

# 한글·조선글 문자부호표준 개선방안



## 목 차

- 1. 서 론
- 2. 한글과 조선글의 차이
- 3. 훈민정음 창제원리
- 4. 정보교환용 부호 분석
- 5. 한글 및 조선글 부호계 개선 방안
- 6. 결 론

변 정 용  
(동국대학교)

## 1. 서 론

북한은 인터넷을 내부에서는 사용하고 있지만 아직 국제적으로 연결되어 있지는 않다. 만약 인터넷이 지금 연결된다고 할 때 이를 이용하여 남북 사이에 전자우편을 이용하여 정보교환을 하는 데는 문제가 있다. 문제의 요인은 기본적으로 남북의 어문체계에서 자모의 순서가 다르다는데 있다. 따라서 남쪽의 한글과 북쪽의 조선글의 문자부호(코드)는 서로 달라지기 때문에 그러한 결과가 일어난다. 자모순서가 다르게 되면 문자에 숫자를 매겨서 부호화를 하게 되므로 같은 글자임에도 불구하고 컴퓨터는 부호가 달라서 다른 글자로 인식하게 된다.

지난 2000년 615공동선언을 낳은 남북 정상회담은 다른 나라와 정상회담을 할 때 들이는 시간과 비교하여 두 배의 단축 효과가 있었다고 한다. 당연한 일이다. 통역을 둘 필요가 없었기 때문이다. 그렇지만 현재 남북 사이에 전자우편을 통한 정보교환을 하려면 통역에 해당하는 문자부호

변환 소프트웨어가 반드시 있어야 한다. 이것은 남북간에 정보교환을 해야 하는 일이 많아지면 상당한 문제를 일으키게 된다. 정보교환은 물론 각종 소프트웨어의 호환성도 함께 사라지게 된다. 이것은 남북간 협력에서 생산성과 경제성을 떨어뜨리는 결과를 낳을 것이다.

본 논문에서는 남북간 어문체계에서 자모순서의 실체를 알아보고, 그 결과로 빚어진 양측의 국가규격인 KS X 1001과 KP S 9566 정보교환용 문자 부호계, 그리고 KS X 1005 (ISO 10646 BMP, 유니코드)의 문제점을 살펴본다. 그리고 문제점에 대하여 한글을 수학적 대상으로 본 연구[1]에 기반하여 훈민정음 창제원리의 적용 등 여러 가지 해결 방안을 찾아 보도록 한다.

## 2. 한글과 조선글의 차이

### 2.1 명칭의 문제

남쪽에선 한글이라고 하지만 북쪽에서는 조선글이라고 달리 부른다. 둘 모두 나름대로 타당성

은 가지고 있다. 둘은 1443년에 창제하고 1446년에 반포한 훈민정음을 뿌리로 하고 있으며, 1527년 최세진의 훈몽자회를 거쳐서 1933년 조선어학회에서 제정한 한글맞춤법 통일안에 뿌리를 두고 있다. 한글맞춤법 통일안은 1948년 대한민국 정부에서 공식적으로 채택하여 한국정서법의 법전이 되었다. 현재 국제문자부호계인 KS X 1005와 ISO 10646 BMP 그리고 유니코드에는 한글(Hangul)이란 용어로 등록이 되어 있다. 조선글 역시 1948년 「조선어 신철자법」, 1954년 「조선어 철자법」, 1966년 「조선말 규범집」, 1988년 「조선말 규범집」으로 이어지면서 북한의 정서법이 되었다.

한글이란 명칭의 기원에 있어서 박승빈은 1910년(융희 말년) 조선광문회에서 최남선이 짓고 주시경이 찬동하여 ‘한글’이란 이름이 확정된 것으로 기록하고 있다. 김민수는 ‘한글’의 ‘한’은 ‘大韓帝國’의 ‘韓’에서 따낸 것이며, 그 의미는 이운재, 최남선과 같이 ‘大’와 ‘韓’의 두 가지를 부여하였다. 그는 또 한글의 사용 연대와 쓰인 사례를 최남선보다 더 분명히 하였으니 “「아이들보이」誌(1913. 9. 창간) 「한글 풀이」란에 처음 보인다”고 말하고 있는 것이다. 이를 근거로 북쪽에서는 한글의 ‘한’이 한국의 ‘한’과 연관되므로 받아 들일 수 없다고 하고, 그에 앞서 사용한 조선글이 옳다고 주장하고 있다. 남쪽의 입장에서도 조선글은 북쪽의 국가명칭과 관계가 되며 한글은 북한이 정부가 세워지기 전부터 이미 널리 사용되어 왔다는 점에서 같은 입장을 가지게 된다. 한글 명칭은 언문, 반절, 본문, 국문 등을 대신하는 명칭이며, 조선글은 나라 이름이 조선일 때 조선의 문자라는 일반명칭이라고 볼 수 있다. 북한은 ISO 10646 BMP에 사용된 Hangul 용어에 대하여 Korean Character로 해야 한다고 여러 차례 개정을 국제표준기구(ISO)에 요청한 바가 있으나 한국의 입장도 있기 때문에 받아 들여지지 않았다.

문자의 호칭을 한글이나 조선글 가운데 어느

것으로 할 것인지는 역사성을 따져서 결정할 문제는 아니며, 남북간의 정치체제와 연관을 가지기 때문에 어려운 문제로 본다.

## 2.2 자모순서

문자의 부호화는 문자에 숫자를 부여하는 것이므로 자모 순서는 부호를 결정 짓는 가장 중요한 요소이다. 한글과 조선글은 자모순서가 다르기 때문에 자연 부호가 다르게 된다. 둘은 모두 1933년 조선어학회가 제정한 한글맞춤법통일안에 근거하며 기본 자모 24자는 같다. 또한 복자음, 복모음의 수도 서로 같지만 순서의 나열 방식에서는 다르다. 1933년 한글맞춤법통일안에는 24자 자모의 순서만 정하고 복모음과 복자음이 이러한 것이 있다고만 했지 그들을 전체적으로 어떻게 나열할 것인지에 대하여 언급이 되어 있지 않다. 여기서 광복 후 분단이 되면서 전체 자모순서를 만들 때 각각의 입장을 다르게 가질 수 있게 된 것으로 본다. 여기서 한글은 주로 글자모양을 기준으로 삼은 자형순서로 정하고, 조선글에서는 음의 세기를 기준으로 하는 음가순서로 나열순서를 정하고 있다.

자모순서를 통일하는 일은 남북의 어문규범을 통일하는 일이 되기 때문에 쉽지 않은 일이다. 무엇보다도 국가차원의 제도에 매여있기 때문이다. 남쪽의 어문규정은 「한글 맞춤법」, 「표준어 규정」, 「외래어 표기법」, 「로마자 표기법」 등이 있으며, 자모순서는 「한글 맞춤법」 제2장 4항에 규정되어 있다. 조선어 맞춤법에서 조선글 자모순서를 규정하고 있는데 자음, 모음, 받침의 경우에 북한은 「조선말 규범집」(1988)에 나와 있는 맞춤법에 따라 자모의 배열 순서를 정하고 있다. 여기서 조선글에서 기본 자모의 순서는 한글 맞춤법의 기본 자모순서와 꼭 같다.

조선글에서 정해 놓은 자모의 숫자와 기본 글자의 배열 순서는 한글과 같다. 하지만 초성 ‘ㅇ’의 경우는 소리가 없다고 해서 자음으로 인정하지 않지만 받침으로 쓰이는 ‘ㅇ’의 경우는 한글과

<표 1> 기본 자모

홀 자 음 (14자)	ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅇ, ㅈ, ㅊ, ㅋ, ㆁ, ㅍ, ㅎ
홀 모 음 (10자)	ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ, ㅣ

<표 2> 조선글 자모 및 받침

자음 (19자)	ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅈ (ㅇ) ㅊ ㅋ ㆁ ㅍ ㅎ ㆁ ㄷ ㅃ ㅅ ㅆ
모음 (21자)	ㅏ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅠ ㅡ ㅣ ㅐ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅠ ㅡ ㅣ ㅐ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅠ ㅡ ㅣ
받침 (27자)	ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅅ ㅈ ㅊ ㅋ ㆁ ㅍ ㅎ ㆁ ㅇ ㅈ ㅊ ㅋ ㆁ ㅍ ㅎ ㆁ ㅃ ㅅ ㅆ

마찬가지로 ‘ㅅ’ 다음에 두고 있다. 한글은 ‘ㄱ, ㄴ, ㄷ, ...’순이지만 조선글은 ‘ㄱ, ㄴ, ... ㅎ, ㄱ, ㄷ, ㅃ, ㅅ, ㅆ’ 순이다. 모음도 겹모음은 뒤에 오고, 받침의 경우 한글과 같지만 된소리들은 마지막에 두고 있다. 예를 들어 한글은 ‘ㄱ, ㄴ, ㄷ, ㄹ, ㅁ... ㅎ’이지만 조선글은 ‘ㄱ, ㄴ, ... ㅎ, ㄱ, ㅆ’순이다[14]. 이러한 차이가 있는데 가운데 남북간의 자모순 통일 노력은 1994년 이래 여러 번 개최된 남북 학술회의에서 북한은 ISO에서 조선글의 입장을 반영하는데 관심이 많았고 우리도 남북간 부호통일의 중요성이 있었기 때문에 아래의 중간 안을 만드는 결과에 까지 이르렀다. 2001년에 진행된 학술회의에서는 “ISO 등 국제기구에 등록하기 위한 글자의 자모배열순서는 1996년도 합의안을 기초로 하며 옛글자와 현대글자를 따로 배열한다.”고 하고 현대어 자모배열순서를 다음과 같이 정한다고 하였으나 그 후 어떠한 실천도 이루어지지 않았다[17]고 북한의 김명규(당시, 조선컴퓨터센터 실장)는 평하고 있다.

<표 3> 자모순 남북공동 합의안

자음 (30)	ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅈ, ㅊ, ㅋ, ㆁ, ㅍ, ㅎ, ㆁ, ㄷ ㅃ, ㅅ, ㅆ, ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ, ㅣ
모음 (21)	ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ, ㅣ, ㅐ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ ㅜ, ㅠ, ㅡ, ㅣ

또한 김명규(과학원 실장)는 북한의 코드 전문가로서 2001년 학술대회에서 합의된 것처럼 우리글을 제대로 지원할 수 있는 1 바이트 조합형 부호계를 공동으로 만들어 실지로 조작체계에 실장하여 이 부호계를 리용하는 경우에 제기될 수 있는 문제점들을 실증시험을 통하여 밝혀내며 그를 국제등록부에 공동으로 등록하는 사업을 실현시켜야 한다[17]고 제안하고 있다.

3. 훈민정음 창제원리

여기서 훈민정음 창제원리에 대한 기존연구[5]를 살펴본다. 해례 서문의 첫 문장은 “有天地自然之聲, 則必有天地 自然之文”이다. 즉 “천지자연의 소리가 있으면 반드시 이를 적을 수 있는 천지자연의 글자가 있다”라는 뜻이다. 다시 말하면 천지자연의 모든 소리를 표현할 수 있는 글자를 창제하겠다는 뜻을 담고 있다. 이 원리는 간단하여 배우기 쉽기로는 “지혜로운 이는 아침 먹기 전에 어리석은 이라도 열흘이면 다 배울 수 있다”고 하였다. 천지자연의 모든 소리를 기록할 수 있는 문자이면서도 쉽게 배울 수 있다는 말은 당시의 주류 문자였던 한자를 생각하면 도저히 불가능한 일이다. 필자는 다음과 같이 해석한다.

- 제자해에서 자음 다섯 자는 사람의 구강구조원리를 파악하여 주 발성기관 다섯 개를 대표하여 아설순치후(ㄱ, ㄴ, ㄷ, ㅌ, ㅇ)로 정하고 이들이 조화를 통하여 모든 소리를 이루어낸다는 원리를 터득한 것이다. 모음 석 자는 소리가 목구멍에서 발원하여 혀 위를 지나서 아래 윗니 사이로 빠져나가는 모양을 천지인(· ㅡ ㅣ)로 나타내었다.
- 초성해에서 자음 5자는 각각 소리의 거침에 따라서 초성자 17자로 확장된다.
- 중성해에서 모음 3자는 같은 원리에서 11자로 확대된다.
- 중성해에서 중성은 초성을 다시 쓴다. 여기서 원래는 45자이지만 초성해와 중성해는 비록 음가는 다르다 할지라도 모양에서 같은 원소이브

로 17자가 축약되어 28자가 된다. 이는 수학의 동치관계 정의 또는 집합의 원소 개념이다.

- 합자해에서는 초성, 중성, 종성의 글자들이 2 또는 3자씩 합용병서 또는 각자병서할 수 있다고 하면서 모두 규칙으로 정의하고 있다.

[연음법] 순음아래 ㅇ 을 붙이면 입술 가벼운 소리가 된다.

[병서법] 자음과 자음을 2 또는 3자 병서로 확대한다.

- 초성글자 |C1| = 17 + 2π17 + 3π17 = 5119자
- 중성글자 |V| = 11 + 2π11 + 3π11 = 1463자
- 종성글자 |C2| = 자음과 동치 = |C1| + {A} = 5220자

[부서법] 가로꼴 모음은 초성 아래 붙여 쓰고, 세로꼴 모음은 초성 오른 옆에 쓰라. 모아쓰기를 하는 방법을 정의한 것이다.

[성음법] 무릇 글자는 초성, 중성, 종성이 모여야 소리를 이룬다. 다시 말하면 소리마디 단위로 모아 쓰기를 하는 이유를 언급한 것이다. 이것은 로마자처럼 풀어쓰기 하는 것 보다 읽기의 효율성을 높여준다.

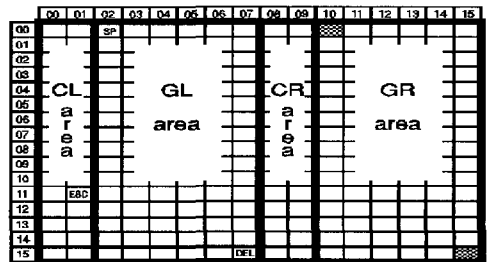
- 훈민정음의 총 음절자 |H| = C1 x V x C2 ≅ 39,856,772,340 (약 399억) 글자

요약하면 훈민정음창제원리는 입안의 5개의 대표 구조가 소리가 목구멍에서 발원하여 입 밖으로 빠져나가는 과정에서 조화를 이루어서 내는 다양한 소리를 모형화하여 만든 기본 기호 45자를 중성부용초성이라는 동치관계에 따라 줄인 28자와 연음법, 병서법, 부서법, 성음법 등의 규칙으로 표현한 위대한 창제로서 약 399억 음절글자를 표현할 수 있다. 이것은 오늘날 수학으로 말하면 전체문자집합을 조건제시법으로 정의한 것이라 할 수 있다.

#### 4. 정보교환용 부호 분석

##### 4.1 한글 정보교환용 부호 KS X 1001

한글 정보교환용 부호는 1974년에 처음 제정되었다. 규격번호는 KS C 5601:1974이다. 이 때는 자음 19자, 모음 21자, 받침 27자를 포함하였다. 부호화 대상은 자음과 모음이므로 한글 자모형 부호계라 한다. 이 부호계는 ISO 646IRV와 ISO 2022 규격에 부합하지만 당시로는 한 음절의 길이가 일정하지 않아서 로마자나 가나를 다룰 때와 비교하여 매우 불편하다고 생각하였다. 그래서 1982년에 자소 정보를 가진 두 바이트의 조합 음절형 부호계로 개정된 바 KS C 5601:1982이다. 이 부호계로는 현대 한글에서 조합 가능한 11172 음절을 모두 표현할 수 있게 된다. 하지만 당시 행정전산망을 추진하면서 국제 규격에 맞지 않는 문제가 심각하게 되어 역시 두 바이트인 완성 음절형 부호계로 개정되니 곧 KS C 5601:1987이다. 이 부호는 나중에 KS X 1001:1998로 개정된다.



(그림 1) ISO 646 제어 및 도형문자 구역

(그림 1)을 설명하면 GL은 왼쪽 도형문자 구역으로 MSB 비트 값이 0인 지역 즉 0번 열에서 7번 열은 128개 문자를 표현할 수 있는데 0,1번 열에 배정된 제어문자 32개를 제하면 96이 되고, 다시 공간 SP와 삭제 DEL을 빼면 94가 된다. GR은 오른쪽 도형문자 구역으로 8번에서 F번 열에 128개 문자를 똑같이 표현할 수 있으며 제어문자와 SP, DEL를 제외하면 역시 94가 남는다. 두 바이트 부호계에서 G0에 로마자를 지정하고 G1에 KS X 1001 두 바이트 부호계를 배정한다

음 GL에 G0를 호출하고, GR에 G1을 호출하면 로마자와 한글을 동시에 사용할 수 있다.

KS X 1001:1998은 ISO 2022 코드 확장규격에 따르기 때문에 두 바이트로 표현하더라도 한 바이트에 도형문자 표현 구역이 최대 94자만을 표현할 수 있기 때문에 8836(94 x 94) 글자 밖에 표현하지 못한다. 한글 완성 음절형 부호를 채택하였기 때문에 현대 한글에서 조합 가능한 11172자를 표현하는 것이 사실상 불가능하다. 여기에 한자를 표현해야 한다는 요구까지 있었기 때문에 제한된 문자 집단을 표현할 수 밖에 없었다. 그래서 일련의 한글 문서에서 빈도수 통계에서 0.001% 이상의 글자 2350음절을 선정하고, 한자는 겨우 지명 정도를 표현할 수 있는 4888자를 선택해서 포함하도록 하였다.

이 부호계는 그 후 ‘똥방각하’라는 드라마 이름에서 ‘똥’자를 표현할 수 없음을 알게 되었는데 전문가의 연구에 따르면 일상생활에서 사용하는 약 329가지 경우의 글자가 부족한 것으로 드러났다[2]. 한자 문자는 더욱 심각하여 행정전산망 시스템에서 한자 4888자가 지원됨으로 인하여 작명을 할 때 이 범위에서 하도록 기술이 문화를 규제하는 비극을 초래하게 되었다. 여기서 소프트웨어 업계의 회비는 ㈜한컴의 워드프로세서가 마이크로소프트의 워드 보다 유리한 입지를 갖게 하였다. 전자는 기존의 조합 음절형 코드를 계속 지원하여 11172 음절을 지원하였고, 후자는 KS C 5601:1987을 지원함으로써 2350음절의 문제로 개별적 기능이 우수함에도 경쟁자에서 불리하였다. 후자는 1995년 8836(94 x 94) 도형문자 구역을 벗어나서 2350음절 이외의 11172에서 빠진 글자를 채워서 이른바 통합형 코드를 만들어서 11172음절을 지원하게 되자 시장이 확대되기 시작하였다. 이것은 가나다순서 정렬을 할 수 없다. 그 후 논란으로 ISO 10646이라고 이름을 바꿔서 부르면서 논란을 눈속임으로 피하긴 하였지만 내막적으로 통합형 체계를 그대로 유지해 오고 있다.

그 후 ISO 10646이 전세계 문자부호계의 통합을 위하여 노력하였고, 때마침 미국에서도 아스키로에서 유니코드로 확장하는 논의가 시작되어 1992년에는 16비트 유니코드와 ISO 10646 BMP (Basic Multilingual Plane)가 통합하게 된다.

그 때 ISO/IEC의 국내 JTC1/SC2에서는 1992년 6월 서울회의에 제안할 한글 부호계의 국가안을 준비 중에 있었다. 여기서 저자는 그 동안 훈민정음 창제원리에 기반하여 연구해 온 정음형 부호계를 1992년 1월에 제안하였다. 그리고 그 해 3월 서울교육문화회관에서 합숙하며 국가안을 준비에 참여하였다. SC2 전문위원 가운데 한 그룹은 완성 음절형으로 정해야 한다고 하고, 다른 한 그룹은 조합형 코드의 당위성을 내세우며 대립하고 있었다. 그리고 후자들은 미국의 유니코드 소위원회 위원인 애플사의 마크 데이비스(Mark Davis, 현재 유니코드 컨소시엄 회장)가 제안한 공간-비공간 문자(Spacing-Non Spacing Character) 방식을 검토하고 있었다. 즉 독일어의 움라우트 문자나 불어의 악센트 표시 문자에서 A는 공간문자이고 움라우트 표시는 비공간 문자로써 겹쳐 쓰기를 한다는 개념인데, 한글의 경우 ‘각’에서 초성 기억은 공간 문자이고 나머지 ‘ㅏ’와 받침 ‘ㄱ’은 비공간 문자로 겹쳐 쓰기를 하면 된다고 하는 제안을 말한다. 마크 데이비스의 의견은 한국측에서 현대한글 11172음절 부호 공간을 요구하고 있었기 때문에 음절형 부호로 하지 말고 자모형 또는 자소형 부호로 하라고 제안한 것이었다. 한국의 전문가들도 그 때까지 한글 문제를 풀기 위하여 훈민정음의 과학성을 도입해야 한다는 생각을 하지 못하였다. 저자는 1987년에 한국전자통신연구원의 행정전산망주전산기 개발팀을 지원하면서 한글코드소위원회에 참여하고 있었다. 당시 정보교환용 한글 완성 음절형 부호에 대한 관련 규정 및 이론 연구를 하면서 훈민정음해례본에 있는 훈민정음창제원리를 접하게 되었고, 한글의 과학적 이론을

규명하게 되었다[3,5]. 당시 국가안 준비 위원들에게 훈민정음창제원리에 관한 세미나를 통하여 데이비스안은 한글의 과학성과 관계가 없다는 점을 이해시켰다. 그 후 한글의 과학성을 담은 정음형 코드가 채택되었다. 그 후 문자 집합이 238자로 확장되긴 하였지만 초성자, 중성자, 종성자라는 골격을 유지한 채 1992년 6월 ISO/IEC JTC1/SC2에서 채택되어 국제표준에 반영되었다. 자소형 부호에 대한 일부 연구는 공병우의 3벌식 타자기에 기초하여 연구된 바가 있는데 훈민정음의 과학적 원리에 기초하지 못하여 과학적 바탕이 미약하고, 자소형 부호를 의미가 없는 ‘새로운 조합형’으로 부르다가 최근 ‘첫가끝’이라 바꾼 것은 나름의 의미가 있다고 본다.

#### 4.2 조선글 정보교환용 부호 KP X 9566

북한의 국규인 조선글 정보교환용 두 바이트 부호계 KPS 9566:1993를 처음 접한 것은 1996년 남북학술회의 때였다. 해설문이 없는 코드 표만 볼 수 있었다. 이 규격에는 한자가 없었고, 조선글 음절은 숫자상으로 한글 음절보다 약 70자 정도가 많아서 2420자였다. 이 부호계로 구현된 소프트웨어가 있는지는 확인되지 않았다. 3차에 걸친 남북간 학술회의는 북한이 국제문자표준에서 남쪽 일방의 국제표준 방영에 대한 불만으로 공동안 마련을 위하여 노력하는 가운데 KS X 1001의 일부 약점을 파악하게 된다. 이러한 내용들이 적어도 상당한 부분 반영되어 개정된 것이 조선글 정보교환용 두 바이트 부호계 KPS 9566:1997이다. 해설에서 조선글 음절글자 부분은 국규 9566-93 정보기술용 조선글자의 부호의 1수준의 글자에 사용빈도가 높은 글자 259자를 추가하여 2679자로 하였다. 이 규격에서는 종전의 제2수준을 쓰지 않는다. 조선한자는 4653자를 해당 한자의 조선말 발음순서에 따라 배열하였으며 소리가 같은 조선한자는 그것들을 부수순, 획순에 따라 배열하였다. 같은 조선한자에 대하여 소리가

여러 가지인 경우 대표적인 소리에 따라 배열하였다. 사용자 정의글자에 대해서는 정보교환에서 완전한 호환성을 담보하지 않는다'[15]고 하였다. 이 부호계는 KS X 1001의 2350음절을 부분 집합으로 포함하고 있으며, 약 329자의 음절이 더 많다. 이 부분은 이미 김중희의 연구[2]에서 336개의 일상생활 용어가 부족하다고 지적한 바가 있는데 숫자 상으로 유사하다. 한자도 포함되어 있는데 KS X 1001에 있는 4888자 가운데 같은 글자 동음이의어 惡은 ‘악’과 ‘오’로 발음되는데 두 번 반복되어 있다. 한자는 남한이 4888자이고 북한은 단순 숫자에서 보면 235자가 적다. 그러나 실제에 있어서 KS X 1001에만 있는 한자는 175자이고 KPS 9566에만 있는 한자는 81자이므로 표면상으로는 256자가 서로 다른 셈이다. 현재 KPS 9566:2000이 출판되어 국제등록부에 등록되어 있다.

<표 4> 정보교환용 한글 부호계와 조선글 부호계의 비교

한글(KS X 1001)		조선글(KP9566-97)	
문자종류	문자이름	문자수	문자수
특수문자	간격문자	1문자	1문자
	기술기호	37문자	37문자
	괄호기호	23문자	46문자
	학술기호	48문자	54문자
	단위기호	93문자	94문자
숫자	일반기호	230문자	105문자
	아라비아숫자	10문자	10문자
	로마자숫자대문자	10문자	10문자
한글자모	로마자숫자소문자	10문자	10문자
	자음	19문자	19문자
	복자음	11문자	11문자
	모음	21문자	21문자
옛글자모	옛글자모	42문자	65문자
한글음절자	완성형	2350문자	2675문자
한자(漢字)	완성형	4888문자	4653문자
알파벳	대문자	26문자	26문자
	소문자	26문자	26문자
그리스문자	대문자	24문자	24문자
	소문자	24문자	24문자
러시아문자	대문자	33문자	33문자
	소문자	33문자	33문자
일본문자	히라가나	83문자	83문자
	가타가나	86문자	86문자

### 4.3 국제문자 부호계 KS X 1005

KS X 1005:1995는 ISO 10646의 기본다국어판 및 유니코드 3.0과 일치한다. 2007년에는 유니코드 표준 5.0이 발표된다. 이 부호계는 ISO 646과 ISO 2022 부호확장법의 불편함을 해결하기 위하여 16비트 공간 전체를 사용할 수 있도록 한 부호계로서 총 65536글자를 표현할 수 있다. 이 공간에는 전세계의 모든 문자들이 포함되어 있다. 한글은 한글자모(Hangul Jamo), 한글호환 자모(Hangul Compatibility Jamo), 한글음절(Hangul Syllable)이 포함되어 있다. 여기서 한글은 구현 수준 1로 규정되어 있기 때문에 한글 자모나 한글 호환 자모로 글자를 모아쓰기 할 수 없다. 엄밀하게 말하면 한글 자모로 조합할 수 있는 글자가 무려 수십만 글자임에도 11172 현대한글만을 쓸 수 있고, 옛한글은 모아 쓰기를 할 수 없는 것이다. 한글자모와 한글 호환자모는 자음과 모음을 표현할 때 사용하거나 한글음절을 구성하는 중간단계 즉 입력자동틀에서 사용할 뿐이다. 호환자모는 저자가 처음 제안한 내용이 반영된 것이며, 한글 자소형 부호계이다. 한글 호환 자모는 한글 자모형 부호계이다. 그리고 한글음절은 현대 한글 11172자를 모두 포함한 한글 완성 음절형 부호계이다. 유니코드에는 한글의 모든 부호계가 존재한다. 특히 현대 한글은 11172개 공간을 차지하게 되어, 훈민정음에서 28자의 기본 자모로 약 399억 음절글자를 조합할 수 있는 과학적인 우수한 문자라는 특성을 모두 없애버린 효과를 가진다. 3가지 문자 부호계를 표현할 수 있는 문자의 범위를 보면 한글 자모만으로도 충분하다. 그런데 한글 음절 부호를 쓰면 구현 수준 1이기 때문에 단순한 면은 있지만 음절의 속성을 알 수가 없다. 물론 현재 역산을 통하여 초성, 중성, 종성의 속성을 알아낼 수 있는 함수를 개발해 두고 있기는 하지만 한글의 기본적 특성을 헤치게 된다.

- 한글음절: U+AC00 - U+D7AF, 11172 현대 한

글 음절

- 한글자모: U+1100 - U+11FF, C1 90 x V 67 x C2 81 = 488,430글자

- 한글호환자모: U+3130 - 318F, 현대 한글 자음 30개+모음 21개+옛한글 자음 35개 + 모음 8개 = 94개

## 5. 한글 및 조선글 부호계 개선 방안

훈민정음 창제원리에 비추어 볼 때 한글 및 조선글 부호계는 자모형 또는 자소형이어야 한다. 그 가운데 자모형이 훈민정음 창제원리에 대한 적합성이 높다. 자모형은 음절구조 정보가 없으나 자소형은 음절구조 정보를 가지고 있다. 자소형은 훈민정음에서 종성부용초성은 집합론이나 동치성에 해당하므로 기본 45자에서 28자가 된다. 자모형의 문제는 합자해에 따라서 2 또는 3자 조합을 하게 되면 예를 들어 모음과 모음 사이에 ㄱ 이 3번 반복되었을 때 받침과 초성 자음의 구별이 어렵다. 하지만 자소형일 때는 명백해진다. 그리고 폰트만 지원된다면 글자의 표현 문제는 거의 해결된다고 할 수 있다. 자모형에서 지금 가장 불편하게 생각하는 것은 한 음절의 길이가 일정하지 않다는 점이다. 이 부분은 관련 함수를 개발해서 해결해야 한다. 한글은 음소문자이면서 동시에 음절문자 특성을 가지고 있기 때문에 로마문자나 가나문자의 입장에서 한글을 보지 말아야 한다. 따라서 훈민정음은 기본 자모 28자 또는 45자와 합자해에서 정의한 규칙으로 음절자를 조합하도록 하고 있다. 한글은 동남아 몇몇 글자처럼 구현수준 2나 3으로 되어야 한다. 즉 조합하는 특성을 유지해야 한다고 본다. 그러면 컴퓨터를 사용하더라도 마치 연필로 글을 쓸 때처럼 원하는 모든 글자를 표현할 수 있을 것이다. 이때 조합형 글자꼴을 지원한다면 구현 가능하다. 따라서 자소형 부호계가 가장 큰 적합성을 가지고 있다. 따라서 유니코드에서 한글 자모만 남기고 나머지는 버려야 한다.

그 다음 문제는 남북간의 자모순 차이에서 오

는 문제가 있다. 완성 음절형일 때 이것은 고착이 되어 해결이 불가능하다고 할 수 있다. 하지만 자소형이나 자모형일 때는 정렬 알고리즘 개발이 가능하다. 훈민정음은 집합론의 원소나열법이 아닌 조건제시법으로 음절 집합을 정의하였다. 그래서 가장 기본 자소만 정의하면 된다. 그러면 초성 17, 중성 11, 종성 17은 남북이 지금도 일치한다. 지금 남북의 KS X 1001과 KPS 9566 그리고 KS X 1005, ISO 10646 BMP, 유니코드 등은 병서법을 없앤 부호화의 결과로 많은 문제점을 빚고 있는데, 이들 부호들과 병서법을 비롯한 연음법, 부서법, 성음법을 소프트웨어 개발로 살리면 남북간의 자모순 문제 뿐만 아니라 표현의 한계 문제는 물론 옛한글은 현대 한글과 같은 수준에서 표현 가능하여 여러 문제가 사라지게 되어 있다.

## 6. 결론

한글 및 조선글이라는 명칭을 비롯하여 자모순서의 차이 등에 대하여 먼저 알아 보았다. 또한 정보교환용 한글 부호계와 조선글 부호계에 대해서도 살펴 보았다. 그리고 한글 및 조선글의 뿌리인 훈민정음의 창제원리를 알아 내고, 현재 한글 부호계가 합자해에서 정한 규칙을 어떻게 사용하고 있는지에 대한 상황을 알 수 있었다. 즉 현대 한글의 표현에는 문제가 없지만 구원 수준1이기 때문에 조합을 할 수 없어서 한글의 고유성 상실은 물론 현재 일부 문서편집기에서는 지원하고 있지만 옛한글을 표현할 수가 없다. 이러한 내용의 남북 및 국제문자표준에서 한글의 표현 범위, 자소정보 표현, 한글과 조선글 사이의 차이 등등을 파악해 보았다.

국제표준에서는 한글 부호계 일방으로 등록이 되어 있어서 북한은 조선글을 사용하기 위하여 자신들이 필요한 소프트웨어는 스스로 개발해야 하는 처지에 있다. 따라서 현재 입력기의 경우도 하나 7.0을 스스로 개발하였다. 이처럼 북한은 국제표준에 반영된 한글이 조선글과 다르기 때문에

자신들이 모든 것을 개발하든지 한글의 어문규정을 따르든지 변환기를 두어서 일일이 변환을 하든지 해야 한다. 우리의 입장에서 완성 음절형 부호를 채택하고 있기 때문에 한글과 조선글의 차이를 극복하기란 더욱 어려운 실정이다. 해결책을 가진다면 훈민정음 창제원리에 따라서 자모형이나 자소형 부호계를 채택한 연후에 북쪽은 정렬 알고리즘을 개발함으로써 가능할 것이다. 이 말은 훈민정음 합자해에서 정의한 규칙을 모두 살려야 일상의 문제가 줄어들고 남북의 요구도 쉽게 해결할 수 있다. 따라서 국제문자표준규격에서 3가지 부호계는 한글자모만 남기고 나머지는 모두 공간을 반납하는 것이 옳다고 본다. 한글 부호계는 한글의 과학성을 저해하지 않고 문자처리나 한국어 처리에서 생산성을 높여주는 역할을 해야 한다. 그렇게 하려면 음절의 속성이나 문자구조 정보 등을 자모형이나 자소형으로 표현함으로써 기본 글자 수준에서 표현해야 한다. 현재와 같이 모든 것을 감추거나 버린 다음 그것을 알아 내기 위하여 그 때마다 함수를 적용하는 것은 문제가 있다. 국제표준에서 현재의 한글 부호계를 바꾸는 일은 쉽지 않은 것으로 안다. 하지만 남북이 현재의 문제를 공동으로 인식하여 훈민정음 창제원리에 입각한 제대로 된 남북 양측이 합의한 문자부호계를 제정하여 제안한다면 변경을 할 수 있는 기회가 될 것으로 판단한다.

공학을 위한 과학은 반드시 수반되어야 한다. 과학적 접근을 등한시 하는 현재의 입장은 반드시 시정되어야 한다. 과학이 없는 공학은 도구의 불완전성을 초래한다. 한글공학은 훈민정음 원리에 입각한 한글과학에 기반을 두고 문제를 근본적으로 해결해야 한다. 과학적 접근이 아니라 몇 가지 현상이나 사례를 가지고 해결하려 한다면 문제는 반복되거나 계속 악화될 수 밖에 없다. 한글 및 조선글 공학은 적어도 이러한 과학적 탐구에 게을리해 왔다. 그 결과 지난 9월에도 유니코드에 뺄릴듯 몇 가지 문자집단을 추가하였다고 한다. 이러



한 일은 앞으로 계속 이어질 것이다. 왜냐면 과학적 연구에 기반하여 한글 부호계를 다루고 있지 않기 때문이다. 이 문제는 남북을 비롯한 모든 한글 과학 연구자들이 힘을 합쳐서 제2의 훈민정음 창제의 정신으로 과학적 연구와 그 결과를 바탕으로 근본적으로 해결해 나가야 할 것이다.

### 참고문헌

[1] 정희성, 수학적 구조로 본 훈민정음의 창제 원리와 구조, 제1회 한글 및 한국어정보처리 학술 논문집, 1989

[2] 김충희, 현행 KS 완성형 한글코드의 문제점, 제1회 한글 및 한국어 정보처리 학술 논문집, 1989

[3] 변정용, 훈민정음 창제 원리와 한글코드 제정 원리, 제3회 한글 및 한국어 정보처리 학술 논문집, 155-158쪽, 1991

[4] 김명규, "정보교환용 조선글자부호의 표준안에 대하여", '94 코리언 정보처리 학술대회 논문집, 1994

[5] 변정용, 훈민정음 창제원리에 기반한 공학화와 한글 코드의 발전 방향, 한국정보과학회지, 1994, 9월

[6] 허배, 본래의 한글 자모 체계, 한국어교육 11.2호, 2001

[7] 한국표준협회, 정보교환용 부호계(한글 및 한자) KS X 1001:2002

[8] 문영호, "국제표준글자부호계에 등록할 조선글자체계의 자모순서문제에 대하여", 제3회 코리언 컴퓨터 처리 국제학술대회 논문집, 1996

[9] 최병수, "컴퓨터 조선글 부호체계에 넣어야 할 옛글자에 대한 연구", 제3회 코리언 컴퓨터 처리 국제학술대회 논문집, 1996

[10] 변정용, "훈민정음 원리에 따르는 한글 코드 제정 방향", 제3회 코리언 컴퓨터 처리 국제학술대회, 1996

[11] 합의문, '96 Korean 컴퓨터처리 국제학술대회, 1996

[12] 서재진, 2001년 남북한 관계 및 북한의 개방 정책 전망, 통일IT포럼 공개 세미나, 2001

[13] 이운재, "한글강의 - 講 한글 말뚝-", 新生 제2권 제9호, 1929

[14] 이승재, "옛 글자의 배열 순서와 북한의 자모 배열 순서", 국어정보화

[15] 국가규격제정연구소, 정보교환용 조선글 두 바이트 부호계 국규 KPS 9566:1997

[16] 김명규, "우리글 부호계의 현실태와 개선방안에 대한 몇가지 고찰" '95 코리언 컴퓨터처리 국제학술대회 발표 논문집, 1995

[17] 김명규, "조선어정보처리 기반기술표준화의 현실태와 당면과제", Proceedings of ICCCKL 2004, Shenyang, China. 15-24쪽

### 저자약력



변 정 용

1980년 2월 동국대학교 컴퓨터공학과(학사)  
 1983년 2월 동국대학교 컴퓨터공학과(석사)  
 1994년 2월 홍익대학교 전자계산학과(박사)  
 1982년~1988년 한국전자통신연구원 선임연구원  
 1988년~현재 동국대학교 컴퓨터멀티미디어학과 교수  
 1986년~현재 한국정보과학회 종신회원  
 1998년~현재 한국정보처리학회 종신회원  
 관심분야 : 한글공학, 데이터베이스, 웹서비스  
 이 메 일 : byunjy@dongguk.ac.kr