klgp175\_2

**Responses to IP feedback to the Korean script root zone LGR**

**Source: KLGP (Korean Language Generation Panel)**

**Date: 2015.09.10.**

**Responses are added below the feedback from IP.**

**This document is a summary of responses to IP feedback (klgp175\_1)**

==========================================================

Considerations concerning the repertoire and variant set of

the Korean script root zone LGR

**Source: Integration Panel**

**Date: June 21st2015**

**2.2 Variant mapping references**

KLGR contains 132 lines with variants definition constituting 66 pairs. Some examples:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **No** | **Glyph** | **Code** | **VP1** | **Code** |  | **VP2** |  |  |
| 5468 | 鑑 | 09451 | 鑑 | 09451 | A | 鑒 | 09452 | A |
| 5469 | 鑒 | 09452 | 鑑 | 09451 | A | 鑒 | 09452 | A |
| 0370 | 减 | 051CF | 減 | 06E1B | A | 减 | 051CF | A |
| 2694 | 減 | 06E1B | 減 | 06E1B | A | 减 | 051CF | A |

No definitions for VP1, VP2, or ‘A’ were provided in the preliminary KLGR draft.

==> VP1: Code position of variant1, VP2: Code position of variant2, A = allocatable

Findings:

KLGR is fully included in MSR-2. This was not the case in its first iteration but is now established, and we are pleased to see this.

KLGR covers completely the Korean basic set: K0 which is also fully included in MSR. This was expected because K0 is much alike the Japanese J0 set. One would expect it to be in any Hanja minimal set such as J0 is in any minimal Kanji set.

The case for the extra characters (840 + 538 + 5 = 1383) still needs to be made. Some of these may be needed for out of repertoire (in a K source meaning) variants only.

==>

- Explanation of marks/signs used in the Venn Diagram.

. K = KS X 1001, P = KPS 9566, I = IICORE/K, T = Hanja Test (한검 시험)

1) K0 4620 -> The number of Hanja chars in K0 (KS X 1001) is 4620.

2) P0 4653 -> The number of Hanja chars in P0 (KPS 9566) is 4653.

3) IK 4743 -> The number of Hanja chars in K column of IICORE is 4743.

4) HT 4641 -> The number of Hanja chars in QTKHP is 4641.

\* QTKHP: Qualifying Test of Korean Hanja Proficiency, 한국 한자 능력 검정 시험, http://www.hanja.re.kr/

5) Note 1. Hanja characters in IICORE/KP is exactly the same as those in P0 (KPS 9566) and therefore IICORE/KP is not shown explicitly in the Venn Diagram.

Note 2. Size of (~K)∩(~P)∩(~I)∩(~T) is 0

Note 3. Size of (K∪P∪I∪T) is 4819 (= The same as the number of Hanja chars in K-LGR-1, v0.3)

- Explanation of regions in the Venn Diagram of four sets:

Example 1: "KPIT 4594" -> The number of characters in K∩P∩I∩T is 4594.

Example 2: "K 2", -> The number of characters in K∩(~P)∩(~I)∩(~T) is 2.

Example 3: "KP 3" -> The number of characters in K∩P∩(~I)∩(~T) is 3.



It is also surprising to find 14 non-K source code points in the KLGR set. One would expect all of them to have a K source value if they are going to express Hanja text. Note however that 8 of these code points have a KP source (North Korea or DPRK).

==> 4819 Hanja characters in K-LGR v0.3 are extracted from four (or five) sources:

|  |  |
| --- | --- |
| Source of Hanja Character Set | # chars |
| 1) KS X 1001 (268 Comptb. Hanja characters excluded) | 4,620 |
| 2) KPS 9566 | 4,653 |
| 3) IICORE - K column marked | 4,743 |
| 4) IICORE - KP column marked  \* This is exactly the same as 4,653 Hanja characters in 2) KPS 9566 | 4,653 |
| 5) QTKHP: Qualifying Test of Korean Hanja Proficiency  (한국 한자 능력 검정 시험)  \* 9 Compatibility Hanja characters not in MSR-1 are excluded. | 4,641 |
| K-LGR-1, Hanja list v0.3 (2015.08.13.) | 4,819 |

1) The following seven code positions are kept in K-LGR v0.3:

05AB4 0741F 074A4 074B8\* 07807 07A66 09ED9

- The above 7 code positions are marked in K column of IICORE.

- The other 7 code positions were in K-LGR v0.2, but NOT in K-LGR v0.3.

|  |  |  |  |
| --- | --- | --- | --- |
| U+5AB4 | 媴 | G, H, T |  |
| U+741F | 琟 | G, H, T, KP |  |
| U+7494 | 璔 | G, H, T, KP | in K-LGR v0.2, but not in K-LGR v0.3 |
| U+74A4 | 璤 | G, H, T |  |
| U+74B6 | 璶 | G, H, T, KP | in K-LGR v0.2, but not in K-LGR v0.3 |
| U+74B8 | 璸 | G, H, T, KP |  |
| U+7682 | 皂 | G, H, T | in K-LGR v0.2, but not in K-LGR v0.3 |
| U+7807 | 砇 | G, T, KP |  |
| U+7A66 | 穦 | G, T, KP |  |
| U+7B30 | 笰 | G, H, T, J, KP | in K-LGR v0.2, but not in K-LGR v0.3 |
| U+7FF6 | 翶 | G, T | in K-LGR v0.2, but not in K-LGR v0.3 |
| U+8CEB | 賫 | G, H, T, KP | in K-LGR v0.2, but not in K-LGR v0.3 |
| U+9667 | 陧 | G | in K-LGR v0.2, but not in K-LGR v0.3 |
| U+9ED9 | 黙 | G, T, J |  |

If we consider basic Hanja at this point, it may be appropriate to limit the repertoire to K0 augmented by any characters in MSR to provide full variant transitivity.

**Conclusion**

It would be sensible to have a KLGR containing K0 (and only K0) if the IP can get reasonable evidence that a Hanja set with such content is being considered as a second level domain repertoire in Korea.

==> We've started a discussion on whether to include Hanja characters in 2nd level domain under .kr/.한국 and hope to deliver Korean internet community decision soon.

Moreover, a final KLGR would be expected to contain the Hangul syllables.

==> K-LGR v0.3 (2015.08.13.) includes 11,172 Hangeul (Hangul) syllables.

\* \* \*