예를 들어서 자음과 모음으로 구성하는 ‘음절 조합 방식’, 초성 중성 종성으로 구성하는 ‘음절 조합 방식’의 그릇된 사용이다. 따라서 부호의 별칭은 부호화 대상에 대하여 붙여야 한다. 그리고 조합형과 완성형이란 별칭은 둘 다 국어학에 기초하지 않고 기술적이고 감각적인 말이므로 음절형이라고 해야 한다. 이 용어가 사람들이 통상적으로 사용하고 있다는 점을 제기하는 이도 있지만 조합형은 상당히 혼란하게 사용하고 있기 때문에 바로 잡아야 한다.

### 3.4 존재하지 않는 부호계 'N 바이트 부호계'

통상적으로 ‘조합형 한글 코드’라고 하는 것에 ‘N 바이트 코드’가 있다. 이것은 KS C 5601-1974의 별칭이다. 그런데 부르는 의도와는 달리 부호계의 원리를 잘 아는 사람에겐 94n로 이해되어 황당 무계한 부호계가 된다. 이것은 ISO 646과 ISO 2022에 근거할 때 1 바이트는 7비트 계에서 도형 문자 지역이 94가 되므로 2 바이트 부호계는 94<sup>2</sup>이 되고 3 바이트 부호계는 94<sup>3</sup>이 된다. KS C 5601-1987이 바로 2 바이트 부호계로 94x94 = 8836개의 도형 문자를 수용할 수 있다. 물론 이것은 통상적으로 자모형 부호계로 한 음절자를 나타내기 위하여 받침이 없을 때는 2 바이트, 없을 때는 3 바이트 또는 4 바이트로 구성되므로 한 음절을 구성하는 바이트 수가 일정하지 않다

는 뜻에서 감각적으로 붙인 별칭이다. 따라서 오해를 낳기 때문에 자모형이라고 해야 한다.

### 3.5 새로운 조합형과 첫가끝

1992년 ISO/IEC SC2 서울 회의에 상정할 부호계의 국가안을 결정하기 위하여 1991년 10월부터 가동되기 시작한 국내 SC2/WG에 정음형 부호계를 제안하였다[15]. 이로써 WG는 초성 중성 종성을 부호화 대상으로 하는 자소형 부호계를 처음으로 채택하였고, 자소의 수는 겹 자소로 하기로 의결하였다. 그래서 확장된 정음형은 국가 안으로써 그 해 6월 ISO/IEC SC2 서울회의에 상정되어 ISO/IEC 10646-1에 반영되었다. 그 후 정음형 부호계를 ‘새로운 조합형’이라고 부르는 학자[7]가 있었다. 여기서 ‘새로운’이라 말은 시간성을 가지고 있기 때문에 별칭으로 삼기에 부적합하다. 그 후 이를 다시 ‘첫가끝’이라고 칭하고 있는데 ‘첫소리’, ‘가운데 소리’, ‘끝소리’의 첫 글자를 따서 붙인 이름으로 설명적인 이름으로는 부적합하다. 별명은 원칙에 어긋나지 않으면서 사람들의 이해를 돕는 방향에서 붙여 주어야 한다. 따라서 그 부호계를 처음 제안한 사람의 뜻을 존중하는 의미에서 ‘확장 정음형’으로 하거나, 국어학에서 선택한다면 ‘겹 음소형’이나 ‘겹 자소형’으로 해야 한다.

## 4. 시행 착오의 고리는 무엇인가?

문자에 관한 한 국민적 자긍심은 대단하지만 현대 과학의 산물인 컴퓨터에서 한글은 그렇지 못하다. 여기서 훈민정음 해례에 관한 여러 연구 [1,5,6,8]로부터 과학성의 실체를 규명하고 국어학에 근거하여 문자의 특성을 파악한다. 마지막으로 컴퓨터를 이용한 국어정보처리의 다양한 응용 분야의 요구에 근거하여 무엇이 문제인지지를 도출한 다음 5장에서 그 개선책을 제시하고자 한다.

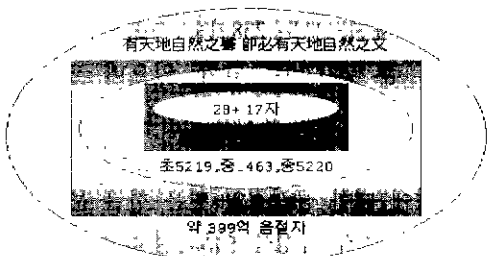
## 4.1 세종임금은 컴퓨터의 출현을 예견하고

현재 한글 맞춤법은 1933년 조선어학회의 한글 맞춤법 통일안을 1980년 한글학회가 일부 수정한 것이다. 이것은 우선 훈민정음 창제 당시에 쓰던 4자(ㄱ, ㆁ, ㅇ, ㅎ)를 제외시킴으로써 부분 집합을 정의한 것이다. 여기에 근거한 부호계로써 훈민정음 창제 후 사용한 모든 글자를 표현하려는 것은 어리석은 일이다. 따라서 옛 글자라고 하는 모든 글자를 자연스럽게 표현하려면 1933년 이전으로 되돌아가야 한다. 훈민정음 창제 원리가 담겨 있는 해례본의 내용은 다음과 같다.

해례에는 먼저 제자해에서 음양오행과 발음기관의 모양에 따라 천지인( 一 )과 아설순치후( ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅇ )를 만들고 소리의 강약 장단에 따라서 확장하여 글자를 만든 원리를 설명하고 있다. 그리고 초성해에서 자음 17자, 중성해에서 모음 11자를 정의하고, 종성해에서 역시 자음 17자를 정의하는데 비록 음가는 다르지만 글자 모양이 같으니 ‘중성부용초성’ 즉 동치관계로 정의함으로써 기본 글자의 수를 45에서 28자로 줄였다. 그 다음에는 합자해로서 순음 아래 세로로 ‘ㅇ’를 나열하여 순경음을 만드는 규칙이 나오는데 이를 연서법이라고 한다. 그 다음은 초성해와 중성해의 날자를 2자 또는 3자를 합용병서하여 자음과 모음의 집단을 확장하는 규칙을 정하니 이것을 합용병서법이라 한다. 그 다음은 음절을 만드는 규칙이 나오는데 가로꼴 모음은 초성 아래 녀에, 세로꼴 모음은 초성 오른 녀에 두며, 종성은 이에 잇대어 적는다라는 규칙이 언급되는데 이를 부서법이라 한다. 그리고 부서법을 만든 이유를 들어서 모든 글자는 모야야 소리를 이룬다고 해서 이를 성음법이라고 한다. 이상의 규정에 따라서 조합할 수 있는 음절의 수는 약 399의 글자이다. 여기에는 깊은 과학이 숨어 있다.

한글 문자가 과학적인 문자라 함은 여러 이유가 있지만 컴퓨터 공학 관점에서 볼 때 과학성은

정인지 선생이 쓴 서문에서 그 대의를 파악할 수 있다. 서문의 첫 문장은 ‘천지 자연의 소리가 있으면 반드시 그 소리를 기록할 문자가 있다(有天地自然之聲 卽必有天地自然之文)’로 되어 있다. 이것은 훈민정음 창제의 목표와 범위를 이야기한 것으로 세상의 모든 소리를 다 적을 수 있는 표음문자를 만들겠다는 의지를 강력하게 보인 것이다. 그렇지만 그 문자를 배우기 쉽기로는 지혜로운 이는 아침 먹기 전에, 어리석은 이라도 열흘이면 깨우칠 수 있다고 하였다. 이것은 극과 극을 달리는 요구로서 과학만이 이를 해결할 수 있음을 알 수 있다. 다시 이를 해석하면 ‘천지자연의 소리를 적을 수 있는 기본 소리 글자와 소리의 구성 및 구조 원리를 만들겠다.’가 된다. 이것이 과학이다. 과학은 어떤 대상에 대하여 일어날 수 있는 가능한 모든 경우를 꿰뚫을 수 있는 하나의 원리를 찾는 것이다. 바로 훈민정음은 한자처럼 모든 글자를 나열하지 않고 천지자연의 소리로부터 기본 소리와 그 구조 원리를 발견해 낸 것이다. 약 399억 음절 글자를 표현할 수 있는 것은 그 소리의 실존함의 여부를 묻기에 앞서서 천지자연의 소리에 필적하는 결과라 할 수 있다.



(그림 1) 훈민정음 원리도

## 4.2 로마자와 가나는 반쪽 한글

컴퓨터 기술자들은 한글의 음절 표현에 집착하고 있다. 그러다 보니 음절형 부호계에 집착하거나 자모형과 자소형 부호계에 대하여 음절 구성

에 중점을 두기 때문에 자모형과 자소형의 중요성을 간과하고 소위 원성형이라는 음절형 부호계의 사용을 고집하고 있다. 한글 문자는 가장 발전된 표음문자에 속하며 또한 음소문자와 음절문자의 특성을 함께 지니고 있다. 이것은 가나가 표음문자이지만 음절문자이며, 로마자 역시 표음문자이지만 음소문자라는 점에서 이들은 반쪽 한글에 지나지 않는다는 사실을 알 수 있다.

그런데 컴퓨터 기술을 수입하는 과정에서 미국에서 들어오는 소프트웨어에서는 로마자의 음소에 대하여 한글의 음절이 문제가 되었고, 일본에서 들어온 소프트웨어는 반대로 가나 음절에 대하여 음소가 거추장스러웠던 것이 사실이다. 자소 없는 음절형(원성형)의 경우에는 음소문자 특성을 버림으로서 스스로 반쪽 한글이 된 것이다. 따라서 이것은 역사적으로 보아 훈민정음 창제 이전으로 후퇴한 결과를 빚게 되었다. 따라서 마치 고려시대의 구결이나 신라시대의 이두를 쓰는 꼴이 된 것이다. 사실 현재 우리가 쓰고 있는 컴퓨터는 글자판(keyboard)에서 입력하는 순간 글자는 반쪽 한글이 되어 고려나 신라시대로 후퇴하는 것이다. 자소형은 음소문자에 대하여 부호화를 하였지만 초성 중성 종성 글자의 속성을 지니고 있어서 음절 구조는 일차원의 풀어쓰기 꼴로 존재하더라도 가지고 있다. 그러므로 자소형으로 표현하게 되면 음소 및 음절 문자 특성을 동시에 가지게 된다. 자모형은 KS C 5601-1974에 따를 때 현대 한글에서는 음소 및 음절 문자 특성을 만족한다. 하지만 여기에 옛 한글 자모를 추가할 경우 음절 구성 정보가 부족하여 음절문자 특성을 만족하지 못한다.

KS C 5601-1982는 자소 있는 음절형으로 둘 다 만족하지만 옛 한글 표현이 어렵고, 부호 확장 규격을 근본적으로 위반하기 때문에 부적합하다.

## 4.3 한글 앞에 우리는 평등해야

한글은 우리의 말뿐만 아니라 바람소리 새소리

등 천지자연의 소리를 적을 수 있는 문자이다. 적어도 우리말을 적을 때는 어떠한 직업이나 신분에서 구애됨이 없이 공평해야 한다. 분야에 따라서 차별적으로 어렵거나 불가능해선 안 된다. 그런데 현행 부호계는 분야에 따라서 불평등하다. 특히 옛 글자를 사용하는 국문학자와 어려운 낱말을 연구하는 사람이나 외국어를 소리나는 대로 적어야 하는 직업을 가진 사람은 표현이 아예 불가능하다. 따라서 표준 부호계를 따르지 않는 (주)한컴의 워드프로세서가 이를 가능하게 함으로써 절대적인 경쟁 우위를 가지고 있음을 기억할 필요가 있다. 그런데 기술 품질원의 문자코드 전문위원회의 구성원의 성향은 아직도 자소 없는 음절형인 완성형의 사용에 편향되어 있다. 하지만 이제 응용의 중심은 문자처리에서 언어처리로 넘어 가고 있다. 자소 없는 음절형 부호계는 언어처리를 지원하지 못한다. 따라서 언어처리를 하려면 겹 자모형이나 겹 자소형이 아닌 완벽한 낱 자소형이어야 한다.

## 5. 한글 부호계가 나아갈 방향

한글 부호계를 제정하거나 개정하는 과정에서 전문가들이 한글 문자를 이해함에 있어서 컴퓨터에서 처리하기 쉬운 방향으로 결정하였으며 분석적이고 과학적인 접근법을 취하지 못하였음이 여러 곳에서 찾아 볼 수 있었다. 특히 훈민정음 원리가 매우 과학적이며 이것을 컴퓨터에 적용할 때 너무나 서로 궁합이 잘 맞다는 사실에 주목해야 한다. 하나는 현재 제정되어 있는 국내 국외 부호계 규격 범위 내에서 최선의 방법을 찾는 것이고, 다른 하나는 훈민정음 원리에 근거하여 만든 정음형 부호계를 적용하는 방안이다.

### 5.1 국내 문자 부호계 규격에서

국내 규격은 현대 한글을 위주로 되어 있다.

KS C 5601-1987만이 94 x 94의 네 번째 줄에 현대 및 옛 한글 자모형 부호계를 배치하고 있을 뿐이다. 여기서는 원칙적으로 자소 없는 음절형만 사용하도록 되어 있기 때문에 자모형을 사용하여 부족한 음절이나 옛 한글을 표현하는 것이 가능하기는 하지만 정보교환에서 일차원으로 나열하여 표현하는 것만 허용하고 있다. KS C 5601-1974는 자모형으로 현대 한글 11172자를 완벽하게 지원한다. 이것은 '한글 맞춤법'에 기준을 두고 있다. 그런데 KS C 5601-1982는 KS C 5601-1992에서 다시 한번 반복해서 표준으로 채택되는데 기본적으로 ISO 646과 ISO 2022를 위반하기 때문에 사용할 수 없는 방법이다. 다시 말해서 현대 한글의 경우에는 자모형이 적합하고 옛 한글을 표현할 경우 적합한 부호계가 없다.

### 5.2 국제 문자 부호계 규격에서

국제 문자 부호계 KS C 5700-1995는 자모형 자소형 음절형 등 세 가지 부호계를 포함하고 있다. 자모형은 KS C 5601-1987에 있는 자모형을 그대로 옮겨 놓은 것이고 음절형은 한글 맞춤법에서 정의한 11172 음절 글자를 모두 지원한다. 자소형(음소형)은 훈민정음 창제 이래 사용한 자소 가운데 1993년까지 발견한 초성 91자, 중성 66자, 종성 81자 등 모두 238자와 중성 채움과 종성 채움 등 총 240자로 되어 있으며 조합하면 499,954 음절 글자를 표현할 수 있다. 자소형 자체는 아직 국내 규격으로 제정된 바가 없다. 240자 자소형에 대한 긍정적인 면은 한글 문자의 음소 및 음절 문자 특성을 어느 정도 반영하고 있다는 점이다. 부정적인 면은 겹 자소를 부호화 대상으로 한 것이다. 238자는 완전한 집합이 아니다. 아직도 자소를 찾고 있는 중이므로 새로운 글자를 뒤에 배정할 때 정렬이 불가능해진다. 그리고 겹 자소를 부호화 대상으로 하는 것은 훈민정음 원리에 위반된다. 일부에서는 한 음절을 6 바

이트로 표현할 수 있기 때문에 경제적이라고 하고 있는데 부호화를 하면서 자료 압축 효과까지 추구한다는 것은 과욕이다. 다시 말해서 언어 처리에 필수인 언어 정보의 표현 기능과 같은 본래 목표를 포기하면서 그렇게 하는 것은 있을 수 없는 일이다. 또한 왜 3 종류의 부호계가 포함되어야 하는지 그 이유가 분명하지 않다. 3 종류의 부호계가 표현할 수 있는 범위를 표 1.에서 보면 자모형은 현대 한글 11172자와 옛 한글 음절의 일부를 지원한다. 자소 없는 음절형은 11172자를 지원하지만 낱자를 표현하려면 자소형이나 자모형과 함께 사용해야 한다. 자소형은 현대 한글과 옛 한글을 포함하여 약 50만 음절 글자를 표현할 수 있다. 그런데 음절형은 구현 수준 1에서, 자소형은 구현 수준 3에서 구현을 허용하며 이 때 자소형은 음절을 조합하지 못한다고 규정하고 있다.

〈표 3〉 한글 부호계의 글자 표현 비교

표준명칭	부호형	자소정보	옛글	음절수
KS C 5601-1974	자모형	30x21	무	11172
KS C 5601-1982	음절형	19x21x28	무	11172
KS C 5601-1987	음절형	없음	무	2350+a
KS C 5700-1995	자모형	64x29	유	>11172
KS C 5700-1995	자소형	91x67x82	유	499954
KS C 5700-1995	음절형	없음	무	11172
정음형-1996-	자소형	17x15x17	유	399억

그래서 음절형은 옛 한글을 지원하지 못한다. 여기서 옛 한글을 지원하려면 규정 24를 없애야 한다. 그렇게 되면 자소형과 자모형으로 음절 구성을 할 수 있기 때문에 혼합 부호계를 사용으로 인하여 혼란은 극심해 질 것이다. 연구[14]에 따르면 어떤 한 음절을 표현하기 위하여 적어도 2에서 13가지의 혼합 부호계 문자열이 존재한다고 한다. 따라서 최선의 개선 방안은 이들 셋 가운데 음절 글자 표현에 있어서 최대 집합을 표현할 수 있는 자소형을 남기고 나머지 공간은 반납해야 한다. 현재 표준 위원회에서는 공간을 많이 차지하고 세 종류의 부호계를 가진 것이 대단한 전리

품인양 부끄러운 줄 모르고 자랑하고 있다. 그들을 반납해야 하는 첫째 이유는 자소형으로 충분히 나머지 둘을 대신할 수 있기 때문이고, 둘째는 세계에서 으뜸가는 과학적인 문자가 한자 다음으로 공간을 많이 차지할 뿐만 아니라 혼란을 자초하는 3 종류의 부호계를 가지고 있다는 사실 때문이다.

- 00 라틴
- .....
- 11 한글 자모(Hangul Jamo) (자소형 240자)
- .....
- 31 호환 한글 자모(Compatibilitv Hangul Jamo)
- .....
- AC 한글(Hangul)
- .....
- FF .....

(그림 2) 국제문자부호계(UCS) KS C 5700-1995

### 5.3 정음형 부호계의 적용

훈민정음 해례에서 정의한 초성 17자 중성 11자를 비롯하여 동치관계로 정의되었던 중성 17자를 포함한 45자를 부호화 대상으로 하는 정음형 부호계[5,6]가 제안되었다. 그리고 자소 확장파 음절 구성 방식에 관한 규정인 연서법, 함용병서법, 부서법, 성음법 등은 입력 출력 편집계의 소프트웨어에 반영하도록 하였다. 여기서 정음형이라 한 것은 훈민정음 원리를 그대로 따른다는 뜻에서 붙인 이름이다. 이는 앞서의 분류에 따르면 자소형에 해당한다.

훈민정음에 바탕한 부호계는 1991년에 처음 제안되었으며 45자를 기본으로 하고 있다. 1992년에는 국내 JTC1/SC2에 필자가 직접 제안하여 ISO/IEC 10646에 들어갈 기본 부호계로 채택되었고, 이것은 처음으로 초성 중성 종성을 부호화 대상으로 하는 부호계가 되었다. 또한 SC2는 자소 집단의 범위를 필자가 낱 자소로 할 것을 과학적 근거에 따라 권유하였지만 결국 다수결로 겹 글자로 의결되어 그 때까지 찾은 훈민정음 반포 이후 사용

되었던 240자로 하였다.

정음형에서 고려할 사항은 현재 단 모음 형태로 익숙하게 쓰고 있는 ㅏ ㅑ ㅓ ㅕ와 정음형 'ㅈ, 등 일부를 부호계에 포함할 것이냐 하는 문제가 있다. 우리는 ㅏ 글쇠를 치는 것에 익숙해져 있어서 ㅑ를 ㅓ로 치는 것을 선뜻 받아들이지 못하고 있다. 연서법에서 비읍 순경음 ㅍ은 'ㅍㅇ'의 세로 나열로 표현한다. 그런데 옛 글자 가운데는 가로 나열 'ㅍㅇ'이 있기 때문에 구별하는 방법이 요구된다.

언어 정보는 낱 자소를 부호화 대상으로 할 때 완전하게 표현된다. 겹 자소를 부호화 하면 언어 현상처리를 하기가 어렵다. 현재 국제문자부호계에 포함된 자소형은 겹 자소를 부호화 하였기 때문에 새로운 자소가 발견되었을 때 정렬 문제가 발생하지만 정음형인 경우에 그런 걱정은 없다. 뿐만 아니라 훈민정음 해례에서 정의한 천지자연의 소리를 적을 수 있는 399억 음절 글자를 표기할 수 있다. 이것은 국제문자부호계의 자소형이 240자이면서 50만 음절 글자를 표기할 수 있는 것과 비교할 수 없다.

정음형 부호계는 음절을 표현할 때 컴퓨터 내부에서는 완전하게 풀어쓰기를 하고 컴퓨터 밖으로 표출될 때만 음절을 구성하여 보여준다. 그러므로 입출력 관련 프로그램에서 해례에 규정된 규칙을 구현하면 된다. 한글에서 음절은 읽기에서 가독성과 변별력을 높이는 기능을 한다고 한다. 결국 컴퓨터 안쪽에서는 음소문자 특성을, 바깥에서는 음절문자 특성을 가지도록 구현하면 된다. 이렇게 하였을 때 일부 컴퓨터 프로그래머들은 한 음절의 길이가 일정하지 않아서 문제가 된다고 한다. 그래서 프로그램 작성하기가 쉽도록 겹 자소의 부호화를 선호하며 현재 그렇게 반영되어 있다. 이것은 이미 지적한 바와 같이 하나를 쉽게 하기 위하여 아홉을 어렵게 만드는 결과를 낳고 있다. 해례에서 규정하고 있는 4가지 법칙을 하나

라도 없애면 그만큼 부호화 대상 문자 집단이 늘어나고 그 만큼 표현에 있어서 제약이 커진다.

정음형은 현재 국내 또는 국제 규격화가 되어 있지 않다. 단지 연구 수준에서 논의되고 있는 실정이다. 국제 문자 부호계에서 3 종류의 부호계 가운데 240자 자소형만 남기고 나머지를 모두 없애는 것이 좋다고 제안하였다. 하지만 더욱 이상적인 것은 그 240자 자소형 대신에 45자에 불과한 정음형을 대치하는 것이 최선의 방법이다. 그렇게 되면 문자처리에서 언어처리에 이르기까지 국어정보처리의 모든 분야가 최상의 조건에서 보다 쉽게 구현이 가능하게 된다. 또한 글자꼴만 지원된다면 399억 음절 글자를 컴퓨터에서 표현할 수 있기 때문에 마치 연필로 글을 쓸 때처럼 자연스럽게 쓸 수 있는 도구가 된다.

## 6. 결 론

현행의 한글 부호계 표준은 문자처리 중심의 관점에서 제정되었기 때문에 3번에 걸쳐 한글 부호계를 개정하고 최근에는 국제문자부호계까지 국내 규격으로 채택하였지만 국어정보처리 전반의 응용에서 필요로 하는 언어 정보를 표현하는 데는 한계가 있음을 보였다. 또한 옛 한글을 표현하고 언어처리를 하려면 훈민정음 해례에서 규정한 과학적 원리를 부호계 제정에 반영하여야 함을 알았다. 또한 각 부호계에 대한 별명은 감각적으로 용어가 아니라 국어학에 근거하여 붙임으로써 기술 보급 및 전수에 혼란을 줄일 수 있음을 보였다. 전반적으로 훈민정음에서 창제 원리에 내재하는 과학성을 적용할 때 한글 및 국어정보처리의 근본 문제들이 풀릴 수 있으며, 특히 정음형 부호계의 해결 능력에 대하여 보였다.

이상에서 한글 부호계 제정에서 반복되는 시행착오를 일으키는 고리의 실체를 규명하고 다음과 같은 결론을 내린다. 첫째, 단일 응용에 국한하여



고려하지 말고 국어정보처리 전반의 요구를 만족하는가를 검증해야 한다. 따라서 현재의 문자처리 중심의 사고에서 언어처리 중심으로 응용이 옮겨감을 감안하여 언어처리를 중심으로 문자처리 응용을 고려해야 한다. 둘째, 한글 부호계 제정에서 훈민정음 해례에서 규정한 모든 규칙을 따라야 한다. 현재처럼 한글 맞춤법과 같은 부분 집합으로 전체를 대신하려는 노력은 시행착오를 낳을 뿐이다. 셋째, 국제 및 국내 부호계 제정 규격에 적합한지를 검토해야 한다. 자소 있는 음절형 처럼 규격에 위반하는 부호계를 표준으로 삼는다든지 국제문자부호계에서 처럼 옛 한글 표현을 못하도록 항목을 모순되게 넣거나 세 가지 부호계를 포함시켜서 혼란을 스스로 초래하는 일이 없어야 한다. 전체적으로 컴퓨터 공학의 입장에서 문제가 없음을 실증적으로 검증한 다음 공청회 등을 통하여 최후 검증을 받도록 해야 할 것이다. 왜냐하면 한 번 잘못된 표준은 산업의 발전은 물론 국가 경제에 심대한 영향을 미칠 수 있기 때문이다.

현행 음절형(완성형) 부호계는 많은 문제를 가지고 있음에도 불구하고 행정전산화와 같은 큰 국책 사업의 덕택에 짧은 시간안에 표준으로 정착하는데 성공하였다. 그런데 지금 유니코드나 ISO/IEC 10646과 같은 새로운 부호계가 표준으로 되면서 다시 부호계를 바꿔야 할 필요성이 대두되고 있다. 그러면 앞에서 지적한 대로 국제문자부호계가 현재의 음절형 보다는 낡지만 아직도 문제점을 안고 있으니 현재 시점에서 반드시 점검해 보아야 할 것이다. 우선 세 가지 부호계 가운데 자소형(한글 자모)만 남기고 나머지는 빨리 공간을 반납해야 한다. 이보다 더 이상적인 것은 자소형 부호계를 대신해서 훈민정음 원리를 그대로 따르고 있는 정음형 부호계를 대치함으로써 현재 가지고 있는 언어처리 및 문자처리에서 가지고 있는 문제의 근본을 일시에 제거해야 한다.

한글과 컴퓨터의 만남은 곧 조선의 과학과 현대 과학의 만남이며, 정보화 시대를 예견하고 창제하신 세종 임금의 높으신 뜻을 오늘에 되살려 제2의 훈민정음을 창제하는 각오로 정보화 시대를 맞아 세계사의 주역이 될 전기를 만들어야 할 것이다.

### 참고문헌

- [1] 강신항, 훈민정음연구, 성균관대학교출판부, 1991
- [2] 강진곤.변정용, “언어정보처리에 기반한 한글 코드의 효율성 평가”, 1996 봄 학술발표논문집, 제23권 1호, 한국정보과학회, 1993, pp953-956
- [3] 김문창, 국어문자 표기론:전용톤과 혼용론 비판, 문학세계사, 1984
- [4] 박혁로.최운천.최기선, “기존 한글 코드의 제 문제와 3바이트 코드의 고찰”, '90 봄 학술발표 논문집, 1990
- [5] 변정용.임해철, “훈민정음 창제원리와 한글 부호계의 제정원리 연구”, 제3회 한글 및 한국어 정보처리 학술발표논문집, 한국정보과학회. 한국인지과학회, 1991, pp155-158
- [6] 변정용, “훈민정음 창제원리의 공학화에 기반한 한글 부호계의 발전 방향”, 한국정보과학회지, 제12권 2호, 1994
- [7] 유영호.조충래.김경석, “X Window 환경에서 국제 표준 한글 부호계의 지원 방안 기초 연구”, 1996 봄 학술발표논문집, 제23권 1호, 한국정보과학회, 1993, pp899-902
- [8] 정희성, “수학적 구조론 본 훈민정음 창제원리와 구조”, 한글 및 한국어정보처리 학술발표논문집, 1989
- [9] 북한(조선인민민주주의 공화국) 국규 9566-93
- [10] 한국공업진흥청, 정보교환용부호계 KS C 5601-1974, 1982, 1987

- [11] 한국표준협회, 국제문자부호계 KS C 5700-1995, 1995
- [12] <http://pantheon.cis.yale.edu/~jshin/qa8.html>
- [13] <http://asadal.cs.pusan.ac.kr/kr-jtc1-sc2/sc2-ksx.txt>
- [14] 변정용, "KS C 5700-1995의 문제점과 개선 방향", '97 춘계 봄 학술발표논문집, 한국정보처리학회, 1997, pp387-392
- [15] SC2 K055, DIS 10646에 대한 SC2의 공식 입장 보고, JTC1/K375, 1992



**변 정 용**

- 1980년 동국대학교 전자계산학과 (학사)
- 1983년 동국대학교 전자계산학과 (석사)
- 1994년 홍익대학교 전자계산학과 (박사)

1982년-1987년 한국전자통신연구소 연구원  
 1995년-1996년 영국 요크대학 전산학과 객원교수  
 1988년-현재 동국대학교 경주캠퍼스전자계산학과 부교수

**'98 추계 학술발표대회 및 정기총회 참석 안내**

- ◎ 일 시 : 1998년 10월 16일(금) ~ 17일(토)
- ◎ 장 소 : 경희대학교(수원캠퍼스)
- ◎ 내 용 :
  - 16일(금) : 논문발표, 초청강연, 튜토리얼, 정기총회
  - 17일(토) : 논문발표
- ◎ 문의전화 : (02)593-2894 팩스 (02)593-2896
- \*자세한 내용은 발송된 팸플렛 참조 바람