

# IDN TLD文字列のルール 策定活動について

第9回 IGCJ

2015年9月29日

日本語生成パネル

チェア 堀田 博文<hotta@jprs.co.jp>

# 背景

- 2012年のICANN新gTLDプログラム
  - 1930件の申請 (そのうち75件が漢字を含むIDN)
  - TLDラベル間の混乱を避けるため、文字列の類似性を含む混乱の危険性を、申請された新gTLD一つずつに対しパネル(人間)が判断
- IDN TLDはさまざまな言語・scriptで申請されるため、ルートゾーンにはさまざまな言語・scriptのラベルが混在
- いくつかの言語・scriptには
  - 異体字(字形・コードポイントは異なるが同じ読み・意味の文字)が存在
  - 異体字同士を同一視すべき場合が存在



文字列の適切さや同一視すべき異体字をもつ文字列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、さまざまな言語・scriptの文字集合および異体字を統一的に取り扱うルール(RootLGR : Root zone Label Generation Rules)を決めておく

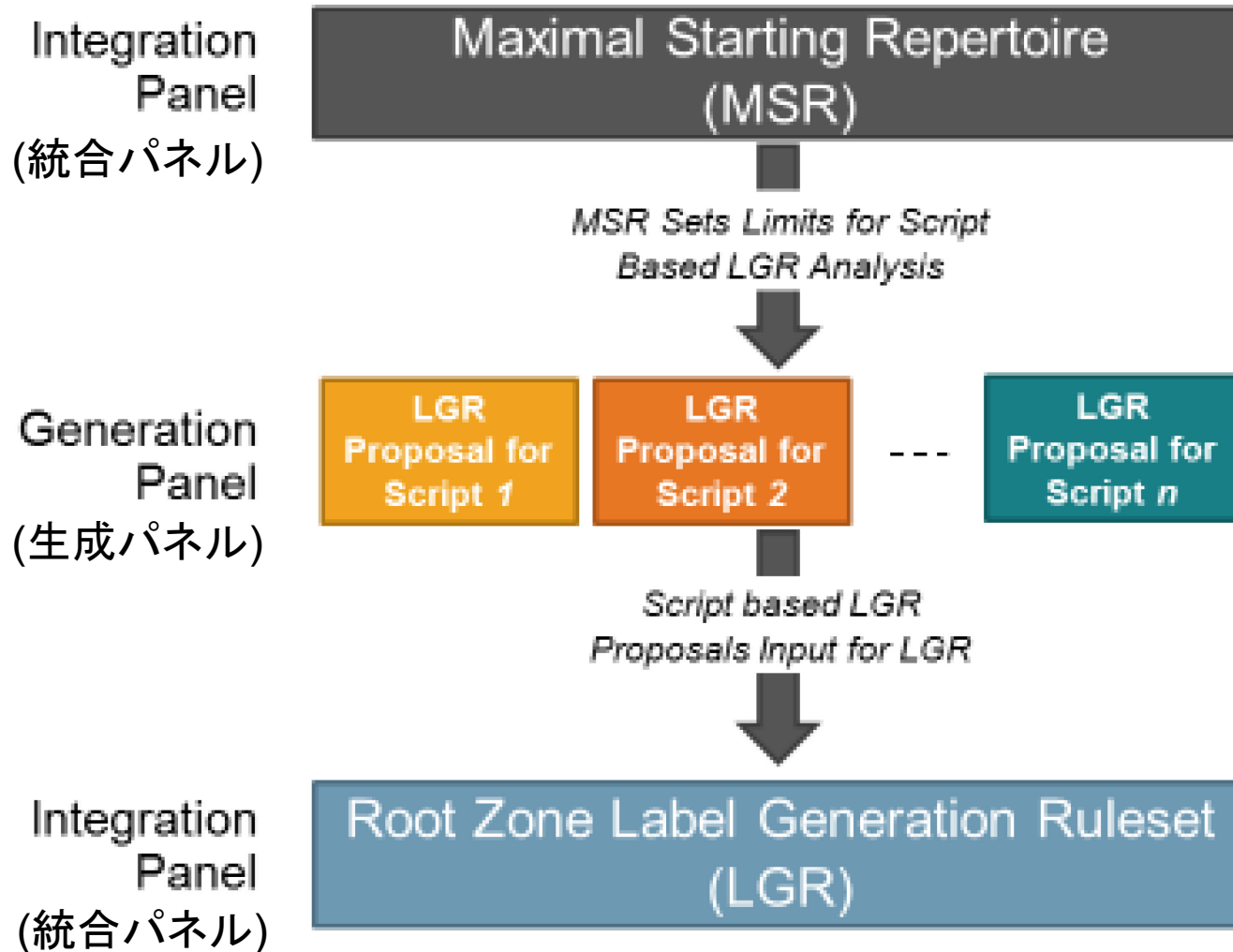
# RootLGRとは

- 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・scriptに対する次の4種のルールからなる
  - 使用可能な文字の集合
    - 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」は異体字
  - どの異体字を文字列内で使えるかの定義
    - たとえば、「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用不可
  - ラベル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

# 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
  - 各scriptで使用可能な文字の集合の最大枠(MSR)を定義
- 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
- IPが、各GPとの情報交換開始

# RootLGR開発プロセス

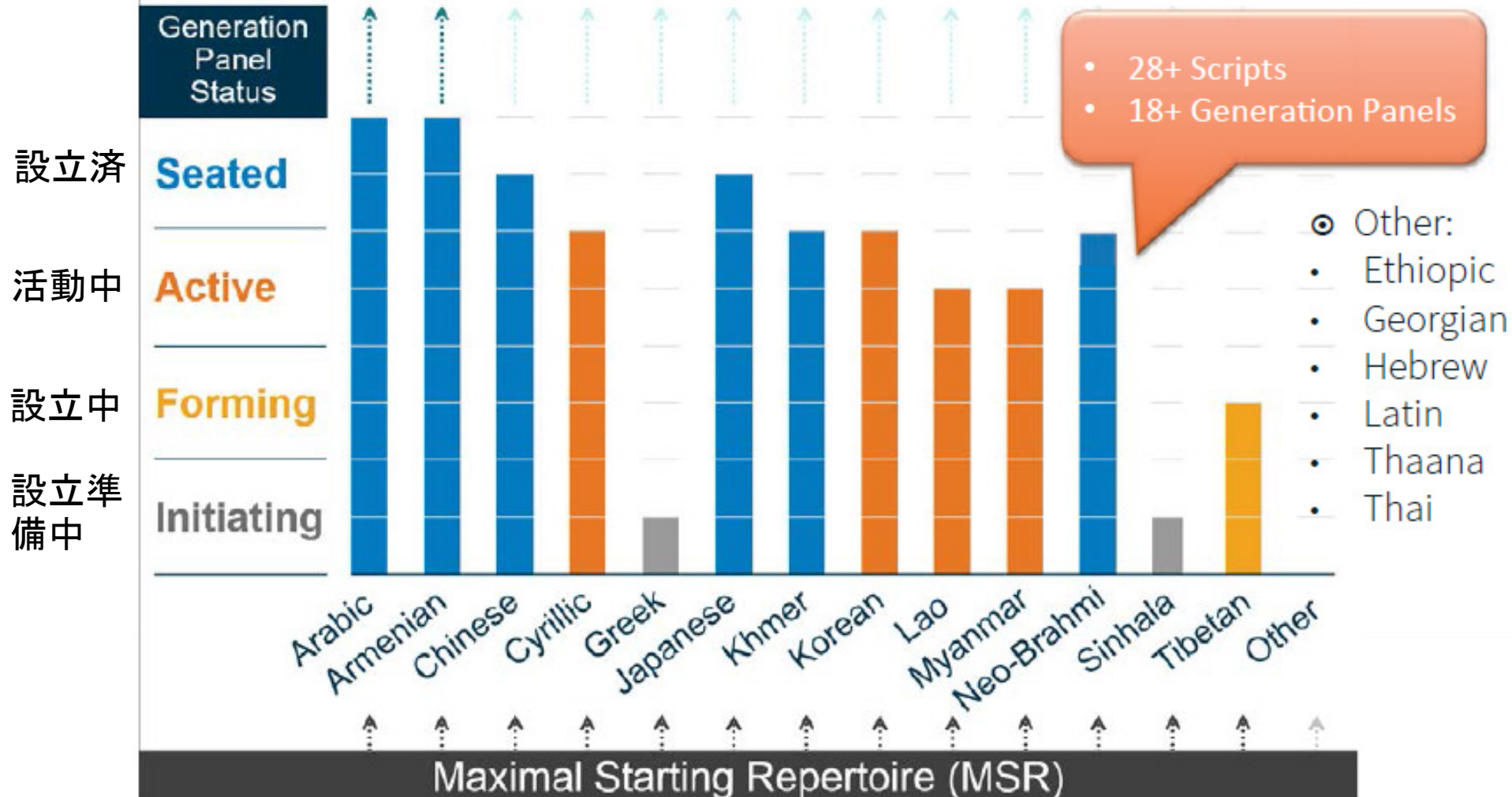


# 各国の言語生成パネルの状況

2015年6月24日ICANN発表

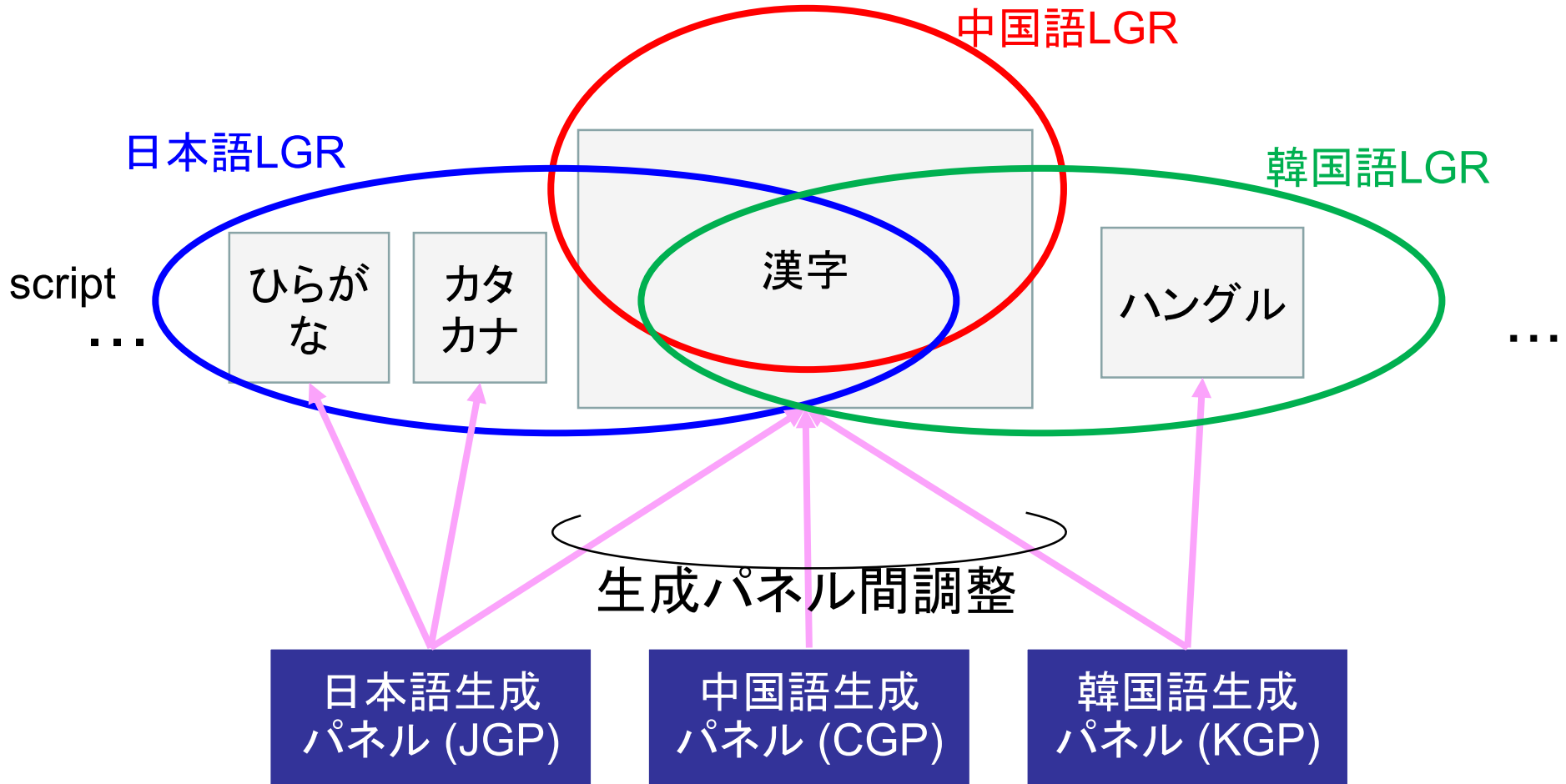
ICANN 53

## Label Generation Rules (LGR)



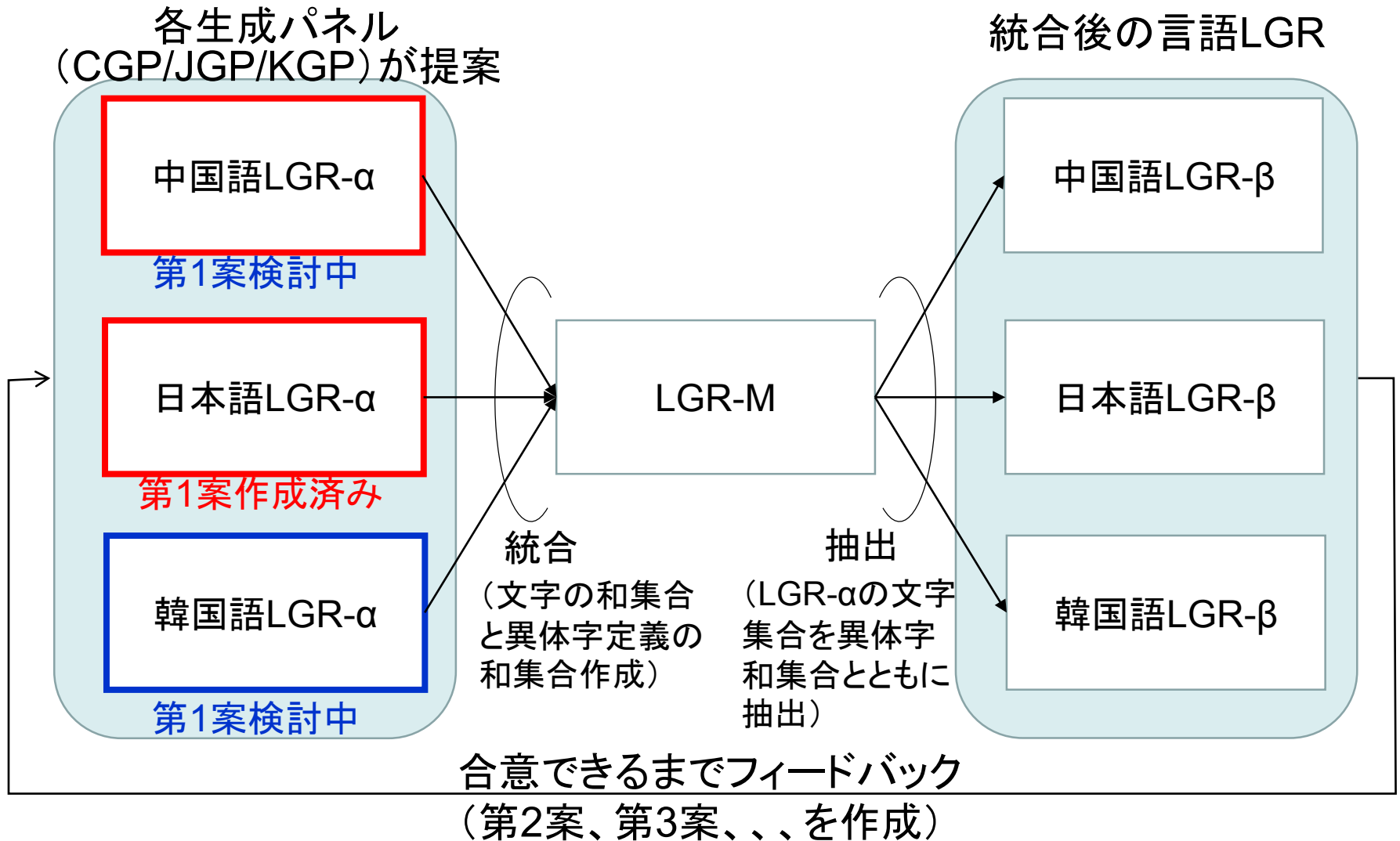
# CJKの言語用LGRとscriptの関係

(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

# 漢字に関する生成パネル間の協調





# 日本語生成パネルの状況

- 日本語LGRを作るチーム
  - 日本語生成パネルJGP (Japanese Generation Panel)の前身が2014年8月より検討開始
  - 2015年3月10日に、ICANNが正式にJGP設立を承認
- 漢字を共通に使うCJK(中国語、日本語、韓国語)間の調整
  - 漢字(の異体字)の取り扱いをCJKで合意した上で各GPからIPに提案することをIPがCJKに要望
  - 各生成パネル(CGP、JGP、KGP)が、協力してドメイン名登録者・利用者に混乱を与えない無矛盾の言語LGRを作成するための3者調整委員会を持つことを合意
- CGP/JGP/KGPの検討状況
  - CJKのccTLDレジストリを中心に言語LGRの検討中
  - CJKのccTLDレジストリを中心にCJKでの共同検討を開始

# 日本語LGR- $\alpha$ 定義の方向性

- 使用可能な文字の集合
  - 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6,358文字)
- 異体字の定義
  - 第1案では定義しない
    - 中国語/韓国語LGR- $\alpha$ での異体字を取り込んだ結果(日本語LGR- $\beta$ )を評価し、必要に応じ第2案では異体字を定義
  - 第1案ではあらゆる文字の組合せをラベルで使用可能
    - 中国語/韓国語LGR- $\alpha$ での異体字を取り込んだ結果(日本語LGR- $\beta$ )を評価し、必要に応じ第2案では異体字とその使用可能性を定義
- ラベル文字列全体に関するルールの定義
  - 定義しない
    - どんな文字列も禁止しない

# CJK間調整状況

## • 会合

- 5月15日(金)～16日(土)にソウルにてCJK会合を開催
- 6月21日(日)～25日(木)のICANN会合中に4度のCJK会合を開催
- 10月18日(日)～22日(木)のICANN会合中にさらにCJK会合開催予定

## • 結果

- 各GPからLGR- $\alpha$ (文字集合と異体字定義)の漢字部分の第1案が出された
  - Cは約19,000文字、約3,000組の異体字グループ
  - Jは約6,000文字、異体字グループなし
  - Kは約5,000文字、37組の異体字グループ
- CJK統合アルゴリズム(スライド8参照)を詳細合意
- Allocatable(割当可能)なラベル数を削減する方向で今後も継続検討

## • 完成スケジュール

- 韓国のLGR- $\alpha$ の第1案確定が2015年12月予定との表明があり、日本国内での広いコメント募集及びCJKのRootLGR完成は2016年以降となる予定
- RootLGR完成時期は2016年末までが望ましいが、ボトムアップ活動なので確定的な期限を設定していない、というのがICANNのスタンス

# JGPの今後の活動予定

- 主な検討課題

- allocatable(割当可能)なラベル数を削減する方向で継続検討
  - ラベルとはドメイン名の「.」で区切られた部分。

例: 中国語LGR- $\alpha$ の現在の案を採用した場合、日本語LGR- $\beta$ における「.国際会議」の異体字ラベルは32種類

国: 国、國、囯、国(4通り)

際: 際、际(2通り)

会: 会、會(2通り)

議: 議、议(2通り)

ラベル:  $4 \times 2 \times 2 \times 2 = 32$ (通り)

32種類すべてがTLDとして割当可能とするのではなく、数を減らす方策が必要

- 今後のスケジュール

- ICANNダブリン会合(10月)でのCGP・KGP・IPとの調整に向け、allocatable(割当可能)なラベル数の削減への対応を検討中
- その他、CGP/KGPの検討進展に伴う中間的LGR- $\alpha$ 案を入手しつつ、CGP/KGPとLGR- $\alpha$ 案を調整

# 日本語生成パネルとコミュニティ

- JGPメンバー
  - IDNに詳しい専門家有志7名で検討開始
  - 現メンバーとICANNが必要と判断すればメンバーを追加
    - gTLDと商標実務に詳しいメンバーを追加 (2015年2月)
- コミュニティからのJGP検討への参加
  - JGPからの(およびICANNからの)意見募集に対する意見送付
  - JGP Webサイトで検討状況を確認
    - <http://j-gp.jp>
  - JGPに質問や意見を送信
    - [info@j-gp.jp](mailto:info@j-gp.jp)
  - 各種説明会・意見交換会での議論
    - IGCJ、ICANN報告会、各種分野別会合、...
  - ICANNレベルでの参加
    - ICANNからの意見募集に対する意見送付
    - ICANN提供のメーリングリスト [japaneseGP@icann.org](mailto:japaneseGP@icann.org)への参加
    - IPや他のGPとの意見交換への参加

# 日本語生成パネルメンバー

- メンバー (2015年9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)