

한글문자의 로마문자 표기에 대한 제안

강 주 상
고려대학교 물리학과

A Proposal for Roman Representation of Hangeul Characters

Joo Sang Kang
Korea University

초 목

한글문자를 기계화하기 위하여 로마문자로 표기하는 방법에 대한 새로운 제안을 한다. 현재 남한안이나 북한안은 서로 이견을 보이고 있어서 국제적 합의에 이르지 못하고 있다. 본 제안에서는 남, 북한 안과는 달리 영문자의 대문자와 소문자를 모두 활용하며 다양한 모음의 구성에서 음성학적으로 표기한다. 이 표기법은 자체의 합리성, 국제표준과의 호환성, 판독의 편리성, 음성인식의 이용성등에서 기존 표기체계보다 유리하다.

I 서 론

정보화 사회로 접어들면서 우리는 정보의 신속한 기계적 처리능력에 크게 의존하고 있다. 그리고 정보는 대부분의 경우에 문자로 표현되므로 문자의 기계변환 코드체계를 세울 필요성을 절실하게 느낀다. 그러나 과거 반세기 동안 급속히 발전한 계산과학의 결과 우리는 기계적 변환을 의미할 때 항상 "컴퓨터"를 연상하게 된다. 문자의 코드체계는 임의로 정의할 수 있겠지만 정보산업의 바람직한 성장을 위하여는 표준화되는 것이 중요하다. 미국의 ASCII는 그 대표적인 코드체계이며 국제적인 표준화를 위하여 ISO (International Standards Organization)는 많은 노력을 기울이고 있다.

우리나라의 정보산업은 무엇보다도 한글코드의 통일성 결여로 인하여 성장의 큰 장애를 안고 있다. 기업

체간의 코드가 다를 뿐 아니라 심지어는 같은 기업내의 계열회사간에도 차이가 있는 경우가 있다. 완성형, 조합형, N-byte형 등 여러 형태가 산재하며 같은 형태 내에서도 여러가지 종류의 코드체계가 존재하는 것이 현실이다. 이는 한국이 정보화 사회로 도약하려는 전략적 측면에서 가장 큰 문제점이며 정부는 최근에 표준화된 KS 코드를 제정하여 발표하였다. 표준화된 코드는 정보산업 발달의 필수적인 요소이기 때문이다. 그러나 KS 코드에 대한 비판적인 시각도 많으며 이는 기존 코드체계에 난맥상을 부채질 할 것인지, 아니면 강력한 정책추진으로 통일화된 표준코드가 될지 두고 관망할 문제라고 판단된다.

한편 한글문자를 로마문자(영문자)로 표기하여 기계화작업을 도울 수 있도록 코드체계를 정립하는 것은 또다른 측면에서 중요하다. 외래어 표기법과도 밀접한 관련이 있기 때문이다. 한편 ISO는 한국어를 포함

하여 일본어, 중국어, 아랍어, 그리스어 등 세계각국의 문자를 로마문자로 표기하는 방법을 국제규격으로 표준화 하려고 추진하고 있기 때문에 그 필요성이 더욱 절실히 지고 있다. 하지만 여기에도 통일성이 없이 시행착오의 시간적 진화를 하고 있으며 결과는 혼돈상태에 있다. 구 문교부안, 새로운 문교부안, 맥퀸-라이샤워 표기법, 심지어는 한걸음 나아가서 남,북한간의 이견등 복잡한 양상을 이루고 있다.

이러한 상황속에서 또다른 한글문자의 로마자표기 방안을 제시한다는 것은 혼란을 가중시킨다는 비난을 받을 수도 있다. 그러나 이 제안에는 지금까지의 표기체계에서 간과된 면이 수용되었다는 점과 음성합성, 음성인식등 정보과학의 다른 응용분야에도 활용될 수 있다는 점에서 깊은 의미를 가질수 있다. 한글문자의 로마문자 표기방법도 임의로 정의할 수 있는 문제이지만 과학적으로 체계를 정립하면 정보처리능력을 향상시키고 결과적으로 대중에게 쉽게 수용될 수 있기 때문에 표준화에 앞서서 심층적 연구를 하여야 할 것이다.

II 한글문자의 로마문자 표기에 대한 고려사항

모든 문자체계에서 공통적인 현상이지만 로마 문자 표기법에도 어떤 원칙이 필요하다. 이를 위하여 ISO는 몇가지 원칙을 정하고 있다. 첫째는 문자간의 자동변환 가능성으로 한글문자를 로마문자 또는 그 반대로 변환할 때 모호성을 배제하는 것이다. 이는 기계적 처리를 용이하게 해 준다. 그러나 역변환을 알 경우 원래의 문자를 재생하기 위하여는 1대1 대응원리에 의한 변환이 바람직하다.

둘째는 정확한 발음표기와의 무관성이다. 전자법(轉字法)체제를 알고 한글을 아는 사람은 누구나 원음대로 발음할 수 있으면 된다. 그렇지만 가급적 자음은 자음에, 모음은 모음에 대응시키는 것이 좋고 그것도 비슷한 음가를 가진 문자끼리 짝지우는것은 전자법 체계를 익히는데 훨씬 도움이 될 것이다. 여기에 큰 문제점이 있다. 로마문자는 모음의 수가 한글문자보다 훨씬 작기 때문이다.

셋째는 자의적 전자(轉字)문자나 기호의 선택을 회피하자는 것이다. 새로운 기호의 도입은 코드체계상 새로운 문자를 창출하는 것과 같으므로 로마문자의 초집합(superset)을 만들게 된다.그리고 이와 함께 음성학적 논리성을 유지하는 것이 바람직한것은 물론이다.

이러한 취지로 ISO는 1985년 남한과 북한에 한글의 로마자 표기법에 대한 규격초안을 제출하도록 요청하였다. 남한은 부처마다 표기법이 달라서 선뜻 초안이 마련되어 있지 않았지만 공업진흥청이 주관하여 한국정신문화연구원, 한국표준연구소등이 세밀한 검토를 거쳐 1986년 ISO에 표기법안을 제출하였다. 그러나 북한은 다른 표기법안을 제출하였으며 1987, 1988년 두 번의 공동회의에서 두 초안간에 이견이 노출되었고 타협점이 이루어 지지 않았다.

남북한 안들의 차이점은 주로 ㄱ, ㄷ, ㅂ, ㅈ 과 이들의 탁음,경음에 대한 표기법과 상당한 수의 모음 표기법이다. 이들에 대한 구체적 표기법은 본 논문의 제안과 함께 도표1에 실려져 있다.

우리는 ISO의 대원칙이외에도 바람직한 요망조건을 몇가지 더 고려할 수 있다. 1대1 대응이라 하더라도 한 문자를 한 문자에 대응시키는 것이 한 문자를 두 문자에 대응시키는 것 보다는 전자법에 대한 알고리즘 작성에 훨씬 편리하다. 예를 들면 북한안은 ㄱ을 KH로 표기하도록 되어 있으나 역변환시에는 좀더 상세한 알고리즘이 필요하다. 또한 대중 수용성도 고려해야 한다. 전자법 체계는 과학적이고 논리적이어서 한글과 로마문자를 아는 사람은 쉽게 이해할 수 있어야 한다. 한편 음성인식, 재생등이 용이하도록 정의되면 한글문화의 정보화 사회 진입에 큰 도움이 될 것이다.

III. 한글문자의 로마문자 표기법

로마문자 표기법에 남,북한 안들과 본 논문의 큰 차이점은 로마문자의 대문자와 소문자를 구별하여 한글문자와 대응시키자는 제안이다. 로마문자의 모음은 기본적으로 A,E,I,O,U 다섯개이지만 한글에서는 복모음까지 합하여 21개가 사용되고 있다. 남,북한안에서는 대문자만 사용되고 있다. 아직도 렐렉스등에서

는 대문자만을 사용하고 있지만 이미 보편화된 대부분
의 전자통신수단에서는 7비트 내지 8비트 코드를 사용
하고 있다. ISO-646, ISO-4873, ASCII, KS 코드등에
서 공통으로 영문자의 대문자와 소문자를 상이한 코드
로 정의하고 있으므로 이를 최대한 활용하는 것이 나
람직하다. 따라서 우리의 제안은 로마문자의 대, 소문
자를 사용하고 대문자 전용 환경하에서는 특별규약을
두자는 것이다.

이미 전술한 바와 같이 여러가지 원칙을 정할 수 있
겠으나 이들을 엄격히 따르게 되면 모순적인 상황에
이룰 수도 있다. 그러므로 우리는 가급적 원칙에 충
실하면서 한글문자의 특수성을 고려하여 융통성있게
표기법을 정의하고자 한다.

우선 자음구조에 있어서 ㄱ, ㄷ, ㅂ, ㅈ 는 남한안과
같이 g, d, b, j 이지만 소문자로 사용한다. 그리고 탁음
ㄱ, ㄷ, ㅂ, ㅈ은 k, t, p, z 로 표기하지만 경음 ㄱ, ㄷ, ㅂ,
ㅈ는 대문자 G, D, B, J 로 정의한다. 따라서, ㅅ은 S 이
다. 자음 ㄹ은 초성(설측음)일 경우에는 r, 종성(설
전음)일 경우에는 l 로 표기한다. 그리고 〇 의 경우
에는 초성은 무성음이므로 ' , 종성은 대문자 N 으로
표기한다.

모음의 경우에는 단모음과 복모음을 구별하도록 최
대한의 고려를 하였다. ㅏ, ㅓ, ㅗ, ㅛ, ㅜ 는 쉽게 로
마문자의 a, e, i, o, u 에 대응시킬 수 있지만 ㅡ,
ㅝ, ㅞ 등이 문제가 된다. 따라서 독일어의 움라우트
(umlaut)를 연상하도록 ㅝ, ㅞ 는 대문자 A, O로 정의
하였고 한글에서는 단모음이 아니지만 기를 U 로 표기
하도록 하였다. 모음 ㅡ는 영문 표기를 찾기 힘들어
서 소문자 w 로, 단모음 ㅜ 는 나머지 모음자 E 로 정
의하였다. 그러면 기타의 복모음은 단모음들의 결합
이므로 표기도 그렇게 되도록 정리하였다. 여기서 한
가지 예외는 복모음 "ㅝ" 이다. 이 규칙에 따르면
"wi" 로 표기하여야 하였으나 한글문장에서 "ㄱ" 의
빈도가 많은 점을 고려하여 "y" 로 정의하였다. 실제
적인 면에서 타전수를 상당히 줄일 수 있는 장점이 있
기 때문이다.

우리의 제안은 도표1에 남, 북한안들과 비교되어 제
시되어 있지만 독자적으로 도표 2 에 자음, 쌍자음,

복자음, 모음순으로 나열되어 있다. 그 외에도 한글
문자 <-> 로마문자 변환을 아무 모호성없이 수행하기
위하여 몇가지 변환 규칙이 필요하며 이들을 다음에
열거하고자 한다.

(1) 단어의 첫글자에 나오는 무성음 〇 는 무시한다.
단어 앞의 빈칸으로써 그 존재를 확인할 수 있기 때문
이다.

보기: 애정 ----> AJEN ('AJEN 이 아님)

(2) 복모음이 아닌 두 모음은 무성음 ' 로 분리한다.

보기: 여자 ----> iEja, 이어자 ----> i'Eja

(3) 두 모음 사이의 한자음은 뒤의 문자에 속하는 것
으로 정의한다.

보기: gagu ----> 가구 (gag'u ----> 각우)

(4) 두 모음 사이의 두자음은 앞, 뒤 문자에 각각 한
자음씩 속하는 것으로 정의한다.

보기: gaggo ----> 각고

(gaGo -> 가꼬, gaG'o -> 각오)

anja ----> 안자 (anj'a -> 얌아)

(5) 두 모음사이의 세자음은 앞문자에 들, 뒷문자에
한자음씩 속하는 것으로 정의한다.

보기: anhgo -> 앵고

(6) 로마대문자만 사용하는 상황에서는 대문자를 소문
자로 간주하고 대문자는 "~" 기호를 문자앞에 넣어서
표현한다.

보기: 연애 ----> I\EN'\A

쉬다 ----> S\EN\GDA

IV. 결 론

본 논문에서 제안한 로마문자 표시법은 한글문자의
기계화 과정에서 이미 코드상으로 분리된 영문자의 대,
소문자를 최대한 활용하였다는 면에서 현재의 남, 북
한의 공식적인 초안보다 편리하다고 판단된다. 단모
음과 복모음을 구별하는 것은 음성인식 연구에 유리할
것이며, 이 표기법에서 미사용 문자가 많은 점은 한글
고어 표기에 큰 도움이 될 것이다. 대문자만 사용하
는 남한의 공식안에서는 미사용 문자가 5개이지만 본
논문의 제안에서는 21개의 여유가 있으므로 고문 연구

에 필요한 문자들을 쉽게 대응시킬 수 있다.

이 표기법은 필자가 외국 여행 중 한국과의 전자통신수단을 통하여 3년간 사용한 경험이 있으며, 문제점을 발견하지 못하였다. 그러나, 이 표기법과 연관된 다른 연구들이 수행되어야 그 효용성이 더욱 분명해지리라고 판단한다.

이 연구에 앞서 한국데이터통신 주식회사가 후원한 한강사랑방 모임에서 여러 회원들과의 의견 교환이 많은 도움이 되었음을 밝혀둔다. 특히 “니”의 “y”표기는 유 경희님의 착상을 최대한 수용하였으며 이에 감사한다. 또한, 한국표준연구소의 박 동순님의 도움으로 남,북한 초안의 배경자료를 얻게 된 것을 고맙게 생각한다.

