

두벌식 한글 입력 방식의 정형적인 기술과 분석

김용목⁰, 김 국

연세대학교, 서경대학교

sebulsik@gmail.com, kimkuk99@daum.net

A Formal Description and Analysis of 2-beolsik Hangul Input Methods

Yongmook Kim⁰, Kuk Kim

Yonsei University, Seokyeong University

요 약

한글 입력 방식은 글쇠배열과 더불어 추가적인 낱자들을 입력하기 위한 결합 규칙이라는 형태로 정의할 수 있다. 그런데 이 규칙을 토대로 입력 방식을 실제로 구현해 보면 겹낱자를 결합하거나 음절이 바뀌는 과정에서 모호성 같은 문제가 발생할 수 있다. 초성과 종성을 문맥에 따라 구분해야 하는 두벌식, 모바일 환경처럼 매우 적은 글쇠, 수십 종류의 낱자들을 조합해야 하는 옛한글이라는 조건이 더해지면 입력 방식을 기술하고 분석하는 난이도가 더욱 높아진다.

본 논문에서는 한글 낱자의 결합 규칙을 대결합과 소결합으로 구분해서 기술하는 체계를 제안하며, 이를 토대로 입력 방식의 예상 동작을 분석해 주는 프로그램을 소개하였다. 그리고 모바일용 삼성 천지인과 KT 나랏글 한글 입력 방식을 동일 프로그램으로 기술하고 분석한 결과를 제시하였다.

주제어: 한글, 옛한글, 훈민정음, 두벌식, 세벌식, 2벌식, 3벌식, 입력기, 오토마타

1. 서론

한글은 초·중·종성에 해당하는 세 낱자가 모여서 한글자를 형성하며, 입력 역시 낱자 단위로 순차적으로 행해진다. 입력된 낱자들은 지정된 규칙에 따라 더 복잡한 낱자로 바뀌기도 하면서 글자 형태로 조합된다. 그런데 한글 낱자 중 마지막에 등장하는 종성은 별도의 새로운 글자가 아니라 초성과 동일한 자음 글자들을 사용하여 표현되며, 때에 따라서는 생략도 가능하다.

이런 이유로 인해 한글 입력 방식에서도 초성과 종성에는 같은 글자에 대해 같은 글쇠를 부여하고, 이 자음이 초성과 종성 중 어디에 속하는지는 입력 상태에 따라 자동으로 결정되게 하는 일명 ‘두벌식’ 체계가 오늘날의 컴퓨터 기반 한글 입력 환경에 정착해 있다[1].

한글 입력 방식을 설계하는 연구자는 자음과 모음 낱자 중에 자주 등장하는 것들을 타자 능력을 최대화할 수 있는 형태로 배당한다[2][3][4][5]. 그리고 글쇠배열에 없는 낱자들은 글쇠배열에 있는 낱자들을 둘 이상 결합해서 입력하게 규칙을 정한다[6].

그런데 그런 입력 규칙을 실제로 구현해 보면, 각각의 낱자를 입력하는 것은 문제가 없으나 그 상태에서 겹낱자를 결합하거나 다른 낱자를 이어서 입력할 때 모호성이 발생하기도 한다. 두벌식은 글쇠에 초성과 종성의 구분이 없기 때문에 입력 방식을 설계할 때 모호성 문제를 더욱 신중하게 생각해야 한다.

그러므로 어떤 한글 입력 방식이 주어졌을 때 글쇠배열 같은 타자 행동 측면뿐만 아니라 그 저변에서 암시적으로 행해지는 낱자의 결합, 이동, 음절 분리 같은 동작의 발생 조건을 기술하는 체계를 만들고 예상되는 문제점을 자동으로 분석하는 도구가 있다면 한글 입력 방식의 연구에 도움이 될 것이다. 이 도구는 PC나 모바일,

현대 한글이나 옛한글 등 어떤 용도의 한글 입력 방식을 고안하더라도 동일하게 활용 가능하다.

2. 기존 두벌식 한글 입력 프로그램들의 사례

한글의 초성과 종성은 동일한 자음 글자를 사용하지만 결합 규칙은 서로 다르다. 예를 들어 현대 한글 기준으로 초성에는 ㄱ이 존재하지 않으며 종성에는 ㅁ이 존재하지 않는다. 그렇기 때문에 한글, Mac OS 등 대부분의 PC용 한글 입력 프로그램들은 두벌식이라 해도 초성과 종성의 문맥이 별도로 분리되어 내부적으로 세벌식의 형태를 갖췄다. 그리고 거기에서 두벌식을 위한 처리가 덧붙는 형태로 두벌식 오토마타를 구현했다.

쌍자음은 초성과 종성 모두 Shift+1타로 입력할 수 있고 Shift 없이 동일 글쇠 연타로도 입력할 수 있다. 한글은 받침 쌍자음도 연타 입력이 가능한 두벌식을 과거에는 ‘한글 2벌식 #2’라고 별도로 제공했으나, 나중에는 둘 다 연타가 가능하게 바꿨다¹⁾. 한글과 달리 Mac OS는 두벌식과 세벌식 모두 초성 쌍자음만 연타가 가능하며 종성의 그것은 그렇지 않다.

유일하게 다소 이례적으로 구현된 프로그램은 Microsoft Windows의 한글 IME이다. 이 프로그램의 두벌식은 두벌식답게 문맥도 정확하게 자음과 모음 둘만 존재하며, 여기서 자음은 ‘ㄷ, ㅁ, ㅂ’을 제외하면 언제나 종성을 의미한다. 그렇기 때문에 종성 없이도 ㅁ 같은 겹받침들을 ‘ㅂ+ㅁ’과 같은 식으로 단독 입력 가능할 뿐만 아니라, 그 상태에서 종성을 입력하면 ‘ㅂㅁ’ 같은 글자가 만들어진다.

1) 그래도 ‘아까’를 입력할 때처럼 초성 쌍자음을 연속 입력하는 것은 물론 언제나 Shift+1타로만 가능하다.

현대 한글이 아닌 옛한글로 가면 상황이 달라진다. 초성도 중성과는 별개의 고유한 대결합이 필요해지며, 모음 역시 홀모음 단위에 대해 소결합과 대결합을 엄격하게 구분해야 결합 규칙의 전체 복잡도를 원활히 제어할 수 있다. 현대 한글에 정의된 초·중·중성의 개수는 각각 19, 21, 27이지만, 유니코드에서 KS X 1026-1 [8]이 규정하는 옛한글 초·중·중성의 개수는 124, 94, 137에 달하기 때문이다. 더구나 세 성분이 모두 삼중 겹 낱자까지도 존재할 수 있기 때문에³⁾ 대결합의 최대 길이는 3:3:3이 된다. 이는 현대 한글의 대결합 최대 길이인 1:1:2를 크게 증가하는 규모이다.

이렇듯, 대결합과 소결합은 담당하는 역할이 서로 다르다. 대결합의 결과값은 소결합의 출발지나 목적지에 존재하지 않아야 한다. 예를 들어 ‘ㅅ+ㅅ → ㅅㅅ’라는 대결합이 존재하는데 ㅅㅅ을 다른 낱자로 바꾸거나 ㅅ 자체를 만드는 규칙이 소결합에 또 있어서는 안 된다는 뜻이다. 이 경우 소결합은 대결합의 부품인 ㅅ과 ㅅ을 생성하거나 변형하는 규칙만 명시해야 한다.

4. 소결합과 대결합의 합성

어떤 한글 입력 방식을 기술하는 글쇠배열과 소결합 및 대결합 규칙이 주어졌다면, 이를 토대로 기대 가능한 한글 입력 동작은 다음과 같다.

4.1. 낱자를 입력하는 모든 중간 과정을 계산

먼저, 대결합에 명시된 모든 중간 낱자들에 대해 소결합(1타만으로 바로 입력 가능하지 않은 경우)을 적용하여 각 단계별 세부 결합 규칙들을 생성한다. 여기서 중요한 것은 생략된 중간 과정도 모두 고려해야 한다는 점이다. 중간 과정은 한 낱자로 표현하는 대결합이 존재하지 않기 때문에 이때는 부득이하게 입력된 낱자들을 한 데 풀어서 표현하게 된다.

KT나랏글 입력 방식을 예로 들면, ㅅ을 입력하기 위해 중간에 거치는 ‘ㄴ+ㅅ’, ㄹ을 입력하기 위해 중간에 거치는 ‘ㄴ+ㅇ’ 등도 모두 임시 낱자로 등록한다. 낱자들을 풀어서 표현하는 것 역시 기존 글자의 위치와 자연스럽게 연계하며 동작해야 한다. 중성이라면 맨 앞의 낱자가 원래 글자의 중성에 들어가고 다음 낱자들이 뒤를 따르지만(예: ‘안ㅅ’), 초성에서는 반대로 앞의 낱자들이 원래 글자의 앞에 먼저 자리 잡고 마지막 낱자가 원래 글자의 초성 자리에 들어가야 한다(예: ‘ㄴ산’).

소결합의 중간 과정뿐만 아니라 대결합 자체의 선행 단계가 통째로 생략될 수도 있다. 이항인 ‘ㄴ+ㅅ → ㄴㅅ’이 없이 삼항 ‘ㄴ+ㅅ+ㅅ → ㄴㅅㅅ’ 결합만 등록될 수 있다는 뜻이다. 이 경우 ‘ㄴ+ㅅ’까지만 진행되었을 때는 이것을 ㄴㅅ으로 합쳐서 표현할 근거가 없으므로 부득이하게 ‘ㄴㅅ’이 풀려져서 ‘갈ㅅ’(중성) 또는 ‘ㄴㅅ’(초성)과 같은 형태로 표현된다. 하지만 이 상태로 현 글자의 조합은 계속 유지되며, ㅅ이 추가로 입력되는 순간 낱자는 ㄴㅅ 하나로 다시 뭉쳐져야 한다. KT나랏글은 ㅅ이 1타로 입

력 가능하지 않고 ‘ㄴ+가획’을 거치므로, ‘ㄴㅅ’ 이전의 ‘ㄴㅅ’까지도 임시 낱자로 처리되어야 함은 물론이다.

4.2. 대결합 단위로 도깨비불 현상을 처리

두벌식 입력 방식에서 중성을 입력한 다음에 중성을 입력하면, 중성에서 마지막으로 입력된 부분이 다음 글자의 초성으로 이동하고 중성은 그 글자의 중성으로 붙는다. 이것은 일명 ‘도깨비불 현상’이라고 불린다.

PC용 입력 방식에서는 단순히 마지막으로 입력된 중성 1타만을 다음 글자로 옮기면 된다. 그러나 모바일까지 생각한다면 도깨비불 현상의 단위는 ‘대결합의 마지막 단계를 구성하는 소결합’ 전체라고 간주해야 한다.

모바일용 한글 입력 방식에서는 ㅅ이나 ㅅ처럼 된소리나 거센소리 자음은 일반적으로 1타 만에 입력 가능하지 않다. 그러나 그런 자음들도 어떤 과정으로 입력되었든 무관하게 중성 ㅅ은 모음이 왔을 때 ‘ㄴ+ㅅ’으로 분리되어야 하며, 받침 ㅅ은 초성 ㅅ으로 한꺼번에 넘어가는 것이 보장되어야 한다⁴⁾. 이럴 때 대결합과 소결합을 분리해서 결합 규칙을 기술하면 도깨비불 현상도 무조건 마지막 1타가 아니라 마지막으로 입력된 흘낱자 전체가 분리되게 할 수 있다.

4.3. 낱자를 완성한 다음에 필요한 임시 낱자를 계산

‘대결합 단위 도깨비불 현상’이라는 단서로 인해, 두벌식 입력 방식은 고려해야 할 임시 낱자가 중성 문맥에 한해서 더 늘어난다. 지금 한 중성이 완성되었더라도 다른 낱자가 소결합 차원에서 추가로 결합 가능하다면 이것들은 모두 현재 글자의 조합으로 임시로 잡아 뒀야 한다. 그래야 나중에 도깨비불 현상이 발생할 때 대결합의 마지막 단위가 모두 뒷글자의 초성으로 넘어갈 수 있다.

그 예로는 KT나랏글을 기준으로 ‘알자’라는 어절을 입력하는 상황을 생각할 수 있다. ㄴ 다음에 ㅅ을 입력하면 ‘알’이 된다. 현대 한글 중성에는 ㄴ 이후로 ㄴㅅ이나 ㄴㅅ, ㄴㅅ 같은 겹받침이 존재하지 않는다.

그럼에도 불구하고 한글 입력 프로그램은 ㅅ으로부터 소결합 차원에서 파생될 수 있는 ㅅ을 염두에 두고, ‘ㄴ+ㅅ’이라는 대결합도 앞서 언급했던 ‘ㄴ+ㅇ’이나 ‘ㄴ+ㅅ’과 마찬가지로 임시로 인정해야 한다. 이렇게 하지 않으면 ‘알’ 이후로는 조합을 강제로 끊지 않으면 입력이 더 진행되지 않으며, ‘알자’는 연속 입력이 불가능해진다.

[표 3]은 삼성 천지인과 KT나랏글이 필요로 하는 임시 낱자들을 프로그램의 분석을 통해 모두 구해서 나열한 결과이다. 겹받침들을 입력하는 중간 과정에서 반드시 필요한 임시 낱자보다는 그 뒤에 도깨비불 현상과 연속 입력의 원활한 처리를 위해 추가된 임시 낱자가 훨씬 더 많음을 알 수 있다.

또한 추상적인 글쇠를 쓰는 KT나랏글보다는 삼성 천지인이 임시 낱자의 사용량이 더 적다는 것도 알 수 있

3) 중성 중에 ㄴㅅ과 같은 삼중모음이 있다.

4) KT나랏글의 경우 자음의 입력은 가획이나 쌍자음이라는 추상적인 글쇠로 끝나는 경우가 많은데, 그 글쇠 하나만을 다음 글자 자음으로 분리하는 것은 이치에 더욱 맞지 않는다.

[표 5]는 삼성 천지인의 모음 입력 체계를 확장해서 옛한글 낱자까지 입력하는 것을 가정했을 때 존재하는 모든 모호성을 프로그램의 분석으로 찾은 것이다. 일례

로 ‘-+ . + . + |’는 현대 한글에서는 ‘ㅏ+ ㅑ’로만 해석되기 때문에 거이지만, 옛한글에서는 ‘ㅓ+ ㅣ’로 해석되어 깨가 될 수도 있다. 대결합으로 ‘ㅏ+ ㅑ’와 ‘ㅓ+ ㅣ’를 명시한 것까지는 문제가 없는데 여기에다 각각의 홀모음을 만드는 소결합을 합성하면 모호성이 생긴다.

모호성이 발견되었을 때는 프로그램은 대결합과 소결합 중 옵션으로 지정된 해석 방식을 먼저 제시하고, 사용자가 별도의 보정 글쇠를 누르면 다른 해석 방식으로 교환하게 하는 것이 바람직하다고 여겨진다.

예를 들어, ㅓ가 되었을 때 보정 글쇠를 누르면 해석 방식이 달라지면서 ‘ㅏ·’로 모음이 갈라지고, 여기서 ㅑ를 누르면 ㅑ가 된다. 그리고 ㅑ가 된 상태에서 보정 글쇠를 누르면 ㅑ가 되게 한다는 것이다. 사용자가 어느 옵션을 선택했느냐에 따라 처음에 ㅓ 대신 ‘ㅏ·’가 먼저 나올 수도 있다. 또한 어느 해석 방식을 선택하느냐에 따라 어느 한쪽은 현대 한글이 나오고 다른 한쪽은 옛한글이 나오는 경우, 그때는 대소 결합 방식과 무관하게 사용 빈도가 더 높은 현대 한글을 먼저 제시하면 번거롭게 수동 보정을 하는 경우를 줄일 수 있다.

6.2. 음절 경계 모호성 (연속 입력 가능성)

한 글자 내부의 낱자 결합 과정에서 발생하는 모호성은 초·중·종성 어느 낱자에나 별식과도 무관하게 존재한다. 그러나 음절 경계 모호성은 두별식 체계에서 종성과 초성 사이에 한해 존재한다.

가장 먼저 초성에 이항 이상의 대결합이 존재한다면 그 결합 결과에 해당하는 자음은 소결합 구조가 어떠한 원천적으로 연속 입력이 불가능하다. 종성 문맥에서 입력된 자음은 단순한 결합 불가로 인한 조합 종료, 도깨비불 현상, 초성 전용 낱자 등 그 어떤 경우에도 마지막 1타 또는 마지막 소결합 묶음에 해당하는 자음만이 뒷글자의 초성으로 넘어가기 때문이다.

이 문제를 피하기 위해 PC용 두별식 한글 입력 방식들은 쌍자음을 Shift를 동원해서 1타만으로 입력하게 해서 초성에서의 대결합 가능성을 제거했음을 앞서 확인한 바 있다. 하지만 옛한글 입력을 고려한다면 초성의 구조가 그렇게 간단할 수 없어진다. 그렇기 때문에 두별식 옛한글 입력 방식은 글쇠 수가 많은 PC 환경이라 해도 조합을 강제 종료해서 음절을 구분하는 글쇠가 별도로 배당되어 있다. ‘값가’와 ‘값까’(일반적인 조합 종료), ‘값시’와 ‘가씨’(도깨비불 현상)를 서로 구분할 수 없기 때문이다.

초성의 대결합 외에 다른 모호성으로는 종성이 소결합에 의해 2타 이상의 타수로 입력되고 그 소결합이 내부에서 두 개의 자음으로 또 쪼개지는 것이 가능한 경우이다. 이것은 삼성 천지인에서 ‘구카’와 ‘국가’, ‘장마’와 ‘자아’를 구분할 수 없는 것이 대표적인 예이며, 연속 입력이 불가능한 자음의 조합이 이 외에도 더 있다.

[표 6]은 연속 입력이 불가능한 모든 경우를 역시 프로그램의 도움을 받아 정리한 것이다. 삼성 천지인에는 현대 한글을 기준으로 총 76가지 경우가 존재한다. 이는 초성과 종성의 전체 조합 가짓수인 $19 \times 28 = 532$ 와 비교했을 때 정확하게 1/7에 해당한다.

[표 6] 삼성 천지인에서 연속 입력을 할 수 없는 종성과 초성 조합

종성	초성	계
ㄱ, ㄲ, ㅋ / ㄹ	ㄱ, ㄲ, ㅋ	12
ㄴ, ㄸ	ㄴ, ㄸ	4
ㄷ, ㄲ, ㅌ / ㄹ	ㄷ, ㄲ, ㅌ	9
ㄹ, ㄹ / ㄹ	ㄹ, ㄹ	6
ㅁ, ㅂ / ㄹ, ㄹ	ㅁ, ㅂ, ㅌ	12
ㅅ, ㅆ, ㅇ / ㄲ, ㄴ, ㄷ, ㄹ, ㅁ	ㅅ, ㅆ, ㅇ	24
ㅈ, ㅊ, ㅊ / ㄴ	ㅈ, ㅊ, ㅊ	9

그에 반해 KT나랏글은 음절 경계 모호성이 전혀 존재하지 않는다. 이는 단독으로 낱자를 시작할 수 없는 ‘가획’과 ‘쌍자음’이라는 글쇠만으로 소결합을 수행함으로써 중간 쪼개짐의 가능성을 제거했기 때문이다.

음절 경계 모호성은 음절 구분 글쇠를 별도로 두거나 종성의 입력이 시작된 뒤부터 조합 종료 타이머를 설정해서 해소하면 된다. 조합 종료란 음절 구분과 동일한 개념이며, 이 동작을 타이머를 두고 자동으로 수행하느냐 사용자가 글쇠를 눌러서 수동으로 수행하느냐의 차이만이 존재한다. 한글 입력 프로그램은 이 입력 방식에 존재하는 모호성의 성격과 실제 예를 사용자에게 보여준 후, 이를 해소하는 방법을 선택받으면 될 것이다.

7. 결론

새로운 한글 입력 방식을 정의한다는 것은 자음과 모음 낱자들의 글쇠배열과 기본적인 결합 순서를 지정하는 것과 같다. 그러나 입력 프로그램의 관점에서는 그 방식대로 모든 겹낱자들의 결합을 실제로 지원하고 음절도 구분하기 위해 내부적으로 고려해야 할 것이 많다. 본고에서는 그런 것을 수행하는 보편적인 체계를 정립했으며, 이를 자동으로 수행하는 프로그램을 통해 삼성 천지인과 KT나랏글이라는 잘 알려진 모바일 입력 방식을 분석하였다. 각 입력 방식들이 내부적으로 임시 조합들을 얼마나 사용하는지 확인했으며, 삼성 천지인이 음절 경계에서 모호성을 일으키는 모든 경우를 살펴보았다. 또한 특별히 모음 체계를 옛한글로 그대로 확장할 때 발생하는 내부 모호성도 확인했다.

본고에서 제안한 절차 중에는 세별식 입력 방식에 같이 적용 가능한 것도 있다. 그러나 ‘낱자의 완성 이후에 필요한 임시 상태’, ‘종성 문맥에서 초성 전용 낱자의 입력’, ‘음절 경계 모호성’은 글쇠 차원에서 초성과 종성의 구분이 없는 두별식 입력 방식에만 특별히 적용되는 요소이다. 낱자 규칙을 대결합과 소결합으로 나눠서 기술한 것이 도깨비불 현상처럼 두별식의 고유한 낱자 처리 동작을 명시하는 데도 도움이 됨을 알 수 있었다.

2010년대부터는 Windows의 기본 글꼴인 ‘맑은 고딕’에 KS X 1026-1을 준수하는 옛한글 자모가 포함되기 시작했으며, 비록 동작 환경이 제한적이지만 옛한글 입력기도 운영체제에 기본 내장되기 시작했다⁶⁾. 옛한글에 대한 접근성이 예전에 비해 크게 개선된 셈이다. 그런

만큼 한글 입력에 관해서는 언어 통계를 기반으로 한 타자 행동의 개선[9][10][11]뿐만 아니라 그 저변에 있는 입력 오토마타의 기본 동작 자체에 대한 연구가 진행될 여지가 남아 있다. 필수와 글쇠 수는 적는데 입력 가능해야 하는 낱자는 많은 입력 방식을 설계할수록 입력 방식을 자동으로 구현하고 분석하는 도구의 유용성은 더 커질 것이다.

참고문헌

- [1] 김용목, “한글 입력·편집기의 통합적 설계와 구현에 관한 연구”, 연세대학교 석사학위논문, 2012.
- [2] 변정용, 이형준, “완전조합 음절 지원 훈민정음 웹 입력기”, 컴퓨팅의 실제 및 레터 제19권 제6호, pp. 371-375, 2013.
- [3] 남경완, “훈민정음의 제자 원리를 통해서 살펴본 핸드폰 한글입력 방식에 대한 연구”, 한국어학 제41권, pp. 325-255, 2008.
- [4] 임양원, 임한규, “모음의 방향성을 이용한 한글 입력 방식에 관한 연구”. 한국 멀티미디어학회 학술대회 논문집, 13(2), pp. 19-22, 2010.
- [5] 김호식, 전재웅, 최윤철, “터치스크린 기반 스마트폰에서의 한글 입력 기법”. 한국 컴퓨터 종합 학술대회 논문집, 38(1), pp. 263-265, 2011.
- [6] 강승식, 한광수, “한글 입력 시스템을 위한 기본 모음 집합의 구성”, 한국정보과학회 가을 학술발표논문집 제31권 제2호(I), pp. 784-786, 2004.
- [7] 최광무, “한글 모아쓰기에 관한 연구”. KAIST 석사학위논문, 1978.
- [8] KS X 1026-1: 2007 <http://www.unicode.org/L2/L2008/08225-n3422.pdf>
- [9] 오길록, 최기선, 박세영, “한글공학”, 대영사, pp. 81-108, 1995.
- [10] 김 국, 유영관, “사용빈도와 표준정합성을 고려한 컴퓨터 한글자판의 개선에 관한 연구”, 대한인간공학회지 제27권 제3호, pp. 7-14, 2008.
- [11] Ken Lunde, “CJKV Information Processing, Second Edition”, O'Reilly Media, pp. 350-361, 2008.

6) 과거에는 Microsoft Office의 한국어 Plus Pack에서만 옛한글 글꼴과 입력기가 제공되었다.