

# Root Zone LGRおよび日本語 生成パネル(JGP)について

- ご意見・コメント募集 -

2019年12月13日

日本語生成パネル チェア  
堀田 博文 <hotta@jprs.co.jp>

# 背景

- ルートゾーンには様々な言語・scriptのラベルが混在
- 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(「異体字」と呼ぶ)が存在
  - 例1:字形は異なるが読み・意味が同じ文字(「国」と「國」と「圀」)
  - 例2:視覚的に同一(もしくは酷似)な文字(漢字「二」とカナ「ニ」?)



ラベルの適切さや同一視すべき文字をもつラベルをできるだけ自動的に判断できるように

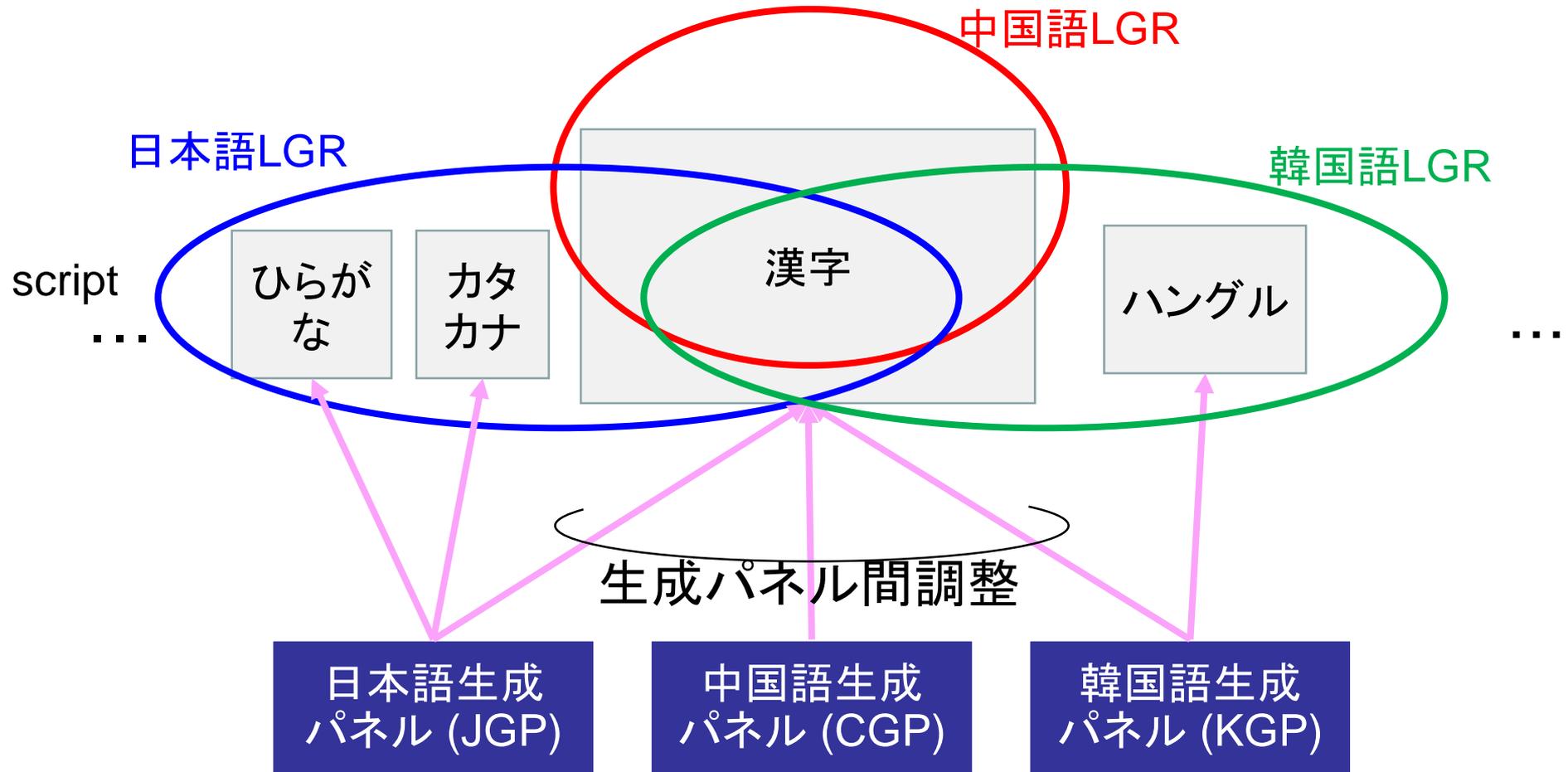
- ルートゾーン用に、様々な言語・scriptで使える文字の範囲およびその異体字を統一的に取り扱うルール(Root zone Label Generation Rules; RootLGR)を決めておく
- scriptを共有する言語同士は統ルールルの作成において調整が必要
  - 例: CJK(中国語/日本語/韓国語)は漢字を共有

# RootLGRとは

- 創設を申請されたIDN TLDラベルをルートゾーンの中でどのように取り扱うかを規定するルール
- 各言語やscriptに対する次の4つのルールからなる
  - 使用可能な文字の範囲(レパートリー)
    - たとえば、JIS第一水準と第二水準の文字?
  - 異体字の定義
    - たとえば、「国」と「國」と「圀」は異体字?
  - どの異体字をTLD文字列内で使えるかの定義
    - 「国」と「國」は異体字であり、「国」が入ったラベルをTLDとして使用する場合、その「国」を「國」に置き換えたラベルは使える?
  - TLD文字列全体に関するルールの定義
    - たとえば、中国語の簡体字と繁体字は1ラベル内では同時使用不可?

# CJKの言語用LGRとscriptの関係

(CJK : Chinese-Japanese-Korean)



各生成パネルは、各言語コミュニティ内での調整を実施

# RootLGR作成の枠組み

- ICANNでRootLGRの作成を開始
  - ICANN会合でRootLGRの解説と実装の呼びかけ
    - 2013年11月以降、各ICANN会合で情報共有会合開催
    - 非ASCIIを使用する主要な国(地域)へアウトリーチ活動
      - 2014年5月8日にJPNICオフィスでLGRのワークショップを開催
- ICANNが2013年10月に各言語・スクリプトのLGR作成を支援し、全LGRを統合して1つのRootLGRを作成するチーム (IP : Integration Panel : 統合パネル) を設立
- 2014年以降、各言語のコミュニティがLGRを作成するチーム (GP : Generation Panel : 生成パネル) を作り、検討開始
  - 日本語生成パネル(JGP)は、2014年より活動開始
  - IPが、各GPとの情報交換開始

# 日本語生成パネル(JGP)メンバー

- メンバー (2019年12月時点)
  - 堀田博文 議長
    - Registry/Registrar (esp. Policy/business aspects)
  - 前村昌紀 副議長
    - Policy (esp. Internet governance and domain name in general)
  - 後藤滋樹
    - Community (esp. Internet technology in general)
  - 小西和憲
    - Community (esp. Internet technology in general)
  - 久保次三
    - Community (esp. Trademarks and DRP)
  - 村上嘉隆
    - Registry/Registrar (esp. Trademarks and gTLD market)
  - 田代秀一
    - Language expert (esp. Character codes)
  - 米谷嘉朗
    - Registry/Registrar (esp. technical aspects of IDN and LGR)

# 日本語LGR定義の方向性

赤字は、JGPの当初の方針からの変更点

- 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
    - スライド2の例1(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
    - スライド2の例2(視覚的に同一(もしくは酷似)な文字(「ニ」と「ニ」))は最小限のものだけ異体字とする(評価中)
  - 中国語/韓国語LGRの異体字
    - 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<sub>の定義</sub>
  - ルートゾーンの肥大化を防ぐため、異体字のうち非常用漢字の使用可能性を抑制

# 視覚的同一/酷似文字の異体字化

- 当初、CGP/JGP/KGPとも視覚的同一/酷似文字は異体字としない方針
    - 視覚的同一/酷似文字は、LGRでは異体字とせず、必要に応じブラウザ等で対処すべきとの意見書をCJK合同でICANNに提出
    - しかし、利用者の混乱やセキュリティ懸念を低減するため各言語にて異体字と定義すべきとのIP/ICANNの見解は変わらず
  - 日本語LGRでは、次の条件でフィールド実験を行い、視覚的同一/酷似文字を判断することとした（近くフィールド実験実施予定）
    - 被験者はランダムに選ぶ（恣意性を持たない）こと
    - 被験者数は30人以上とすること
    - フォントは3種以上使うこと
- CGP, KGPも同条件でフィールド実験することとなった(2019年11月)

# 異体字である非常用漢字使用の抑制

- TLDを申請し承認を得たTLD申請者は、
  - JGPは当初、申請文字列の各文字を異体字に交換してできる文字列(「異体字文字列」と呼ぶ)すべてを使用(委任)可能とする方針であった
  - しかし、IPの意見を受け、ルートゾーンの肥大化を防ぐため、申請文字列以外で使用可能な異体字文字列は常用漢字のみの文字列とした
- 例：
  - 次の3文字は異体字同士とする
    - 国 常用漢字
    - 國 常用漢字
    - 圀 常用漢字でない
  - 次の2文字は異体字同士とする
    - 云 常用漢字でない
    - 雲 常用漢字
  - 「.国云」を申請し承認を得たTLD申請者は、異体字の全組合せ(6つ)に対する権利を持つが、実際にTLDとして使用可能なのは3つとなる
    - .国云 ○ 申請文字列
    - .国雲 ○ 常用＋常用
    - .國云 × 常用＋非常用
    - .國雲 ○ 常用＋常用
    - .圀云 × 非常用＋非常用
    - .圀雲 × 非常用＋常用

# Opinions? Questions?