

# 다국어 입력기에서 한글 입력의 최적화 방안

유정원\*, 변정용\*

\*동국대학교 컴퓨터멀티미디어학과

e-mail: {ryujw96, byunjy}@dongguk.ac.kr

## An Optimization of Hangeul Input in Multilingual Input Method

Jeong-Won Yoo\*, Jeong-Yong Byun\*

\*Dept of Computer & Multimedia, Dongguk University

### 요 약

다국어 입력기에서 한글 입력의 최적화를 위하여 문자별 보편성과 개별성을 밝혀내어서 어떤 문자의 개별성이 다른 문자의 보편성을 해치지 않도록 충돌 요인을 최대한 낮추어야 한다. 특히 한글은 표음 문자로서 음절 및 음소문자의 특성을 가지고 있고, 한글전용을 하고 있다. 반면에 일본어와 중국어는 음절문자에 해당하며 가나 또는 병음을 입력하여 마지막에는 한자로 바꾸어야 한다. 여기서 훈민정음 창제원리를 기본으로 삼아 최적화에 적용하고자 한다.

핵심낱말: 다국어입력기, 훈민정음창제원리, 입력방법 엔진, 보편성, 개별성 최적화

### 1. 서론

유니코드 (Unicode)[5]는 아스키(ASCII)의 확장으로 다국어입력에 대한 표현을 지원하며 동시에 입력할 수 있는 문자집합이 확대되고 있다.[9,10]

문자마다 가진 보편성과 특히 개별성을 제대로 파악하지 못하게 되면 다국어입력기에서 다른 문자를 입력할 때 많은 장애를 일으킬 수 있다.

본 논문은 한글 또는 조선글을 입력할 때 지리적으로 가까운 나라인 일본 및 중국 문자 입력을 많이 하고 있다는 점에서 충돌 요인을 최소화 하는 방안을 제안하고자 한다. 또한 훈민정음 창제원리를 논의의 기준으로 삼아서 한글부호의 잘못된 제정으로 빚어지는 문제가 무엇인지를 밝혀내서 그 해결책을 제안하여 한글입력을 최적화 하고자 한다.

훈민정음의 과학성에 대하여 연구가 이루어져 왔고, 기존의 국가규격 문자부호계에 대하여 개정하여야 한다는 지적과 연구 결과가 있었다.[1,2,3]

남쪽 국가규격인 정보교환용 한글부호(KS X 1001-2002)[12]와 북쪽 국가규격인 정보교환용 조선글부호(KPS 9566-1997)[4,7,11]는 부호화 대상 음절자의 집단과 자모의 순서가 서로 다르며, 훈민정음

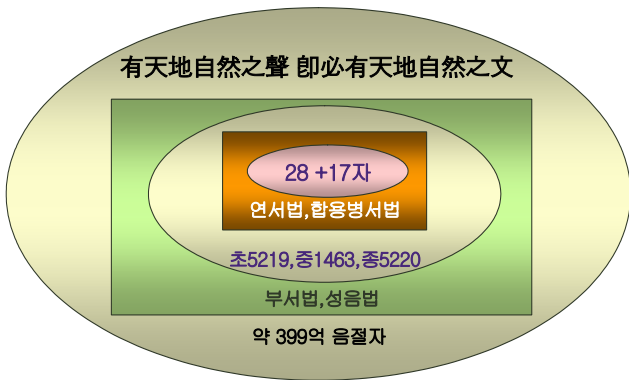
창제원리에 따라서 자동틀(automaton)을 저마다 잘 개발하여 사용하고 있다. 한글의 입력기는 기존에 많은 제품이 문서편집기에 탑재되어 활용해 왔고, 마이크로소프트 제품의 경우 기존의 연구결과를 바탕으로 보급하고 있고, 리눅스의 경우 많은 제품이 있지만 “나비”[8] 등이 인기가 있다. 북쪽도 최근 조선컴퓨터센터와 평양정보센터에서 각각 개발한 입력기를 “하나”[7]라는 이름으로 통합되었다. 또한 다국어 입력기의 경우 최근에는 일본어와 중국어 등을 기본으로 하여 유니코드에서 지원하는 모든 문자를 입력할 수 있도록 그 기능을 확대해 왔다.

일본어와 중국어 입력기는 가나(히라가나, 가타가나)를 입력하여 한자로 변환하거나, 병음 또는 주음으로 입력하여 간체나 번체로 바꾼다. 이들은 가나, 병음 입력 후 한자로 변환해야 하기 때문에 임시 버퍼에서 편집상태로 있다가 (화면에서는 밑줄표시 글자로 보임) 공간막대글쇠를 눌러서 대상 한자 목록이 나타나면 공간막대 또는 방향글쇠를 이용하여 대상 한자를 지정한 다음 넣기글쇠(enter key)를 치므로 최종 문자부호로 확정되어 완료된다. 한글이나 조선글의 경우에는 한 음절자가 완성이 되면 자동으

로 완료상태로 되어 응용 프로그램에 전달되지만, 항상 마지막 음절은 입력상태로 있게 된다. 이것은 현재 완성음절형 부호계 문제로 발생하는 결과이다.

### 3. 한글의 과학성 규명

한글은 과학적인 문자라고 한다. 하지만 한글이나 조선글의 부호를 모두 완성음절형으로 하고 있다. 이점은 이미 여러 논문[2,3]에서 훈민정음 창제원리에 어긋나는 일이라고 지적하였다. KS X 1001[12]은 2350음절, KPS 9566-1997[11]은 약 2670음절을 포함하고 있다. 국제문자부호계인 ISO 10646 기본다국어판(BMP)(유니코드와 같음)에는 현대 한글(조선글) 음절(1933년 조선어학회의 맞춤법)인 11172 음절글자를 표현하는 완성음절형 부호와 240자 자소형(영문으로 JAMO로 되어 있음) 부호와 자모형 부호 등 세 가지 종류가 등록되어 있다.



<그림 1> 훈민정음 문자생성원리도

훈민정음 해례에는 제자해에서 글자를 만든 원리를 설명하고, 초성해, 중성해, 종성해에 이어 합자해에서 초성자, 중성자, 종성자는 둘 또는 석자 조합할 수 있다고 했다. 그래서 약 399억 음절자를 조합할 수 있는 근거 원리를 제공하고 있다.

하지만 앞에서 본 바와 같이 자소형 및 자모형 부호는 1수준 구현에서는 쓰지 못하게 규정하고 있기 때문에 옛한글 지원은 사실상 불가능하다. 그래서 현재 옛한글 지원은 편법으로 사용하고 있는데 개선되어야 한다. 근본적인 해결 방법은 훈민정음 창제원리를 따르는 것이다. 훈민정음 창제원리는 기본 문자집합만 정하고 나머지 글자는 규칙으로 표현하였다. 이것은 인쇄기술의 발달에 있어서 목판, 활자, 타자, 컴퓨터의 문서편집기로 발전되어 오는 과정에

서 한글이 점점 더 적합성이 극도로 높아진다는 점에서 그렇다.

### 4. 한글 입력의 최적화

한글문자 입력의 최적화는 훈민정음 창제원리[1,2]를 따를 때 이루어 질 수 있다. 한글을 표현할 때 내부적으로는 로마자처럼 1차원으로 나열하고, 모든 소리는 합해야 소리를 이룬다(凡字必合而聲音)라고 하므로 화면표시 때만 음절을 보여주면 된다. 훈민정음은 문자론적으로 보아 표음문자라는 점은 같으나 좀 더 세분화해서 보면 로마자는 음소문자이고 가나는 음절문자이며, 한글은 음소문자이면서 음절문자이다. 이점은 한글의 가장 중요한 특성 가운데 하나이다. 한자는 음절문자에 속한다. 전술하였듯이 일본어는 표기체계가 가나와 한자로 되어 있기 때문에 가나에서 한자로, 또는 로마자에서 가나로 그리고 다시 한자로 변환하는 과정을 반드시 거치게 된다. 중국어는 현재 병음 또는 주음을 기본으로 입력해서 한자로 변환하고 있다.

입력기에서 가나나 병음을 입력한 후 공간막대를 치면 변환 대상 한자목록이 나타나고 다시 공간막대 글쇠 또는 방향글쇠를 가지고 글자를 지정하고 넣기 글쇠를 치면 입력이 마무리(완료) 상태가 된다. 여기서 마무리된다고 함은 최종 한자부호가 결정된다는 뜻이다. 이점은 한글 입력에서는 다르다. 한글은 입력할 때 자동등을 사용하기 때문에 한 음절이 완성되면 사람의 개입없이 자동으로 마무리가 된다는 점이다.

일본어나 중국어 입력 중에 실수로 마우스를 다른 글자가 있는 곳에 짚각하게 되면 현재 편집중에 있는 글자가 입력 마무리가 됨으로써 원하지 않는 글자가 입력되게 된다. 그러면 지우고 다시 긴 기본 문자열을 입력하는 번거로움을 들기 위하여 현재 편집중인 글자들은 마우스 위치가 바뀐 곳에 그 글자들을 나타나게 하는 방법이 등장하였다.

예를 들어 '한글'을 입력할 때 '한'은 입력이 완료되고, '글'은 입력중인 상태에서 다음에 오는 글자에 따라서 또한 음절을 완성하게 된다. 그렇기 때문에 '글'은 우리의 입장에서 보면 마우스가 짚각하게 되면 입력마무리 상태로 바뀌어야 하지만 일본어 입력의 경우에는 한자 변환을 계속하는 편리를 취하기 위하여 '글'을 현재 마우스 위치로 가져가서 나타나게 하는 일이 벌어진다. 이것은 우리의 입력방식에

서는 허용할 수 없는 일이다.

이와 같은 기능은 일본어 또는 중국어 입력기에서는 편리함을 제공하지만 한글의 경우에는 한자로 바꾸는 일이 거의 없기 때문에 그리고 현재 남북의 우리말 표기 방식이 한글 또는 조선글 전용이 주류이기 때문에 다르게 구현되어야 한다.

### 5. 최적화 구현방안

구현방안은 두 가지 방향에서 고려한다. 하나는 마우스 짤깍, 방향글쇠, 공간막대글쇠 등을 누르게 되면 한글입력 자동틀에서 인식하여 음절구성을 마무리 상태로 바꾸어 주는 방식이고, 다른 하나는 한글 문자부호계를 고치는 방식이다.

#### 5.1 이벤트 추가 방식

일본어와 중국어의 편집중 상태에서 공간막대글쇠, 엔터글쇠, 방향글쇠의 입력에 대하여 이벤트 발생 여부를 파악하여 한글입력 자동틀에서 음절 구성을 완료시켜서 현재 편집중인 또는 입력중인 음절의 부호를 최종 결정하여 마무리 처리한다.

#### 5.2 한글부호계 개선 방식

이것은 한글입력을 로마자 입력 수준과 같게 하는 일이다. 이것은 훈민정음 창제원리에 따라서 현재 유니코드에 등록된 세 가지 부호계를 45자로 된 초성자, 중성자, 종성자의 낱자 부호로 대체하는 일이다. 그렇게 되면 여러 가지 좋은 점을 얻게 된다.

첫째, 입력하는 과정에서 초성자, 중성자, 종성자가 각각 입력될 때마다 음절부호로 바꿀 필요 없이 글자꼴만 해당 음절로 바꾸어 주면 된다. 다시 말해서 로마자는 입력되는 글자마다 글자꼴을 갖지만 한글은 초성자만 꼴을 가지고 중성자와 종성자는 꼴이 없이 글자꼴만 계속 바꾸어주므로 로마자입력 수준과 거의 같게 된다. 더군다나 입력 자판을 현재의 자모형에서 안마태 자판[6] 같은 자소형으로 고친다면 입력기의 복잡성은 현재의 삼분의 일 수준으로 줄일 수 있다. 또한 문자처리에서도 거의 절반 수준으로 줄어 든다.

둘째, 옛글처리는 현대 한글과 같은 수준에서 입력 및 처리할 수 있다.

셋째, 자소 정보를 모두 가지게 되어 한글의 음운

현상처리는 물론 훈민정음 창제원리에서 규정한 음절글자인 약 399억 음절자를 글자꼴이 지원되면 모두 표현할 수 있게 된다.

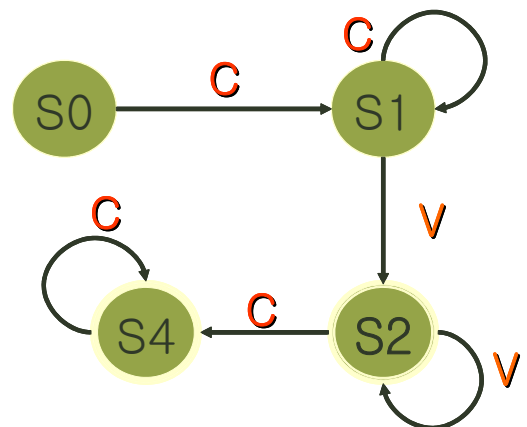
반면에 약점도 있다. 여기서 언급하는 약점은 실제로 약점이 아니라 훈민정음창제원리의 특성에 따라서 필연적인 부분으로 봐야 한다. 훈민정음 창제원리는 다른 연구[2,3]에서도 이미 지적하였듯이 컴퓨터 내부에서 표현은 로마자처럼 일차원으로 풀어쓰기를 하고 화면에 보일 때는 음절자 형태로 보여준다는 것이다.

그리고 컴퓨터 내부에서 완전 풀어쓰기는 공간 낭비를 초래한다는 점이다. 그러나 앞서 언급한 바와 같이 풀어쓰기가 원칙이라고 할 때 사실 공간낭비는 언급할 가치가 없다고 본다. 설사 공간 낭비를 현재 기준으로 따진다고 하더라도 65% 정도 증가하는 것으로 판단한다.

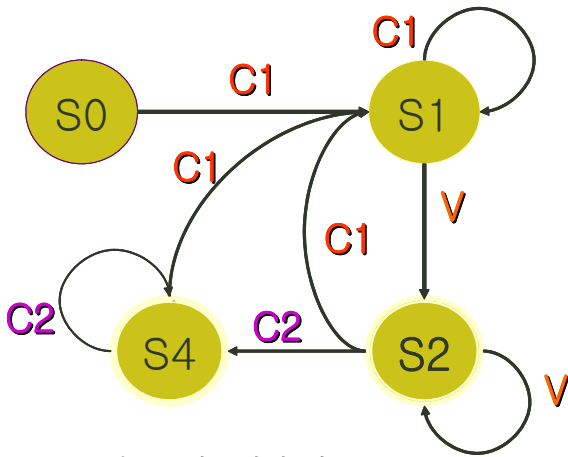
또 다른 방법은 현재 자소의 수가 240인데 이를 45자 범위로 축소해서 U+0500-05ff 판에 히브리 앞쪽 또는 뒤쪽에 배치하면 UTF-8으로 표기하더라도 낱자의 크기를 3바이트에서 2바이트로 줄일 수 있기 때문에 공간효율을 개선할 수 있다.

<표 1> UTF-8

U-00000000 - 0000007F:	0xxxxxxx
U-00000080 - 000007FF:	110xxxxx 10xxxxxx
U-00000800 - 0000FFFF:	1110xxxx 10xxxxxx 10xxxxxx
U-00010000 - 001FFFFF:	11110xxx 10xxxxxx 10xxxxxx 10xxxxxx
U-00200000 - 03FFFFFF:	111110xx 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx
U-04000000 - 7FFFFFFF:	1111110x 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx 10xxxxxx



<그림 2> 한글입력 자동틀-자소



<그림 3> 한글입력 자동틀-ㅂㅈㅇ

<그림 2>는 자모형이고 <그림3>은 자소형이며 둘 다 밀리형 방식을 따른다. <그림3>에 안마태[6]의 자소형 자판을 붙이면 가장 최적화된다. 이 자동틀은 음절인식기이며 한 음절이 완성이 되면 자동으로 완료를 시켜준다. 상태 1에서 초성자, 2에서 중성자, 3에서 종성자의 조합범위를 제한시킬 수 있다. 또한 여기서 허용한다고 하더라도 글자꼴 일치에서 실패하면 없는 것으로 간주하여 제한할 수 있다.

## 6. 결론

다국어 입력기에서 각 문자들의 입력방식 차이에서 나타나는 문제점에 대하여 분석하고 한글문자 입력의 최적화에 대하여 검토해보았다.

다국어 입력기에서 문자간 개별성에 의하여 보편성이 훼손된다면 그로 인하여 입력에 혼란이 생기게 되고, 결국은 생산성 저하로 이어지게 된다. 따라서 두 가지 방식의 해결책을 제시하고, 이로 인하여 빚어지는 장단점을 살펴해보았다.

첫 번째 이벤트 추가 방식은 현재 상태에서 문제를 해결한다는 점에서 무난한 방법으로 볼 수 있다. 하지만 이것은 근본적인 해결책이 될 수 없고 훈민정음 창제 원리에도 크게 어긋나는 부호계 방식이라는 점에서 근본적인 문제가 있다. 두 번째 방식은 훈민정음 창제 원리에 따라서 근본적으로 해결한다는 점에서 가장 바람직한 방법이다. 하지만 현행 부호계계를 바꾸어야 한다는 점에서 이해력을 높이고 관련 전문가의 의견을 수렴하여 효율적으로 해결하도록 결집해야 한다는 점이 지적된다. 결론적으로 말하면 단기적으로는 전자의 방식으로 현재의 문제를 해결하고 장기적으로는 지금부터 두 번째 방식에 대하여 꾸준히 노력하여 완벽한 방식으로 발전시켜서 부호

계를 고치는 일이라고 본다.

추후 연구로는 일본어 및 중국어 입력방식의 전반을 파악하여 두 언어 및 문자의 개별성을 파악하여 이를 보편성을 해치지 않도록 개별성을 표준에 반영하도록 노력할 것이다. 또한 한글의 경우 현재의 자모방식 입력 자판에서 자소방식 입력기로 대체한다면 입력하는 소프트웨어의 난이도를 로마자 수준으로 내리고 언어처리 관련 응용에서 좋은 성과를 낼 수 있을 것으로 생각한다.

## 참고문헌

- [1]강신향, 훈민정음연구,성균관대학교출판부,1991
- [2]변정용, "훈민정음 창제원리의 공학화에 기반한 한글보호계의 발전방향", 한국정보과학회, V12, N2, 1994
- [3]변정용, "국제문자부호계에서 한글부호계의 개선방안", 가을 학술발표논문집, 제25권 제2호, 한국정보과학회, p192-194, 1998
- [4]김명규, "정보교환용 조선글자부호의 표준안에 대하여", '94 코리언 정보처리 학술대회 논문집, 1994
- [5]The Unicode Consortium, The Unicode Standard Worldwide Character Encoding, Ver. 1.0, Vol.1, 1991
- [6]안마태, "컴퓨터에서의 과학적 한글 사용 방법", KAIST공동연구보고서, 2004
- [7]조선글 입력기 '하나 7.0' 입력기 사용 안내서, 2004
- [8]최환진, 한글입력기 '나비' 사용안내서, 2004
- [9]한컴 한글2004 다국어입력기, 2004
- [10]마이크로소프트 워드2002 다국어입력기, 2002
- [11]국가규격위원회, KP S 9566-1997
- [12]기술표준원, KS X 1001-2002