

한글 출력코드 체계에 관한 기초 연구

이 강섭, 이 기성
단국대학교 전산통계학과

A basic study on the output code system of the Hangeul

Kang Sup Lee, Ki-Sung Lee
Dept. of Computer Science and Statistics, Dankook University

요 약

알파뉴메릭 사용 언어권과는 달리 한글에서는 처리코드와 출력코드를 구별할 필요가 있다. 한글의 처리코드로는 완벽한 한글 처리가 가능한 조합형이 옳다는 데에 재론의 여지가 없다. 그러나 전자출판의 경우 출력코드는 조합형과 완성형을 모두 사용하는 절충형이어야 한다. 절충형 코드를 구현하기 위한 기초 연구로서 국민학교 국어교과서를 분석하여 완성형꼴(음절)의 범위를 예측하였다. 1400자는 완성형으로, 나머지 9700자 가량은 조합형을 사용하는 방식이 한글 출력코드 체계로서 적합할 것이다.

I. 서론

2000년이 넘게 종이책을 사용하다가, 개인용컴퓨터의 대중화에 따라서 새로운 형태의 책인 비종이책(non-paper book)이 출현하게 되었다. 대표적인 예가 전자책이라고도 불리는CD-ROM, CD-G, CD-I 등 콤팩트디스크 형태의 디스크책이다. 종이책이건 비종이책이건 출판에서 사용하는 출력용 한글 글자꼴은 그 기준이 되는 본그림(원도)이 있어야 한다. 이 본그림의 수준에 따라서 이를 사용하여 제작된 활판 인쇄용 활자(금속활자)나 윗셋 인쇄용 활자(사진식자용 활자 음판) 또는 컴퓨터용 디지털 활자의 품질이 결정된다. 또한 전자출판(CAP : Computer Aided Publishing)에서는 본그림을 그대로 아날로그 형태로 사용하던 방식에서 점차 디지털 형태로 사용하는 방식으로 바뀌고 있다.

본 연구에서는 한글의 출력코드 체계를 연구하기 위한 첫 단계로서, 현행 국민학교 국어(읽기, 말하기) 교과서에 나타나는 한글 음절과 그 빈도를 조사하였고 몇가지 기초자료를 제시하였다. 이를 바탕으로, 한글 출력코드로서는 가독성과 변별력은 물론 글자꼴 자체의 균형과 미려함을 특성으로 갖는 절충형 코드 사용을 제안하였으며 후속 연구과제를 밝혔다.

II. 한글 코드와 한글 글자꼴

컴퓨터를 이용하는 출판방식은 여러가지가 있는데, 금속활자로 인쇄하는 대신 글자를 컴퓨터가 사용하는 한글코드로 바꾸어 프린터나 인화지출력기로 출력하는 데스크톱 출판방식(Desk Top Publishing: DTP)과 전산사식시스템(Computerized Typesetting System Publishing: CTS) 방식이 주종을 이룬다. 이들 방식은 종이책을 출판할 때 많이 사용된다. 출력매체가 종이가 아닌 디스켓인 경우에는 디스크책 출판(Disk Book Publishing: DBP)방식과 화면책 출판(Screen Book Publishing: SBP)방식이 주로 사용된다. DTP는 출력용 글자꼴의 품위에 따라, 다시 사무자동화(OA) 전용과 CTS 호환용으로 구별된다. 본 논문에서는 CTS용으로 사용될 수 있는 출력용 한글 글자꼴에 한하여 논의하기로 한다.

1. 한글 코드의 종류

글자에는 내부용글자(Inner-letter)와 외부용글자(Outer-letter)가 있다. 내부용글자는 사람의 머리속에서 생각하는 글자꼴을 말하는데, 머리속글자라고도 한다[1]. 외부용글자는 눈에 보이도록 손으로 쓰거나 인쇄한 글자를 말한다. 이 글자를 컴퓨터에서 표현하는 것이 코드(code)인데, 출판용 한글코드는 입력용 코드, 내부처리용 코드, 출력용 코드의 3가지로 구별된다. 입력용 한글코드는 초성/중성/종성이 조합되어 입력되는 형태이다. 내부처리용 한글코드 역시 초성/중성/종성을 조합하여 사용한다. 입력용과 내부처리용은 반드시 조합형이어야 하지만, 출력용 한글 코드는 다음 표1에서 보는 바와 같이 조합형, 완성형 모두 가능하다. 또한 본 논문에서 제안하는 절충형(일부의 글자만 준비해놓고 준비되지 않은 글자는 출력용 자소를 조합하는 방식)도 가능하다.

표 1. 코드별 현대 한글 음절의 표현 가능성

코 드 종 류	내 부 처 리 용	종 이 책 출 력 용	디 스 크 책 출 력 용	정 보 교 환 용
조 합 형	100% 가능	100% 가능	100% 가능	100% 가능
완 성 형	일부만 가능	일부만 가능	일부만 가능	일부만 가능
절 충 형	—	100% 가능	100% 가능	—

2. 한글 글자꼴 분류

글자꼴(typeface)은 몇가지 원소를 조합한 추상적인 디자인에 의해서 어떻게 글자(letters)가 표현되었느냐 하는 것으로 정의할 수 있다. 한글 글자꼴은 다음과 같이 다섯가지로 크게 분류할 수 있다[4,5].

- 1) 본문체 = 붓체, 조화체, 정조체, 바탕체, 명조체 : Body text type, Serif.
- 2) 네모체 = 훈민정음체, 반포체, 강조체, 고딕체 : San Serif, Grotesk, Gothic.

- 3) 제목체 = 헤드라인체, 타이틀체 : Heading type.
- 4) 디자인체 = 그래픽체 : Graphic type.
- 5) 서예체 : Calligraphic type.

한글의 경우는 '공간의 글자'라고 불릴 정도로 공간이 중요하다. 본문체를 확대하여 사용하면 되지 왜 제목체가 필요한가 하는 의문을 가질 수 있으나, 확대시킨 글자와 제목체로 쓴 글자를 비교하면 한 글자내에서 자소의 배열과 크기가 다르다는 것을 알 수 있다. 글자꼴은 종이책에 쓰이느냐, 비종이책에 쓰이느냐에 따라서, 가로쓰기용이냐 세로쓰기용이냐에 따라서도 변화가 있어야 한다.

3. 기계가독형 글자꼴

워드프로세서(전용기)와 개인용컴퓨터는 문장(text)을 작성하고, 수정하고, 검색하고, 프린트할 수 있다. 이렇게 작성된 내용(content of documents)이 정보로 사용된다면, 즉, 컴퓨터로 이 내용을 처리하려면, 디지털 방식으로 문장이 작성되어야만 한다(digital text preparation). 이런 환경은 전통적으로 글자꼴을 만들던 아날로그(analog) 방식의 금속활자 만들거나 수동사진식자기의 사식활자 원도와는 다른 방식의 활자를 제작하게 만들었다. 일반적으로, 활자(font)를 출력코드(output code)와 같은 뜻으로 사용하나, digital font는 output code에 attribute가 추가된 것으로 보아야 정확할 것이다.

III. 한글 음절 분석

1. 국민학교 국어 교과서 분석

미려한 출력결과를 얻기 위하여는 조합형과 완성형의 장점을 살릴 수 있는 절충형 코드가 적합하다. 완성형 글자꼴의 범위를 알아보기 위하여 현행 국민학교 국어(읽기, 말하기) 교과서 24권(1학년-6학년, 1,2학기)을 분석하였다. 그 결과는 다음 표2, 표3, 표4, 그림1과 같다.

표 2. 국민학교 국어 교과서의 한글 음절

학년	책	한글 음절 수	총음절수
1학년	읽기/말하기	726/398	16,718/ 7,854
2학년	읽기/말하기	774/470	33,879/11,279
3학년	읽기/말하기	909/584	42,484/17,218
4학년	읽기/말하기	1,010/697	48,791/26,890
5학년	읽기/말하기	1,008/737	51,424/28,920
6학년	읽기/말하기	1,043/756	57,339/30,522
계		5,469/3,642	250,635/122,683
총 출현 음절		1,405	373,318

표 3. 국민학교 교과서에 나타난 한글 음절(출현빈도순)

습부알적더재더쓰박월스골향랏규응곤뚝벽긋특엔뿌듬텐든세덱신너긋남칩실넉흙새훑름옥형덜법심진타
 한정상차길씨도워쪽현약봉뒗험공님퍼긋뚝름력싱때뒗킨댓샅넉셔꼭갓넙짱셈난형쌌헛렷웃헛맷습순지귀
 들구세느날목작렸준실병향등죽요뽕덜데대참랜맨닥닷조늬맷길붓긋갓갈춤뻗공흙산헛램윈햇닝맑차친괴
 로인르엇금난랑끝빠결후악광릉첫라료수흙좁말렴능늬중넛륙혹벨균혁갓젓뻗겐뚝습핑필웨필닐릿속재광
 시마제스누점을희창태운술집렸채곤녀메케빗덤늬군넙취궁렬헛민흙평긋잉뻗각풍뽕뒗땀땀땀땀땀땀땀땀땀땀땀땀
 계과위림양늘렁설언녀닭는낭곰잡탄흙맛벤칭늬늬가름숯땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀
 그까거히치년함에종왕체웠허흔옥점픽섯름옥징휴뒗쳐뽕땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀땀
 여것되데속종왔거럼졌늬뽕좁름습뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 를수용간경맛침행남잡갈럼애칠볼커겹춤말연땀앗척빙낭엇길검정음교흙선통흰참뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 기무임방찾린갓임석결형극필습뽕진뽕총욕름긋술짖뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 은면님두영국불건웃파키엿평찬쟁밀뒗본땀땀뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 아우물른뽕유관열큰외돌풀여차북릴닐끝뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 보각등들실섯표줄뜻슴삼천빛색므낫넛금넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛
 리대였버친들민외별독힘쇠베력물거고코취출혜피랍껍덜랑속명퀴덜자첩뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 자주모분확파디열등농죽직호흐몇헤졌업점음푹홍뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 니요람미살울런례도질감나벼홀담땅취삭셔득꽃들근되음긋뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 지으글신호졌는룩배피봄개엄숯취척갑흔킨샅괘갓쏘삭캐뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 서라바앗산명추루끼레매복황발득착브윤획홍섯죄웬샅라주니끝앙더보없긋긋긋긋긋긋긋긋긋긋
 의해선잘개화드참머퍼작쳐뜨겨포련압십혹빈섯염살습덜억깍개셋결바섯맹겐춤뽕뽕뽕뽕뽕뽕뽕
 가내발입네늘답강크첼빨즐짓냥트훈문룻햇막릿섯술다괴의품흰채귀유쇄를경춷뽕뽕뽕뽕뽕뽕뽕
 고소러래했공원회뒗뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 어야대학교단노은당밤근녕송홍깊곧팔뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 에일할저달식많꾸본김굴넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛넛
 는생만머중진된견번군귀채불쳤낭섬잎갓잇안른티돌즈잡밧압뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 하있와날없성곳반란풍났먼밀립겨웃찰줄웬데타답담좁이필빛츠뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 을사머안압타합너저든씨긴품덕떨카령익지꺼벗달낫원뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 이도장전심새연편절활슬뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 다음오비갈던반앞통책뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕뽕
 표 3. 국민학교 교과서에 나타난 한글 음절(출현빈도순)

IV. 결론 및 앞으로의 과제

한글에서 처리코드와 출력코드는 서로 다를 수 있다. 처리코드는 조합형이어야 하지만, 출력코드는 조합형, 완성형, 절충형의 3종류가 가능하다. 이 가운데 CAP용 출력 방식에는 절충형이 적당하다. 그 이유는 11,172개의 모든 현대 한글 음절을 전부 표현할 수 있고, 글자꼴도 미려하기 때문이다. 또한 11,172개의 자모를 전부 미리 그려놓는 완성형의 방식은 메모리를 많이 필요로 하나 1,400자 정도의 완성형 글자꼴만 그려놓고 나머지 9,700개 정도는 903개의 자소 모양에 코드를 부여하고 이를 조합하여 출력시키는 절충형은 메모리가 훨씬 적게 소요되기 때문이다. 자모를 조합하여 출력시킬 때 조합시키는 시간이 오래 걸릴 것으로 생각하기 쉬우나 메인 메모리의 증가로 여기에 소요되는 시간은 별로 문제가 되지 않는다. 아울러, 절충형은 현대 한글을 전부 표현할 수 있으므로, 통일뒤에도 사용할 수 있는 출력방식이다.

결론적으로, 전자출판 시스템에서 한글 출력 방식으로는 절충형 방식을 택하는 것이 '모든 현대 한글 음절을 나타낼 수 있고, 한글 글자꼴의 미려함 유지'라는 2가지 목표를 모두 해결할 수 있는 방안이라고 생각한다.

본 논문에서는 제안한 절충형 출력코드 체계를 완성하기 위하여는 첫째, 비교 연구로서 중 고등학교 교과서를 분석하여야 하며 둘째, 실제 코드 체계를 확립하고 셋째, 절충형 출력코드로 작성된 결과에 대한 미려도 검사가 이루어져야 한다. 이들을 다루지 못한 것이 본 논문의 제한점이라 생각하며, 앞으로의 과제로 남겨둔다.

참고 문헌

- [1] 송 현, 한글자형학, 27, 디자인사, 1985.
- [2] 이 기성, 한글 글자꼴과 컴퓨터 출판매체, 월간코스마, 통권 13호, 126-133, 1990.
- [3] 이 기성, 한글 글자꼴과 전자출판, 한글글자본 제정 경위보고서, 23-27, 1991.
- [4] 이 기성, 사무자동화, 218-221, 장왕사, 1991.
- [5] 이 기성, 사진식자개론(중보판), 23, 장왕사, 1991.