

## 한글 로마자 자동 표기 시스템

### 설계 및 구현

#### ( 轉寫法 기계적 변환에 관한 연구 )

김홍섭\*, 박종섭, 이현걸  
동국대학교 교육대학원 전산교육과

The design and implementation of  
automatic translation system for  
hangul's romanization

( A study on mechanic conversion using  
transcription )

Kim. Hong-Söp , Pak. Jong-Söp, I Ilyön-Köl  
Computer science education graduate course of Dong-kuk University

### 요 약

국제 협약에 따라 한글에 관한 정보는 로마자로 표기해야 한다. 우리나라는 1959년 2월 로마자 표기법'을 제정, 수차례 개정을 통해 1983년 6월 문교부(Ministry Of Education)안을 발표했으나, 표지판, 역명, 교과서, 공공문서등에서 활용되었지만, 영자신문, 외국 학술지등은 M-R( McCune-Reishauer ) 표기법을, 인명, 신문 및 방송매체등은 혼합표기방식을 사용함으로써 人.地名, ID, 대표어등 정보 검색시 혼란을 야기시켰고, 개정안에 따른 표기 정정으로 수백억원에 가까운 예산을 낭비하였으며, 최근 ISO(International Standard Organization)에서는 남북한 단일화 및 기계적 변환을 요구하고 있으나, 반달표(~)표기곤란, 편리성 결여, 북한의 주장등의 사유로 제정등을 거론하고 있는 실정에 있다. 잘 쓰지 않는 ASCII 코드 중에서 반달점 폰트를 제작하고 단어, 문장, 문서를 STRING으로 받아 알고리즘화된 음운법칙을 적용하여 소리글자로 변환하고 MOE 테이블에서 대응글자를 참조하여 기계적 변환이 가능하도록 하였으며, 세련된 디자인, 풀다운 팝업방식을 채택, 2HD 1장으로 국내최초로 개발하였다.

# I. 서론

## 1. 연구의 필요성

정보화 사회의 도래로 우리는 정보의 홍수 시대에 접하여 살고 있다. 이에 컴퓨터의 활용은 더이상 언급하지 않아도 될만큼 보편화 되어 있는것이 사실이다. 특히 정보의 검색에 있어서 컴퓨터의 이용가치는 두드러지게 되는데 이러한 정보는 대부분이 문자로 표현된다. 문자는 컴퓨터에 의해 처리되어질때 코드로 처리되게 되어 있다. 아스키 코드가 대표적인 예이다. 우리나라는 한글이란 우수한 문자를 가지고 있고 국제적인 교류에 있어서는 국제적인 의사소통의 매개체로 통용되는 로마자로 표기 되어야 한다. 현재 우리나라 로마자 표기법은 도로표지판, 교과서, 정부의 공식 문서등에는 문교부 개정안(1984)으로 표기되어 있고 국내의 일부언론과 학술자들의 인명과 지명, 미국을 비롯한 외국의 각종 자료등에는 McCune-Reischauer(1939) 방식또는 이를 준용하고, 기업체의 명칭과 상품명 또한 개개인의 인명과 특정한 고유명사 표기에는 임의적인 혼합방식, 즉 관습적인 표기방식(송기중 1988:109)으로 쓰고 있어서 각종자료를 읽고 쓰고 있어서 엄청난 혼란과 시간을 낭비하고 있다. 예를들어 Kyonggi-Do, Kyung Ki Do, Gyeonggi-do 등으로 여러가지로 표기 되었다. 이로한 연유로 로마자로 표기된 국어의 혼란을 방지하고 한글을 로마자로, 전사처리, 송수신, 복원하는 일련의 과정에서 모호성이 전혀 발생하지 않는 로마자 전사법의 표준화와 그 과정을 기계화할 수 있는 소프트웨어의 개발은 국제 정보화 시대에 국가적으로 시급히 해결해야할 문제들이다.

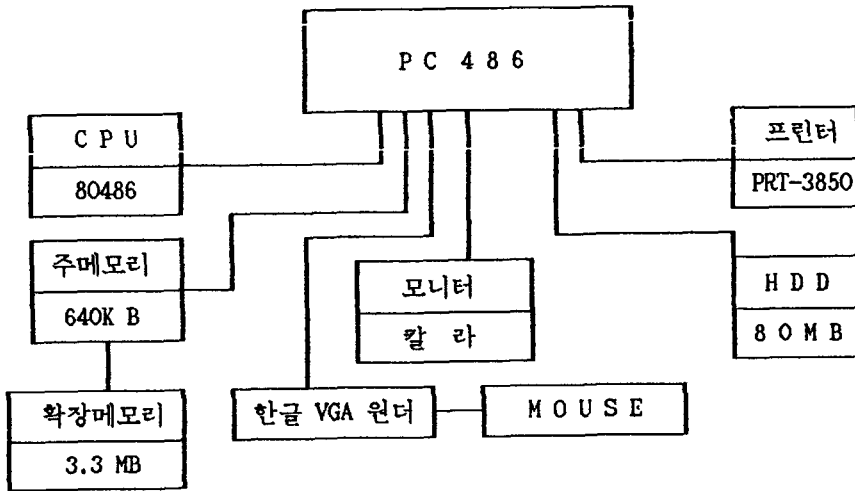
## 2. 연구내용및 기대효과

본 로마자 자동변환 시스템은 전사법의 기계적 변환에 초점을 맞추었다. 로마자표 기법에 관한 표기법에는 전자법과 전사법이 있는데 이 두가지를 비교하고 전자법 소프트웨어는 연구된바 있으나 전사법의 소프트웨어는 연구가 없었던 이유에 대한 문제점을 파악하고 전사법의 문제방안을 해결하고자 한다. 이러한 문제방안을 해결하면 다음과 같은 효과를 얻을수 있다. 기대효과로는 로마자 변환시 수작업 및 시간의 최소화할 수 있고 다량의 변환및 표시가 가능해야한다. 예를 들어 행정기관이나 학교, 지도제작이나 신용카드회사와 같은 로마자를 자주사용하는 기관의 시간을 최대한으로 줄일수 있다. 또한 연간 수백억대의 막대한 도로 표지판도색 작업은 이시스템을 CAD/CAM 장비와 연결하면 표준화와 동시에 많은 예산절감 효과가 있으리라 생각된다. 이러한 시스템구현을 위해 먼저 사용자 요구사항을 알아보자. 우선 표준화 되고 정확한 로마자로 변환이 되어야하고 단어, 문장, 문서, 이름의 로마자 표기및 인쇄가 가능해야 한다. 요즘 유행하고 있는 HYPERTEXT HELP 기능을 지원하여 신속하고 편리한 도움말이 지원이 되어야 한다. 오래 지속해서 소프트웨어를 쓰려면 사용하기에 단조롭거나 지루해서는 않되기 때문에 세련된 디자인이 바탕이 되어야 하겠다. 이러한 요구사항들을 최대한 반영하여 한글로마자 표기법을 만들었다.

## II 시스템 구성

### 1. H/W

「 한글로마자 자동표기 시스템 」은 IBM PC/486를 사용하여 개발하였으며, 호환기종인 XT, AT, 386에서도 무난히 수행된다. 주메모리는 640KB정도면 충분하고, VIDEO CARD는 EGA, VGA 를 지원한다. 프린터는 한글(완성형, 조합형 코드)을 지원하는 제품이면 사용 가능하며 사용자의 편의를 위해 마우스를 지원하고 2D한장이면 모든 시스템이 설치가 가능하므로 범용적으로 쓸수있다.



< 그림 1 > 하드웨어 구성도

### 2. S/W

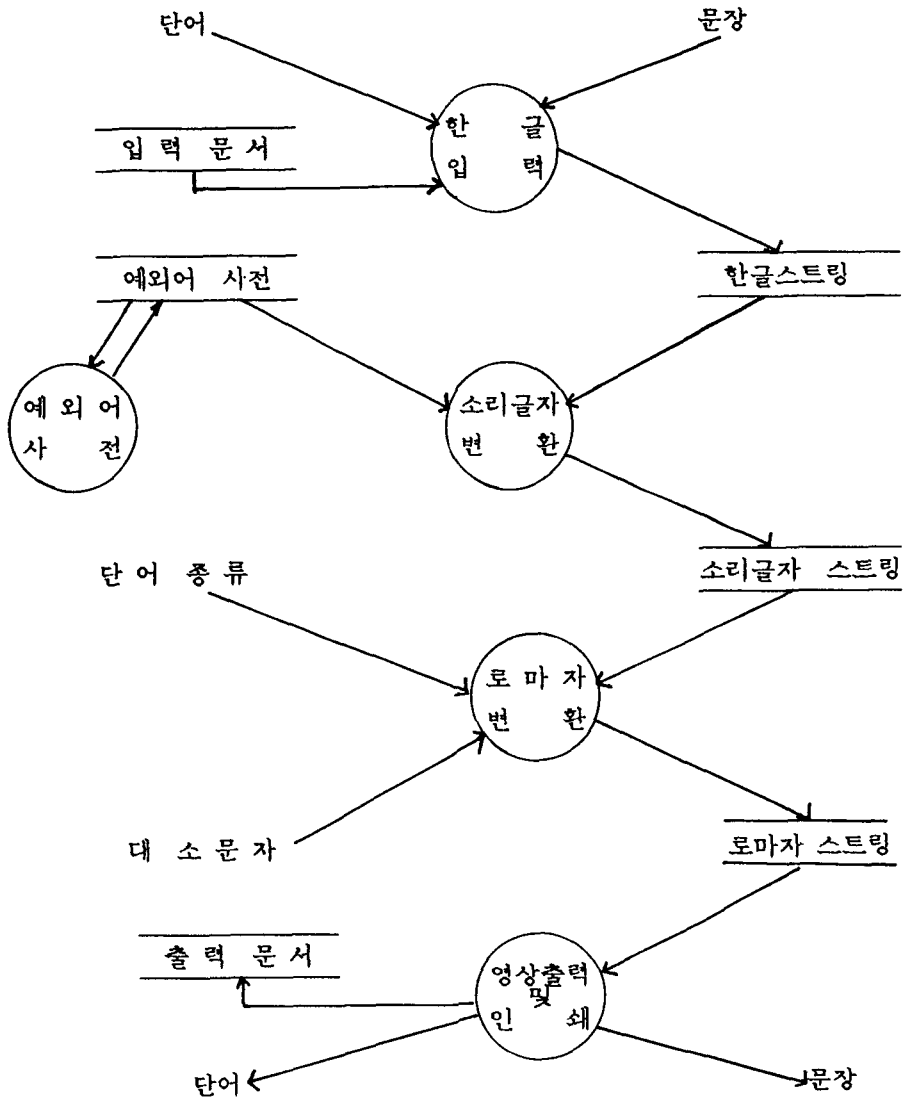
개발에 사용한 S/W 로는 OS 는 MS-DOS Ver 6.0이 있으며 개발 환경으로 Borland 사의 Borland-C++ 3.1을 이용하였다. 그래픽 처리를 위해 한글 라이브러리를 자체 개발하여 사용하였으며 폰트제작을 위해 폰트제작도구인 폰트 에디터를 사용하고특전사 마크그리기 위해 Image256과 Image Scanner 가 사용 되었다.

구 분	사 양
운 영 체 제	MS-DOS Version 6.0
사 용 언 어	BORLAND C++ Version 3.1
영 상 출 력	GUI(Graphic User Interface) 환경
라이브러리구축	한글 라이브러리 자체제작
인 쇄 양 식	다양한 폰트제공
	크기와 모양조정
	미려한 인쇄 제공

< 표 1 > 소프트웨어 특징

### 3. 자료흐름도

아래의 그림은 단어, 문장, 입력문서를 스트링으로 받아서 한글스트링을 만들어내고 음운법칙을 적용하여 소리글자로 변환하고 변환과정에서 예외어를 참조하고 소리글자 스트링은 단어의 종류를 참고하고 대.소문자를 선택여부를 참고하여 로마자스트링을 만들어내고 영상으로 출력할것인지 프린터로 출력할것인지 화일로 출력할것인지를 사용자로부터 선택을 받고 출력하게 되는 자료흐름도를 나타낸다.



< 그림 2 > 한글 로마자 자동표기 시스템의 자료흐름도

### III 시스템 설계

#### 1. 개정된 한글 로마자표기법의 특징

- (1) 표음주의 원칙을 채택하여 국어의 표준 발음을 소리나는 대로 표기  
보기 독립문 : Doglibmun => Tongnimmun
- (2) '한글의 로마자 표기법'을 '국어의 로마자 표기법'으로 명칭을 변경함
- (3) 모음 ㅏ, ㅑ와 파열음, 파찰음을 외국인의 실제 발음에 접근 되도록 조정  
보기 거북선 : Geobugseon => Köbüksön  
김 포 : Gimpo => Kimp'o  
부 산 : Busan => Pusan
- (4) 고유명사 표기 방식의 통일을 위한 규정을 신설  
보기 김정호 : Kim Chǒng-ho  
전라북도 : Chöllabuk-do  
속리산 : Songnisan
- (5) 국제 관계 및 종래의 관습적 표기를 고려해서 최소 한도의 관용을 허용  
보기 서울 : Seoul
- (6) 반달점(·)와 어깨점(˘)은 인쇄나 타자의 어려움이 있을 때는 의미의 혼동을 초래하지 않는 범위 내에서 생략을 허용

과 같이 되어 있다.

1번 사항을 기계적으로 변환하기 위해서 모든 음운법칙을 알고리즘화 하는데 많은 시간과 노력을 하였다. 초성, 중성, 종성을 분리해야 하는 번거로움과 많은 음운법칙을 적용해야하므로 시간문제를 고려하여 많이 쓰이는 음운법칙부터 가장 잘 쓰이지 않는 음운법칙별로 소트하여 음운법칙을 적용 하였다. 2번 사항은 기계화로 변환하는과정에서는 별로 큰 의미를 두지 않으므로 크게 어려움을 겪지 않았다. 3번과 4번사항은 임명변환과 고유명사변환규정을 별도로 적용하여 처리 하였다. 5번 사항은 예외어처리 사전을 별도로 운영하여 예외어의 증가와 소멸에 별문제 없게 작성하였다. 변환과정에 있어서 예외어가 발견이 되면 예외어처리를 할것인지 그렇지 않을것인지에 관하여 의사결정을 물은후에 처리 할수 있도록 하였다. 6번 사항은 반달표의 폰트를 특별히 제작하여 사용하였다.

#### 2. 국어의 로마자 표기법 원칙

##### 제 1장 표기의 기본 원칙

제 1항 국어의 로마자 표기는 국어의 표준 발음에 따라 적는다.

제 2항 로마자 이외의 부호는 되도록 사용하지 않는다.

제 3항 1음운 1기호의 표기를 원칙으로 한다.

##### 제 2장 표기 일람

제 1항 모음은 다음과 같이 적는다.

단모음: ㅏ ㅑ ㅓ ㅕ ㅗ ㅛ ㅜ ㅠ ㅡ ㅣ ㅞ ㅟ ㅠ ㅡ ㅣ

a ö o u ū i ae e oe

중모음: ㅝ ㅞ ㅟ ㅠ ㅠ ㅡ ㅣ ㅤ ㅥ ㅦ ㅧ ㅨ ㅩ ㅪ ㅫ ㅬ ㅭ ㅮ

ya yö yo yu yae ye ūi wa wo wae we wi

(붙임) 장모음의 표기는 따로 하지 않는다.

제 2항 자음은 다음과 같이 적는다.

파열음: ㄱ ㅋ ㆁ ㄷ ㅌ ㅍ ㅂ ㅃ ㅍ  
 k, g kk k' t, d tt t' p, b pp p'

파찰음: ㅈ ㅊ ㅅ  
 ch, j tch ch'

마찰음: ㅅ ㅆ ㅎ  
 s, sh ss h

비 음: ㅁ ㄴ ㅇ  
 m n ng

유 음: ㄹ  
 r, l

(붙임1) 'ㄱ, ㅋ, ㆁ, ㅅ'이 모음과 모음사이, 또는 'ㄴ, ㄹ, ㅁ, ㅇ, '과 모음 사이에서 유성음으로 소리날 때에는 각각 'g, d, b, j'로 적고 이 외에는 각각 'k, t, p, ch'로 적는다.

(붙임2) 'ㅅ'은 '시'의 경우에 'sh'로 그 외에는 's'로 적는다.

(붙임3) 'ㄹ'은 모음앞에서는 'r'로 적고, 자음앞이나 낱말의 끝에서는 'l'로, 'ㄹㄹ'은 'll'로 적는다.

### 제 3장 표기상의 유의점

제 1항 음운변화가 일어날 때는 변화의 결과에 따라 다음과 같이 적는다.

1. 자음 사이에서 동화작용이 일어나는 경우
2. 'ㄴ, ㄹ'이 덧나는 경우
3. 구개음화가 되는 경우
4. 'ㄱ, ㅋ, ㆁ, ㅅ'이 'ㅎ'과 어울려 나는 경우

(붙임) 형태소가 결합할 때 나타나는 된소리는 따로 표기하지 않는다.

제 2항 발음상 혼동의 우려가 있을 때나, 기타 분절의 필요가 있을 때는 '-'(짧은줄표)를 써서 따로 적는다.

(붙임) 인명과 행정구역 단위명 표기에서 '-'(짧은줄표) 앞뒤에서 일어나는 동화작용은 표기에 반영하지 않는다.

제 3항 고유명사는 첫소리를 대문자로 적는다.

제 4항 인명은 성과 이름의 순서로 쓰되 띄어쓰고, 이름 사이에는 '-'(짧은 줄표)를 넣는다. 다만, 한자식의 이름이 아닌 경우에는 '-'를 생략할 수 있다.

제 5항 제2항 붙임의 규정에도 불구하고 '도, 시, 군, 구, 읍, 면, 리, 동'의 행정구역 단위와 '가'는 각각 'do, shi, gun, gu, ūp, myŏn, ri, dong, ga'로 적고 그 앞에는 '-'(짧은줄표)를 넣는다.

(붙임) 특별시, 직할시, 시, 군, 읍 등의 행정구역 단위명은 생략할 수 있다.

제 6항 자연지물명, 문화재명, 인공축조명은 '-' 없이 붙여쓴다.

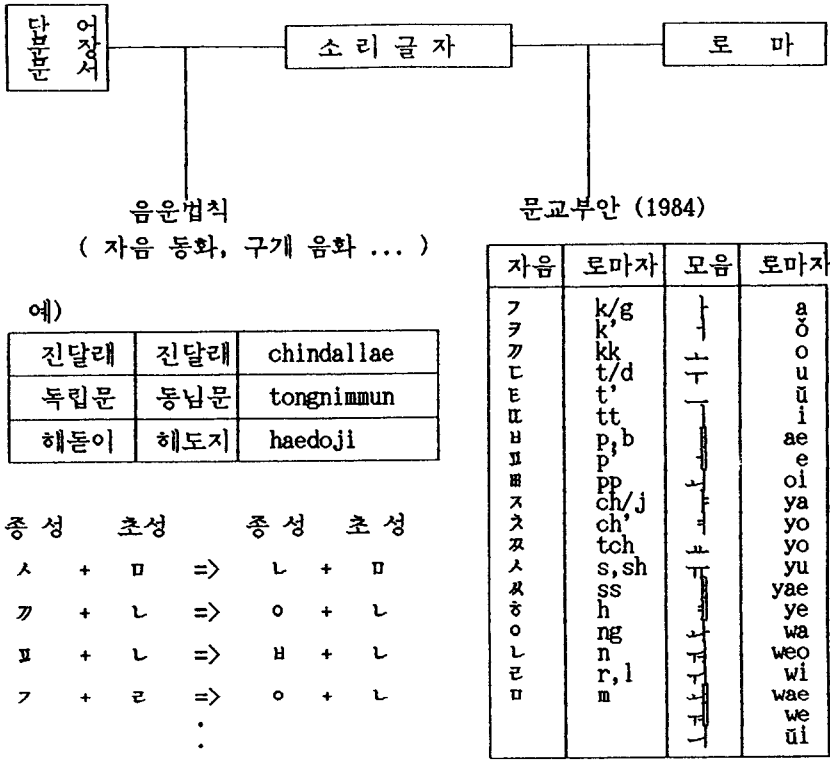
(붙임) 5음절 이상일 경우에는 낱말 사이에 '-'를 쓸 수 있다.

제 7항 고유명사의 표기는 국제관계 및 종래의 관습적 표기를 고려해서 갑자기 변경할 수 없는 것에 한하여 다음과 같이 적는 것을 허용한다.

제 8항 인쇄나 타자의 어려움이 있을 때에는 의미의 혼동을 초래하지 않을 경우 ö, ū, yŏ, ūi 등의 '˘'(반달표)와 k', t', p', ch' 등의 '˘'(어깨점)을 생략할 수 있다.

위의 열거된 사항들을 모두 알고리즘화 시키는 과정에 관하여 논의 해보기로 하자. 단어, 문장, 문서를 스트링으로 입력받아 초성, 중성, 종성으로 분리된 한후 자음동화, 구개음화, 두음법칙, 예외어규정, 일반단어, 대소문자, 인명, 고유명사, 행정구역의 변환법칙을 적용하여 소리글자 스트링을 만들어내고 소리글자의 스트링을 1984 문교부안의 표기 TABLE을 참조시켜 로마자로 생성한다.

### 3. 변환과정 및 반달점 폰트제작



<그림 3> 한글 로마자 변환 과정

반달점 폰트	아스키 폰트	아스키 번호	참 고
ö	⓪	64	폰트 에디터로 반달점 폰트를 작성하여 아스키 테이블에 대치 하여 사용
ü	-	95	
ø	'	96	
ū	ctrl-backspace	127	

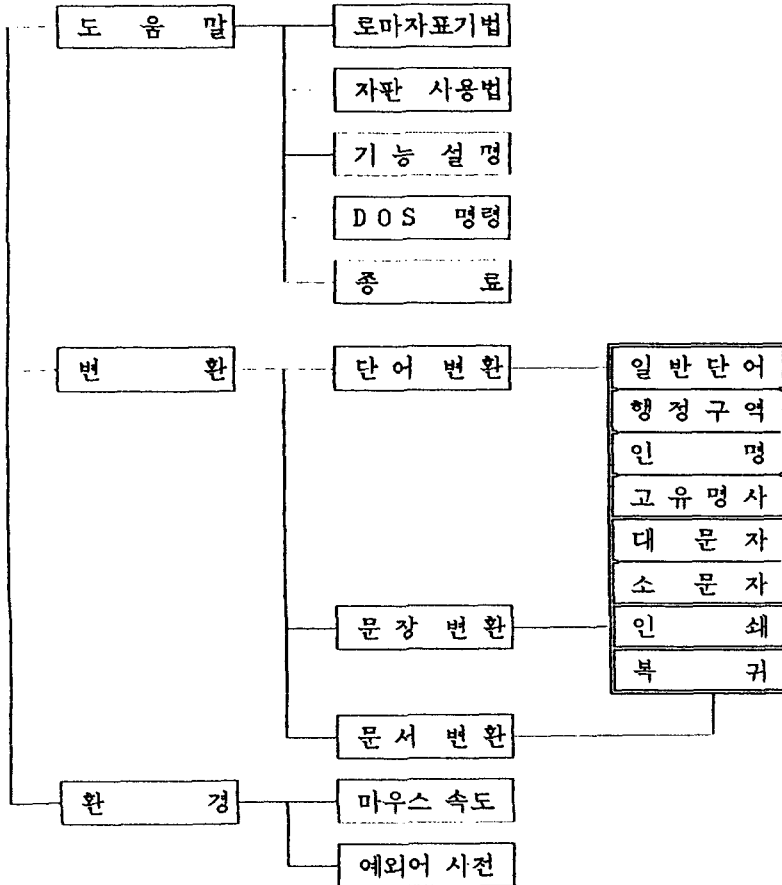
< 표 2 > 반달점폰트의 기계적 변환을 위한 폰트주소

본 로마자 표기시스템은 한글라이브러리를 설계하여 운영하므로서 텍스트상에서의 문제점으로 지적되었던 반달점 표기를 해결할수 있었다. 그리고 음운법칙을 모두 알고리즘화 하므로 사용자로 하여금 최대한의 편리를 제공하고 도움말속에 로마자 표기법의 개정안을 볼수 있음으로 인해 사용자로 하여금 로마자의 관심을 증대시켜 로마자의 통일을 꾀할수 있음을 자랑으로 삼으수 있다. 한글 라이브러를 사용하고도 메모리를 고려하여 XT 급에서 부터 486에 이르기 까지 고루 쓸수 있게 지원하고 있다. 특히 2D 디스켓 한장에 본시스템을 모두 설치 할수 있으므로 초,중,고,대학생까지 별부담없이 모두 사용할수 있다.

## IV. 시스템 구현

### 1. 시스템구성도

한글 로마자 표기법은 크게 도움말, 변환, 환경의 메뉴로 구성되어 있다. 도움말에는 로마자 표기법의 개요 및 개정역사가 설명되어 있고 자판사용법에는 반달점폰트를 사용할수 있는 자판설명이 되어있다. 기능설명은 한글로마자 자동표기 시스템의 기능키 및 운영방법이 설명이 되어 있다. 변환기능은 단어,문장,문서 변환이 있다. 단어변환을 선택하면 단어종류및 대소문자를 선택하시수 있고 고유명사,인명, 행정구역의 단어를 변환할수 있다. 문장변환을 선택하시면 주소,일반문장등을 변환할수 있고 예외어 사전에 등록되어 있는 단어나 문장이면 대치의 여부를 물어 본다. 문서변환을 문서를 한번에 변환할수 있고 환경메뉴에는 마우스의 속도 변경및 예외사전의 등록및 삭제를 시킬 수 있다.

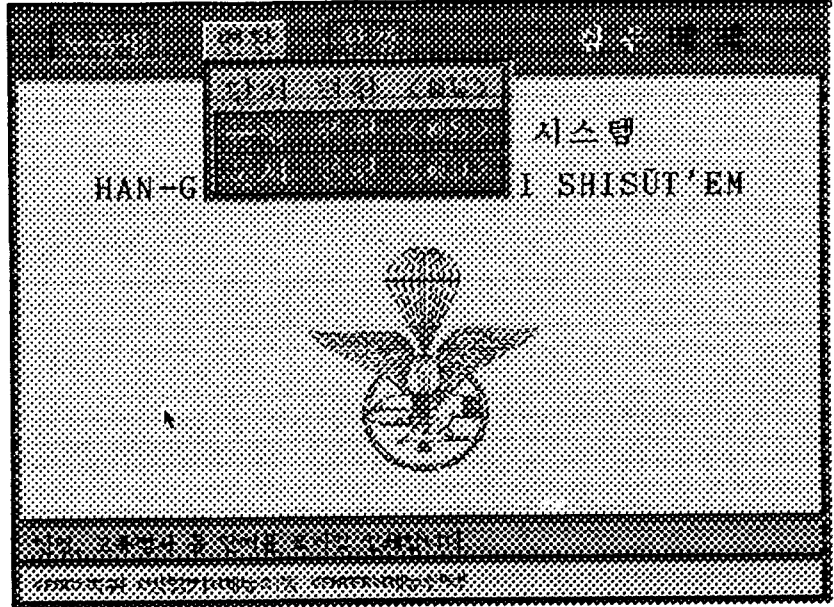


<그림 4> 한글 로마자 자동변환 시스템의 구성도

### 2. 사용자 지침

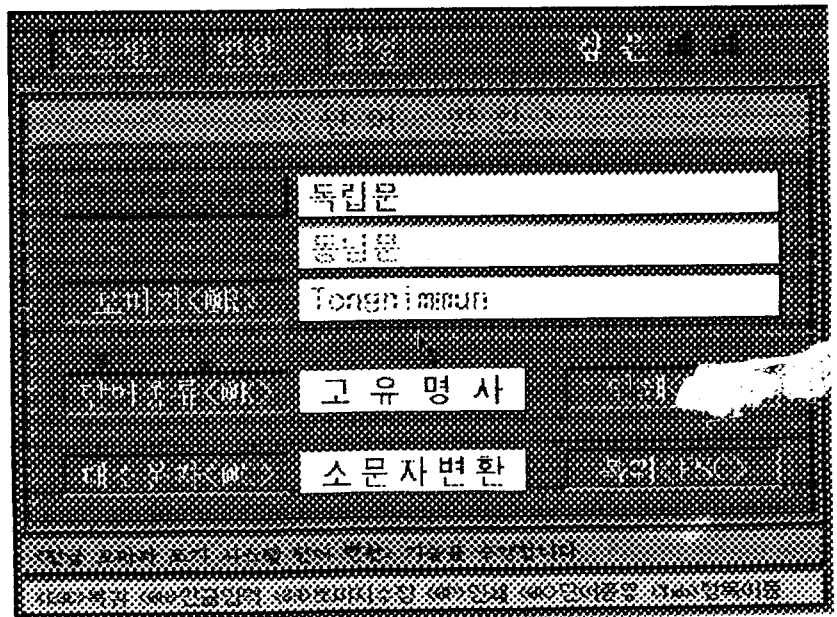


오른쪽화면은한글로마자 자동변환 시스템의 초기화면이다.그래픽 한글라이브러리로 구현이 되었으므로 세련되고 사용하기가 간편하다. 도움말에는 한글로마자 자동포기 시스템의 사용 설명 과 운영법이 잘 설명되어 있다. 그리고 마우스를 지원하므로 마우스 조작과 한글만 입력하면 모든 기능을 수행할수 있다. 마우스가 없는 사용자 를 위하여 핫키 기능및 풀다운메뉴구성으로 사용하고



고 익히는데 별어려움이 없으리라 생각된다. 아래의 그림2는 변환과정의 일부분의 예이다. 변환에는 단어,문장,문서변환이 있다.그림2는 단어변환의 예로서 독립문을 변환단어로 입력하면 음운번칙을 적용하여 소리글자스트링을 만들어 내고 이에 로마자 테이블에 저장되어 있는 로마자 폰트와 대응되어 변환하게 되어 있다.

단어변환의 종류에는 인명,일반단어,고유명사,행정구역이 있다. 로마자로 대응대는 과정에 소문자로 변환할것인지 대문자로 할것인지의 여부도 선택을 할수 있다. 또한 변환된 단어를 직접 프린터로 출력할수도 있게 배려 하였다. 화면의 아래쪽에는 그 당시 화면에 알맞은 설명말이 나오므로 사용자는 어려울때는 도움말을 참조하면 된다. 그리고 고유명사와 인명은 소문자 변환



이더라도 맨앞자는 대문자로 쓰는 규칙을 적용할수 있고 한글단어가 예외어일 경우에는 예외어 사전에서 자동을 찾아내어 이를 로마자 표기법을 따를것인지 아니면 관습적으로 쓰는 관행에 따를 것인지를 결정을 할수 있게 시스템을 설계 하였다.

## V. 결 론

표기법 특징	원음표기력	원어환원성	실 용 성	비 고
전 자 법 Transliteration	하	상	상	음소단위로 로마자 변환(환원성 우수)
전 사 법 Transcription	상	하	하	발음단위로 로마자 변환(청각효과우수)

<표 3> 전자법과 전사법의 특징 비교표

위 표는 서울대학교 이현복 교수의 로마자 표기법 개정시안의 문제점에서 발췌한 것이다. 전자법의 경우는 언어의 환원성과 실용성면에서는 전사법을 능가하지만 외국인의 발음을 들어보면 전혀 엉뚱한 발음으로 들리기 때문에 오해의 소지가 많다. 하지만 전사법은 한국어 발음에 근접한 방식이기에 이러한 문제가 발생하지는 않는다. 그러므로 실용성과 원어환원성의 효과를 높이기만하면 전사법이 더욱 우수하게 된다. 이러한 약점을 한글로마자 자동변환 표기 시스템에서는 타이프나 컴퓨터 자판에 없는 반달점 표기가능과 또다른 개정법이 나온다하더라도 쉽게적응할수 있음으로 실용성을 증가시키고 복잡한음체계의 전산처리로 언어의 환원성및 실용성을 증가시킬수 있었다.더나아가 한글로마자 자동표기시스템의 S/W 개발로 임의사용법이 감소하고 쉽게배울수 있음으로 인해 로마자 통일의 가능성을 증가시키게 되었다. ISO의 남,북한 단일화요구에도 실용성,편리성,으로 요구에 만족시키고 있다. 그러므로 다음과 같은 효과를 얻게 되었다.

표기법 특징	원음표기력	원어환원성	실 용 성	비 고
전 사 법 한글로마자 자동표기 시스템	상	중	상	

< 표 4 > 한글로마자 자동표기 시스템 특징

### <참고 문헌>

- [1]. Nielsen, Jakob, Hypertext & Hypermedia, Academic Press Inc, 1990.
- [2]. Shneiderman, Ben, Designing the User Interface, Addison Wesley Publinsing Company, 1992
- [3]. Shneiderman, Ben & Grey Kearsley, Hypertext Heads-on, Addison Wesley Publishing Company, 1991.
- [4]. 도형수, 한글의 로마자 표기에 관한 연구, 이화여대, 1992
- [5]. 기계화를 위한 한글의 로마자 표기법 소프트웨어 개발, 과학기술처, 1990
- [6]. 이현복, 로마자 표기법 개정시안의 문제점, 서울대학교, 1979
- [7]. 이상익, 국어의 로마자 표기법 문제 종합 검토, 고려대학교, 1981
- [8]. 송기중외, 기계화를 위한 한글의 로마자 표기법, 정신문화원, 1986
- [9]. 유경희, 한글문자의 라틴문자 전자법의 국제표준화에 관한 소고, 1988
- [10]. 최광부, 한글 모아쓰기에 관한 연구, 한국과학기술원, 1978
- [11]. 김복문, 한글의 로마자 신 표기법(안)과 그용도, 우성문화사, 1982

- [12]. 이상억, 한국어 로마자 표기 '82, 고려대학교, 1982
- [13]. 이상억, 북한의 로마자 표기법, 서울대학교, 1982
- [14]. 문화제안내표시판영자표지지원칙, 문화공보부 문화제관리국, 1978
- [15]. 김충배, 우리말 로마자 표기문제, 「언어」, 1978
- [16]. 송기중, 북한의 로마자 표기법, 「국어생활」, 1988
- [17]. 최현배, 문교부 재정의 한글을 로마자 삼기화 로마자의 한글삼기, 1961

< 출력의 예 >

인텔은 1992년초 해가 가기 전에 486을 이을 다음 세대의 X86 계열 CPU를 내놓겠다고 발표했다. 이 때는 486 DX2가 공식 발표되기 전으로, 그것이 펜티엄이라 이름지어지기 전까지 인텔 내부의 프로젝트 코드명 P5로 불리면서 무성한 소문이 나돌 때이다. 그러나 어떤 이유에선지 인텔은 당초의 발표 일정을 거듭 연기하면서 해를 넘긴 올해 3월 22일에야 펜티엄을 공식 발표하였다.

Int'erün 1992nyönch'o haega kagi chöne 486ül iül taüm sedaeüi X86 kyeyöl CPUrül naenok'ettago palp'yohaetta. i ttaenün 486 DX2ka kongshik palp'yodoegi chönüro, kügöshi p'ent'iömira irümjiöjigi chönkkaji int'el naebuüi p'ürojaekt'ü podümyöng P5ro pullimyönsö musöngghan somuni nadol ttaeida. küröna öttön iyuesönji int'erün tangch'oüi palp'yo iljöng-ül ködüp yön-gihamyönsö haerül nömgün olhae3wol 22ireya p'ent'iömül kongshik palp'yohayötta.

이렇게 어렵게 탄생된 펜티엄이 전세계 컴퓨터 사용자의 구미를 어느 정도 만족시킬 수 있을 지 현재로서는 미지수이지만 지금 이 시점에서 그 소문의 진가가 어느 정도 가치를 가진 것인지 간접적으로나마 분석해볼 수 있다는 것은 다행스러운 사실이다. 지금 이 자리에서는 그러한 정보들을 상세히 분석해 봄으로써 펜티엄의 장래를 예측해 보자.

Irök'e öryöpke t'ansaengdoen p'ent'iömi chönsegye k'ömp'yut'ö sayongjaüi kumirül önü chöngdo manjo kshik'il su itsül chi hyönjaerosön mijisuijiman chigüm i shijö mesö kü somunüi chin-gaga önü chöngdo kachöirül kajin köshinji kanjöpchögüronama punsök'aebol su ittanün kösün tahaengsüröun sashirida. chigüm i chariesönün küröhan chöngbodürül sangsehi punsök'ae pomürossö p'ent'iömüi changnaerül yech'ük'ae poja.