

# 국제 표준 틀 안의 한글 부호계 운용 방안 및 한 바이트 한글 부호계 안

김 경석, 정 연정, 조 충래, 부산대학교 전자계산학과  
kskim@asadal.cs.pusan.ac.kr

## A Proposal on operating ISO-conforming Hangul codes and on a new one-byte Hangul code

Kyongsok Kim, Yeonjeong Jeong, Chunglae Cho  
Dept. of Computer Science, Pusan National University

### 요약

이 글에서 살펴본 바는 다음과 같다. 1) ISO 10646-1 한글 부호계의 운용 방안은, 앞으로 얼마동안은 2,350 완성형 소리마디를 쓸 수도 있겠지만, 길게 보아서 첫가끝 부호계 238 글자로 통일하는 것이 바람직하다고 본다. 2) ISO 2022 틀을 따르면서 한글을 완벽하게 지원하는 한 바이트 한글 부호계를 개발하여, 국제 등록부 (International Register) 에 등록하여, Escape Sequence 를 받아서, ISO 2022 를 따르는 분야에서 쓸 수 있도록 하여야 하겠다. 3) ISO 10646-1 을 개정하여 4,516 소리마디를 넣자는 안은, 완성형으로만 한글을 지원하게 되어 바람직하지 않다고 본다. 모아쓰기 글자를 지원하는 세계 추세가 조합 방식이므로, 10646-1 을 현재대로 두고, 첫가끝 방식의 238 글자를 쓰는 것이 바람직하다고 본다. 4) 10646-1 을 제대로 지원할 수 있는 세벌식 자판을 정부의 복수 표준으로 채택할 것을 제안한다.

### 1. 들머리

#### 1.1. 이 글에서 다룰 내용

이 글에서는 다음의 네 가지를 다루고자 한다.

- 1) 국제 표준 ISO 10646-1 에 있는 한글 부호계의 운용 방안
- 2) ISO 2022 틀을 따르면서 한글을 완벽하게 지원하는 한글 부호계 제안
- 3) ISO 10646-1 을 개정하여 4,516 소리마디를 넣자는 안에 대한 검토
- 4) ISO 10646-1 을 제대로 지원할 수 있는 한글 자판에 대한 검토

#### 1.2. ISO 10646-1 (및 Unicode 1.1) 에 있는 국제 표준 한글 부호계

10646-1 에는, 한글이 완성형과 “첫가끝 한글 부호계 (SYL-IPF: syllable-initial-peak-final)” 의 두가지 방식이 모두 들어가 있다.

한편, 완성형은, KSC 5601 의 2,350 소리마디 [ISO 10646-KSC 5601], KSC 5657 의 1,930 소리마디 [ISO 10646-KSC 5657], 그리고 나머지 2,376 소리마디 [ISO 10646-KSC 0000] 등 모두 6,656 소리마디가 있다. 완성형이 10646 에 들어갔다고 하나, 부호값 (codeword)

는 모두 바뀌어 들어가 있다.

첫가끝 부호계는 첫소리, 가운데소리, 끝소리 글자 238 개로 옛한글까지 완벽하게 지원하는데, 이른 바 두 바이트 조합형 (또는 상용 조합형) 과는 다르다. 이 때까지 국내에서 논란이 되어왔던 완성형대 두 바이트 조합형 사이의 대결 문제를 제 3 의 안으로 한꺼번에 해결한 것으로 본다.

### 2. 앞으로 한글 부호계의 전반적인 운용 방안

앞으로 한글 부호계의 전반적인 운용 방안에 대해서 살펴보기로 하자.

1) KSC 5601 완성형: 과거와의 호환을 위해서 필요한 데서는 쓰되, 앞으로 점점 줄여 나가도록 하는 것이 바람직하다고 본다. 한편, ISO 2022 를 따르는 분야에서는 아래의 4) 에 나오는 새로운 한 바이트 한글 부호계를 써서 한글을 제대로 (요즘 한글 11,172 소리마디) 지원하는 것이 바람직하다고 본다.

2) 두 바이트 조합형: 과거와의 호환을 위해서 필요한 데서는 쓰되, 앞으로 점점 줄여 나가도록 하는 것이 바람직하다고 본다.

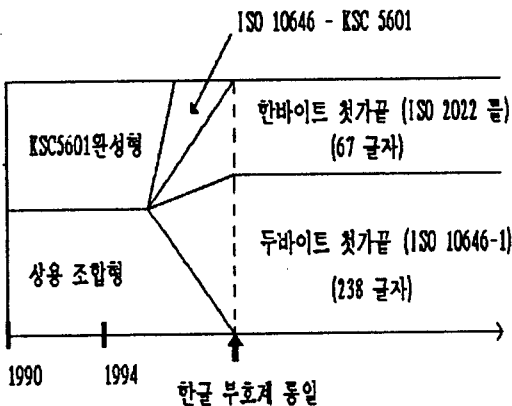
3) 10646-1: 10646-1 에는 첫가끝 방식과 완성형이 있으며, 더욱이 완성형에는 10646-5601, 10646-5657,

10646-0000 (KSC 가 아닌 것) 의 세 가지가 있으므로, 이 가운데 어느 부호계를 지원하는가에 따른 여러 가지 운용 방안이 있다.

글쓴이는, 이미 여러 군데서 밝힌 바와 같이, 얼마동안은 10646-5601 의 2,350 소리마디를 지원할 수도 있겠지만, 길게 보아서는 첫가끝 부호계만 쓸 것을 제안한다. 한편, 10646-5657, 10646-0000 은 완성형으로 지원된 적이 없으므로, 이를 쓰지 말 것을 제안한다. 그런데, 이와는 아주 다른 방향의 제안 (4,516 안)도 있는데 이에 대해서는 아래의 4 에서 따로 살펴보겠다.

4) ISO 2022 를 따르는 분야: 한글을 제대로 지원하면서 ISO 2022 를 따르는 새로운 한 바이트 한글 부호계를 만들어 쓰는 것이 바람직하다고 본다. 이에 대해서는 아래의 3 에서 따로 자세히 살펴 보겠다.

그림: 한글 부호계 운용 방안



### 3. ISO 2022 를 따르는 분야를 위한 새 한글 부호계

ISO 2022 를 따르면 한글을 완성형으로 지원할 수 밖에 없고, 따라서 11,172 소리마디를 제대로 지원할 수 없는 것처럼 생각하는듯 한데, 이는 사실과 다르다고 본다. 앞으로, 한글을 제대로 지원하면서 ISO 2022 를 따르는 한 바이트 한글 부호계를 이른 시일 안에 새로 만들어, ISO 2022 를 따르는 분야에서 쓸 수 있도록 해야 하겠다. 참고로, 10646-1 은 ISO 2022 를 따르지 않기 때문에, ISO 2022 를 따르는 분야에서는 쓸 수 없다.

#### 3.1. ISO 2022 를 따르는 응용 분야

ISO 2022 를 따르는 응용 분야를 들어보면, 먼저 Teletex service, Facsimile service (digital), Message Handling service, Videotex 등과 같은

Telematic Service 분야인데, 여기에 관해서는 ITU-T Recommendation T.52 (1993 년에 확정) 에서 규정하고 있다. 그 밖에, Teletext (문자 다중 방송), digital VCR, Directory Service, 유닉스에서 여러 나라 글자를 지원하기 위한 EUC (Extended Unix Code), EDI (electronic document interchange) 등도 ISO 2022 를 따른다. 특히 유닉스 운영 체제가 국산 주전산기에서 쓸 뿐만 아니라, 아주 많은 샘플에서 쓰고 있는 현실을 감안할 때, 유닉스 분야의 한글 지원 문제가 심각하다.

이런 분야에서는 앞으로 적어도 10-15 년 동안은 ISO 2022 를 따를 것으로 보이며, 따라서 이 분야에서 한글을 제대로 지원할 수 있는 방안이 현재 없으므로, 이에 대한 해결 방안을 아래에서 제시하겠다.

#### 3.2. ISO 2022 를 따르는 응용 분야에서 본 여러 한글 부호계의 문제점

ISO 2022 를 따르는 응용 분야에서 한글을 제대로 지원하는 한글 부호계는 아직 없는데, 그 이유를 살펴 보자.

- 1) KSC 5601: ISO 2022 를 따르고 있긴 하지만, 2,350 소리마디밖에 지원하지 못한다 (보기, 퓌방 각하).
- 2) ISO 10646-1: ISO 2022 를 안 따른다.
- 3) 상용 (두 바이트) 조합형: ISO 2022 를 안 따른다.

#### 3.3. 제안: ISO 2022 를 따르면서 한글을 제대로 지원하는 새 한글 부호계

10646-1 의 첫가끝 한글 부호계 (SYL\_IPF) 기본 정신은 그대로 따르되, 238 글자 가운데서 옛글자를 뺀 요즘한글 67 글자로 된 한글 부호계는, 한글 11,172 소리마디를 지원하면서도 ISO 2022 에 따른다.

이 새 부호계를 ISO 2375 에 따라 국제 등록부 (International Register) 에 등록하여 Escape Sequence 를 받아, ISO 2022 를 따르는 응용 분야에서 쓸 것을 이미 여러 군데서 제안했으며, 1995 년에는 구체적으로 추진하고자 한다. 이 안에 대해서 북한쪽이 비공식적으로 관심을 많이 보였다.

한편, 옛한글을 위해서는 흘림자만 넣은 한 바이트 한글 부호계를 따로 만드는 것이 바람직하다고 본다.

#### 3.4. 새 한글 부호계 등록 추진 방향

ISO 2022 를 따르는 응용 분야에서는, 지금 당장 한 바이트 한글 부호계가 있어야 한다고 본다. 이와 같은 상황을 고려해서, 한 바이트 한글 부호계를 KS 나 ISO 표준으로 만드는 것은 좀 천천히 하더라도, ISO 2022/2375 에 따라서 International Register 에

등록하여 Escape Sequence 를 받는 것은 되도록 이른 시일 안에 추진하는 것이 바람직하다고 본다.

#### 4. ISO 10646-1 을 고쳐서, 한글을 완성형으로만 지원하자는 안에 대한 검토

##### 4.1. 10646-1 에 4,516 소리마디를 더 넣어서 완성형만으로 지원하자는 안

10646-1 에는 현재 6,656 소리마디가 완성형으로 들어가 있는데, 여기에 4,516 소리마디를 더하여, 요즘한글을 완성형으로만 지원하자는 안이  $\square$  회사를 중심으로 나온 바 있다 (아래에서는 4,516 안이라고 부르겠다).

글쓴이가 볼 때, 이 안은 개발 회사의 입장에서 단기적인 안목으로 본 것으로 보인다. 다시 말해서, 샘플에서 한글을 제대로 지원해야 한다는 사명감이나, 왜 이 안이 현재의 국제 표준보다 나은지에 대한 기술적인 검토없이, 회사의 개발 부담 경감을 통한 회사의 경영 측면만 너무 고려한 것으로 보인다. 따라서, 이 안은 국제/국가 표준으로 적절하지 못한 안으로 보이므로, 10646-1 을 현재대로 유지하는 것이 바람직하다고 보는데, 이 안에 대한 기술적 검토를 해 보자.

##### 4.2. 4,516 안에 대한 기술적 검토

###### 4.2.1. 11,172 소리마디가 10646-1 에 들어가더라도, 어차피 238 (또는 67) 글자의 첫가끝 방식을 쓸 것으로 보이는 분야가 있다.

실사 11,172 소리마디가 완성형으로 들어간다 해도, 다음의 분야에서는 어차피 완성형보다 238/67 글자의 첫가끝 방식을 쓰게 될 것으로 보인다.

1) ISO 2022 를 쓰는 응용 분야에서는 10646-1 완성형을 쓸 수 없다. 2022 를 따르는 부호계에 11,172 완성형 소리마디를 넣기가 현실적으로 어려우며, 그 대신, 67 글자를 넣은 한 바이트 한글 부호계를 써야 한다고 본다.

2) 형태소 분석 등 우리말 처리를 하는 분야에서는 완성형을 거의 안 쓰며, 조합하는 방식을 쓴다.

3) 옛한글을 처리하는 분야에서는 조합하는 방식을 쓸 수 밖에 없다.

4) 북한은 한 바이트 및 두 바이트 한글 부호계에서 모두 완성형 보다는 조합하는 방식에 관심을 기울이고 있다.

###### 4.2.2. 앞으로 한글 부호계는 첫가끝 방식으로 통일해야

한글 부호계의 장래 그림은, 크게 나누어 다음과

같은 두 가지 가능성을 생각해 볼 수 있다.

1) 4,516 안에 따라 완성형으로 지원할 경우: 한글 부호계의 이분화 현상이 고착될 것으로 보인다.

가) 일부 분야는 완성형으로 고착

나) 또 어떤 분야 (2022, 우리말 처리, 옛한글, 북한 등) 에서는 238 (ISO 10646-1), 또는 67 글자 (ISO 2022 방식) 첫가끝 한글 부호계를 쓸 것임.

2) 10646-1 을 현재대로 놔 둘 경우: 한글 부호계가 하나로 통일될듯

가) 일정 기간 (3-5 년?) 은 완성형과 첫가끝 방식을 같이 쓰다가,

나) 길게 보면 모든 분야에서 238/67 첫가끝 한글 부호계로 나가야

따라서, 앞으로 한글 부호계는 첫가끝 방식으로 통일을 추진할 경우 성공할 수 있을 것으로 보지만, 완성형으로는 한글 부호계의 지속적인 분단 (혼란)이 고착될 것으로 보인다.

###### 4.2.3. 표준은 길게 보고 그 정책 방향을 잡아야 한다.

완성형이 첫가끝 방식에 비해서 프로그램하기가 좀 쉽다. 일반적으로, 나라의 표준 또는 국제 표준을 정할 때, 회사의 건의를 고려하는 것은 바람직하지만, 너무 회사의 단기적인 이익이나 편의 때문에 표준이 왜곡되어서는 안 된다고 본다. 4,516 안은 지금 당장은 쉽지만, 길게 볼 때 한글 부호계에 지속적인 혼란을 줄 것으로 보인다.

###### 4.2.4. 한글과 비슷한 특성을 가진 글자를 조합형으로 지원하는 것이 세계적 추세이다.

아랍 글자가 한글보다 훨씬 지원하기가 어려운데도, 어떤 회사에서는 아랍 글자 지원 기술을 이미 개발했다고 한다. 애플을 비롯하여 어떤 회사에서는 이런 조합 방식 지원 기술을 이미 개발해 두고 있는 반면, 어떤 회사는 그렇지 못한 것으로 보인다.

238 글자의 첫가끝 한글 부호계 지원이 기술적으로 어렵다고 한다면, 그와 비슷한 상황에 있는 글자는 세계에 아주 많다는 점을 고려해야 한다. 보기를 들어, 히브리, 아랍, 인도의 10 여개의 문자들, 타이, 라오스, 크메르 등이다. 따라서, 한글만이 모아쓰고 그래서 한글만이 특별한 대우를 받아야 한다고 생각한다면, 그것은 사실과 다르다고 본다. 비슷한 특성을 가진 글자가운데 10646-1 에서 완성형으로 한 나라의 글자는 어떤 것이 있는가? 아랍이 presentation form 이라고 하여 완성형이 좀 들어있지만 (그것도 일반 구역이 아니라 presentation form 만 들어가는 특별한 구역에),

그 말고는 거의 모두 조합형 방식이다. 구체적인 사항을 좀 살펴보기로 하자.

1) 타이 글자: 비슷한 상황의 타이에서는 완성형과 조합형을 검토한 결과, 조합형을 쓰기로 국가 표준 기관에서 결정하여, 타이 국가 표준 및 10646-1 에 모두 조합형이 들어가 있다. 또한, ISO 2022 에 따르는 한 바이트 타이 부호계 초안을 만들어서 현재 국제 표준화를 위하여 JTC1/SC2/WG3 의 CD (committee draft) 가 나와 있는 상황이다. 한글이 11,172 인데 비해, 타이는 아마 3,000 - 4,000 개이며, 한글보다 낫기도 쉬울텐데도 완성형으로 하지 않았다.

들리는 말로는, 일본쪽에서 타이 사람들에게 타이 글자를 완성형으로 지원하는 것이 훨씬 낫다는 식으로 개발 방식을 제안하자, 제안은 고맙지만 타이 글자는 그런 식으로 지원하는 것이 바람직하지 않으니, 받아들일 수 없다는 입장을 확실하게 보였다고 한다. 현재 타이는 조합하는 방식이 정착되었으며, 사실 완성형은 별로 호응을 얻지 못했다.

2) 인도 글자: 인도의 열가지 정도의 주요 문자에 대한 부호계도 모두 조합하는 방식으로 10646-1 에 들어가 있으며, 또한, ISO 2022 에 따르는 한 바이트 부호계 초안을 만들어서 현재 국제 표준화를 위하여 JTC1/SC2/WG3 의 CD 가 나와 있는 상황이다.

3) 히브리과 아랍 글자: 히브리와 아랍은 벌써 오래 전에, ISO 2022 를 따르는 한 바이트 한글 부호계를 만들어 Escape Sequence 를 받아서 쓰고 있으며, 물론 10646-1 에도 조합 방식으로 들어가 있는 상황이다.

#### 4.2.5. 11,172 소리마디가 다 들어가더라도 가나다순으로 간추리는 문제가 아주 곤란하다.

이 문제는 완성형의 본질적인 문제는 아니다. 그렇지만 4,516 소리마디를 더 넣더라도, 11,172 소리마디를 가나다순으로 다시 배열할 수 없기 때문에 현실적으로 생기는 문제이다.

따라서, 11,172 소리마디를 쓸 경우, 가나다순으로 간추리기, 비교, 색인 등에서 어려움이 있다. 특히, 현재 대부분의 관계형 데이터베이스에서 사용자가 정의한 가나다순을 지원하지 않기 때문에 문제가 심각하며, 색인의 경우에는 그 해결 방안이 거의 없다. 이에 비해서 ISO 10646-KSC 5601 만 쓰거나 238 글자만 쓰면, 적어도 요즘 한글에선 가나다순으로 간추리는데 문제가 없다. 북한에서도 완성형의 이런 문제점을 지적하고 있다.

#### 4.2.6. 완성형의 실수를 다시는 되풀이하지 말아야 앞으로는 소리글자인 한글과 뜻글자인 한자를

지원하는 틀 자체를 다르게 해야 한다고 본다. KSC 5601 이, 한자를 쓰는 일본의 것을 잘못 흉내내는 바람에 한글 전산화에 큰 문제를 일으켰던 사실을 거울로 삼아야겠다. 4,516 안은 KSC 5601 완성형의 실수를 다시 한 번 하지는 것으로 보인다. 특히, 이런 기회를 통하여 일본에 대한 기술 종속을 조금이나마 벗어날 수 있는 좋은 계기가 될 수도 있다고 본다.

#### 4.2.7. 4,516 안데 예한 현실적인 용급 처방안

지금 당장 238 글자 방안을 그대로 지원하기 어려운 경우, 10646-1 을 고치지 않고도, 한 소리마디를 여섯 바이트로 나타내는 방안을 "임시로" 쓸 수 있다고 본다. 끝소리 글자 (받침)가 없는 것을 나타낼 때, Hanguil Compatibility Jamo 에 있는 0x3164 (채움 글자) 를 쓰면 된다고 본다. 4,516 안 제안자들은 6 바이트 안에 대하여 검토해보기를 바란다.

#### 5. 국제 표준 한글 부호계를 지원하는 자판

현재 10646 에는 첫소리과 끝소리 글자가 따로 들어가 있기 때문에 (보기, 7 과 -), 두벌식 자판으로 지원하기에 어려움이 있다. 따라서, 10646-1 을 제대로 지원할 수 있는 세벌식 자판을 정부의 복수 표준으로 채택할 것을 제안한다.

#### 6. 맺는말

이 글에서 살펴본 바를 줄이면 다음과 같다. 첫째, 한글 부호계의 전반적인 운용 방안: 앞으로 얼마동안은 10646-5601 의 2,350 소리마디를 지원할 수도 있겠지만, 길게 보아서는 첫가끝 부호계로 통일할 것을 제안한다.

둘째, ISO 2022 틀을 따르면서 한글을 완벽하게 지원하는 한글 부호계가 현재 없어서, 유닉스를 비롯하여 2022 를 따르는 분야에서는 한글 지원에 어려움이 많다.

ISO 2022 를 따르면서 한글을 제대로 지원하게 하려면, 한 바이트 한글 부호계를 새로 만드는 것이 가장 바람직하다고 본다. 한 바이트 한글 부호계를 개발하여, 국제 등록부 (International Register) 에 등록하여, Escape Sequence 를 받아서, ISO 2022 를 따르는 분야에서 쓸 수 있도록 하여야 하겠다. 한 바이트 한글 부호계는 성급하게 KS 나 ISO 로 만들 필요는 없다고 보며, 요즘한글과 옛한글 부호계를 따로 만들고, 가능한한 북한과 공동으로 추진하는 것이 바람직하다고 본다.

셋째, ISO 10646-1 을 개정하여 4,516 소리마디를 넣자는 안이 있는데, 이 안을 검토해본 결과, 이렇게 할 경우 완성형으로만 한글을 지원하게 되어 바람직하지

않다고 본다. 특히 이 안은, 회사의 개발비 부담을 줄이고자하는 면이 너무 강해서, 국제 표준에 반영시키기에 적절하지 않다고 본다. 모아쓰기 글자를 지원하는 세계 추세는 조합 방식이므로, 10646-1 을 현재대로 두고, 첫가끝 방식의 238 글자로 조합하는 것이 바람직하다고 본다. 다만, 10646-1 을 고치지 않고도, 한 소리마디를 6 바이트로 나타내는 용급 처방안을 "임시로" 쓰는 가능성은 있다고 본다.

특히, 우리말 처리와 관련된 분야에 있는 사람들이, 완성형보다는 조합하는 방식이 낫다는 점을 한글 표준 제정 관련 담당자들에게 전달하여, 앞으로 한글 부호계가 통일될 수 있기를 바란다.

넷째, 10646-1 을 제대로 지원할 수 있는 세벌식 자판을 정부의 복수 표준으로 채택할 것을 제안한다.

## 7. 참고 문헌

[김 경석 93 다] 고속 중형 컴퓨터 (주전산기 III) S/W 환경에서의 표준 한글 처리 연구. 김 경석 (책임 연구원). 부산대학교 정보 통신 연구소. 1993.1 - 1993.12. 145 쪽.

[김 경석 94 가] 관계형 데이터 베이스에서 국제 표준 한글 부호계 수용 방안 기초 연구 최종 보고서. 김 경석 (책임 연구원). 1993.4 - 1994. 3.

[김 경석 94 나] 앞으로의 한글 부호계 국제 표준화 추진 방향 연구. 김 경석. '94 Korean 컴퓨터 처리 국제 학술 대회, 중국, 연길 (연지).

[김 경석 94 다] 북한, 남한 및 국제 표준 한글 부호계의 비교 및 앞으로의 추진 방향, 김 경석. 국어 정보 학회 가을 학술 대회 논문집, 31-37 쪽. 1994.10.21-22. 충남 아산.

[ISO 93a] ISO/IEC 10646-1:1993(E). International Standard. 1st edition. Information technology - Universal Multiple-Octet Coded Character Set (UCS) - Part 1: Architecture and Basic Multilingual Plane, May 1, 1993. ISO (the International Organization for Standardization).

[KimK 92] "A Common Approach to Designing the Hangeul Code and Keyboard", Kyongsok Kim. Computer Standards & Interfaces.

그림 1. 왜 ISO 2022 를 따르면서 한글을 제대로 지원하는 한 바이트 한글 부호계가 필요한지를 보여주는 그림.

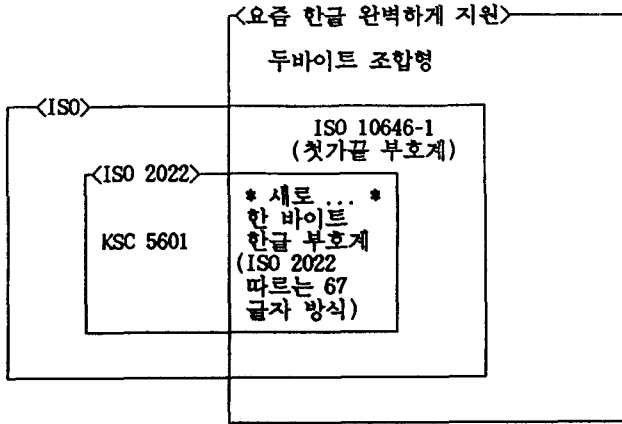


그림 2. 요즘한글용 한 바이트 한글 부호계 (94.10.06)

첫소리 글자 (19), 가운데소리 글자 (21), 끝소리 글자 (27), 쉼 글자 (syllable-break) 1 개가 있다. 겹글자도 모두 들어 있다.

		x(b7b6b5)							
		00	01	02	03	04	05	06	07
y(b4b3b2b1)	00					ㅈ	ㅊ	ㅋ	ㆁ
	01				ㅊ	ㅊ	ㅊ	ㅊ	
	02				ㅋ	ㅋ	ㅋ	ㅇ	ㅇ
	03				ㄴ	ㄴ	ㄴ	ㄴ	
	04			syll break	ㄷ	ㅌ	ㄷ	ㄷ	ㄷ
	05				ㄷ	ㅎ	ㄷ	ㄷ	ㄷ
	06				ㄷ		ㄷ	ㄷ	ㄷ
	07				ㄷ	ㅌ	ㄷ	ㄷ	ㄷ
	08				ㅌ		ㄷ	ㄷ	ㄷ
	09				ㅌ	ㅌ	ㄷ	ㄷ	ㄷ
	10				ㅌ	ㅌ	ㄷ	ㄷ	
	11				ㅌ	ㅌ	ㄷ	ㄷ	
	12					ㄷ	ㄷ	ㄷ	ㄷ
	13				ㅇ	ㄷ	ㄷ	ㄷ	
	14					ㄷ	ㄷ	ㄷ	
	15				ㄷ	ㄷ	ㄷ	ㄷ	ㄷ

그림 3. 옛한글용 한 바이트 한글 부호계 (94.10.06): 생략