

Registries Stakeholder Group Statement

Issue: **Reference Japanese Label Generation Rules (LGR) for the Second Level.**

Date statement submitted: **March 31, 2017**

Reference URL:

<https://www.icann.org/public-comments/japanese-lgr-second-level-2017-01-27-en>

Background

Label Generation Rules for Japanese:

<https://www.icann.org/sites/default/files/packages/lgr/lgr-second-level-japanese-30aug16-en.html>

The Registries Stakeholder Group (RySG) welcomes the opportunity to comment on the proposed reference Japanese Label Generation Rules (LGR) for the Second Level.

The proposed Japanese LGR introduces visually similar variant blocking for two variant pairs:

“ There are two variant pairs:

U+30FC (一) : U+4E00 (一) and U+30FD (\) : U+4E36 (\) that correspond to code points that have confusable glyphs which may be rendered without distinction. These are single stroke characters. In addition, they resemble punctuation characters. All variant mappings have been assigned the variant type "blocked" (see [RFC7940]). “

For over a decade, the Japanese language IDN Registration Service for the Second Level under the .jp ccTLD has demonstrated operational usage for JPAN without variants while ICANN has indicated that it has no operational experience with Japanese language IDNs on the second level. **It is the RySG's opinion that ICANN should not create second level LGRs that are inconsistent with real local experience developed by local expertise, and should continue to consult with relevant language and operational communities to refine them.**

Moreover, the criteria used for identifying the two variant sets are not included in any RFC or based on any documented principles and introducing these two variant sets opens up a slippery slope to have to deal with many other similar characters in Japanese, such as カカ, オオ, 口口, 八八, トト, 二二, or 工工.

The RySG is concerned about the introduction of variant blocking based on visually similar strings in LGRs. While in some cases blocking might seem appropriate, as a general rule such a decision should remain up to the registry.

The fact that two glyphs are visually confusable does not automatically mean that they should be treated as variants. Conversely, visual distinctiveness does not rule out equivalence. In Latin script for instance, the letter O and the digit 0 are confusingly similar, but they are not variants. The same

is true for the upper case i (I) with respect to lowercase L (l). The Latin script also offers many examples of equivalence despite dissimilarity: lowercase and uppercase Latin characters are often highly dissimilar but are perceived as equivalent, so strongly so, in fact, that the DNS protocol was designed from the start to provide case folding in ASCII. Latin letters with and without diacritic marks are also often treated as equivalent (as in Etat vs État).

Cases of confusingly similar non-equivalents like

U+30AB 力 KATAKANA LETTER KA

U+529B 力 Han ("power") (See <https://en.wiktionary.org/wiki/%E5%8A%9B>)

can by no means be treated as variants. In some cases blocking can be appropriate, but the decision should be up to the registry. Automatic activation would be inappropriate.

In cases of equivalence despite dissimilarity like

U+56FD 國 ("country")

U+570B 國 ("country")

the registry may want to consider actions like blocking or even automatic activation. The decision on such policies should be up to the registry.

(This is actually done on TLD level for the .中国 / 中國 IDN ccTLDs of China, see

<https://www.iana.org/domains/root/db/xn--fiqs8s.html> and <https://www.iana.org/domains/root/db/xn--fiqz9s.html>).

We believe that given the above issues, the LGR requires further discussion before being re-submitted to the broader community to the review. We suggest that ICANN continue to consult with the relevant language community as well as registry operators currently offering Japanese language registrations at the second level as a next step in this process.