# Root Zone LGRおよび日本語生成パネル（JGP）について

- 検討状況報告 & ご意見募集 -

2021年5月13日

日本語生成パネル チェア
堀田 博文 〈hotta@jprs.co.jp〉

# IDN TLDのルール作成に向けて

- ルートゾーンには様々な言語・scriptのラベルが存在
- いくつかの言語・scriptには同一視すべき(/した方がいい)文字同士が存在(「異体字」と呼ぶこととする)
  - 例1:字形は異なるが読み・意味が同じ文字(「応」と「應」、「学」と「學」)
    例2:視覚的に同一(もしくは酷似)な文字(漢字「二」と片仮名「ニ」?)

TLDラベルとしての文字列の適切さや同一視すべき
ラベル群を機械的・統一的に判断できるように

- 様々な言語・scriptのTLDラベルで使える文字の範囲および異体字、異体字ラベルの使用可否を標準ルール(Root Zone Label Generation Rules; RZ-LGR)として決めておく
- scriptを共有する言語同士は標準ルール作成にて調整が必要
  - 例 : CJK(中国語/日本語/韓国語)は漢字を共有
    - 中国語生成パネル(CGP)、日本語生成パネル(JGP)、韓国語生成パネル(KGP)が調整

# 日本語LGR提案v0.15 (2020.10.12版) の概要

2020年12月3日のICANN報告会で意見照会させていただいた版

- 使用可能な文字の範囲(レパートリー)
  - JIS X 0208:2012の第一水準・第二水準の範囲（漢字、平仮名、片仮名、漢字および仮名に準ずる一部記号文字からなる約6,300文字）
- 異体字の定義
  - 日本語独自の異体字
    - 字形は異なるが読み・意味が同じ文字（「国」と「國」と「圀」）は異体字としない
    - 視覚的に同一（もしくは酷似）な文字同士は最小限のものだけ異体字とする
      - Unicodeコンソーシアム公開の「錯視が起こりやすい文字同士リスト」に基づく10組
      - 酷似する1ストロークの記号2組: U+ [30FC(ー) 4E00(一)]、[30FD(ヽ) 4E36(丶)]
  - 中国語/韓国語LGRで定義された異体字
    - それらを日本語LGRに取り入れる（日本語LGRでも異体字として扱う）
- ラベル文字列全体に関するルールの定義
  - 申請されたラベル(申請ラベル)に加え、そのラベルの異体字ラベルのうち常用漢字だけからなるものはTLDとして使用可能とする

# 本語LGR提案v0.15 (2020.10.12版) の概要

2020年12月3日のICANN報告会で意見照会させていただいた版

LGR全体を統括しているパネル(IP)からのコメント(2021.12. 16)

- 使用可能な文字の範囲(レパートリー)
  - JIS X 0208:2012の第一水準・第二水準の範囲(漢字、平仮名、片仮名、漢字および仮名に準ずる一部記号文字からなる約6,300文字)
- 異体字の定義
  - 日本語独自の異体字
    - 字形は異なるが読み・意味が同じ文字(「[illegible]異体字としない
    - 視覚的に同一(もしくは酷似)な文字同士は最[illegible]だけ異体字とする
      - Unicodeコンソーシアム公開の「錯視が起こりやすい文字同士リスト」に基づく10組
      - 酷似する1ストロークの記号2組: U+ [30FC(ー) 4E00(一)][illegible]4E36(丶)]
  - 中国語/韓国語LGRで定義された異体字
    - それらを日本語LGRに取り入れる(日本語LGRでも[illegible]
- ラベル文字列全体に関するルールの定義
  - 申請されたラベル(申請ラベル)に加え、そのラベルの異体字ラベルのうち常用漢字だけからなるものはTLDとして使用可能とする

① 10組以外にないという根拠が欲しい (才と才とかも酷似では？)

② TLDとして使用可能な異体字ラベルがまだ多すぎる

# ①への対処の概要:提案v0.17 (2021.4.6版)

## - 視覚的同一/酷似文字 -

- 権威ある機関(Unicodeコンソーシアム)の「錯視が起こりやすい文字対リストに基づく10組(*)以外に錯視の経験をした文字同士は？」というサーベイを実施した
  - 176人にアンケート配布
  - 73人が回答 (多様：技術職～事務職、年齢分布、性別)
  - 結果、当該10対以外の文字同士では、錯視があったとしても、その文字同士を錯視したという回答者は全回答者の3%以下であった
    - たとえば、「オ」と「才」を錯視した経験ありと回答したのは73人中2人
- 視覚的同一/酷似文字は当該10組とするのが適切と判断

(*) UNICODEコンソーシアムの「錯視が起こりやすい文字対リスト」に基づく10組

- へ へ
- ペ ペ
- カ 力
- ハ 八
- ロ 口
- エ 工
- ベ ベ
- ニ 二
- ト ト
- タ タ

# ②への対処の概要:提案v0.17 (2021.4.6版)

## - TLDとして利用可能なラベルの削減 -

- 常用漢字同士が異体字である場合、その漢字もしくはその異体字が申請ラベルに入っていると、異体字ラベルが多くなる
  - たとえば、「乾」「干」「幹」及び「復」「複」「覆」はそれぞれ中国語LGRで異体字と定義されており、日本語LGRではすべて常用漢字
    - 「被覆式乾電池復刻版」が申請ラベルの場合、それ自身も含め3x3x3=27個の常用漢字だけからなる異体字ラベルを持つ
    - 論理的には、これら文字を多く含む長いラベルでは、その異体字ラベル数は何十桁にもなりうる
  - 登録済みの日本語JPドメイン名を調査したところ、自身も含め2個以上の常用漢字を異体字として持つ文字を4個以上持つラベルはなかった。
- 次のルールが適切と判断
  - 申請ラベル内に、常用漢字を異体字として持つ文字が
    ・4個以上であれば、申請ラベルのみ使用可能
    ・3個以下であれば、申請ラベルに加え、常用漢字だけからなる異体字ラベルすべて使用可能

# ②への対処の概要:提案v0.17 (2021.4.6版)

## - TLDとして利用可能なラベルの削減 -

LGR全体を統括しているパネル(IP)からの更なるコメント(2021.5.6)

- 常用漢字同士が異体字である場合、その漢字もしくはその異体字が申請ラベルに入っていると、異体字ラベルが多くなる
  - たとえば、「乾」「干」「幹」及び「復」「複」「覆」はそれぞれ中国語LGRで異体字と定義されており、日本語LGRではすべて常用漢字
    - 「被覆式乾電池復刻版」が申請ラベルの場合、それ自身も含め3x3x3=27個の常用漢字だけからなる異体字ラベルを持つ
    - 論理的には、これら文字を多く含む長いラベルでは、その異体字ラベル数は何十桁にもなりうる
  - 登録済みの日本語JPドメイン名を調査したところ、自身も含め2個以上の常用漢字を異体字として持つ文字を4個以上持つラベ
- 次のルールが適切と判断
  - 申請ラベル内に、常用漢字を異体字として持つ文字が
    ・4個以上であれば、申請ラベルのみ使用可能
    ・3個以下であれば、申請ラベルに加え、常用漢字だけからなる異体字ラベルすべて使用可能

②'
TLDとして使用可能な異体字ラベルがまだ多すぎる。異体字ラベル使用は本当に必須？

# ②’への対処の概要

## - TLDとして使用可能なラベルのさらなる削減 -

- [対案] 異体字ラベルが存在する場合でも、申請ラベルのみを使用可能とする
  - 例：「慶應義塾大學」や「慶応義塾大学」を申請した場合、使用可能なのは下の〇のもの

| | [v0.17] | | [対案] | |
|---|---|---|---|---|
| ↓異体字ラベル | 慶應義塾大學 | 慶応義塾大学 | 慶應義塾大學 | 慶応義塾大学 |
| • 慶應義塾大學 | 〇 | × | 〇 | × |
| • 慶応義塾大学 | 〇 | 〇 | × | 〇 |
| • 慶應義塾大学 | × | × | × | × |
| • 慶応義塾大學 | × | × | × | × |

- 本制約が日本語TLDの実用性を大きく損ねることがないと考え、この[対案]をICANNに提案する方針

# ご質問?
# ご意見?

意見受付) info@j-gp.jp

JGPのweb) https://j-gp.jp/