

# CAI 응용프로그램 작성시 자료공유를 위한 한자 코드 체계 정의에 관한 연구<sup>1)</sup>

고 대 곤

대구교육대학교 전산교육과

## 요 약

CAI 프로그램 작성시 자료 공유를 위한 한자 공통코드 정의를 위하여 주요 한자문화권 국가의 한자 코드에 대한 고찰 결과, 한자코드 정의시 발음순과 부수순을 혼용 및 동일 한자에 대한 간체자(중국 GB)와 이체자(일본 Shift JIS)의 개발, 사용이 많아 정체자(KSC 한국)간의 자료 호환시 적절한 변환 장치가 요한다. 발음순 원칙은 동자 이음어가 다른 코드로 인식되어 코드 영역의 낭비가 있고, 부수순 원칙은 자국의 두음법칙에 맞지 않으나 중복 코드를 방지할 수 있어 자료의 복원시 유리하다. KSC 제1수준과 제2수준 한자 영역은 학계 및 업계의 요구를 수용할 수준으로 확대시킬 필요가 있다. 유니코드는 시스템의 호환, 확장에 유리하고, 다양한 문자 표현이 가능하여 교육목적 코드의 일시적인 대안이 될 수 있다.

## A Study on Data Sharing Codes Definition of Chinese in CAI Application Programs

Daeghon Kho

Taegu National University of Education, Dept. of Computer Education

### ABSTRACT

Writing a CAI program containing Chinese characters requires a common Chinese character code to share information for educational purposes. A Chinese character code setting needs to allow a mixed use of both vowel and stroke order, to represent Chinese characters in simplified Chinese as well as in Japanese version, and to have a conversion process for data exchange among different sets of Chinese codes. Waste in code area is expected when vowel order is used because heteronyms are recognized as different. However, using stroke order facilitates in data recovery preventing duplicate code generation, though it does not comply with the phonetic rule. We claim that the first and second level Chinese code area needs to be expanded as much as academic and industrial circles have demanded. Also, we assert that Unicode can be a temporary measure for an educational code system due to its interoperability, expandability, and expressivity of character sets.

1) "이 논문은 1996년도 한국학술진흥재단 자유공모 과제 연구비에 의하여 연구되었음"

## 1. 서론

CAI 코스웨어 작성시 한글(고한글 포함), 한자, 영 어 및 특수문자 등 다수 문자가 사용되고 있지만, 이들은 교육목적으로 정의된 표준 코드가 아니어서 코드간에 표현범위와 형식 등 일괄성에 많은 문제점이 있다. 우리 나라에는 국가 수준의 행망용 표준코드(KSC 6501)가 있어, 대부분 이를 교육목적으로 채용하는 실정이나, 행정문서 정보교환 목적으로 정의된 문자는 그 목적부터가 교육목적과 일치하지 아니하며, 정의된 문자수와 자형 등이 절대 부족하고, 고(古)한글, 고문서 및 학술 한자 등에 대한 코드의 여지가 없어, 업계 및 학계의 절실한 요구를 수용할 수 없고, 뿐만 아니라 코드 제정 원칙은 한글 창제 원리에 맞지 아니하여 실제 운영상 상당히 많은 문제점이 있는 것으로 판단된다. 또한 각국 나라의 문자의 표기, 도형 및 특수문자의 표기에도 한계성이 드러난 간주되어 광범위한 교육목적 표준 코드 정의에 대한 논의가 요구되어 진다.

아시아권 자료에서 한자 없이는 이야기되지 않을 정도로 한자는 여전히 매우 중요한 요소이다. 우리나라뿐만 아니라 중국, 일본 및 대만 등을 비롯한 동남아시아 국가들의 많은 자료가 한자로 기술되어 있다. 한자 문화권 국가간의 원활한 정보교환 데이터베이스 구축을 목적으로 공통 한자 사용 인식의 필요에 의거, 유니코드에 약 22,000자의 한자코드를 배정하기에 이르렀다. 대부분의 고전이 한자로 되어 있는 우리 나라도 차체에 한자 코드에 대한 명확한 인식이 필요할 때이다.

본 논문은 CAI 프로그램 작성시 자료 공유를 위한 한자 공통코드 정의를 위한 기초 연구에 도움이 될 목적으로 현재 동남아시아 주요 한자문화권 국가인 한국, 중국, 일본, 대만 등지에서 각기 정의하여 사용중인 한자 코드에 대하여 비교, 검토하고, 교육자료 공유를 위한 적절한 전시환경 등에 대하여 고찰하였다. 본 연구는 계속사업의 성격을 띤 것으로서 이후 이에 대한 계속적인 연구가 요한다.

## 2. KSC 5601 한자 코드에 대한 고찰

## 2.1 한국, 중국, 일본, 대만 한자 코드

동남아시아 한자 문화권 각국은 자국의 필요에 따라 정보의 표현 및 교환을 위한 표준 한자 코드를 정의하여 사용하고 있다. 한국, 중국, 일본 및 대만 등지에서 현재 사용되고 있는 이들 한자 코드들은 서로 다르다. 인터넷이 정보사회의 기반으로 자리함에 따라 정보의 유통량이 폭증하는 시기에, 문화권이 다른 국가간 정보교환의 수월성 추구 및 효율성 제고를 위하여 단일 공통코드의 사용이 절실하다. 세계 여타 지역 및 민족에 두루 통용 가능한 공통코드인 유니코드는 오늘날 이를 수용하는 국가들이 빠르게 증가하고 있다. 그러나 이러한 추세에도 불구하고 유니코드가 기존의 모든 컴퓨터에 통용되고 있는 코드들을 당장 대신할 수 있는 것은 아니어서, 유니코드는 당분간 기존코드들과 공존을 모색할 필요가 있으며, 이 와중에 코드간 적잖은 충돌도 예상된다. 표 2.1은 한국, 중국, 일본, 대만 등 4개국에서 표준코드로 정의된 한자 코드를 비교한 것이다.

중국, 일본 및 대만에서는 한자가 중요한 글자이기 때문에, 코드 영역의 대부분을 한자에 할애한다. 우리나라에서는 한자가 사용되고 있지만, 주 사용은 현대 한글이라는 점에서 이들 국가와 구분된다. 그러나 코드표상 한자가 차지하는 영역은 한글에 비하여 더 크므로, 한자의 중요도는 여타 국가에 비하여 결코 뒤지지 않는 것으로 간주된다. 이는 <표 2.1>에서와 같이 4개 나라중 대만이 가장 많은 한자 수를 정의해 놓았고, 그 다음이 한국, 일본, 중국 순으로 되어 있어, 정작 한자가 기본문자로 사용되고 있으며, 한자의 종주국인 중국이 이들 4개 나라 중 가장 적은 수의 한자 코드를 배정한 것은 아이러니이다.

<표 2.1> 한국·중국·일본·대만의 한자 코드수 비교

코드	KS-5601 (한국)	GB-2312 (중국)	Shift JIS (일본)	Big5 (대만)
제 1 수준	4888 자 (한글발음순)	3755 자 (부수순)	3489 자 (일본음순)	5401 자 (부수순)
제 2 수준	2865 자 (부수순)	2914 자 (부수순)	3388 자 (부수순)	7652 자 (부수순)
총 한자수	7753 자	6669 자	6877 자	13053 자

중국의 경우 상용 생활 문자 중심으로 표준 코드가 간체자에 한하여 정의된 것으로 보아, 제정 당시 데이터베이스 구축, 및 고문서를 포함한 학문연구분야를 위한 학술한자에 대하여는 별도의 복안을 가지고 있던지 아니면 정보화의 의식이 아직 높지 않은 것으로 여겨진다. 대만 한자의 전체 정의 코드 수는 한국 한자의 코드 수보다 무려 5300개나 더 많은데, 이는 한자가 그 나라의 기본 문자로 자리하고 있음을 나타내며, 동시에 이에 걸맞은 코드 연구가 병행된 것으로 여겨진다. 한자 코드수가 많은 것으로 보아 대만의 경우, 교육목적으로 사용시 학계와 업계의 표현 요구를 상당한 정도까지 수용할 수 있을 것으로 보아진다. 코드표상 한국 한자의 정의 코드 수는 중국 한자에 비하여 1084개, 일본에 비하면 896개 더 많다. 이처럼 한국이 중국이나 일본 보다 표준 한자 코드수가 많은 것은 흥미로운 것일 수도 있는 바, 그 원인은 우리의 문화가 한자와 관련이 깊고, 발전이 다양하였던 것으로 추정되는 반면, 표준코드의 선정시 한자 출현 빈도수에 근거하여 선정할 방식을 채택하므로써, 중국과 일본에 비하여 코드 수를 많게 한 원인이 될 수도 있다.

한자 문화권 4개국의 한자 배열방식을 검토하여 본 결과 중국과 대만은 철저히 부수순을 고수하였으나 한국과 일본은 발음순과 부수순을 혼용하는 방식을 채택하였다. 이는 한자 종주국인 중국과 대만에서는 한자 본래의 속성에 근거하여 그 방식을 채택하였는데 비하여 한국과 일본처럼 각기 자국 고유의 문자와 한자의 혼용하는 나라에서는 자국 고유문자의 중시와 한자의 속성을 고려한 적절한 선에서의 타협이라고 여겨진다. 그러므로 한자를 취급함에 있어 철저히 한자 본래의 속성만을 고려하여 부수순을 주장하기에는 대중적인 합의를 도출시 어려움이 상존할 수 있으며, 그렇다고 한자에 철저히 자국 고유 문자발음의 법칙을 그대로 적용하기에는 무리가 있었던 것으로 사료된다. 한국과 일본은 공히 제1수준 한자 코드의 설정시 자국 문자의 발음규칙을 적용하였는데 반하여, 제2수준 한자 코드 설정시에는 제1수준의 적용 원칙을 버리고 중국이나 대만식 부수순을 채용한 점의 일치는 단순히 제1수준 적용 원칙의 한계성 때문인 만은 아니며, 오히려 주변국의 형세 변

화에 대한 상황론 우위를 가정해 보지 않을 수 없다.

일본이 한국보다 앞서 제1 및 제2 수준 한자 코드를 정의한 점만을 감안한다면, 이들 두 나라의 처한 상황의 유사성 고려, 우리의 한자 코드 배정 원칙 설정시 다소간 일본국의 선행 연구사례를 근거로 주장하는 의견이 상당히 있었을 것으로 추측된다.

## 2.2 한국표준코드 제1수준 한자(KSC 5601-1987)

한자에 대한 국가 표준 코드는 KSC 5601-1987에 처음으로 정의되었다. 이는 생활 한자 중심으로 이루어진 정의이다. KSC 5601에 정의되어진 한자코드 수는 총 4888자이다. 이 숫자는 중국의 GB 2312보다 1133자, 일본의 Shift JIS보다 398자 많으며, 대만의 Big5 보다 513자 적은 것이다. KSC 5601은 한글 발음 순에 의거 한자 코드를 배정하였다. 한자 코드 표 정리시 단순히 한글 발음순이라는 준칙을 기계적으로 적용할 경우, 한글 고유의 두음법칙에 따라 한자의 수순이 정하여 지므로, 이는 필연적으로 동일 한자가 중복 정의 결과를 얻게 한다. 이러한 현상은 한글의 두음법칙 현상에 의거 동일 한자라 할지라도 글자의 위치에 따라 다르게 발음되는 경우가 있기 때문이다. 동일 한자의 중복 정의는 코드영역의 낭비, 자료의 변환 및 코드의 복원시 정확한 결과를 얻기가 어렵고, 시간이 소요되며, 별도의 처리 프로그램의 개발 등 여타 문제를 발생시킴으로써 시스템의 신뢰성과 효율성을 저하를 가져오게 하여 그만큼 경쟁에 불리한 요인이 된다.

제1 수준 한자 코드가 정의되어진 KSC 5601에서 두음법칙 영형을 받아 동일 한자가 이중으로 발음됨으로써 중복 표기된 곳은 총 205개로 지적되어졌다 (표 2.2 참조).

KSC 5601-1987 정의 규정을 준수할 경우, 가장 흔하게 발생할 수 있는 문제는 특정 한자를 찾을 때 두음법칙을 고려해서 중복으로 찾지 않으면 원하는 결과를 기대하기 어렵다. 예를 들면, “국립(國立)”과 “입법(立法)”에서 쓰인 “立”자는 동일한 모양을 동일한 글자이나 그 글자가 단어의 첫 가 끝인가(여기에서는 “입”으로 발음되는 경우) 아닌가(여기에서는 “립”으로 발음되는 경우)에 따라 그 코드 값을 달리

<표2.2> KSC5601-1987 동자 이음 정의 예

喇[나]	喇[라]	怒[노]	怒[로]	亮[량]	亮[양]	零[령]	零[영]	熾[린]	熾[인]
懶[나]	懶[라]	擲[노]	擲[로]	兩[량]	兩[양]	靈[령]	靈[영]	璘[린]	璘[인]
擘[나]	擘[라]	櫛[노]	櫛[로]	涼[량]	涼[양]	領[령]	領[영]	蘭[린]	蘭[인]
癩[나]	癩[라]	爐[노]	爐[로]	梁[량]	梁[양]	例[례]	例[예]	隣[린]	隣[인]
羅[나]	羅[라]	盧[노]	盧[로]	糧[량]	糧[양]	禮[례]	禮[예]	鱗[린]	鱗[인]
蘿[나]	蘿[라]	老[노]	老[로]	良[량]	良[양]	醴[례]	醴[예]	麟[린]	麟[인]
螺[나]	螺[라]	盧[노]	盧[로]	諒[량]	諒[양]	隸[례]	隸[예]	林[림]	林[임]
裸[나]	裸[라]	虜[노]	虜[로]	量[량]	量[양]	了[료]	了[요]	淋[림]	淋[임]
邏[나]	邏[라]	路[노]	路[로]	勳[러]	勳[여]	僚[료]	僚[요]	臨[림]	臨[임]
樂[악:d1e2]		露[노]	露[로]	呂[러]	呂[여]	寮[료]	寮[요]	立[립]	立[입]
樂[악:d5a5]		魯[노]	魯[로]	廬[러]	廬[여]	料[료]	料[요]	笠[립]	笠[입]
樂[악:e4c5]		鷺[노]	鷺[로]	旅[러]	旅[여]	燎[료]	燎[요]	粒[립]	粒[입]
樂[악:e8f9]		礮[노]	礮[로]	濾[러]	濾[여]	療[료]	療[요]	蘭[란]	蘭[란]
洛[낙]	洛[락]	祿[낙]	祿[락]	礪[러]	礪[여]	蓼[료]	蓼[요]	鸞[란]	鸞[란]
烙[낙]	烙[락]	綠[낙]	綠[락]	間[러]	間[여]	遼[료]	遼[요]		
珞[낙]	珞[락]	素[낙]	素[락]	驪[러]	驪[여]	龍[롱]	龍[용]		
落[낙]	落[락]	錄[낙]	錄[락]	麗[러]	麗[여]	劉[류]	劉[유]		
諾[낙]	諾[락]	鹿[낙]	鹿[락]	黎[러]	黎[여]	柳[류]	柳[유]		
酪[낙]	酪[락]	論[노]	論[로]	力[력]	力[역]	流[류]	流[유]		
駱[낙]	駱[락]	壘[노]	壘[로]	曆[력]	曆[역]	溜[류]	溜[유]		
亂[난]	亂[란]	弄[노]	弄[로]	歷[력]	歷[역]	琉[류]	琉[유]		
卵[난]	卵[란]	籠[노]	籠[로]	轢[력]	轢[역]	留[류]	留[유]		
欄[난]	欄[란]	壘[노]	壘[로]	憐[린]	憐[인]	硫[류]	硫[유]		
爛[난]	爛[란]	牢[노]	牢[로]	戀[린]	戀[인]	類[류]	類[유]		
蘭[난]	蘭[란]	磊[노]	磊[로]	漣[린]	漣[인]	六[육]	六[육]		
鸞[난]	鸞[란]	賂[노]	賂[로]	煉[린]	煉[인]	戮[육]	戮[육]		
嵐[남]	嵐[람]	雷[노]	雷[로]	璉[린]	璉[인]	陸[육]	陸[육]		
濫[남]	濫[람]	尿[노]	尿[요]	練[린]	練[인]	倫[윤]	倫[윤]		
藍[남]	藍[람]	壘[누]	壘[우]	聯[린]	聯[인]	崙[윤]	崙[윤]		
檻[남]	檻[람]	屢[누]	屢[우]	蓮[린]	蓮[인]	淪[윤]	淪[윤]		
拉[남]	拉[람]	樓[누]	樓[우]	輦[린]	輦[인]	輪[윤]	輪[윤]		
臘[남]	臘[람]	淚[누]	淚[우]	連[린]	連[인]	律[윤]	律[윤]		
蠟[남]	蠟[람]	漏[누]	漏[우]	鍊[린]	鍊[인]	慄[윤]	慄[윤]		
廊[낭]	廊[랑]	累[누]	累[우]	列[릴]	列[열]	栗[윤]	栗[윤]		
朗[낭]	朗[랑]	縷[누]	縷[우]	劣[릴]	劣[열]	率[윤]	率[윤]		
浪[낭]	浪[랑]	陋[누]	陋[우]	烈[릴]	烈[열]	隆[용]	隆[용]		
狼[낭]	狼[랑]	勒[낙]	勒[락]	裂[릴]	裂[열]	利[리]	利[이]		
郎[낭]	郎[랑]	肋[낙]	肋[락]	廉[림]	廉[임]	吏[리]	吏[이]		
來[내]	來[래]	凜[눔]	凜[릉]	殮[림]	殮[임]	履[리]	履[이]		
冷[냉]	冷[랭]	凌[눔]	凌[릉]	簾[림]	簾[임]	李[리]	李[이]		
女[녀]	女[여]	稜[눔]	稜[릉]	獵[림]	獵[임]	梨[리]	梨[이]		
年[년]	年[연]	綾[눔]	綾[릉]	令[령]	令[영]	理[리]	理[이]		
燃[년]	燃[연]	菱[눔]	菱[릉]	囹[령]	囹[영]	異[리]	異[이]		
牽[년]	牽[연]	陵[눔]	陵[릉]	寧[령]	寧[영]->녕	痢[리]	痢[이]		
念[념]	念[념]	泥[니]	泥[이]	嶺[령]	嶺[영]	罹[리]	罹[이]		
捻[념]	捻[념]	匱[닉]	匱[익]	伶[령]	伶[영]	裏[리]	裏[이]		
寧[녕:d2bb]		溺[닉]	溺[익]	玲[령]	玲[영]	裡[리]	裡[이]		
寧[녕:d6b8]		樂[락]	樂[악]->낙	矜[령]	矜[영]	里[리]	里[이]		
寧[녕:e7ac]		掠[락]	掠[악]	聆[령]	聆[영]	離[리]	離[이]		
勞[노]	勞[로]	略[락]	略[악]	鈴[령]	鈴[영]	恣[린]	恣[인]		

하는 모순을 가지고 있다. 그러므로 "立(립)"자를 찾을 경우 "입법(立法)"에 쓰인 "立(입)"자를 찾지 못한다.

표준 한자의 코드를 정함에 있어 한글 발음 순 원칙에 대하여 한글을 공공 생활문서로 사용하는 국가의 경우 적용할 만한 나름대로의 타당한 근거로 인정할 수 있다 하더라도, 결과론적으로 볼 때 동일 한자에 대한 이중 코드 할당은 소중한 코드영역의 낭비라는 비난을 면하기 어렵고, 시스템의 효율성을 떨어뜨리는 결과가 된다. 그러므로 한자 코드 정의 시 한글 발음순이란 규정을 강요한 경우, 예외 사항에 대한 보완책이 제정 당시부터 강구되어야 함은 당연한 귀결이 아닐 수 없다. 더구나 오늘날처럼 유니코드에 대한 세인들의 관심이 고조되고 있는 시점에서, 약 20,000여자의 한자 영역을 배정해 놓은 유니코드는 한국, 중국, 일본, 대만 등 4개국 공통 통합 성격을 지니고 있는 점을 감안해 볼 때, 한글 발음 순 한자코드 배정 방식은 우리 나라를 제외한 다른 나라의 경우, 별다른 의미를 줄 수 없다는 것도 발음 순의 주장을 퇴색하게 한다.

두음법칙에 의한 동일한자의 이중 발음 현상은 한글 표기법에 관한 문법으로 해결할 수 있기 때문에, KSC5601-1987의 한자 코드의 배정은 한자 발음에 기초하여 분류하기보다는 한자의 부수에 근거한 분류·배정이 보다 설득력 있을 것으로 간주된다.

표준코드에 정의되어진 한자 중에는 두음법칙의 영향을 받지 않는, 단어에 따라 다르게 발음되는 한자를 중복 정의한 경우가 있어, 이 또한 코드영역의 효율적 사용이 아닌 것으로 사료된다.

이처럼 동일 한자의 이중 발음 현상은 두음법칙 등 발음의 편의성, 그 사회 문화의 복잡한 발달에 따른 낱어스의 적절한 표현, 구현이 쉽고 무한한 표음 문자 한글에 대한, 의미가 같은, 표의문자 한자 대응 시 어쩔 수 없이 나타나는 현상으로 풀이된다. 이 경우의 예는 다음과 같다.

'不'자는 단어에 따라 '부'와 '불'로 발음된다. 예를 들어 "不正"은 "부정"으로 발음되어 지는데 반하여 "不可"는 "불가"로 발음된다. 위의 예는 두 단어에서 모두 '不'자가 앞에 위치하였기 때문에 발음상 앞 글자의 영향을 받지 아니하여 두음법칙에 해당되지 아

니함에도 불구하고, 전자는 '부'로 발음되어지고, 후자는 '불'로 발음된다.

이들의 경우 겉모양이 동일한 글자임에도 불구하고 한국표준코드에 의하면 전자인 '不'(부)의 코드값은 (dcf4)이고, 후자인 '不'(불)의 코드값은 (ddd5)로 되어 있어 있다. '樂'자의 경우는 위의 경우보다 중복의 정도가 많아서 훨씬 복잡하다. '樂'자는 발음에 따라 '낙', '락', '악' 및 '요' 등으로 소리남에 따라 이들에 대한 4개의 국가표준코드는 "樂[낙:d1e2]", "樂[락:d5a5]", "樂[악:e4c5]", "樂[요:e8f9]" 등이다.

동일 한자에 대한 중복 코드 배정은 코드간 변환 시 최초 정보의 복원이 제대로 이루어질 수 없기 때문에 정보교환에 상당한 문제점으로 예상된다. 추후 전세계적으로 사용되어질 유니코드가 실용화 수준으로 사용되어진다면 기존 사용 코드인 KS와 유니코드간에 빈번하게 정보 변환이 이루어질 것인 바, 올바른 정보 교환이 이루어지지 않는 코드체계를 표준으로 고수함에 따라 이로 인해 발생하는 사용자 불편사항과 업계의 부담 또한 적지 아니할 것으로 예상되어 이에 대한 대응마련이 요구되어 진다.

### 2.3 한국표준코드 제2수준 한자(KSC 5601-1992)

KSC 5601-1987에 정의된 현대 한글 코드는 제정 과정과 공용코드로서의 사용상 많은 문제점들이 지적되어 이의 타당성에 논란은 지정 후에도 끊임없이 일었던 코드이다. 이에 대한 보완책으로 1992년에 제2수준 코드를 확장하게 되었는데, 제2수준 코드에는 한글 2,350자, 한자 2,865자를 추가 배정함으로써 제1수준 코드의 미비점이 어느 정도 보강하였다. 동 기간중 중국의 GB 2312는 한국의 KSC보다 49자 더 많은 2904 자를 추가하였고, 일본의 Shift JIS와 대만의 Big5는 KS보다 각각 523자, 4787자 많은 3388자와 7652자를 확정하였다.

이처럼 각국의 제1수준 코드 확장은 제1코드 제정 당시 예측했던 과학 정보통신기술 발전 속도가 상상 이상으로 급성장 하였으며, 그만큼 정보사회로 진입 속도가 상당히 빨라지고 있음을 시사하는 것이어서 그 의미가 크다 하겠다.

KS의 경우 제1수준 코드로 정의된 4888자의 한자

는 한자를 전문적으로 다루는 학계나 출판계의 수요를 충족하기에는 턱없이 모자랐던 점을 감안한다면, 제2수준 표준코드의 확장은 환영할 만한 일임에 의 없으나, 민간업체인 (주)한글과 컴퓨터 사의 제2수준 확장 한자수가 10,800자 인 점을 감안하여 볼 때, 국가 표준 제2수준 확장 한자는 여전히 학계와 업계의 요구를 수용하기에 역부족이 것으로 사료된다. KSC 5601-1992 확장세트 제2수준 한자의 선정 원칙을 살펴보면 다음과 같다.

① 일상생활, 학문, 기술, 인명, 지명 등에서 자주 쓰이는 한자 중에서, KSC 5601에 포함되지 않은 자를 추가한다. 특히 성씨 중 누락된 자는 꼭 추가한다.

② 신문, 도서관, 출판, 고전 등에서 자주 쓰이는 한자 중에서 KSC 5601에 포함되지 않은 자를 추가한다.

③ 글자체는 KSC 5601과 적합성을 고려하여 강회 이전에 준한 정자체를 택한다.

④ 확장한자의 배열은 동자이음자, 두음법칙에 의한 중복배열은 허용하지 않고, 자형당 하나의 코드를 부여한다.

⑤ 약자, 속자 및 이체자는 특별한 원칙을 설정하지 않고, 본 연구의 선정기준에 의해 추출된 글자는 모두 수용한다.

KSC 5601-1992 확장세트 제2수준 한자의 배열방법으로 한자 속성에 근거한, 현실적으로 상당히 설득력을 가지는 것으로 평가되어 진다. 제2수준 확장세트 정리 원칙에 대한 논의는 대략 2가지로 압축되어 졌는데, 첫째는 현재 사용되고 있는 코드체계를 그대로 존속시키는 방법, 즉 KSC 5601의 제1수준 한자의 경우와 똑같이, 동자 다음을 허용하는 발음별 코드를 부여하는 방식이고, 둘째는 제1수준에서 채용한 방법과 전혀 기본을 달리하는 새로운 방법, 즉 동자에 대한 중복 코드 부여를 허용하지 않고 하나의 문자에 대하여 오직 하나의 코드만을 부여하는 부수순 또는 대표음에 의한 배열방식이었다.

다양한 양케이드 조사와 걸친 전문가들의 의견 수렴을 근거로 “ 대표음에 의한 배열 방식”을 최종적으로 확정이 되었다. 이는 코드표상의 배열은 제1수준에서와 같이 “가나다 . . .” 대표음 순으로 배열하

되, 동음에 대해서는 “부수 획”순으로 한다는 것이다. 이렇게 제정하는 가장 큰 이유로는, 첫째 코드영역의 낭비를 막을 수 있고, 둘째 동자 이음 한자에 대하여 그 음을 하나로 고정해 버리게 되어 현실과의 거리가 있음을 알 수 있다. 이는 한국에서의 한자 표기는 2중의 서로 다른 원칙이 적용된 표준 코드체계를 고수함으로써 한자처리에 대한 일관성이 결여되고, 제1수준 코드 위주로 개발된 각종 한자변환사전 및 소프트웨어에는 추가의 개발 부담이 지워지게 되어 논란의 여지가 여전히 존재한다고 말할 수 있다.

### 3. 중국, 일본, 대만, 유니코드 한자에 대한 고찰

#### 3.1 GB코드(중국)의 한자

중국의 경우 제1수준의 한자로 3755자를 채택하고 이들은 부수순에 기준을 두어 정리하였다. GB코드 한자는 간체자로 불리우는 것으로, 이들 간체자는 정체자에서 획을 단순화시킨 글자들이다. 이처럼 중국에서 대대로 사용되어 내려오던 정체자의 사용을 지양하고 대신 그 글자에 대한 간체자의 채택은 난해한 문자는 자국민의 문자 해독율을 낮게 함으로써 현대 산업사회로의 발전에 큰 저해요인으로 지적되어 오랜 숙고 끝에 문맹률을 낮추기 위한 목적에서 출발한 것이었다. 현재 중국의 대부분의 공문서는 간체자로 기록되고, 학교 교육도 간체자 위주의 교육이 이루어지고 있는 실정이라서 정체자는 실생활에 거의 쓰이지 않고 있다. 언어에 대한 실용성 추구하고 정통성 고수간의 상호 이율배반적인 논리가 공존하고 있는 현장에서 변질된 한자의 이용은 주변국과의 정보교환시 큰 걸림돌이 될 것으로 간주된다.

유니코드에서는 중국의 간체자를 모두 별도의 글자로 등록하고 있는데, 이는 이들 간체체를 전혀 새로운 문자로 정의함을 의미하기 때문에 한자를 알고 있는 사용자라 하더라도 별도의 간체자 이해 없이는 중국 문서 해독이 가능하지 않는 것이어서, 간체자와 정체자간의 코드 변환 작업이 요한다. 이처럼 다단계 변환작업 공정은 경비의 부담 및 시간의 지연을 초래하는 것이어서, 정보의 수집, 가공, 분배와 신속,

정확을 생명력으로 하는 정보사회에서의 발전을 느끼게 하는 요인중의 하나가 될 수도 있을 것으로 사료된다.

### 3.2 Shift JIS(일본) 한자

일본의 제1수준의 한자는 3388자로 일본식 발음을 기준으로 일본식 사전순서로 되어 있다. 즉 “히라가나”와 “가타가나” 발음 순서로 되어 있다. 하지만 나중에 추가된 2수준 확장 한자 코드는 부수 순으로 코드 값을 할당함으로써, 한자 코드 배정시 한국과 마찬가지로 발음순과 부수순이라는 이중 잣대를 적용하여 사실상 완전한 한자코드 체계를 가지고 있다고 볼 수는 없다.

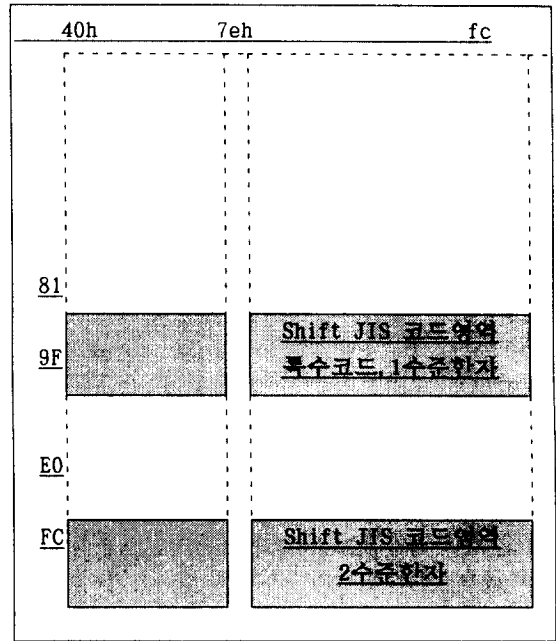
일본의 Shift JIS 제1수준 한자 코드수는 3489자로 같은 시기의 4888자를 정의한 KS 코드의 72%에 해당되는 것이어서, 한자의존도가 우리 나라보다 훨씬 높은 나라에서 이 정도의 제1수준 한자로 만족해하는 것은 의외의 현상으로 여겨진다. 일본 한자는 중국 한자를 기반으로 사용하는 것은 사실이나 경우에 따라 일본식 약자를 나름대로 만들어 내는 경우가 허다하여, 자료 호환시 이러한 이체자의 처리에 각별한 주의가 요청되는 것으로 간주된다.

이체자의 예는 KS의 “나라국”은 ‘國’인데 반하여 Shift JIS에서는 “입 구(口)자 속에 “옥 옥(玉)”자를 넣은 만든 약자 형태의 “國(나라국)”자가 쓰인다. 이러한 일본식 한자 약어는 KS에 정의된 정체자와 같은 글자로 취급하여야 하는지에 혹은 다른 글자로 간주하여야 하는지에 대한 뚜렷한 가이드가 없는 실정인바 본격적인 데이터베이스 구축 및 자료교환을 앞두고 이에 대한 심도있는 논의가 있어야 하겠다. <표 3.2>는 Shift JIS 코드의 제1수준 한자 및 제2수준 한자 영역을 나타낸 것이다.

### 3.3 Big5 코드(대만)의 한자

Big5 대만 한자 코드는 제1수준 및 제2수준에서 한국, 중국 및 일본 등에 비하여 가장 많은 한자를 정의하여 놓고 있어, 한자 코드 정의를 위한 많은 남다른 사전 연구가 있었음을 알 수 있다.

<표3.2> Shift JIS 코드 영역



Big5 코드는 대만의 5대 기업이 연대하여 만든 통합코드로서, 제1수준 5,401자, 제2수준 7,652자, 총 13,053자를 정의하여 놓고 있다. 실제 각 대 메이커 별로 표준한자 이외에 자회사의 편의에 따라 이에 부가하여 더 많은 코드를 추가하여 활용하는 실정인바 실제 정의되어 사용되고 있는 한자수는 이 숫자를 훨씬 능가하는 것으로 잠정 조사되었다. 대만의 Big5 코드 개발방식은 기업의 중복 투자 개발 비용을 줄이고, 동시에 최대의 효과를 얻게 하였을 뿐만 아니라, 공통 코드를 개발, 표준코드로 활용하게 함으로써 기업의 경제효과 사용자의 편의를 십분 고려한 조치라 하겠다. 또한 각 기업은 표준코드에 더하여 자회사의 특성을 살릴 수 있도록 추가 코드를 허용한 것으로 보아 정보화 시대에 대한 탄력적 대응 능력이 높게 평가되어진다. 특히 최근에는 Big5코드를 대폭 보강한 “Big5 플러스”라는 새 코드체계를 마련중인 바, 이는 기존 Big5코드체계를 그대로 유지하면서 여기에 다시 약 8,000 자의 한자 추가가 이루어지고 것으로 되어 있다. 한국의 표준코드는 오랜 논란 끝에 처음 완성형으로 결정되었지만, 이는 조합형과 완성형에 대한 오래 동안의 논란에도 불구하고,

결론은 두 주장의 타협의 여지없이 반쪽의 의견만을 수용하는 결론에 내림으로써, 오랜 동안의 논쟁이 공론으로 끝났고, 제정된 국가 표준코드 제구실을 하지 못한 채 학계나 업계로부터 외면 당하는 수모를 겪어야 했다. 이후 상당한 시간이 지난 다음에서 완성형-조합형 양쪽을 모두 허용하는 코드의 제정이 있어 환영할 만한 일이나, 이는 기존코드에 맞추어 개발된 각종 장비의 재투자 및 기존 소프트웨어에 대한 수정 등의 개발 부담을 전제로 하는 것이어서 여전히 중복 투자 및 비효율적이라는 지적이 감수하여야 했다.

상대방의 주장을 외면하는 논쟁은 아무리 길어도 무익하다. 대만의 경우, 표준코드에 각 주장의 논쟁보다는 공조 협력을 통하여 더 좋은 코드를 만들어 냈고, 정보 산업 현장의 세계시장에서 팔목할 만한 성장을 이루게 되었는데, 기 이면에는 정보사회에 걸맞은 열린 시민 의식이 이를 이룩하게 한 원동력이 되었을 것으로 여겨진다. 이후 대만의 한자코드는 한국, 중국, 일본 등 주변국의 한자코드 제정시에 많은 참조가 되었던 귀중한 코드라는 점에서 높이 평가된다.

### 3.4 UNI(유니)코드의 한자

다양한 나라, 다양한 프로그램에서 서로 동일하거나 비슷한 문자를 각기 다른 인코딩 방식을 사용하는 자료 및 프로그램, 시스템의 호환 및 확장성에 상당한 낭비와 효율성을 저하시키게 되므로 이를 하나의 문자세트를 사용함으로써 이러한 불필요한 중복과 낭비적인 요소를 방지할 수 있다. 이러한 목적하에 등장한 유니코드는 하나의 프로그램 내에서 다른 언어권의 문자 등 여러 가지 문자세트를 처리하거나, 자료로 표현, 저장이 이루어질 수 있게 하며, 전송에서 자료의 복원 시 모호 현상이 없어지게 된다. N개의 기존 문자세트 표준사이의 변환을 위해서는  $N*(N-1)$  개의 변환 테이블을 준비하여야 하나, 유니코드를 사용한다면, 일단 유니코드로 변환한 다음 다른 문자세트 표준으로 변하게 함으로써  $2*N$  개의 변환테이블만 필요로 하는 것이어서 문자세트 변환 프로그램이 상당히 간단해 질 수 있다.

오늘날 세계 표준 코드로서의 유니코드는 빠르게 자리를 잡아가고 있으며, 이를 응용하는 소프트웨어의 개발이 증가 추세에 있다. 현재 유니코드에는 약 22,000 이상의 한자가 배정되어 있으며, 계속 추가 중이다(부록1 참조). 이것은 한국, 중국, 일본, 대만의 한자를 조사하고, 동의 문자라 하더라도, 외형이 다르면 이를 다른 문자로 간주하여 추가하는 형식을 취하였다. 유니코드에서는 한국, 중국, 일본, 대만 등 4개국에서 사용되는 한자 중, 동의 이형 문자, 즉 일본식 약자와 중국의 간체자는 별도의 배정 영역을 확보하도록 하였으며, 이후 고문헌에서 새로이 발견되거나 혹은 새로 만들어지는 한자들을 위한 여지를 확보, 한자들이 계속 추가가 가능하도록 되어 있다.

유니코드의 한자 배정 원칙은 기본적으로 부수순이지만, 이는 이후 추가되는 글자들에 대해서까지 확대된다는 의미는 아니다. 코드 할당이 이루어진 다음에 새로이 추가되는 한자의 경우, 기 배정 원칙의 예외 사안으로 취급되어 질 수밖에 없다. 이는 추가되는 한자의 배열 순위가 기존 문자의 맨 끝에 위치하기보다는 대부분 기존 코드의 중간에 삽입하여야 한다는 난점을 지니고 있기 때문이다. 각 나라의 발음 순서를 국제코드 순서로 채택하기에는 설득력이 없다. 예를 들면 발음을 기준으로한 “가나다 . . . .” 순서나 “히라가나 . . .” 순서는 해당국 한국이나 일본에서는 중요한 의미를 갖는 것임에는 틀림이 없으나 해당국 이외의 다른 나라에서는 별 의미를 가질 수가 없다. 그러므로 국제 통일 코드에서의 한자 배정 원칙은 부수 부수순에 의한다.

인터넷 시대를 맞아 정보사회에서의 데이터베이스의 구축, 자료의 교환 등을 위한 목적으로 국경을 넘어 이민족간에 통일된 하나의 코드체계를 갖는다는 것은 효율성 측면에서 대단한 획기적인 업적으로 간주될 수 있다. 그러나 유니코드의 통일성, 공통성, 효율성, 수월성의 지나친 강조는 우월성을 앞세운 문화의 흡수, 종속을 초대하여 고유성과 특수성이 희석될 수 있다는 우려도 있다. 현재 유니코드는 실용에 앞서 몇 가지의 보완책이 마련이 요구된다. 기존 아스키 코드와의 충돌 및 아스키 파일 사이즈의 2배 증가 등이다. 한국, 중국, 일본 및 대만 등은 같은 한자



문화권 국가들이라 하더라도, 오랜 세월을 두고 한자는 각기 지역에서 발전하여 오늘날 의미는 같으나 외형이 달라 의사 소통이 이루어지기 힘든 부분도 상당수 있다. 일본식 약자나 중국의 간체자는 정체자와 그 외형이 달라 한국이나 대만 등에서 사용하기엔 곤란하다. 이를 가능하게 하기 위하여는 동일 한자에 대한 정체자, 간체자, 약자 등에 대하여 두루 습득하거나 혹은 전문가의 수정작업을 통한 후야야 의미소통이 가능할 것이다. 그러므로 본격 시행을 앞두고 각국간의 정보변환을 상당한 수준으로 해낼 수 있는 소프트웨어의 개발 등 장기적인 보완책이 요구된다. 사전 준비없이 무리한 유니코드의 적용은 정보교환을 위한 모처럼의 시도가 본래 의도가 빛 바래거나 그 의의가 반감될 수 있음을 감안하여야 한다. 그러나 단일 코드체계의 사용은 고문헌 연구에 크게 도움이 될 것으로 사료된다.

### 3.5 워드프로세서 한글 한자

현재 우리 나라에서 가정 넓은 사용자층을 가지고 있는 (주)한글과 컴퓨터사의 워드프로세서 “한글”은 제2수준한자(확장한자세트)는 무려 10,800자에 해당되는 엄청난 숫자의 한자를 정의하여 놓고 있다. 이는 국가수준에서 1.2차에 걸쳐 정의한 한자 총수보다 3,047자 많고, 국가수준 2차시기 정의한 숫자의 3.77배에 해당된다. 이는 한 회사의 업적이 국가수준보다 더 컸음을 의미하는 것이어서 시사하는 바가 크다 하겠다. 워드프로세서 “한글”의 한자영역은 0x4000부터 0x7D97까지 배정되어 있어 그 영역의 크기는 15,768자에 이른다. 워드프로세서 “한글”에서의 한자 정의 코드는 한국사료사연구회 조사 자료에 기초한 것으로, 한국사료사연구회에서 현재까지 조사, 집계된 한자의 수는 15500자인 것으로 알려져 있다. 이 숫자는 이후 계속되는 한국 고전 연구, 불교 관련 문헌 연구 등을 통하여 계속 추가될 가능성이 많아, 이를 유치할 여유공간의 확보 및 일괄적인 코드 부여 방식이 요구된다.

그러나 이후 추가되는 단순히 기존 문자의 다음에 순서대로 차례로 배정되는 것이 아니라 기존 코드의 중간에 삽입하여야 하는 경우가 대부분이기 때문에

추가 문자들에 대한 일관성의 유지는 사실상 곤란한 것으로 판단되어, 한자 코드 부여 방식에 대한 예외 규정을 두지 않을 수 없을 것이다.

### 4. 결론

한국, 중국, 일본 대만 등 주요 한자 문화권 국가의 국가수준 한자표준코드를 비교, 고찰한 결과 다음과 같은 결론을 얻었다.

1. 각국은 일반목적 한자 표준코드를 가지고 있는 반면, 교육목적 코드를 갖고 있지 않다. 체계적인 교육용 데이터베이스구축을 위해 이를 지원할 교육목적 한자코드 정의가 요한다.

2. 중국의 GB2312 한자코드는 간체자 중심으로 정의된 것이어서 한국의 정체자와 간체자 자료 호환시 별도의 코드변환 프로그램을 가질 필요가 있다.

3. 일본의 Shift JIS 한자코드에는 이체자의 정의가 많아 한국의 정체자와 이체자 간의 자료 호환을 위한 별도의 변환 시스템을 가질 필요가 있다.

4. KSC 5601 제1수준 및 제2수준 한자영역은 학계와 업계의 요구를 충분히 수용할 수준으로 확대할 필요가 있다.

5. 대만의 Big5 플러스 코드는 21,053자의 한자영역이 확보되어, 교육목적 한자 대안 코드로 가장 적절한 것으로 사료된다.

6. 유니코드를 수용하되 우리 문화의 유일성을 유지하는 국가 수준의 별도 코드의 마련과 이들간의 호환이 요구된다.

7. 교육목적 표준 한자 정의는 한글 발음규칙을 적용하되, 동자 이음어 처리 등을 위하여 부수순 원칙을 준용할 필요가 있다.

### 참고 문헌

- [1] 고대곤, 한글저작시스템의설계와특성연구연세대학교 대학원 박사학위 청구논문, 1989
- [2] 고대곤, 우리 나라의 초·중등 교육 환경에서 요구되는 고(古)한글을 포함하는 한글의 코드 체계의 구성에 관한 고찰, 한국정보교육학회 1권 2호, 1997, 12

- [3] 고대곤, CAI프로그램 작성시 자료 공유를 위한 한글 코드 체계 정의에 관한 연구, 한국정보교육학회 제2권 1호, 1998, 5
- [4] 양왕성, 유니코드와 한글 통합형코드, 월간 마이크로소프트, 1995
- [5] 한국표준연구소, 정보교환용 부호 확장 시안 확장한글 처리방식 제안서, 1991
- [6] 한국표준연구소, 정보교환용 부호 확장 시안 확장한글 처리방식 제안서 표준 한글·한자 코드확장 및 처리방식 연구 별책
- [7] 한글사용 환경하의 소프트웨어 개발 기술에 관한 연구, 과학기술처, 1986
- [8] 한글정보처리 표준화 연구, 과학기술처, 1988, 7
- [9] 한글한자 정보처리기술 핸드북, 과학기술처, 1988, 7
- [10] 현규섭외, 한국 표준 확장한자세트 제정 방안연구, 출판연구, 1994, 2
- [11] 현규섭외, 한자코드 표준화 개선 방안에 대한 기초 연구; 국내외 표준화 대응을 위한 확장한자세트 선정을 중심으로, 출판연구, 1992, 6
- [12] <http://camars.kaist.ac.kr/~dtkim/java/encoding.html>
- [13] <http://trade.chongbuk.ac.kr/~lees/code/1557exp.html>
- [14] <http://www.standard.nca.co.kr/activity/sgenda/d1.html>
- [15] <http://www.unicode.org/unicode/uni2book/u2.html>-Unicode Standard Book
- [16] INTERNATIONAL STANDARD ISO/IEC 10646-1/First edition, 1993
- [17] ISO/IEC 10646-1 Information technology Part 1 Architecture and Basic Multilingual Plane
- [18] ISO 2022 Information Processing- ISO77-BIT AND 8-BIT Coded Character sets-Code extension Techniques, Third edition, ISO, 1986, 5
- [19] KSC 5602 정보교환용 부호의 확장법, 한국규격협회, 1977, 12

부록1. 유니코드(unicode) 문자 세트

분 류	코드값 (16진수)	내 용
일반 언어	0000 ~ 007F	<u>상위 바이트가 0인 아스키 문자 (ASCII)</u>
	0080 ~ 00FF	<u>상위 바이트가 0인 ISO-8859-1 (유럽)</u>
	0100 ~ 024F	<u>확장 라틴계 언어 (유럽) (Latin Extended-A , Latin Extended-B )</u>
	0250 ~ 02AF	<u>국제 음성 표기 기호 (IPA Extensions)</u>
	02B0 ~ 02FF	<u>공간을 차지하는 수정 문자 (Spacing Modifier Letter)</u>
	0300 ~ 036F	<u>결합형 구별 표식 (악센트등) (Combining Diacritical Mark)</u>
	0370 ~ 03FF	<u>그리스어 (Greek)</u>
	0400 ~ 04FF	<u>소련어 (Cyrillic)</u>
	0530 ~ 058F	<u>아르메니아어 (Armenian)</u>
	0590 ~ 05FF	<u>히브리어 (Hebrew)</u>
	0600 ~ 06FF	<u>아랍어 (Arabic)</u>
	0700 ~ 08FF	예약됨 (512 자)
	0900 ~ 097F	<u>벵어 (Devanagari)</u>
	0980 ~ 09FF	<u>방글라데쉬어 (Bengali)</u>
	0A00 ~ 0A7F	<u>편잡어 (인도 북부) (Gurmukhi)</u>
	0A80 ~ 0AFF	<u>Gujarati (인도 북부)</u>
	0B00 ~ 0B7F	<u>Oriya (인도 북부)</u>
	0B80 ~ 0BFF	<u>Tamil (인도 남부)</u>
	0C00 ~ 0C7F	<u>Telugu (인도 남부)</u>
	0C80 ~ 0CFF	<u>Kannada (인도 남부)</u>
	0D00 ~ 0D7F	<u>Malayalam (인도 남부)</u>
	0D80 ~ 0DFF	예약됨 (128 자)
	0E00 ~ 0E7F	<u>타이어 (Thai)</u>
	0E80 ~ 0EFF	<u>라오어 (Lao)</u>
	0F00 ~ 0FBF	<u>티베트어 (Tibetan)</u>
	0FC0 ~ 109F	예약됨 (224 자)
	10A0 ~ 10FF	<u>그루지아어 (Georgian)</u>
	1100 ~ 11FF	<u>3벌식 한글 자모 (240 개)</u>
1200 ~ 1DFF	예약됨 (3072 자)	
1E00 ~ 1EFF	<u>구별 표식 (악센트등)이 붙은 라틴계 언어 (Latin Extended Additional)</u>	
1F00 ~ 1FFF	<u>구별 표식 (악센트등)이 붙은 그리스어 (Greek Extended)</u>	
기호	2000 ~ 206F	<u>구두점 (General Punctuation)</u>
	2070 ~ 208E	<u>어깨글자, 밑글자 (Superscripts, Subscripts)</u>
	20A0 ~ 20AB	<u>화폐 기호 (Currency Symbols)</u>
	20D0 ~ 20E1	<u>결합형 구별 표식 (Combining Marks for Symbols)</u>
	2100 ~ 2138	<u>라틴, 그리스, 히브리어 유사 기호 (Letterlike Symbols)</u>
	2153 ~ 2182	<u>로마 숫자, 분수 (Number Forms)</u>

분 류	코드값 (16진수)	내 용
	2190 ~ 21EA	<u>화살표 (Arrows)</u>
	2200 ~ 22F1	<u>연산자 (Mathematical Operators)</u>
	2300 ~ 237A	<u>기타 기술적 기호 (Miscellaneous Technical)</u>
	2400 ~ 243F	<u>제어 문자 기호 (Control Pictures)</u>
	2440 ~ 245F	<u>광학 판독 기호 (Optical Character Recognition)</u>
	2460 ~ 24FF	<u>원, 점, 괄호로 묶인 알파벳 및 숫자 (Enclosed Alphanumerics)</u>
	2500 ~ 257F	<u>박스를 그리기 위한 기호 (Box Drawing)</u>
	2580 ~ 259F	<u>무늬 (Block Elements)</u>
	25A0 ~ 25FF	<u>도형 (Geometric Shapes)</u>
	2600 ~ 26FF	<u>다양한 그림 (Miscellaneous Symbols)</u>
	2700 ~ 27BF	<u>Zapf Dingbat 폰트 (Dingbats)</u>
	27C0 ~ 2FFF	예약됨 (2112 자)
한중일	3000 ~ 304F	<u>한중일 기호</u>
음성 표기 및 기호	3040 ~ 30FF	<u>일본어 (히라가나, 카타카나)</u>
	3100 ~ 312F	<u>중국 발음 기호 (Bopomofo)</u>
	3130 ~ 318F	<u>KSC5601 호환 2벌식 한글 자모 (94 자)</u>
	3190 ~ 319F	<u>Kanbun</u>
	3200 ~ 32FF	<u>원, 괄호로 묶여진 한중일 문자 (Enclosed CIK Letters and Months)</u>
	3300 ~ 33FF	<u>여러 문자가 묶여진 문자</u>
예약 영역	3400 ~ 4DFF	(예약됨) (6656 자)
한자 영역	4E00 ~ 9FFF	<u>한중일 통합 한자 (부수순) (20902 자)</u>
예약 영역	A000 ~ ABFF	(예약됨) (3072 자)
한글 음절	AC00 ~ D7A3	<u>현대 한글 음절 (11172 자)</u>
Surrogate 영역	D800 ~ DB7F	<u>예약된 상위 코드 (High surrogates) (896 개)</u>
	DB80 ~ DBFF	<u>개별 용도 상위 코드 (Private Use high surrogates) (128 개)</u>
	DC00 ~ DFFF	<u>하위 코드 (Low surrogates) (1024 개)</u>
개별 영역	E000 ~ F8FF	<u>개별 용도 (특정 응용 프로그램에서만 사용) (6400 자)</u>
호환 영역	F900 ~ FAFB	<u>기존 표준 호환 한자 (302 자)</u>
	FB00 ~ FB4F	<u>라틴, 아르메니아, 히브리어 관련 프리젠테이션용 문자 (Alphabetic Presentation Forms)</u>
	FB50 ~ FDFB	<u>아랍어 프리젠테이션용 문자 (Arabic Presentation Forms-A)</u>
	FE00 ~ FE1F	예약됨 (32 자)
	FE20 ~ FE2F	<u>결합형 반쪽 구별 표식 (Combining Half Marks)</u>
	FE30 ~ FE4F	<u>수직방향의 한문에 사용되는 기호 (CIK Compatibility Forms)</u>
	FE50 ~ FE6F	<u>작은 크기의 아스키 기호 (Small Form Variants)</u>
	FE70 ~ FEFF	<u>아랍어 프리젠테이션용 문자 (Arabic Presentation Forms-B)</u>
	FF00 ~ FFEF	<u>전각 아스키, 전각 기호, 반각 한중일 문자 및 기호 (Halfwidth and Fullwidth Forms)</u>
	FFFD	<u>대치 문자 (유니코드로 변환되지 못한 문자를 나타냄) (Replacement Character)</u>