

Proposal for a **Korean** Script Root Zone LGR

LGR Version K_LGR_v2.3

Date: 2021-05-01

Document version: K_LGR_v23_20210501

Authors: Korean script Generation Panel

1 General Information/ Overview/ Abstract

The purpose of this document is to give an overview of the proposed Korean Script LGR in the XML format and the rationale behind the design decisions taken. It includes a discussion of relevant features of the script, the communities or languages using it, the process and methodology used and information on the contributors. The formal specification of the LGR can be found in the accompanying XML document below:

- proposal-korean-lgr-01may21-en.xml

Labels for testing can be found in the accompanying text document below:

- korean-test-labels-01may21-en.txt

In Section 3, we will see the background on Korean script (Hangul + Hanja) and principal language using it, i.e., Korean language. The overall development process and methodology will be reviewed in Section 4.

The repertoire and variant sets in K-LGR will be discussed in Sections 5 and 6, respectively. In Section 7, Whole Label Evaluation Rules (WLE) will be described and then contributors for K-LGR are shown in Section 8.

Several appendices are included with separate files.

2 Script for which the LGR is proposed

ISO 15924 Code: Kore

ISO 15924 Key Number: 287 (= 286 + 500)

ISO 15924 English Name: Korean (alias for Hangeul + Han)

Native name of the script: 한글 + 한자

Maximal Starting Repertoire (MSR) version: MSR-4 [241]

Note. 'Korean script' usually means 'Hangeul' or 'Hangeul'. However, in the context of the Korean LGR, Korean script is a union of Hangeul and Hanja.

3 Background on Script and Principal Languages Using It

This section introduces the background on Hangeul and Hanja (Han script). The main focus of this section is to review Hangeul and Hanja from different aspects and to conclude that both Hangeul and Hanja must be included in K-LGR repertoire.

Section 3 is organized as follows. Section 3.1 covers Korean script, Hangeul and Han script (Hanja). Significant User Communities for the Korean Script is described in Section 3.2. Korean Language and Korean Script, Hangeul, are explained in Section 3.3. Finally, in Section 3.4, summary is given.

More information about Hangeul is described in Appendix G and more detailed information about Hanja usage in Republic of Korea is described in Appendix H.

3.1 History of Korean language and script¹

(1) Origin of the Korean Language

The origin of the Korean language remains unclear. In the early 20th century, when comparative linguistics prevailed, certain linguists presented a theory that Korean belonged to the putative Ural-Altaic language family. This theory was later discredited as others argued that even if Korean originated from the Altaic family, the two were barely inter-related, posing a mystery about the origins of the language.

Korean textbooks used in Korean high schools write that "it is likely that Korean belongs to the same proto-language from which Manchu, Mongolian and Turkish are derived. In other words, it is possibly a member of the Altaic language family." [221]. Old Korean had two geographically distinct dialects, the North and South Korean dialects. This can be verified in ancient Chinese texts such as *Sanguozhi* (三國志) (The Records of Three Kingdoms) and *Hou Han Shu* (後漢書) (The History of the Later Han).

(2) Old Korean

From a linguistic perspective, Korean can be classified into three groups: Old Korean, Middle Korean, and Modern Korean.

Old Korean refers to the Korean language that was used from the Three Kingdoms Period (B.C. 57~) to the Unified Shilla Period, and whose usage during this time can be verified in various ancient records. The unification of the three kingdoms by the Shilla Dynasty led to the consolidation of the dialects of the three kingdoms into the dialect of the conqueror Shilla (AD 676, 7th century). At that time, there was no writing system in Korea, and *Hanja* or adapted Chinese characters were used instead to stand for Korean words.

Chinese characters were used to denote the meaning or to represent the sound of Korean words in such writing systems as *idu*, *hyangchal*, or *gugyeol*, all of which remained in use until the end of the Early Modern Korean era. A Korean high-school textbook writes, "During this period, people spoke Korean, but there was no writing system. People had to use Chinese characters for the purpose of writing, which

¹ This information was provided by National Institute of Korean Language, Republic of Korea

is significantly different in terms of structure and syntax [221].

(3) Middle Korean

Middle Korean was the language used from the early 10th century, when the Goryeo Dynasty was founded, until the end of the 16th century when Japan invaded Korea. Unlike Old Korean, Middle Korean was largely consolidated and standardized. Middle Korean is divided again into two periods: The first half of the Middle Korean era corresponds to the period of the Goryeo Dynasty, while the second coincides with the early period of the Joseon Dynasty.

During the first half of the Middle Korean period, the influence of Goguryeo's language continued to linger, but in general, the Shilla language was predominant, and both languages were still written in Chinese characters. The creation of *Hunminjeongeum* (The Proper Sounds for the Instruction of the People) or *Hangeul* marked the beginning of a new era in the latter half of the period, as Korea now had its own writing system.

(4) Early Modern Korean

Modern Korean was used from the early 17th century to the late 19th century when the Joseon Dynasty began to open up to the outside world and ushered in the period of enlightenment.

Jae-il Kwon, professor of Seoul National University and former Director General of the National Institute of the Korean Language, explains, "From the 17th century on, the Korean language began to go through distinctive changes in its own phonology, vocabulary and grammar, compared to the previous periods, and its vowels underwent structural changes as well.

Many grammatical phenomena disappeared altogether, and new ones were introduced into the language [232]. During this period, unlike the previous period of Middle Korean, Hangeul was expanding its ground as the prevailing writing system in all aspects of people's daily life.

A school textbook explains, "King Gojong declared the Gabo Reforms in 1894 and ordered that 'the Korean versions of laws and regulations shall prevail as the primary and official forms of all documents, and their translation into Chinese characters shall be provided, or Korean and Chinese characters shall be mixed in such documents'" [221]. This period is also characterized by rapid inflows of Western culture into the country via China and Japan. As a result, there was a large influx of new vocabulary to express new cultural phenomena, items, and concepts.

(5) Modern Korean

Modern Korean refers to the Korean language that has been used since the period of enlightenment to the present day. In the early stage of Modern Korean, Hangeul and *Hanja*, the Korean name for the Chinese characters used in the Korean language, were mixed in writing [221], but Hangeul gradually took over Hanja.

A school textbook writes, "Hangeul took over Hanja as the dominant script in Korean writing after Korea gained its independence from Japanese rule in 1945, but Hanja remained in use, mixed with Hangeul, in newspapers and technical publications until the 1980s. Even today, Hanja is sometimes provided in parentheses, next to Korean words such as homonyms, contracted words, academic or technical terms, people's names, etc., when it is deemed that further clarification is required because Korean words alone may not be sufficient to deliver the intended meaning or may be misunderstood and confused with other words that have the same sounds."

However, Korean is written mostly and only in Hangeul, and *Hanja* are simultaneously added for additional information only in exceptional cases where it is deemed necessary by the author to clarify the meaning or to avoid potential confusion or vagueness in understanding the meaning of the words.

Korean language is mainly used in Republic of Korea (South Korea) and Democratic People's Republic of Korea (North Korea). Korean language, on the other hand, is also recognized and used by those who have Korean origin living abroad under their Korean ancestors' influence or those who learn Korean as their second language in many parts of the world including China, USA, Japan, Europe, Brazil, Russia, Vietnam, and so on.

3.2 Significant User Communities for the Korean Script

There are three groups of people who use Korean around the world. According to Jae-il Kwon [232], professor of Seoul National University and former Director General of the National Institute of the Korean Language,

the first group of people speaks Korean as their first language and the official language, and this group includes South and North Koreans.

The second group speaks Korean as a second language and includes Koreans who live in other countries than Korea, such as China, Japan, the U.S.A., and Central Asia, as well as Koreans who speak the language of the host country as their first language while speaking Korean in their homes and the ethnic community. Korean immigrants and their descendants living in other countries belong to this group.

Meanwhile, the "Joseon language" is a variant of Korean that is spoken by the so-called "Joseonjok" or Joseon people living in the Yanbian Korean Autonomous Prefecture in Jinlin Province and other parts of the three Northeastern Provinces of China, while the variant of Korean used by the descendants of Goryeo people living in Central Asia including Kazakhstan, Uzbekistan, and Kyrgyzstan is called, "Goryeo-mal" or the Goryeo language.

Foreigners who learn and speak Korean as a foreign language, but not as a first or second language, make up the third group.

A considerably large number of people speak Korean. According to Professor Kwon [232],

In the latest edition of *Ethnologue* (<https://www.ethnologue.com/country/KR>), which provides a wide variety of information on languages around the world, 77.2 million people spoke Korean as of 2010, making the language the 13th largest linguistic group in terms of the number of speakers of the language.

According to data released recently by an Internet-related international organization, Korean is the world's 10th most frequently used language on the Internet.

In addition, Korean was adopted as a language of international publication at the 43rd General Assembly of the WIPO held in September 2007, allowing applications for international patents to be submitted in Korean and international patents to be viewed in Korea.

3.3 Korean Language and Korean Script, Hangeul

Until the first half of the Joseon Dynasty (equivalent to the later period of Middle Korean), Chinese characters were used to write Korean as no Korean writing system existed at that time. Koreans developed "idu" and "gugyeol" as writing systems, in which Chinese characters were used to represent the sounds of Korean words. After the creation of Hunminjeongeum (or Hangeul), both Hangeul and Hanja were used together, while only Hangeul is used in modern Korean.

Today, Korean is written in Hangeul. Hanja are sometimes provided in parentheses next to Korean words, but only when the word in Hangeul alone may be misunderstood due to its multiple meanings or when further clarification of a specific meaning is necessary. In rare cases, words are written in Hanja and the particles and suffixes associated with the words are left in Hangeul.

Under "The Writing System of the Korean Language" the Doosan Encyclopedia [231] explains that "*Hanja* consists of ideographs while Hangeul is a phonemic orthography, so they are two very distinct scripts that are used together in the Korean language. Hanja was imported from China as a product of growing interaction with China. It is presumed that Hanja became the dominant writing system in Korea from the 4th to the 6th century, during which each of the three kingdoms that were competing for hegemony on the Korean Peninsula established a centralized system of government in their respective territories, and many historical books were published during this period.

Consequently, Chinese characters dominated written texts in the three old kingdoms, and this state of affairs continued until the end of the 19th century. So the Korean people spoke Korean but wrote their speech in Chinese characters, the script of another country, creating a discrepancy between speech and script, which caused interferences between the two languages over time. As a result, elements of the Chinese language were incorporated into Korean, and many Chinese words made their way into the Korean vocabulary, replacing the original Korean words."

Huminjeongeum was invented in 1443 during the early days of Joseon Dynasty and promulgated in 1446. Finally, Koreans had their own writing system to record the sounds of their speech. The section of Doosan Encyclopedia titled "The Korean Language" describes Hangeul "as an independent and complete system of writing that is unique and highly scientific."

Most countries adapted existing writing systems and adopted them as their own, but Hangeul is an original script that is isolated in the sense that it has not been derived from any other previous script. The graphemes and phonemes of Hangeul are directly and systematically related [231].

3.4 Summary: both Hangeul and Hanja are needed in K-LGR repertoire

In Republic of Korea, Hangeul is mainly used for writing Korean. Therefore, Hangeul syllables must be included in K-LGR repertoire. There are 11172 modern Hangeul syllables and they are included in K-LGR repertoire. Hangeul is already allowed at the second level domain under .kr and .한국.

Although Hangeul is used much more than Hanja in daily life, Hanja is still used. The Hanja usage in Republic of Korea is described in detail in Appendix H. Therefore, both Hangeul and Hanja must be

included in K-LGR repertoire. 4758 Hanja characters which is a union of KS X 1001 and IICORE characters with source K are included in K-LGR.

Having a separate Hanja repertoire in K-LGR (different from that in C-LGR or J-LGR) currently implies the following:

- 1) The repertoire of K-LGR is different from that of C-LGR or that of J-LGR.
- 2) The variant sets in K-LGR are different from those in C-LGR.

As shown in Appendix H from different angles, Hanja is still critical and still used quite a lot in Republic of Korea. Therefore, it is necessary that Hanja shall be included in K-LGR. A brief summary of issues discussed in Appendix H is given below:

- Many Korean words have a corresponding Hanja spelling. Hangul-Hanja mixed writing had been popular since the end of 19th century. It has decreased in recent years.

Cheol Huh (2010) [222] found that Hanja-based Korean words are approximately 66% in the Korean corpus research done in 2002 and 2005, purely Korean words around 26%, loan words originating from other languages around 4%, and others 4%. In addition, Hanja-based and Hangul-Hanja mixed words are approximately 69.7%.

Considering the history of Korean language, Hanja, and Hangul, it is natural that many Korean words are directly coming or derived from Chinese ones. Korean dictionaries show entry words in Hangul. When it can be written in Hanja, a Korean entry word is listed in Hanja as well as Hangul.

Around the end of the 19th century, a Hangul-Hanja mixed writing system appeared and, in the 20th century, this Hangul-Hanja mixed writing system became quite popular.

Sample pages of law books showing Hangul-Hanja mixed writing are given in Appendix E.

- The Framework Act on Korean Language regulates government documents only.

Framework Act on Korean Language was announced in 2005. Article 14, item 1 of the Act (see English version at http://elaw.klri.re.kr/eng_mobile/viewer.do?hseq=28092&type=part&key=17) is shown below:

Article 14 (Preparation of official documents)

Public institutions, etc. shall prepare official documents in the Korean language in accordance with language norms: Provided, That Chinese characters or other foreign letters may be entered in parentheses, in cases prescribed by Presidential Decree.

Historically, there was a very similar act before the Framework Act on Korean Language was announced in 2005. In Oct. 1948, "Act on using Hangul only" was announced. It says:

Official documents of Republic of Korea are written in Hangul. However, for the time being, if necessary, Hanja can be used in parallel.

It should be noted that these two acts are concerned with the use of Hangul in "official documents"

of public institutions (governments and government-related institutions) only. However, newspapers, magazines, non-governmental organizations' documents, etc., have not been regulated by the authority.

- Most place names, personal names, and traditional Korean family clans have their own corresponding Hanja characters and these are good candidates for use in Hanja domain names.
- There are many Hanja or Hangul-Hanja-mixed registered Trademarks and store signboards.
- KGP maintains that K-LGR must include both Hangul and Hanja in its repertoire based on the discussions in Section 3 and Appendix H.
- KGP decided that either Hangul only or Hanja only labels will be allowed in Korean language TLD while no Hangul and Hanja mixed labels are permitted.

4 K-LGR: Overall Development Process and Methodology

The overall development process and methodology of K-LGR are explained in this Section.

The creation of KGP (Korean script Generation Panel) and the public hearings in Korea for K-LGR are briefed in 4.1 and 4.2, respectively. 4.3 explains how Hangul repertoire and Hanja repertoire are created.

The challenge of incorporating the DPRK's voice in K-LGR is described in 4.4. The main activities of the Korean script Generation Panel (KGP) are covered in 4.5.

4.1 KGP (Korean script Generation Panel)

The Korean script Generation Panel (KGP) was initially formed in December 2013 and reorganized in January 2015. The panel includes two persons in charge of managing the state domain registration system and registration agencies, two Hanja experts, one Korean language expert, three community members and three industry representatives.

4.2 Public Hearings on K-LGR

A draft K-LGR repertoire and variant sets were presented for feedback at the Workshop #1 of the 5th Korea Internet Governance Forum in September, 2016: <http://igf.or.kr/1163>. The KriGF home page is <http://igf.or.kr>. The title of the workshop was Domestic Public Review of Generation of Hanja Top-level Hanja domain.

In September, 2017, there was a similar workshop. The title of Workshop #9 of Korea Internet Governance Forum 2017 was Opening Policy of Internet Address Resources.

In July, 2018, there was another similar workshop. The title of Workshop #3 of Korea Internet Governance Forum 2018 was Domestic Public Review regarding Top-level Hangul and Hanja domain.

After the ICANN public comments period for K-LGR v1.0 in 2018, the following KGP meetings were open to the general public:

KGP meetings #28 (2018.03.08.), #29 (2018.03.29.), #30 (2018.04.26.), #31 (2018.05.18.), #32 (2018.06.15.), #33 (2018.08.10.), #34 (2018.12.17.), #35 (2019.04.12.), #36 (2019.05.09.), #37 (2019.07.05.), #38 (2019.08.22.), #39 (2020.08.26.).

Issues raised were discussed and consensus was reflected in K-LGR v2.0.

4.3 Hangul repertoire and Hanja repertoire

For Hangul repertoire, no special coordination was necessary since the set of 11172 Modern Hangul syllables is commonly accepted by Korean-speaking people.

For Hanja (Han script) repertoire, no special coordination was needed. K-LGR v1.0 includes 4758 Hanja characters which is a union of KS X 1001 and IICORE characters with source K.

4.4 Challenge of Consultation with DPRK

Although it would be desirable for K-LGR to incorporate the voices of DPRK, one of the key stakeholders of Korean communities, it is almost impossible to communicate with the experts in DPRK regarding K-LGR issues as of today. As a result, the repertoire of K-LGR includes Hanja characters reflecting the needs of Republic of Korea (South Korea) only.

4.5 Main Activities of Korean script Generation Panel (KGP)

Dec. 2013, Korean script Generation Panel set up.

Jun. 2014, KGP meeting #1

Jan. 2015, KGP meeting #3 (Korean script Generation Panel reorganized), monthly meeting

May 2015, KGP meetings #6, #7; K-LGR v0.1 published;

Jun. 2015, KGP meeting #8; K-LGR v0.2 published

Aug. 2015, KGP meeting #10; K-LGR v0.3 published

Feb. 2016, The Korean Community "formally" Forms Korean script Generation Panel for Developing the Root Zone Label Generation Rules (LGR), 2016-02-01;

Mar. 2016, K-LGR v0.4 published

Sep. 2016, K-LGR v0.5 published

Nov. 2016, K-LGR v0.6 published

Mar. 2017, K-LGR v0.7 (repertoire and variant sets) published

Dec. 2017, K-LGR v1.0 (repertoire and variant sets) published

Jan. ~ Mar., 2018, ICANN Public comments period for K-LGR v1.0

Mar. 2018 ~ Aug. 2020: More than 10 KGP meetings (open to the general public) were held to revise K-LGR v1.0

Feb. 2020, K-LGR v2.0 (repertoire and variant sets) published

Sep. 2020, K-LGR v2.1 (repertoire and variant sets) published

Dec. 2020, K-LGR v2.2 (repertoire and variant sets) published

Feb. 2021, K-LGR v2.3 (repertoire and variant sets) published

5 Repertoire

The repertoire includes 11,172 Hangul syllables and 4,758 Hanja characters for a total of 15,930 code points (11172 + 4758). The LGR also includes 3 out-of-repertoire characters that are used in variant sets involving Hangul syllables. See section 6. Variants for more details.

Hangul repertoire is presented in 5.1 and Hanja repertoire is presented in 5.2.

5.1 Hangul repertoire

11172 Hangul syllables of Unicode code points are from U+AC00 to U+D7A3 [1, 3].

In modern Hangul, there are 19 syllable-initial (leading consonant) letters, 21 syllable-peak (vowel) letters, and 27 syllable-final (trailing consonant) letters, including complex letters. A Hangul syllable block is composed of a syllable-initial letter, a syllable-peak letter, and optionally a syllable-final letter. The number of syllable blocks without a syllable-final letter is $19 * 21 = 399$ and the number of syllable blocks with a syllable-final letter is $19 * 21 * 27 = 10773$. If we add 399 and 10773, the total number of syllable blocks becomes 11172. We can get the same number this way: $19 * 21 * (27 + 1) = 11172$ where 1 is added for syllable blocks without a syllable-final letter.

Though some might argue that not all 11,172 syllables are actually used in our daily lives, there is no proper process for extracting a meaningful stable subset. The full repertoire should be accepted as the general industry practice for the Hangul syllables.

5.2 Hanja repertoire

The set of 4758 Hanja characters is a union of the KS X 1001 (= K0) and IICORE characters with source K (see Table 1).

1) KS X 1001 (previously KS C 5601) = K0: This used to be very widely used. The Hanja set was first included in KS X 1001 in 1987. The repertoire can be identified in [101].

2) The set of 4743 Hanja characters which are IICORE characters with source K. The repertoire can be identified in [102].

Table 1. The source and number of characters for Hanja in K-LGR v1.0

Source of Hanja Character Set	# chars
1) KS X 1001 (268 compatibility chars excluded)	4620
2) IICORE characters with source K.	4743
K-LGR v1.0 (2017.12.10.): Hanja repertoire (Union of 1) and 2))	4758

Note: While the Hanja character corresponding to code point U+9ED9 does not have a K source in Unicode 6.3 (corresponding closely to ISO/IEC 10646:2012 in terms of repertoire) which is the reference for all root zone LGRs, a K source (K6-1021) was added in Unicode 12.0 and will be added in ISO/IEC 10646:2020 shortly. The other 6 KLGR Hanja characters without K source (U+5ABA, U+741F, U+74A4, U+74B8, U+7807, and U+7A66) are expected to get K sources in the near future as part of the planned KS X 1027-5 revision. These K sources can then be added in future versions of Unicode and ISO/IEC 10646. This process is called 'horizontal extension' and is a common and non-controversial process in character encoding.

6 Variants

This LGR contains 289 variant sets, which can be classified as follows:

- 1) 283 variant sets consist of variants between two or more in-repertoire Hanja characters (i.e. part of this LGR repertoire). One of these variant sets also includes a Hangul syllable.
- 2) 4 variant sets consist of variants between one or more in-repertoire Hanja character and a Hangul syllable. This includes one variant set already included in the previous group.
- 3) 3 variant sets consist of variants between one out-of-repertoire Hanja character and a Hangul syllable. One variant set also includes one in-repertoire Hanja character.

Because one variant set is classified in both group 1 and 2, the total of variant sets is $283+4+3-1= 289$.

In addition to the variants listed in this LGR, integration into the Root Zone LGR will result in many additional variants to out-of-repertoire code points as result of variants defined in other LGRs, including the effects of transitivity. However, the list of in-repertoire variants in this LGR is exhaustive.

In 6.1, variant issues with Hangul are reviewed. In 6.2, variant issues between Hangul syllables and Hanja characters are reviewed. Variant issues with Hanja are reviewed in 6.3.

Lastly, in 6.4, we explore how each of variant character in a variant set can be allocated or blocked in a label.

6.1 Variant issues with Hangul

There are no variant issues with (or between) 11172 Hangul syllables.

6.2 Variant issues between Hangul syllables and Hanja characters

Most Hanja characters are too complicated to be confused with Hangul syllable blocks. There are, if any, just a few Hanja characters that could be possibly confused with Hangul syllable blocks for those who do not know Hangul and/or Hanja well.

KGP investigated this issue systematically. In summary, it is found that the following 17 Hanja characters could be possibly confused with a Hangul "letter" (not a syllable block).

1-1) 8 Hanja characters that could be possibly confused with a Hangul consonantal letter:

U+4E5A ㄴ (U+1102 ㄴ); U+4EBA 人 (U+1109 ㄴ); U+5165 ㄴ (U+1109 ㄴ);
U+531A ㄴ (U+1103 ㄴ); U+5338 ㄴ (U+1103 ㄴ); U+53E3 ㄴ (U+1106 ㄴ);

U+56D7 □ (U+1106 □); U+5DF1 己 (U+1105 己)

1-2) 9 Hanja characters that could be possibly confused with a Hangul vowel letter:

U+4E00 一 (U+1173 一); U+4E04 丄 (U+1169 丄); U+4E05 丅 (U+116E 丅);
U+4E0C 丌 (U+1172 丌); U+4E28 丨 (U+1175 丨); U+4E29* 乚 (U+116C 乚);
U+4E85 丿 (U+1175 丿); U+4EA0 丮 (U+1169 丮); U+535C 卜 (U+1161 卜)

* Not included in MSR-4. All the other 16 characters are included in MSR-4.

However, since only syllable blocks are included in K-LGR, KGP checked whether all combinations of 8 * 9 = 72 could be possibly confused with any Hangul syllable blocks and found that only the following 8 Hanja characters fall into that category²:

2-1) 4 Hanja characters that are in MSR-4 and also in K-LGR:

U+C2A5 𪛗 - U+4ECA 今 (sources: G, H, T, J, K, V)
U+C2B4 𪛘 - U+5408 合 (source: G, H, T, J, K, V)
U+D2BD 𪛚 - U+9577 長 (sources: G, H, T, J, K, V)
U+C870 𪛜 - U+4E15 丯 (sources: G, H, T, J, K, V)

2-2) 3 Hanja characters that are in MSR-4, but not in K-LGR:

U+C4F0 𪛞 - U+4E1B 𪛟 (source: only G)
U+B9C8 𪛠 - U+535F 𪛡 (sources: G, H, T)
U+B258 𪛢 - U+723F 𪛣 (sources: G, H, T, J, K2)

2-3) 1 Hanja character that is not in MSR-4 and not in K-LGR:

U+C2A4 𪛔 - U+4EBC 𪛕 (sources: G, H, T, J)

² In order to explore if there exist such confusions, KGP conducted a pilot experiment with sample labels ('가슴-가슴' & '슴井-슴井'), participated by 162 users some of which were not aware of Hanja. The result was 38% became confused while 55% did not.

KGP obtained the above result

a) by eye-checking each Hanja character in MSR-4, and also

b) by systematically checking Hanja characters which contain both one of 8 characters in 1-1) and one of 9 characters in 1-2).

KGP added the following four variant sets (see table 2-2 in Appendix F worksheet 'K_LGR_v2.3_20210201_vg_exp') each of which contains one Hangul syllable and one or two Hanja characters in 2-1) to K-LGR:

(Hangul syllable U+C2A5 **승**, Hanja U+4ECA **今**)

(Hangul syllable U+C2B4 **슴**, Hanja U+5408 **合**, Hanja U+95A4 **閤**)

(Hangul syllable U+D2BD **脹**, Hanja U+9577 **長**)

(Hangul syllable U+C870 **조**, Hanja U+4E15 **丕**)

KGP added the following three variant sets (see table 2-3 in Appendix F worksheet 'K_LGR_v2.3_20210201_vg_exp') each of which contains one Hangul syllable and one or two Hanja characters in 2-2) to K-LGR:

(Hangul syllable U+C4F0 **쓰**, Hanja U+4E1B **叢**, Hanja U+53E2 **叢**)

(Hangul syllable U+B9C8 **ㅁ**, Hanja U+535F **ㅁ**)

(Hangul syllable U+B258 **누**, Hanja U+723F **𠂇**)

It needs to be noted that three Hanja characters are out of K-LGR repertoire and are out-of-repertoire-var.

6.3 Variant Issues with Hanja

1) Prior to integration with C-LGR, there were 37 KLGR variant sets between Hanja characters.

Among these 37 variant sets, 36 sets each consist of 2 Hanja characters and the last one consists of 3 Hanja characters.

The list of 37 Hanja variant sets is shown in Table 1-1 of Appendix F worksheet 'K_LGR_v2.3_20210201_vg37+4+3'.

2) The additional number of variant sets between Hanja characters required by integration with C-LGR increased the total value from 37 to 283.

Among these 283 variant sets, 264 sets each consist of 2 Hanja characters, 18 sets each consist of 3 Hanja characters, and the last one consists of 2 Hanja characters and 1 Hangul syllable.

The list of 283 Hanja variant sets is shown in Table 2-1 of Appendix F worksheet 'K_LGR_v2.3_20210201_vg_exp'.

Prior to integration with C-LGR, there is a variant set (Hangul syllable U+C2B4 습, Hanja U+5408 合). Then, for integration with C-LGR, variant set (Hanja U+5408 合, Hanja U+95A4 閤) is added to expanded variant sets. As a result, variant set (Hangul syllable U+C2B4 습, Hanja U+5408 合, Hanja U+95A4 閤) is listed both in Hanja-Hanja variant sets and in Hangul-Hanja variant sets.

6.4 Variant Policy in a label: allocatable vs. blocked

Because mixed Hangul/Hanja labels are disallowed using a WLE rule (see 7), and no two Hangul code points are variants of each other, the only variant labels that can exist for a Hangul label would be all Hanja (with every code point having a Hangul variant). For a Hanja label, variants could be either all Hanja or all Hangul.

1) Consider a variant set composed of Hangul1 and Hanja1 as explained in 6.2.

When a Hangul syllable (Hangul1) in an applied-for label belongs to a K variant set [a situation we may represent as (Hangul1, Hanja1)], the label containing Hangul1 is allocated but the other label containing Hanja1 in the variant set is blocked. Actually, the label allocated consists only of Hangul1 and other Hangul syllable(s), while the label blocked is composed exclusively of Hanja character(s).

Conversely, when a Hanja character (Hanja1) in an applied-for label belongs to a K variant set [a situation we may represent as (Hangul1, Hanja1)], the label containing Hanja1 is allocated but the other label containing Hangul1 in the variant set is blocked. Actually, the label allocated consists only of Hanja1 and other Hanja character(s), while the label blocked is composed exclusively of Hangul syllable(s).

2) Consider the case where a variant set is composed of 2 or 3 Hanja characters as explained in 6.3.

When a certain Hanja character (Hanja1) in an applied-for label belongs to a K variant set, [a situation we may represent as (Hanja1, Hanja2) or (Hanja1, Hanja2, Hanja3)], only the label containing Hanja1 is allocated: all other labels that contain either of the characters Hanja2 or Hanja3 in the variant set are blocked.

7 Whole Label Evaluation Rules (WLE)

A label may consist of Hangul syllables only or Hanja characters only.

8 Contributors

Contributors include two technical experts, one Korean linguist, two Hanja experts, two policy makers, four persons from community, three persons from registry (registration institute), and three persons from registration agency.

Field	Name	Organization
Technical experts	Kyongsok KIM (Chair)	Pusan National University
Technical experts	Dongman LEE	Korea Advanced Institute of Science and Technology
Linguist (Hangul)	Jeongdo CHOI	The National Institute of the Korean Language
Linguists (Hanja)	Sanghyun SHIN	Korea University
Linguists (Hanja)	Sungduk CHO	Dankook University
Policy Makers	Youngeum LEE	Korea National Open University
Policy Makers	Youn Jung PARK	The State University of New York, Korea
Community	Eungjun JEON	You Me Patent & Law Firm
Community	Boknam YUN	Hankyul Law Group
Community	Byeongil OH	Jinbonet
Community	Kyuhong BYUN	Skelter Labs
Registry	Jungsup PARK	Korea Internet & Security Agency
Registry	Minjung PARK	Korea Internet & Security Agency
Registry	Boyoung KIM	Korea Internet & Security Agency
Registrar	Geonwoo KIM	inames
Registrar	ChangKi JANG	Gabia Inc.
Registrar	Myungsoo LEE	MEGAZONE

References

[0] Unicode 1.1

[1] Unicode 2.0

[3] Unicode 3.0

[14] The Unicode Consortium. The Unicode Standard, Version 6.3.0, (Mountain View, CA: The Unicode Consortium, 2013. ISBN 978-1-936213-08-5)

<http://www.unicode.org/versions/Unicode6.3.0/>

[101] 'kIRG KSource' values starting by K0 in the Unihan database

[102] 'kIICore' VALUES including the letter 'K' (indicating source) in the Unihan database

[221] Kim Jungshin et. al, Section 3 of High School Korean II, The Past and Present of the Korean Language and the Writing System, taken from the website of Kyohaksa,
http://www.kyohak.co.kr/2014textbooks/h_kor1.html

[222] Huh Cheol, 'An Analysis of the Use of Hanja, based on "The Survey of Frequently Used Modern Korean Vocabulary 1-2" ', <Korea Education Research, Vol. 34, (June 2010), pp. 221-244, ISSN 1225-1321 KCI>.

[231] <http://terms.naver.com/entry.nhn?docId=1189683&cid=40942&categoryId=32978> The Writing System of the Korean Language (Doosan Encyclopedia)

[232] Jae-il Kwon, The History of Speech, Stories of the Korean Language, National Institute of the Korean Language, retrieved from
<http://news.korean.go.kr/index.jsp?control=page&part=view&idx=9171>

[241] Internet Corporation for Assigned Names and Numbers, "Maximal Starting Repertoire — MSR-4 Overview and Rationale", 25 January 2019. Available from
<https://www.icann.org/en/system/files/files/msr-4-overview-25jan19-en.pdf>

[242] Internet Corporation for Assigned Names and Numbers, "Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels." (Los Angeles, California: ICANN, March, 2013)
<http://www.icann.org/en/resources/idn/variant-tlds/draft-lgr-procedure-20mar13-en.pdf>

[243] Integration Panel "Requirements for LGR Proposals from Generation Panels" available online as
<https://www.icann.org/en/system/files/files/Requirements-for-LGR-Proposals-20150424.pdf>

[244] Out of Repertoire Variants in Root-Zone LGR and Proposals, Version 2017-09-25a:

[Appendix A.](#) Several figures showing usage of Hanja in ROK

[Appendix B.](#) Hangul-Hanja mixed trademarks

[Appendix C.](#) Hangul-Hanja mixed store signboards

[Appendix D.](#) A sample of Corporate Register of Samsung Mulsan containing
Hanja in corporate name

[Appendix E.](#) Law books showing Hangul-Hanja mixed style

[Appendix F.](#) Variant sets in K-LGR v2.3

[Appendix G.](#) More Information about Hangul

[Appendix H.](#) More Information about Hanja and Hangul-Hanja in ROK