

*SICOPS '96: The first Seoul International Conference  
on Phonetic Sciences, Seoul, Korea, Oct. 24-26, 1996*

**Some Characteristics of Hanmal and Hangul  
from the viewpoint of Processing Hangul Information on Computers**

**KIM, Kyongsok**

Department of Computer Science, College of Natural Sciences,  
Pusan National University, Pusan 609-735, South Korea

Internet: kskim@asadal.cs.pusan.ac.kr

**Abstract**

In this paper, we discussed three cases to see the effects of the characteristics of Hangul writing system. In applications such as computer Hangul shorthands for ordinary people and pushbuttons with Hangul characters engraved, we found that there is much advantage in using Hangul. In case of Hangul Transliteration, we discussed some problems which are related with the characteristics of Hangul writing system.

Shorthands use 3-set keyboards in England, America, and Korea. We saw how ordinary people can do computer Hangul shorthands, whereas only experts can do computer shorthands in other countries. Specifically, the facts that 1) Hangul characters are grouped into syllables (syllabic blocks) and that 2) there is already a 3-set Hangul keyboard for ordinary people allow ordinary people to do computer Hangul shorthands without taking special training as with English shorthands. This study was done by the author under the codename of "Sejong 89". In contrast, like QWERTY or DVORAK, a 2-set Hangul keyboard cannot be used for shorthands.

In case of English pushbuttons, one digit is associated with only one character. However, by engraving only syllable-initial characters on the phone pushbuttons, we can associate one Hangul "syllable" with one digit. Therefore, for a given number of digits, we can associate longer words or more meaningful words in Hangul than in English.

We discussed the problems of the Hangul Transliteration system proposed by South Korea and suggested their solutions, if available. 1) We are incorrectly using the framework of transcription for transliteration. To solve the problem, the author suggests that a) we include all complex characters in the transliteration table, and that b) we specify syllable-initial and -final characters separately in the table. 2) The proposed system cannot represent independent characters and incomplete syllables. 3) The proposed system cannot distinguish between syllable-initial and -final characters.

**1. Introduction**

In this paper, we will discuss three cases to see the effects of the characteristics of Hangul writing system. In Section 2, we will discuss how ordinary people can do computer Hangul shorthands, whereas only experts can do computer shorthands in other countries. In Section 3, we will investigate engraving Hangul syllable-initial characters on phone pushbuttons. In these two cases, we found that there is much advantage in using Hangul.

In Section 4, we will discuss the problems of the current Hangul transliteration system and their solutions, if available. Conclusions are given in Section 5.

**2. A Syllable-typing Hangul keyboard for ordinary people - Sejong 89**

**2.1 We need to adopt both 3-set and 2-set Hangul keyboards as a standard**

Although a 2-set Hangul keyboard is the current standard, a 3-set keyboard has definite advantages in some cases:

1) A 3-set keyboard can successfully support the Hangul code in ISO 10646-1 [ISO 93], whereas the current 2-set keyboard cannot support it. For example, a 2-set keyboard cannot distinguish between the syllable-initial character ㄷ and the syllable-final character ㅏ.

2) A 3-set keyboard can support Old Hangul much more efficiently than a 2-set keyboard. The most serious problem with a 2-set keyboard is that, since there are lots of complex consonantal characters in Old Hangul, a 2-set keyboard has problems in finding out syllable boundaries from a stream of keystrokes. For example, with a 2-set keyboard, the following

character stream can be syllabized into two different words:

ㅁ 꺾 ㅓ ㅓ ㅓ ㅓ ---> 뺨근 ???  
 ---> 뺨곶 ???

The main reason of this ambiguity is that we do not know whether each of ㅓ, ㅓ, and ㅓ is syllable-initial or syllable-final. In contrast, with a 3-set keyboard, we clearly indicate at the time of typing whether each of ㅓ, ㅓ, and ㅓ is syllable-initial or syllable-final, and, therefore, there is no such ambiguity as shown below:

ㅁ 꺾 ㅓ ㅓ ㅓ ㅓ ---> 뺨근 (ㅓ, ㅓ: syllable-final; ㅓ: syllable-initial)  
 ㅁ 꺾 ㅓ ㅓ ㅓ ㅓ ---> 뺨곶 (ㅓ: syllable-final; ㅓ, ㅓ: syllable-initial)

3) Syllable-typing is possible with a 3-set Hangul keyboard, but not with a 2-set. In the remaining Subsections of this Section, we will investigate this topic in more details.

## 2.2 Shorthands use 3-set keyboards in England, America, and Korea

For shorthands, we could use hands, machine, or computers. In England, Palantype adopted a 3-set keyboard by arranging s, t, p, and h both in the left hand side (syllable-initial characters) and in the right hand side (syllable-final characters). Syllable-peak characters in the middle of keyboards. Similarly, in America, Stenotype adopted a 3-set keyboard by duplicating s, t, p, and r in the same manner. In contrast, ordinary people use QWERTY or DVORAK keyboards, which are 2-set. There is no 3-set English keyboards for ordinary people.

In Korea, we have both 3-set and 2-set keyboards for ordinary people. The most popular 3-set keyboard for ordinary people is called "390", meaning that a 3-set keyboard whose arrangement was decided in 1990. A 3-set keyboard was originally invented by late Dr. Kong, Byungwoo, in 1950 (yes, when the Korean war occurred!). Hangul computer shorthands for experts adopted a 3-set keyboard, although its arrangement is different from the 390 keyboard for ordinary people.

With machine or computer shorthands, we type a syllable at a time, not a character at a time as with ordinary typing. As a result, we cannot use QWERTY or DVORAK for English shorthands, since they are 2-set. Therefore, only highly-trained experts can do English shorthands. This is a natural outcome considering that there is no 3-set English keyboard for ordinary people.

## 2.3 Two reasons why ordinary people can do computer Hangul shorthands

However, in case of Hangul, it is highly possible for ordinary people to do computer shorthands for two reasons explained below. These two facts come from the characteristics of Hangul writing system.

### 1) Hangul characters are grouped into syllables (syllabic blocks)

In English, people need to take some training to be able to syllabize words fast, since ordinary people cannot easily recognize syllable boundaries. In contrast, with Hangul, students even in the second grade of an elementary school can easily recognize Hangul syllable boundaries.

Furthermore, in English shorthands, we need to derive sounds from written words (e.g., enough, knife), after which we type the sounds of the words, not the written words themselves. In contrast, with Hangul shorthands, we just type the written words.

### 2) There is already a 3-set Hangul keyboard for ordinary people

Although there is no 3-set English keyboard for ordinary people, there has been a 3-set Hangul keyboard for almost 50 years.

The two facts allow ordinary people to do computer Hangul shorthands without taking special training as with English shorthands.

## 2.4 The author's proposal: Sejong 89

Under the codename "Sejong 89", the author investigated a computer Hangul shorthands for ordinary people for one and half year, starting at the summer in 1989 (89 in Sejong came from

the last two digits in year 1989). The idea is that ordinary people can use a 3-set Hangul keyboard without any modification for computer Hangul shorthands. Let's call this method layman's shorthands in contrast with expert's shorthands. Although layman's shorthands is less efficient than expert's, there is one big advantage with layman's: that is, no special training is needed to do layman's shorthands as long as he/she uses a 3-set Hangul keyboard. With software supporting layman's shorthands, ordinary people can just try to press (usually) two to four keys simultaneously to enter a syllable.

### 2.5 Efficiency of layman's shorthands

There is no concrete data since software/hardware supporting layman's shorthands is not widely available yet. The author guesses that the input speed will be increased by about 30 % to 50 %.

Since there is a limit on the speedup with layman's shorthands, the author suggests that, for more speedup, we utilize mnemonics from expert's shorthands (e.g., shortening common phrases into short mnemonics). Note that we just utilize mnemonics, not the expert's Hangul keyboard.

### 2.6 A comparison of 3-set and 2-set keyboards in the context of shorthands

A 2-set keyboard cannot enter a syllable which has syllable-final character. For example, when we type "ㄱ, ㅏ, ㄴ" simultaneously, we do not know whether it is meant for "간" or "낙". As another example, we cannot enter "각", since there is one and only one key for "ㄱ".

### 2.7 A 3-set keyboard goes well in harmony with the characteristics of Hangul writing system

As we saw above, we cannot avoid using some form of 3-set keyboard for shorthands. 3-set Hangul keyboards go well in harmony with the characteristics of Hangul writing system. Since 3-set keyboards allow ordinary people to do layman's shorthands without much training, we need to adopt a 3-set keyboard for ordinary people as a standard.

## 3. Pushbuttons with Hangul characters

### 3.1 Pushbuttons with Hangul syllable-initial characters

People from North America and Western Europe may be quite familiar with the following arrangements of 10 digits and 24 or 26 characters. For example, a numeric equivalent for characters RENT is 7368.

1	2 ABC	3 DEF
4 GHI	5 JKL	6 MNO
7 PRS	8 TUV	9 WXYZ
*	0	#

1	2 ABC	3 DEF
4 GHI	5 JKL	6 MNO
7 PQRS	8 TUV	9 WXYZ
*	0	#

1 QZ	2 ABC	3 DEF
4 GHI	5 JKL	6 MNO
7 PRS	8 TUV	9 WXYZ
*	0	#

Similarly, we can think of arranging Hangul characters on pushbuttons. There can be a few options:

- 1) only syllable-initial characters are arranged;
- 2) both consonantal and vowel characters are arranged (a 2-set approach);
- 3) syllable-initial, -peak, and -final characters are arranged (a 3-set approach);
- 4) in each of the above three options, we can either include or exclude complex characters.

In case of English, one digit is associated with only one character. In a "RENT" example shown above, 7 is associated with one character R. However, in case of Hangul, by choosing an option 1), we can associate one Hangul "syllable" with one digit. For example, in the arrangement shown below, we can associate a four-syllable word "부산 대학 (Bu-san-dae-hag)" with 5630 (note that a numeric equivalent of a syllable-initial character "ㅂ" in syllable "부" is 5). Therefore, for a given number of digits, we can associate longer words or more meaningful words in Hangul than in English.

This comes from the characteristics of the Hangul writing system: namely, Hangul characters are grouped into syllabic blocks, which are then grouped into words. Therefore, Hangul words can be easily decomposed into syllables (We must admit that, sometimes, due to a liaison phenomenon, the syllable boundary based on the written form does not coincide with the phonetically correct boundary though. For example, “잠이” is actually pronounced as “자미”.) In contrast, with English, it is somewhat hard and it takes some time to find out the syllable boundaries.

Pushbuttons with Hangul syllable-initial characters:  
a Proposal 2 (1996.04.13) by Kim, Kyongsok

1 ㄱ	2 ㄴ	3 ㄷ
4 ㄱ ㅁ	5 ㅅ ㅆ	6 ㄴ ㄷ
7 ㅇ	8 ㅅ ㅆ	9 ㄷ ㅈ
*	0 ㅇ ㅎ	#

### 3.2 Pushbuttons with Hangul characters

We can also think of another type of arrangement, which allows entering Hangul syllables without any ambiguity. For example, in the above arrangement, when we press digit 4, we do not know whether it is meant for ㄱ (rieul) or ㅁ (mieum). Furthermore, we cannot enter syllable-peak characters.

For the purpose of entering Hangul syllables, we need to adopt either option 2) or option 3) shown above. Based on the author's previous researches, he feels that a 3-set approach (option 3)) is better than a 2-set approach (option 2)), since it is much easier to find syllable boundaries with a 3-set approach than with a 2-set approach. A separate paper on this issue is under preparation.

## 4. Transliteration and Transcription of Hangul/Hanmal

### 4.1 Problems of the proposed Hangul Transliteration system

Currently, South Korea has a proposed (NOT a national standard) transliteration system for the ISO. We will see its three problems and their solutions, if available, here.

#### 1) The framework

In Hangul transcription, we need to represent 19 consonantal sounds: 14 simple sounds and 5 complex sounds (ㄱ, ㄴ, ㅁ, ㅅ, ㅆ). In the proposed Hangul transliteration, we include 14 simple consonantal characters and 5 complex consonantal characters (ㄱ, ㄴ, ㅁ, ㅅ, ㅆ), excluding the remaining 11 complex consonantal characters (ㄱ, ㄴ, ㄷ, ㄱ, ㄴ, ㄷ, ㄱ, ㄴ, ㄷ, ㄱ, ㄴ, ㄷ). It seems to imply that those complex characters can be transliterated using two constituent simple characters in sequence. For example, ㄱ (ㄱ) can be transliterated as L (ㄱ) followed by G (ㄱ). Then we can apply the same reasoning in transliterating 5 syllable-initial complex characters (ㄱ, ㄴ, ㅁ, ㅅ, ㅆ)? Therefore, rules for ㄱ, ㄴ, ㅁ, ㅅ, ㅆ seem redundant.

When why do we have such redundant rules? The answer is that we are incorrectly using the framework of transcription for transliteration. In transcription, 5 syllable-initial complex sounds need to be represented, whereas there is no such sound as ㄱ or ㄴ. Therefore, it is correct that we include only five syllable-initial complex sounds, excluding the remaining 11 complex consonantal characters. However, in transliteration, we need to specify how to write down all (i.e., 5 plus 11) complex characters (ㄱ, ㄴ, ㅁ, ㅅ, ㅆ, ㄱ, ㄴ, ㄷ, ㄱ, ㄴ, ㄷ, ㄱ, ㄴ, ㄷ, ㄱ, ㄴ, ㄷ), not just 5 syllable-initial complex characters.

To solve the problem, the author suggests that 1) we include all complex characters in the transliteration table, and that 2) we specify syllable-initial and syllable-final characters separately in the transliteration table, as shown in Appendix. (The transliteration table was used for naming 11,172 Hangul syllables in ISO 10646-1.) By doing so, we can clearly specify how to transliterate complex characters (e.g., ㄴ, ㄱ) and how to transliterate ㅇ (i.e., NULL for syllable-initial ㅇ and NG for syllable-final ㅇ) and ㄱ (i.e., R for syllable-initial ㄱ and L for

syllable-final  $\text{ㄴ}$ ), depending on whether the character is syllable-initial or syllable-final.

### 2) The scope of application of the Hangeul Transliteration

The proposed system can only represent complete syllables, which are composed of syllable-initial and syllable-peak characters, and, optionally, syllable-final characters. It cannot represent independent characters (e.g.,  $\text{ㄱ}$ ,  $\text{ㄴ}$ ,  $\text{ㄷ}$ ,  $\text{ㄹ}$ ,  $\text{ㅁ}$ ,  $\text{ㅂ}$ ,  $\text{ㅅ}$ ,  $\text{ㅇ}$ ) or incomplete syllables (e.g., a syllable composed of syllable-peak and syllable-final characters:  $\text{ㄴ}$ ). As a concrete example, both "ㅇ" and "ㄴ" are represented as "a" and, therefore, there is an ambiguity. The same problem occurs with "ㅇ" and "ㄴ".

In the field of information technology, we sometimes need to represent independent characters and incomplete syllables. We need to make a further investigation in this regard.

### 3) Syllable-initial vs. syllable-final characters

The proposed system cannot distinguish between syllable-initial and syllable-final characters. Both are represented the same. For example, both syllable-initial "ㅁ" and syllable-final "ㅁ" are represented as "m". We need to make a further investigation in this regard.

## 4.2 Standardizing Hangeul Transliteration within ISO TC46/SC2/WG4

Within ISO, TC46/SC2/WG4 is in charge of standardizing Hangeul transliteration. On June 16-17, 1992, in Paris, South and North Korea "tentatively" agreed on a proposal. However, it did not proceed any further. Later, two Koreas proposed their own versions, which will be included in the ISO Technical Report 11941. It is supposed to be published in October 1996. Two proposals have the same assignment of Latin characters for vowels.

## 4.3 Names of Hangeul syllables and characters in ISO/IEC 10646-1

ISO/IEC 10646-1 is a code accommodating most major scripts in the world. In ISO/IEC 10646-1, two different Hangeul transliteration systems are used to name Hangeul syllables and characters. The 1992 version was used for writing down the following items:

- 1) names of 238 Hangeul characters for SYL-IPF (syllable-initial-peak-final) Hangeul code (code positions 0x1100-11ff),
- 2) names of 6,656 Hangeul syllables (code positions 0x3400-4dff), and
- 3) names of other Hangeul-related characters

The system was not even a CD at that time and, therefore, the author objected to using the system. However, it was decided to use the system by majority. As we saw above, the system did not progress any further and was dropped. I still think that it was a mistake to have used the system at that time.

Now a proposal to remove 6,656 syllables from, and to add 11,172 syllables to, ISO/IEC 10646-1 is being processed. We used South Korea's proposal [see Appendix] in naming those 11,172 syllables. Therefore, if and when the proposal is passed, we have two different transliteration systems in one and the same ISO/IEC standard.

## 4.4 WWW home page containing information for Hangeul Transliteration

- URL: <http://asadal.cs.pusan.ac.kr/han-tl-ts/> (Note. asadal.cs.pusan.ac.kr is the computer used by the author and his graduate students).
- Major contents
  - (1) Transcription: by Korean Language Society (1984) (in Hangeul)
  - (2) Transcription: by Ministry of Education (1984) (in Hangeul)
  - (3) Transcription: by Ministry of Education (1984) (in English)
  - (4) Transliteration tables: used for naming Hangeul syllables and characters in ISO/IEC 10646-1 (in English)
  - (5) a link to ISO TC46/SC2/ home page
- \* Items (1)-(4) are image files in GIF format.

## 5. Conclusions

In this paper, we discussed three cases to see the effects of the characteristics of Hangul writing system. In applications such as computer Hangul shorthands for ordinary people and pushbuttons with Hangul characters engraved, we found that there is much advantage in using Hangul. In case of Hangul Transliteration, we discussed some problems which are related with the characteristics of Hangul writing system.

### 5.1 A Syllable-typing Hangul keyboard for ordinary people - Sejong 89

Shorthands use 3-set keyboards in England, America, and Korea. We saw how ordinary people can do computer Hangul shorthands, whereas only experts can do computer shorthands in other countries. Specifically, the facts that 1) Hangul characters are grouped into syllables (syllabic blocks) and that 2) there is already a 3-set Hangul keyboard for ordinary people allow ordinary people to do computer Hangul shorthands without taking special training as with English shorthands. This study was done by the author under the codename of "Sejong 89". In contrast, like QWERTY or DVORAK, a 2-set Hangul keyboard cannot be used for shorthands.

A 3-set keyboard goes well in harmony with the characteristics of Hangul writing system. Specifically, 1) a 3-set keyboard can successfully support the Hangul code in ISO 10646-1; 2) a 3-set keyboard can support Old Hangul much more efficiently than a 2-set keyboard; and 3) ordinary people can do computer Hangul shorthands with a 3-set, but not with a 2-set. Therefore, we suggest that we adopt both 3-set and 2-set Hangul keyboards as a standard

### 5.2 Pushbuttons with Hangul syllable-initial characters

In case of English pushbuttons, one digit is associated with only one character. However, by engraving only syllable-initial characters on the phone pushbuttons, we can associate one Hangul "syllable" with one digit. Therefore, for a given number of digits, we can associate longer words or more meaningful words in Hangul than in English. This comes from the characteristics of the Hangul writing system.

### 5.3 Transliteration and Transcription of Hangul/Hanmal

We discussed the problems of the Hangul Transliteration system proposed by South Korea and suggested their solutions, if available.

1) We are incorrectly using the framework of transcription for transliteration. To solve the problem, the author suggests that a) we include all complex characters in the transliteration table, and that b) we specify syllable-initial and syllable-final characters separately in the transliteration table.

2) The proposed system can only represent complete syllables. In the field of information technology, we sometimes need to represent independent characters and incomplete syllables.

3) The proposed system cannot distinguish between syllable-initial and syllable-final characters. Both are represented the same.

We need to make a further investigation for points 2) and 3).

There is a Web home page containing information for Hangul Transliteration. The URL is "<http://asadal.cs.pusan.ac.kr/han-tl-ts/>".

\* written on August 3, 1996 (4329).

## 6. References

[KimK 88] A New Proposal for a Standard Hangul (or Korean Script) Code: How to accommodate both Databases and Word Processing, Kyongsok Kim & et al. Report No. UIUCDCS-R-88-1447, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL., U.S.A., August 1988, 54 pages.

[KimK 90a] "A New Proposal for a Standard Hangul (or Korean Script) Code", Kyongsok Kim. Computer Standards & Interfaces, Vol. 9, No. 3, pp. 187-202, 1990.

[KimK 90b] "Databases supporting Hangul (or Korean Script)", Kyongsok Kim. 1990 IEEE International Conference on Systems, Man and Cybernetics. pp. 485-490, Nov. 4-7, 1990. Universal City, CA., U.S.A.

[KimK 92a] "A Future Direction in Standardizing International Character Codes - with a special reference to ISO/IEC 10646 and Unicode", Kyongsok Kim. Computer Standards & Interfaces, Vol. 14, No. 3, pp. 209-221, May 1992.



- 1) At the beginning of a word, it is not transliterated.
- 2) At other places, it is transliterated as an apostrophe.
- 3) In naming Hangul syllables in ISO 10646-1, syllable-initial character ieung always comes at the beginning of a word, and, therefore, is not transliterated (e.g., ㅇㅈ: HANGUL SYLLABLE A).