

한글의 열린조합과 이를 위한 오토마타

이균하
인하대학교 전자계산학과

An Open Combination and Automata of Hanguk

Kyoon Ha Lee
Department of Computer Science, Inha University

요 약

본 연구는 컴퓨터에서 한글표기의 제한현상을 설명하고 한글표기의 제한을 없애어 훈민정음 반포당시와 같이 모든 가능한 표기를 위한 방안과 이를 위한 오토마타를 제안하여 컴퓨터 내에서 한글이 문화적 측면에서 보다더 풍부한 표현이 가능하도록하여 충실한 문화의 매개체가 되도록 하는 방안을 제시하였다.

1. 서론

한글이 소리글이며 24개의 자모로 만들어졌음은 너무나도 잘 아는 사실이다. 마찬가지로 영문도 소리글이며 26개의 자모로 만들어졌다. 그러나 영문권을 비롯한 우리가 알고있는 세계 대부분의 소리글이 필기 작성방법에 있어서 수평으로만 나열되어 대단히 간단하다. 한편 한글은 독특하게도 단음절을 단위로 자모들을 모아서 하나의 네모꼴 글 틀에 작성한 다는 것은 누구나가 다 아는 사실이며 세계적으로도 달리 찾아볼 수 없는 유일한 소리글 조합이다.

한글 조합 방식은 가변조합[1]이 두드러진 특성이라고 할 수 있으며 2~7개의 자모가 하나의 네모꼴 틀에 조합되면서 각 자모의 수평 및 수직 크기와

위치가 다른 자모의 종류와 모양에 따라 가변되기 때문에 세계 각 국의 뜻글자 및 소리글자들이 항상 일정한 크기와 형태를 가지고 있는 것과 비교할 때 분명한 특색이라고 할 수 있다. 따라서 한글을 깊이 알지 못하는 외국의 학자들에게는 경이스럽게까지 보여질 수 있다.

한글이 갖는 이러한 특질이 과거 수작업에 의한 필기 작성시에는 불편한 점이 될 수 없었다. 그러나 한글을 전혀 모르는, 그리고 관심도 없었던 영문권에서 영문 위주의 정보처리를 위한 컴퓨터를 개발하고 국내에 보급되면서 컴퓨터에 의한 한글 표기는 방향을 잃기 시작하였다. 가변조합방식의 한글을 가변조합방식이 아닌 영문권에서 개발된 컴퓨터에 맞추기 위하여 한글 표기에 제한이 가해졌으며 결국 완성형이라는 2350자의 표준을 만들어 24개 자모로 구성할 수 있는 수 많은 음들을 2350개의 단음절의 범위로 제한하게 되었다. 이러한 표현 범위의 제한은 한글이 2차원적으로 가변조합된다는 특질을 간과한 상태에서 다른 문자 문화권에서 개발된 컴퓨터의 하드웨어 및 소프트웨어에 한글을 조급하게 적용하기 위한 근시안적 발상에서 시작되었다고도 볼 수 있다. 이 결과로 한글문화는 컴퓨터의 기능을 무시하고 재래의 방식에 의존하거나 그렇지 않고 컴퓨터를 이용할 경우 표현영역의 제한을 받게 되었다.

"동아 세계 대 백과사전"의 "훈민정음"항을 찾아 보면 "훈민정음의 체계는 당시 국어의 전면적 표기를 위하여 마련되었다. 따라서 우리나라 고유요소와 외래요소까지를 고려에 넣은 체계였던 것이다."라는 부분을 볼 수 있으며 "문자는 문화의 소산이며 문화의 매개체"란 점을 주시할 때 우리는 우리의 古語를 포함한 모든 가능한 표기를 제한 없이 컴퓨터에서 다룰 필요가 있다. 뿐만 아니라 국제화가 활발히 진행되고 있는 현실을 볼 때 古語의 표기 뿐 아니라 외래어의 표기도 주어진 자모(24자 또는 훈민정음 반포 당시의 28자)의 범위 안에서 자유로이 작성할 수 있어야 한다. 자유로이 작성할 수 있다 함은 우리가 또는 영문권이 아닌 외국인이 자기의 이름을 영문 알파벳 26개로 마음대로 표기하여 영국 또는 미국인이 읽기에는 낯 설은 표기도 허용 될 수 있는 것과 마찬가지로 이어야 한다. 그러나 현재의 한글은 소리글임을 주장하면서도

한글에는 초보 지식만을 가지고 있는 외국인이 주어진 24개 자모를 이용하여 외국의 고유명사 또는 외래어를 표기할 때 이해하여 주기 힘든 제약조건들에 당혹하게 될 수도 있다. 한 예로 미국의 "Spring creek"을 외국인이 단음절 단위의 한글로 표기할 때

- (1) **스프링 크릭** - 수작업 필기의 경우
- (2) **스프링 크릭** - 조합형 한글
- (3) **스프링 크릭** - 완성형 한글

의 3가지로 표현이 점차로 제약됨을 알게 될 때 한글이 24개 자모로 만들어진 소리글이라는 주장이 공허해 질 수 있으며 오히려 2350개의 단음절 글자로 만들어진 복잡한 소리글로 이해될 수 있다.

따라서 한글의 체계가 우리나라 고유요소와 외래요소까지를 고려에 넣은 체계였던만큼 컴퓨터에 의한 한글의 표시기능은 현대 국어와 고어는 물론 현재 일반 대중이 쓰지않는 음절인 외래어 까지도 훌륭한 소리글의 특징을 살려 자유롭게 표기할 수 있도록 컴퓨터 하드웨어 및 소프트웨어를 우리 손으로 개발할 필요가 있다. 이에 본 연구에서는 이를 위한 오토마타를 제안하였다.

2. 한글의 열린조합

한글을 창제. 반포하기 이전에는 집현전의 많은 학자들이 사람의 발음기관을 여러모로 연구분석하여 발음할 수 있는 모든음을 표기하기 위한 자모를 창제하였고 또한 창제된 자모의 타당성 및 실용성을 입증하기 위하여 "용비에 천가", "석보상절"등과 같은 다수의 작품들을 통한 응용실험 후에 28개의 자모가 반포된 것으로 우리는 알고 있다.

이와 같이 과학적으로 창제.응용실험 후에 반포할 당시에는 28개 자모를 사용하여 소리 나는대로 적으라고 하였으니 얼마나 자유롭고 개방적이었는가 다시 한번 감탄 할 뿐이다. 이에 따라 "ㅅㅅ, ㅅㅅ, ㅎㅎ, ㅅ" 등과 같이 조선조 당시의 발음에 가까운 여러가지 표기들이 고어문에 흔히 나타나고 있다는 것은 잘 아는 바이며 표기할 수 있는 영역이 혼란스러울 정도로 풍부하였다는

질도 충분히 이해할 수 있다.

그러나 1933년에 조선어학회에서 제정한 한글 맞춤법 통일안이 발표되면서 초성 자음으로는 홑자음 이외에 "ㄱ ㄲ ㅋ ㆁ"의 쌍자음만으로 표기의 제한을 하였고 "ㅇ, △, ㅅ, ㅈ"도 문장에서 사라지고 비교적 사용빈도가 높은 자모만이 남게 되었다. 이에 따라 문자 위주로 한글을 익히고 발음을 할 경우 발음의 영역이 줄어들게 되었으며 소리나는대로 표기할 수 없는 주어진 조건에서의 표기만이 가능한 닫힌 조합의 길로 접어들기 시작하였다.

필자의 개인적인 판단으로 볼 때 1900년 초기부터 활자에 의한 인쇄기술이 국내에 널리 보급되기 시작하였고 이 때 제정한 한글 맞춤법 통일안이 인쇄에 필요한 활자의 수요를 상당히 줄일 수 있어 경제성을 높였다고 생각된다.

3. 닫힌조합으로서의 한글

소리글인 한글을 "소리 나는대로 써라" 라는 자유로운 표기의 틀에서 1933년에 조선어학회에서 제정한 한글 맞춤법 통일안으로 표기의 제한을 받은 한글이 한글과는 전혀 관계가 없고 특성도 전혀 다른 영문권에서 개발된 컴퓨터로 인하여 더욱 더 닫혀지게 되었다. 한글이 소리글로서 단음절을 단위로 자모들을 모아서 하나의 네모꼴 글 틀에 작성한 다는 것은 누구나가 다 아는 사실이며 세계적으로도 달리 찾아볼 수 없는 유일한 소리글 조합이다. 어느날 컴퓨터가 발명되고 이것의 기억용량 및 처리속도가 점차 증가되어 수치정보의 계산, 처리 뿐 아니라 문자정보의 처리에 까지 응용되면서 또 다른 문제가 제기되었다.

영문권에서 개발된 컴퓨터에서는 모든 문자가 2진 부호로 기억되며 기억용량 한 바이트당 영문 한자를 기억하고 영문자의 크기는 특별한 경우를 빼고는 일정한 크기로 통일하여 쓰고 있었다. 이것을 출력장치에 출력해 낼 경우 기억장치의 점유길리와 출력된 글자열의 길이가 비례하였다. 이를 바탕으로 탭 조절 기능 및 위치 제어기능 등의 출력조절 기능들이 소프트웨어화 하였으니 2~7개의 자모로 구성된 하나의 단음절이 하나의 네모 글 틀안에 작성

되는 한글에서 기억장치의 점유길이와 출력된 글자열의 길이사이의 비례가 깨어지게 되었고 따라서 한글의 구조를 염두에 두지 않은 소프트웨어들을 이용하기 힘들게 되었다. 이에 따라 원고지에서 영문 2 글자를 하나의 한글 글자에 끼워넣는 과거의 경험을 살려 영문 2자를 표기할 수 있는 2 바이트에 한글 단음절 하나를 표기하는 방안이 이르게 되었으며 이것이 2 바이트 조합형 부호의 기본이 되었다.

조합형 한글부호에서는 2 바이트를 구성하는 16 비트중 1 비트는 한글 표시자로 사용하고 15 비트를 초성자음, 중성모음 및 종성자음으로 구분하니 5 비트씩 할당이 되고 각 각은 32개의 기본자모를 표기할 수 있게 되었다. 이러한 컴퓨터 환경의 요구에 따라 초성은 19, 중성 21 및 종성 28(무음 포함)의 홑자모와 복합자모로 제한이 되었으며 조합 가능한 한글의 수는 11,172자로 제한이 되었다. 이 방식은 현대 국어의 범위를 완전히 포함하였다고 볼 수도 있으나 역시 고어와 외래어의 자유로운 표기에는 제한이 있었다. 한 예로서 ["왕"이라는 글자를 "우왕"이라고 발음하는 학생이 10명이나 되었다]라고 컴퓨터로 표기하는데 문제가 될 수 있다.

3. 더욱 단힌 완성형 한글

컴퓨터가 처음 개발된 영어권과 그 이후 첨단에 들어선 일본의 한자어권에서 컴퓨터의 부호는 곧 문자 도형을 의미하며 컴퓨터 내의 부호 열은 대응하는 문자도형의 1차원 열로 바꾸면 곧 출력이 될 수 있는 1차원적 출력특성을 가지고 있다. 단지 영어권의 문자들은 1 바이트/문자로 표기되는 반면 한자어권에서는 방대한 수효의 문자를 표기하기 위하여 2 바이트/문자로 표기하는 것이 다르다.

조합형 한글도 2 바이트/문자로 표기 되기 때문에 사용상 큰 지장은 없었으나 컴퓨터 통신시에 필요한 제어부호와 중복이 되어 컴퓨터 통신망을 이용한 정보의 전송에 문제를 느끼게 되었으며 한자를 수용할 수 없었다. 이러한 이유로 한글 단음절 중 현대 국어에서 쓰이는 2350자만을 선정하여 조합형이 아닌 완성형이란 이름으로 표준안을 마련하였다[2]. 이 표준안은 컴퓨터의 부

호체계만으로 볼 때 24개의 기본자모를 갖고 가변조합되어지는 소리글이라는 특징을 무시하고 2350가지의 음을 갖는 2350개의 글자도형으로 구성된 완성형 한글로 변모되어 훈민정음 반포 당시의 정신을 무시하였을 뿐 아니라 한글을 훌륭한 문화의 매개체에서 문명만의 매개체로 전락을 시켰다.

컴퓨터의 보급이 계속 확대되어 활용 계층이 훨씬 넓어지고 이의 이용이 일상 생활화 하는 한편 국제적인 교류가 더욱 활발해 질것에 대비할 때 그리고 "문자는 문화의 소산이며 문화의 매개체"란 점을 생각할 때古語의 자유로운 표기는 물론 외래어의 제한없는 표기가 가능할 수 있는 열린조합으로의 한글부호와 입·출력 시스템을 개발하고 계속 발전시켜 나아가야 한다고 생각되어 본 연구에서는 이를 위한 오토마타를 제안하고자 한다.

4. 열린조합을 위한 오토마타

우리나라의 고유명사를 영어로 표기할 때 26개의 알파벳들 중 마음대로 선택 조합하여 표기하여도 영어권 컴퓨터 처리에 문제가 없는것 처럼 한글이 24개의 기본자모로 구성된 소리글이라고 하면 당연히 24개의 기본자모를 이용하여 외래어를 마음대로 표기할 수 있어야 마땅하다. 물론 이러한 자유로운 표기가 문화적으로 또는 국어학적으로 문제가 있던 없던간에 우리가 종이 위에 마음대로 작성할 수 있는것처럼 컴퓨터에서도 기억이 되고 통신망을 통하여 전송이 될 수 있으며 또한 출력될 수 있도록 컴퓨터 종사자들은 노력하여야 할 것이다.

본 연구에서는 현재의 컴퓨터가 24자의 현행 한글 자모 뿐 아니라 훈민정음 반포 당시 28자의 기본 자모까지 충실히 가변조합해 낼 수 있다고 전제하였다. 이러한 전제는 현재의 기술로 충분히 해결할 수 있는 것이며 또한 한글의 문화적 발전을 위해서도 반드시 실용화 해야할 것으로 믿는다. 반포 당시 28자의 기본 자모를 제한없이 가변조합할 경우 古語의 표기는 물론 외래어까지 보다 더 실제음에 가깝도록 표기할 수 있을 것이다.

이 때에 음절과 음절의 경계를 구분하기 위한 조치가 필요하며 1933년 이후의 현대 국어에서는 초성으로 14개의 홑자음 이외에는 "ㄱ ㅋ ㆁ ㅅ ㅆ" 만의

쌍자음을 사용하므로 이를 하나의 자모로 임의 취급하여 모음 바로 앞에 쓰여지는 자음을 한 음절의 경계자로 사용하였다. 그러나 古語에서는 초성으로 서로 같은 쌍자음 뿐 아니라 복자음도 사용을 하고 연속된 자음의 수효도 2자의 경우는 물론 3자인 경우도 있다. 이러한 古語의 입력을 위해서 어떤이는 3벌식 한글 입력이 적합하다고 주장하기도 한다.

3벌식 한글 입력을 이용할 경우 한글 입력을 위한 오토마타는 극히 간단하게 되어 본 연구에서 거론할 필요 조차 없을 것으로 판단되지만 2벌식 입력에서는 면밀한 검토가 필요하다. "갈배대"를 입력할 경우 "ㄱ ㅏ ㄹ ㅂ ㅅ ㄷ ㅏ ㅓ"로 되며 "갈ㅓ대"와 "갈ㅓ 대"로도 표기 되어질 수도 있다. 그렇다고 음절마다 블랭크를 두는것도 입력 효율이나 기억장치 이용율의 측면에서 볼 때 바람직스럽지 않으며 한글 단음절의 평균자모수가 2.5정도인 점을 고려할 때 현재보다 40% 정도의 부담이 더 추가된다. 따라서 본 연구에서는 17개의 기본자음(현행 14개의 기본자음 + ㅏ, ㅓ, ㅓ)을 일상자음과 초성복자음을 의미하는 자음으로 2중 표기가 가능하도록 초성복자음부호 방안을 제안하여 2벌식 한글 오토마타를 설계하였다.

이 제안에 의하면 기본 자모가 일상자모 28개에 초성복자음 17개가 더하여져 45개의 자모로 구성되어 영어의 대.소문자가 모두 52개인것과 비교할 때 많은 수효가 아님을 알 수 있다. 또 한 영어에서 대.소문자는 컴퓨터 내부에서의 부호는 물론 이를 의미하는 글자도형이 서로 다르지만 본 연구에서 제안하는 일상.복자음은 컴퓨터 내부에서의 부호가 다름은 물론 글자도형상에서는 선행자음과 함께 초성임을 의미하게 된다. 그림 1의 오토마타에서 보이는 바와 같이 초성 첫 자음은 일상자음으로 표기하고 두번째 자음은 모음과 같이 취급하여 선행하는 자음이 새로운 음절의 시작임을 알 수 있도록 하였다. 일반 한글에서 초성이 쌍자음일 확율이 2~3%인 점을 감안할 때 한글 자판에서 일상자음과 초성복자음의 식별은 동일한 字키를 사용하되 초성복자음의 경우 shift를 동시에 누르도록 하면 현재의 2벌식 입력과 큰 차이 없이 고어와 자유스러운 표기의 외래어 까지 입력할 수 있게 된다.

그림 1에 의한 오토마타에서

"가 | 러 | 비 | 스 | 디 | |"
 ↑
 shift

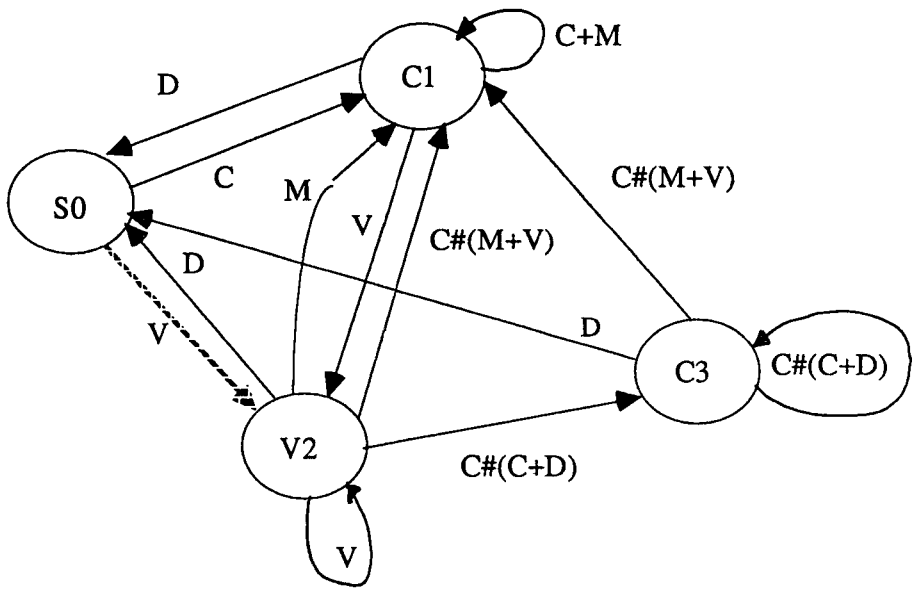
과 "가 | 러 | 비 | 스 | 디 | |"
 ↑
 shift

은 "가 | 러 | 비 | 스 | 디 | |"
 ↑
 shift

은 "가 | 러 | 비 | 스 | 디 | |"
 ↑
 shift

은 "가 | 러 | 비 | 스 | 디 | |"
 ↑
 shift

은 "가 | 러 | 비 | 스 | 디 | |"
 ↑
 shift



- C : 일상자음
- M : 초성복자음
- V : 모음
- D : 기타 문장부호
- #?: Look ahead with?

그림 1 열린조합을 위한 한글 오토마타

5. 결론

한글은 훈민정음 반포 당시에 국어의 전면적 표기를 위하여 마련되었으며 따라서 우리나라 고유요소와 외래요소까지를 고려에 넣은 체계였던 것으로 모든음을 28개의 자모로 소리 나는대로 표기하도록 하였고 이 당시에는 모든 글자의 작성이 필기에 의존하므로 특별히 제한을 가할 이유도 없었다고 판단된다. 그러나 1933년의 한글 맞춤법 통일안은 한글표기의 범위를 제한하였다고 볼 수 있으며 그 결과로 그 때 당시에 한참 보급되던 한글 인쇄사업에 활자의 수요를 줄이는데 기여하였다고도 추측할 수 있다. 컴퓨터가 등장하면서 한글은 닫힌 조합으로 표기되어 표기 가능한 문자의 범위를 11,172자로 제한하게 되었으며 다시 원활한 정보 교환을 이유로 완성형 한글을 2350자로 제한하게 되었다. 따라서 한글은 24개의 자모로 구성된 소리글이라는 점에 강한 의문을 던질 정도가 되었고 외국인이 고유명사 등을 한글 표기로 하고자 할 때에는 24개의 자모가 아닌 2350개의 완성형 글자 중에서 선택하는 식으로 후퇴하고 컴퓨터로 고어나 방언을 표기하는데는 다른 방안을 찾아야만 하는 형편이 되었다. 이러한 점들은 한글을 문명적인 면에서만 다루어 왔고 문화적인 측면에는 손상을 주었다고 판단된다.

이제는 한글 맞춤법 통일안과는 별도로 컴퓨터에서만은 모든 한글 자모의 조합을 열린 조합으로 발전시켜 외래요소까지를 수용할 수 있도록 할 필요가 있으며 본 연구에서는 이를 위한 오토마타를 제안하여 모든 자모들을 컴퓨터에서 자유롭게 입력하고 출력할 수 있는 기반을 마련 하였다.

이 결과로 45개의 한글 자모 부호로 통신과 모든 조합이 가능할 수 있으며 이와 관련된 한글 입·출력 시스템들이 등장하여 한글이 문명적인 측면 뿐 아니라 문화적인 측면에서도 계속 발전하기를 바라는 바이다.

참고문헌

[1] 이주근, 이균하 "가변조합방식의 문자 Display에 관한 연구," 전자공학회지 11권 1호, pp. 27~33, 1975.

[2] 박동순 외, 한글·한자 코드 표준화에 관한 연구, 과학기술처, 1987.