

# 국제 문자 부호계에 포함된 한글 부호계의 개선 방안

변정용

동국대학교 전자계산학과

## A Method of Improving Hangul Code Systems Embedded in UCS

Byun, Jeongyong

Dept. of Computer Science, Dongguk University at Kyongju

byunyj@mail.dongguk.ac.kr

### 요 약

현행 국제문자부호계(KS C 5700-1995)에는 세 가지 종류의 한글 부호계인 완성 음절형, 자소형, 자모형이 포함되어 있으며, 이들을 적용할 때 구현 수준에 따라서 표현이 불가능하거나 혼합 부호계 표현으로 혼란이 일어날 소지가 매우 크다. 뿐만 아니라 완성 음절형 부호계를 주된 부호계로 삼고 있기 때문에 이런 기반 위에서 국어정보처리를 하는 것은 불가능하며, 옛 한글 표현은 수준을 달리하였을 때 가능하다. 본 논문은 국제문자부호계에 세가지 한글 부호계를 분석 평가하여 이들이 함께 존재할 필요성이 있는 지 여부와 제정자의 관점 그리고 각각의 한계점과 파생될 문제점을 검토하고 국제문자 부호계의 바른 나아갈 바에 대하여 2 가지 방안을 모색한다.

### 1. 서론

국제문자부호계 KS C 5700-1995에서 한글 부호계는 진통을 거듭한 끝에 현재 완성 음절형(완성형)을 중심으로 자소형 및 자모형과 함께 세 가지 부호계를 반영하였다. 여기서 완성형의 국어정보처리 적용은 이제까지 많은 연구들[1,3,4,5,8,9]에서 그 어려움이 지적되어 왔다. 이제 UCS-2를 지원하는 운영체제와 언어가 나타나고 있으며 이들 부호계를 바르게 구현하려면 한글 부호계의 성격과 제정자의 의도를 파악하는 것이 매우 중요하다. 현재 국어정보처리 응용은 문서 편집과 같은 문자정보처리에서 정보검색을 필두로 한 기계번역과 같은 언어정보처리 시대로 넘어 가고 있다. 본 논문은 이상의 몇 가지 중요한 요구에 대하여 국제문자 부호계인 KS C 5700-1995를 한글 문자가 가진 특성과 국어정보처리 각 분야의 공통적인 요구를 기반으로 분석하고, 부호계 제정 과정에서 간과한 한글 특성의 반영 부분과 구현상에서 일어날 한글 표현의 혼란과 소프트웨어 생산성에 있어서 문제점을 제기하고 이들에 대한 개선 방안을 제시하고자 한다.

### 2. 한글 문자 부호화의 분석적 검토

#### 2.1 한글의 특성과 부호화 대상

기존 연구[3,6,9]에 따르면 훈민정음 해례는 기본 28자와 몇 가지 규칙으로 되어 있는데 이것은 무려 약 399의 음절자를 조합할 수 있다고 한다. 이것은 기본 문자 집합과 음절 구성에 대한 5개의 생성 규칙을 규정하고 있다. 하지만 현재 우리는 1930년 조선어학회가 제정한 한글 맞춤법에 인식의 바탕을 두고, 여기서 규정된 초성 자음 19개, 모음 21개, 중성 자음 27+1(받침 없음

)계로부터 조합한 11172음절자에 집착하고 있다. 이것은 훈민정음 해례에서 정의한 규칙을 일부 없애고, 옛글자 4자를 배제함으로써 훈민정음 원리를 깨뜨리고 있다. 현재 상태로 보면 부호화 대상은 자음과 모음, 음절(조합, 완성), 음절자 요소인 초성, 중성, 종성 등 4가지가 된다. 여기에 합용병서의 경우를 더하면 6가지가 된다. 여기서 우리가 깨달아야 할 중요한 사실은 한글 맞춤법이 훈민정음의 부분 집합이란 점이다. 그러나 우리는 훈민정음 이래 사용한 모든 글자를 표현해야 한다는 점에서 훈민정음 해례에 기준을 두어야 한다.

#### 2.2 부호화 대상의 변천과 표준 규칙

1974년 첫 정보교환용 한글 부호계인 KS C 5601-1974는 조선어학회의 한글 맞춤법에 충실하게 자음과 모음을 부호화 대상으로 삼았다. 하지만 문제는 당시 한글 맞춤법이 훈민정음의 문자 구조 원리의 부분 집합을 정의한 점에 있다. 1982년 개정에서는 초성,중성,종성 정보를 가진 조합 음절을 대상으로 해서 KS C 5601-1982를 제정하였다. 이 부호계는 ISO 2022 확장 규격에 따르지 않고 있다. 따라서 1987년 정부는 행정전산화 계기로 2350음절만을 지원하는 완성 음절을 대상으로 한 부호계 KS C 5601-1987를 제정하여 채택함으로써 현재 국내 표준으로 완전하게 정착되었다. 반면 언어 처리 응용 분야에선 자소 정보를 얻을 수 없었기 때문에 기존의 자모형이나 조합 음절형 부호계를 사용하거나 새로운 부호계 만들어서 사용하고 있으며, 한국 과학기술원의 '3바이트 부호계[8]'와 훈민정음 원리를 그대로 담고 있는 '정음형 부호계[3]'가 그 예이다. 이어서 1992년에는 조합 음절형 부호계가 복수 표준이 되었다. 1995년 MS사는 통합형 한글 부호계를 윈도

95에 구현하여 논란이 되었고, 그 후 정부는 UCS를 수용하여 KS C 5700-1995를 제정하게 되었다. 그 부호계에는 자모형과 240자 자소형, 그리고 11172자 완성 음절형이 포함되어 있다. 이 사실은 결정적으로 표준화의 근본 취지에 어긋난다. 까닭은 부호계의 단일화보다 오히려 부호계의 다중화로 유도하는 꼴이기 때문이다. 부호화 대상의 변천 과정에서 파악되는 제정자의 의도는 완성 음절을 기반으로 한 문자처리에 집착하고 있다. 특히 완성형은 자소 정보를 제공하지 못하며 이것은 한글문자의 음소문자 특성을 없애는 것이며, 옛 한글 표현을 배제하고 있다. 한글 부호화 대상의 이러한 변천은 문자론적 통찰력의 결여에 기인한다. 한글은 음소문자이면서 동시에 음절문자이기 때문에 로마자와 같은 음소문자나 가나와 같은 음절문자의 문화권에 편향된 시각으로 한글 문자를 보기 때문에 혼민정음의 과학적 완전성을 깨뜨리고 있다.

**3. 국제문자부호계에 포함된 한글 부호계**

**3.1 문자의 표현 범위**

정부는 요구가 있을 때마다 그 요구에 맞는 여러 한글 부호계를 제정하여왔다. 그 결과 전체 요구를 만족하는 최적 부호계를 제정하지 못한 채 국제문자 부호계에 그대로 반영되어 있다. 또한 KS C 5700-1995의 규격 해설에는 편집기나 워드프로세서의 요구인 문자 처리 응용이 중심이 되어 있다. 이것은 정보산업 인종이 다양화되고 있으며, 국어정보처리 응용의 중심이 문자처리에서 언어처리로 그 중심이 이동하고 있다는 사실을 간과하고 있다. 더구나 언어 정보 표현의 완전성을 위하여 날 자소 정보를 부호계로부터 제공하는 것은 매우 중요하다. 그동안 선진 소프트웨어를 도입하여 한글화 하는 과정에서 로마자권이나 가나권에서 개발된 소프트웨어를 한글화하는데 한글의 개별성을 적용하지 못하고, 반대로 한글 문자를 그들 문자 특성에 맞추려는 데 노력한 결과로 완성형을 제정하였다.

표 1. 한글 부호계의 글자 표현 비교

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

**3.2 구현 수준 3과 한글 표현**

한글 문자 표현에서 현대 한글과 옛한글을 함께 다루어야 한다. (주)한컴의 워드프로세서 HWP가 MS 워드에 대하여 시장 점유율이 높은 이유는 옛한글을 현대 한글과 같은 수준에서 지원을 하며, 한글 맞춤법에서 지원하는 모든 문자 집단을 지원하기 때문이다. 그런데 최근에 제정된 국제문자부호계에서 옛한글은 구현 수준 3의 조합 방식을 사용할 때만 표현 가능하다. 구현 수준 3에서는 조합 방식을 허용하므로 자소형과 자모형은 물론 완성형을 포함한 혼합된 부호계로 음절 구성이 가능

하다. 예로 ‘한’은 다음과 같이 세 가지 부호계로 표현할 수 있다.

- 11172자 완성형(2바이트): D55C
- 240자 자소형(6바이트): 1122 1161 11AB
- 51자 자모형(6바이트): 314E 314F 3134

이들 문자열에 대하여 출력 오토마톤(automaton)에서는 모두 ‘한’이라는 글자표를 화면에 표현해 주어야 한다. 그리고 문자열 비교에서 이들이 서로 같은 글자임을 인식하고 그렇게 처리되도록 해 주어야 한다는 점이다. 이것은 바로 세 가지 부호계를 사용하고 있기 때문에 일어날 수 있는 경우이다.

**3.3 한 음절자에 대한 혼합 부호계 표현**

구현 수준 3에서 최악의 경우 한 음절자를 표현함에 있어서 세 가지 부호계를 혼합하는 경우를 보자. 첫째, 완성형의 받침 없는 음절자와 자모형 또는 자소형의 받침이 조합되는 경우가 있다. 둘째, 경우는 자모형과 자소형을 혼합해서 한 음절자를 조합할 수 있다. 복자음 또는 복모음을 사용하는 경우에는 더욱 더 복잡해 질 수 있다. 이러한 경우를 모두 따져 볼 때 UCS에서 세 가지 부호계를 표현하는 것은 한 편으로 모든 가능성을 부여해 놓은 것 같지만 잘 못 사용하였을 때 그 혼란은 매우 심각해 진다. 여기서 받침이 없는 경우에는 ‘하’를, 받침이 있는 경우 ‘한’의 경우를 보자.

(1) 받침 없는 음절의 경우

- 자소형 ㅎ + 자모형 ㅏ : 1122 314F
- 자모형 ㅎ + 자소형 ㅏ : 314E 1161

(2) 받침 있는 음절의 경우

- 완성형-자모형: 완성형(하), 자모형 ㄴ: D558 11AB
- 완성형-자소형: 완성형(하), 자소형 ㄴ: D558 3134
- 자모형-자소형 :
  - 자소형 ㅎ, 자모형 ㅏ, 자모형 ㄴ: 1122 314F 3134
  - 자소형 ㅎ, 자소형 ㅏ, 자모형 ㄴ: 1122 1161 3134
  - 자소형 ㅎ, 자모형 ㅏ, 자소형 ㄴ: 1122 314F 11AB
  - 자모형 ㅎ, 자모형 ㅏ, 자소형 ㄴ: 314E 314F 11AB
  - 자모형 ㅎ, 자소형 ㅏ, 자소형 ㄴ: 314E 1161 11AB
  - 자모형 ㅎ, 자소형 ㅏ, 자모형 ㄴ: 314E 1161 3134

이처럼 세 가지 부호계를 혼합해서 음절자를 구성하면 한 음절에 대하여 8 종류의 표현이 가능하며, 여기에 세 가지 부호계 각각의 표현과 완성형과의 조합 표현까지 합하면 무려 13 종류의 표현이 가능하게 된다.

**3.4 언어 처리**

완성형을 사용하는 것은 언어 현상을 처리하거나 분석하는 언어 처리 분야 응용에서는 부적합하다 물론 완성형 부호계로부터 자소 정보 분해 함수를 적용해서 알 수 있다고는 하지만 역순 사전이라든지 ㄱ, ㅋ, ㆁ으로 시작하는 낱말을 찾는 것과 같은 정규식 적용을 알 수 없다. 언어 처리에서는 형태소 분석과 같은 경우 언어 정보가 글자의 원소 단위에 숨어 있기 때문에 부호계 부여 대상이 낱자소가 되어야 한다. UCS에서 설정 자모형이나 자소형을 사용한다고 해도 이들이 모두 복자소 또는 복자모를 부호화 대상으로 하였기 때문에 낱자소 또는 낱자모 정보를 알아 내기 위하여 부가적인 정보를 부호계 밖에서 제공해 주어야 한다.

따라서 정음형 부호계는 낱자소를 부호화 대상으로

삼고 있다. 따라서 정음형과 같이 이를 완전하게 지원하  
는 부호계가 있음에도 불구하고 단순히 문자 처리의 편  
이성에 편향되게 부호화를 하는 것은 문제가 아닐 수  
없다. 더구나 이것은 28자를 정의하고 음절 구성은 규  
칙으로 정의한 훈민정음 원리에도 크게 어긋난다. 결국  
옛 한글 표현이나 언어처리를 하려면 240자 자소형의  
사용이 불가피하고, 구현 수준 3을 써야 한다.

**4. 개선 방향**

현재의 문제를 해결하려면 두 가지 관점에서 볼 수  
있다. 하나는 현재 국제문자부호계 KS C 5700-1995 내  
에서 개선안을 모색해 보는 것이고, 다른 하나는 근본적인  
개선 방안으로 한글의 과학성에 바탕하여 모색하는 것  
이다.

**4.1 국제문자 부호계에서 개선 방안**

국제문자 부호계에는 표2에서 보는 바와 같이 세 가  
지 한글 부호계가 존재한다. 완성형은 현대 한글 11172  
자는 만족하지만 자소 정보가 없으며 또한 옛 한글을  
표현할 수 없다. 호환 한글 자모라 명명된 자모형 역시  
자소형의 부분 집합이며 현대 한글과 옛한글의 정렬과  
표현이 불완전하다. 완성형은 낱자 표현을 위하여 자소  
형을 쓰지만 구현 수준 1에서 자소형으로 조합하지 못  
하도록 하려고 구현 수준 1에서는 한글 조합을 하지 못  
하도록 규격 24항을 두고 있으며, 옛 한글 지원은 구현  
수준 3에서 허용하고 있다. 또한 완성형이 자소 정보가  
없음을 보완하기 위하여 자소분할 수식을 추가하고 있  
다[9]. 하지만 그것은 과학적인 훈민정음의 원리를 무시  
하는 처사이다. 따라서 문제의 해결을 위하여 다음과 같  
은 개선안을 제안한다.

- 첫째, 한글 완성형 부호계와 호환용 한글 자모 부호계  
를 없애고 공간을 반납한다.
- 둘째, 한글 자모로 되어 있는 자소형 부호계를 유일한  
부호계로 두고 조합 방식을 사용하도록 한다.
- 셋째, 한글은 구현 수준 1을 사용하도록 한다.

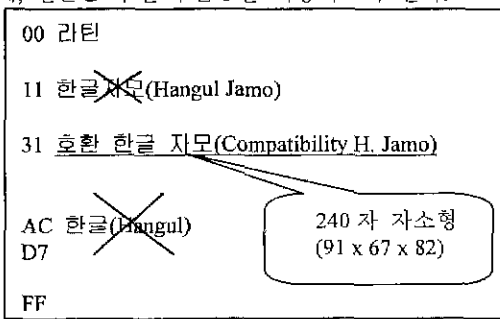


표 2 국제문자부호계의 다국어문자판(BMP)

**4.2 정음형 부호계의 적용**

훈민정음 해례에서 정의된 문자 구성 원리는 매우  
과학적이다. 즉 기본적인 문자 28자만을 정의하고 나머  
지 음절자 구성은 모두 규칙으로 표현하고 있다. 한글  
부호계는 이러한 과학성을 간과하고 한글 맞춤법에 집  
착하여 여러 가지 부호계를 제정하였다. 정음형은 ISO  
646과 ISO 2022 규격에 매우 적합하며 또한 ISO  
10646 BMP에서 구현 수준 3의 조합 문자 군으로 등록

될 수 있다. 다음은 정음형 부호계의 구현 결과이다.

첫째, 훈민정음 원리에 따라서 399억 음절자를 구성  
하므로 모든 부호계의 수피셋이다.

둘째, 모든 한글 부호계의 전체집합이므로 부호간 변  
환시 호환성이 완벽하게 보장된다.

셋째, 정음형은 낱자소 조합으로 음절을 구성하기 때  
문에 현대 한글은 물론 옛 한글을 자연스럽게 지  
원한다.

넷째, 문자집합이 작으므로 국내외 규격을 만족한다.  
다섯째, 낱자소에 대한 부호계를 가지고 있기 때문에  
언어 처리를 하는 한글 및 국어정보처리 전반에  
걸쳐 충분히 언어정보를 표현한다.

정음형 부호계는 유닉스 시스템에서 xterm과 vi에  
적용하여 정음형 플랫폼 hunterm과 정음형 편집기  
hunvi를 구현하였고, 윈도우 95에서 정음형 편집기를  
구현하여 정음형이 자소 정보를 완벽하게 표현하고 훈  
민정음 해례에서 정의한 모든 음절을 표현할 수 있음을  
입증[7]한 바 있다.

**5. 결론**

KS C 5700-1995에 포함된 세 가지 한글 부호계의  
부호화 대상과 한글 문자구조 원리를 분석하였다. 그리  
고 규격 해설집에서 구현 수준 1과 3에 대하여 발생하  
는 문제점을 검토하고 개선 방향을 모색하였다. 옛 한글  
을 표현하려면 구현 수준 1에서는 불가능하기 때문에 구  
현 수준 3을 사용해서 한글 음절 조합을 해야 한다. 그  
러면 부호계의 혼합 가능성을 낮게 되고 결과 혼란을  
일으킬 수 있음을 보았다.

따라서 국제문자 부호계에서 한글 자모(자소형)만을  
남기고 나머지는 없앤 다음 그 공간을 반납하고, 구현  
수준 1에서 조합 방식을 기본적으로 채택해야 한다. 하  
지만 한글 자모도 복자소에 대하여 부호화를 하였기 때  
문에 자소 정보를 상실하고 있으므로 보다 근본적으로  
개선한다는 차원에서 훈민정음 원리에 따르는 정음형  
부호계를 국제 문자 부호계의 한글 자모와 대치시켜서  
국어정보처리와 한글 문자 정보처리의 모든 응용 분야  
에서 만족할 수 있도록 해야 한다.

**참고문헌**

- [1] 한국표준협회, 국제문자부호계 KS C 5700-1995, 1995
- [2] 변정용, "훈민정음 창제원리의 공학화에 기반한 한글  
부호계의 발전방향", 한국정보과학회지, V12, N2, 1994
- [3] 변정용, 강진곤, "한글문자의 음소 및 음절 문자 특성  
의 구현 방안", 제 6회 한국어정보처리 학술대회, 1994
- [4] 강진곤 함경수, 변정용, "훈민정음이 살아있는 hunterm",  
한국정보과학회, '95 봄 학술대회, V22, N1, 1995
- [5] 강진항, 훈민정음연구, 성균관대학교출판부, 1990
- [6] 변정용, 한글정보처리를 위한 입력 라이브러리 연구,  
1,2,3 차년도, 시스템공학연구소, 1995, 1996, 1997
- [7] <http://www.unicode.org/unicode/reports/tr8.html>, The  
Unicode Standard Version 2.1, Technical Report #8, 1998
- [8] 박혁로 최운천, 최기선, "기존 한글코드와 제문제와 3  
바이트 코드의 고찰", '90 봄 학술발표논문집, 1990
- [9] 정희성, "수학적 구조론 본 훈민정음 창제원리와 구  
조", 한글 및 한국어정보처리 학술발표논문집, 1989