

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

**Publication Date:** 28 April 2017

**Prepared By:** Sarmad Hussain

## Public Comment Proceeding

Open Date: 27 January 2017

Close Date: 31 March 2017

Staff Report Due Date: 28 April 2017

## Important Information Links

[Announcement](#)

[Public Comment Proceeding](#)

[View Comments Submitted](#)

**Staff Contact:** Sarmad Hussain

**Email:** sarmad.hussain@icann.org

## Section I: General Overview and Next Steps

ICANN has developed reference Internationalized Domain Name (IDN) tables in [machine readable format](#), called [Reference Label Generation Rules \(LGRs\) for the Second Level](#), to facilitate Pre-Delegation Testing (PDT) and the Registry Service Evaluation Process (RSEP) for the relevant gTLDs. During the [initial public comment](#) on the release of these reference IDN tables, there were specific comments received by the Japanese community.

Following these comments, ICANN organization worked with the Japanese community and had released the updated version of the reference Japanese LGR for the second level.

ICANN organization has received feedback through the second public comment period, suggesting further discussion with the language community and the registry operators. ICANN organization will continue to work with the relevant stakeholders to address the comments received in the current round to finalize the reference LGR for Japanese language.

## Section II: Contributors

*At the time this report was prepared, a total of [number] (n) community submissions had been posted to the forum. The contributors, both individuals and organizations/groups, are listed below in chronological order by posting date with initials noted. To the extent that quotations are used in the foregoing narrative (Section III), such citations will reference the contributor's initials.*

### Organizations and Groups:

| Name                         | Submitted by        | Initials |
|------------------------------|---------------------|----------|
| Google Registry              | Nick Felt           | GR       |
| Registries Stakeholder Group | Stéphane Van Gelder | RySG     |

### Individuals:

| Name | Affiliation (if provided) | Initials |
|------|---------------------------|----------|
|      |                           |          |

### Section III: Summary of Comments

*General Disclaimer:* This section intends to summarize broadly and comprehensively the comments submitted to this public comment proceeding but does not address every specific position stated by each contributor. The preparer recommends that readers interested in specific aspects of any of the summarized comments, or the full context of others, refer directly to the specific contributions at the link referenced above (View Comments Submitted).

The comment from GR supports the reference LGRs as a helpful resource for registry operators. The comment provides specific feedback on the following cases, where the proposed reference LGR for Japanese diverges from the existing IDN tables implemented by GR “in ways that create a potential for variant conflicts or user confusion”. First, it is suggested that the current rule restricting U+30FC to always follow another Japanese code point (Han, Hiragana, Katakana, or U+30FC itself) be further restricted, where U+30FC must only follow Hiragana or Katakana code points. Second, additional cases are identified as candidates for variant code points:

U+3078 (ゝ) is confusable with U+30D8 (ゝ)  
U+3079 (ゞ) is confusable with U+30D9 (ゞ)  
U+307A (ゝ) is confusable with U+30DA (ゝ)  
U+30CB (ニ) is confusable with U+4E8C (ニ)

A concern is also raised that variant sets in the LGR could result in blocking legitimate, distinct variant labels.

RySG suggests that ICANN organization should keep consulting with relevant operational and language communities while developing the reference second level Japanese LGR to keep it consistent with operational experience. It is pointed out that the two variant code point sets, U+30FC (一) with U+4E00 (一) and U+30FD (ゝ) with U+4E36 (ゝ), are not considered as variant code points in the current practice. Further, it is suggested that there are no criteria set by any RFC or documented principles for defining variant sets and the analysis may lead to a slippery slope for other cases like カカ, オオ, ココ, ハハ, トト, ニニ, or エエ. RySG states concern on variant blocking due to string similarity, and suggests that this should be determined by the registry. RySG advises to have further discussion, especially to have consultation with the relevant language community and relevant registry operators.

### Section IV: Analysis of Comments

*General Disclaimer:* This section intends to provide an analysis and evaluation of the comments submitted along with explanations regarding the basis for any recommendations provided within the analysis.

The two comments received differ in their analysis, where GR advocates for a stricter analysis of variants and rules to prevent user confusion and RySG suggests an approach which allows registries to make such decision based on operational practice.

The current reference LGR has been developed in consultation with the Japanese language community. It has been designed so that similarity cases should be handled on the registry

level. This includes the code points identified as part of GR's comment as well as in the RySG comments, which states that there are "many other similar characters in Japanese, such as カカ, オオ, ロロ, ハハ, トト, ニニ, or エエ."

However, based on the input from GR and to lay out the examples of identical or near-identical code points more clearly, the description section of the reference LGR can be updated to explicitly include such variant sets. Though this explicit mention in the description section will not be a formal variant relation, it will allow the registries to note these and take the appropriate action of their choice to mitigate any impact on users. This change will be discussed with the stakeholders, especially with the Japanese language community for finalization.

As a general principle, in cases of confusability with code points that closely resemble punctuation, or punctuation-like code point, an increased scrutiny is warranted from a security point of view (see RFC 6912). In this context, it is also noted that one of the code points, U+30FC (—), is subject of another comment received, asking for stricter restrictions in its allowable placement. As U+30FC (—), U+4E00 (一) and U+30FD (ゝ), U+4E36 (丶) are single stroke glyphs with resemblance to punctuation, these code points require such special attention and, therefore, have been defined as variant code points.

For these two pairs that fit the criteria, as the meanings of these code points are unrelated, there appears no prima facie case requiring them to be available to the same applicant. Therefore, these are presented here as "blocked". This type assignment should be understood as a default, which a registry could change as needed.

Related to the GR's concern that "variant sets in the LGR could result in blocking legitimate, distinct variant labels", it should be noted that variant set is the preferred mechanism which prevents visually indistinguishably labels to be delegated simultaneously to different registrants, therefore reducing the risk of confusion also mentioned by GR. Furthermore, the variant mechanism only affects otherwise identical labels, still leaving other label alternatives open for registration.

Regarding the placement of U+30FC (—) to follow only Hiragana or Katakana, during the previous round of consultation with the Japanese community, it was revealed that there exist examples for usage of U+30FC that cannot be accommodated by the proposed restriction. Therefore, the question must be asked whether the requested restriction is too strict. As noted in GR's comment, the motivation for the tighter restriction is in part a reduction in possible confusion with U+4E00. The reference LGR handles this issue robustly with the variant mapping, which would allow relaxing the restriction on placement without compromising security. However, a registry may always choose a more conservative handling of U+30FC, for example by restricting U+30FC to only follow Hiragana or Katakana.

The current solution presented in the reference Japanese LGR had been agreed by the Japanese language community. ICANN organization will update the reference LGRs, as discussed, and will work with the Japanese community, GR and RySG to finalize the reference LGR for Japanese language.