



## Re: Reference Japanese Label Generation Rules (LGR) for the Second Level

Via Electronic Mail:

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

March 31, 2017

Charleston Road Registry d/b/a Google Registry (Google Registry) believes that reference Label Generation Rules (LGRs) can be a helpful resource for registry operators, but expresses reservations about the Japanese LGR<sup>1</sup> as proposed. Google Inc., the parent company to Google Registry, employs several experts in Unicode and internationalization; we consulted a number of these experts in the development of these comments.

Our concerns arise primarily from cases where the LGR diverges from existing IDN tables for Google Registry's Japanese IDN tld `.みんな` and for other Japanese IDN-supporting TLDs or from Google Chrome's IDN policy in ways that create a potential for variant conflicts or user confusion.<sup>2</sup> Further, while not detailed below, we are also generally concerned about the possibility that the proposed variant sets in the LGR could result in blocked variant labels that have a distinct and legitimate semantic meaning from the registered label.

We have the following specific concerns about the strictness of the proposed Japanese LGRs:

### **U+30FC (ー) should follow only Hiragana or Katakana**

The LGR currently proposes a contextual rule restricting U+30FC to always follow another Japanese codepoint (Han, Hiragana, Katakana, or U+30FC itself). We believe that it would be more conservative to restrict U+30FC to only follow Hiragana or Katakana codepoints since, in typical word usage (as opposed to stylistic or typographical usage), it would follow only Hiragana or Katakana. This rule also reduces the overall potential for variant conflicts with U+4E00 (Variant Set 1 in the LGR) by blocking usage of U+30FC following Han codepoints, where U+4E00 is more likely to occur. This revised restriction would match Google Registry's current practice for `.みんな`; current practice for Japan Registry Services (JPRS), the backend provider for three TLDs that allow Japanese IDN registrations `.sakura`, `.ntt`, and its `.brand` TLD `.jprs`;<sup>3</sup> and intended practice for Google Chrome (current practice for Google Chrome is even stricter but this is likely to be changed).

### **Other potentially confusable characters**

We have identified several other potentially confusable pairs of characters currently within the LGR repertoire and note that there may be more that we have not yet identified. Two of these

---

<sup>1</sup> <https://www.icann.org/sites/default/files/packages/lgr/lgr-second-level-japanese-30aug16-en.html>

<sup>2</sup> <https://www.chromium.org/developers/design-documents/idn-in-google-chrome>

<sup>3</sup> [https://www.iana.org/domains/idn-tables/tables/jprs\\_ja\\_1.0.txt](https://www.iana.org/domains/idn-tables/tables/jprs_ja_1.0.txt)

pairs are listed in the latest version of UTS #39 confusables.txt,<sup>4</sup> while two of them are not (which may be an unintentional omission). The pairs are as follows:

1. U+3078 (へ) is confusable with U+30D8 (へ)
2. U+3079 (ゑ) is confusable with U+30D9 (ゑ)
3. U+307A (え) is confusable with U+30DA (ゑ)
4. U+30CB (二) is confusable with U+4E8C (二)

The first three pairs represent Hiragana (U+307X) and Katakana (U+30DX) versions of the letters HE, BE, and PE, and they are represented by very similar glyphs in many fonts. Chrome currently places a restriction on the appearance of the U+307X codepoints in a label that is otherwise entirely Katakana, and likewise restricts the appearance of the U+30DX codepoints in a label that is otherwise entirely Hiragana. We do not necessarily endorse this specific restriction, which may be either overbroad or incomplete, but believe it highlights a set of concerns that a reference LGR may need to address.

The fourth pair represents the Katakana letter NI (U+30CB) versus the CJK UNIFIED IDEOGRAPH-4E8C meaning "two". While in many fonts these are distinguishable, in some they are more easily confused, and may also represent a confusability risk that should be considered.

### **Conclusion**

We request further community deliberation to arrive at a consensus on the desired rules. We welcome further engagement on the issues raised herein, as well as the general concern over whether variant sets in the LGR could result in blocking legitimate, distinct variant labels.

Sincerely,



Nick Felt  
Software Engineer, Google Registry

---

<sup>4</sup> <http://www.unicode.org/Public/security/latest/confusables.txt>