

RootLGR における JGP の具体的課題定義(案)

1 背景

1.1 歴史

国際化ドメイン名(Internationalized Domain Name、以降 IDN と記載)は、英語以外の言語を母語とする人々にとって利用しやすいアドレスを提供するものとして、1998 年にアジアを起点として検討が開始された。2000 年にインターネットプロトコルの標準化を行っている IETF(Internet Engineering Task Force)で WG が設立され、2003 年に技術標準が RFC 3490、3491、3492 として制定された¹。

IDN の標準化の際に、中国語コミュニティから IDN の利用には異体字²問題の解決が必要だとの主張が行われた。中国語における異体字問題とは、「中国本土で使われている簡体字は台湾や香港などで使われている繁体字に基本的に対応するものであり、簡体字を使用する地域と繁体字を使用する地域間でのコミュニケーションのためには簡体字のアドレスとそれに対応する繁体字のアドレスは、それらを表現する文字列は違えども同一のアドレスとしてアクセスできる必要がある」というものである。

このような異体字の有無、また異体字がある場合の異体字に関するルールは、言語ごとに異なる。たとえば、漢字を共通的に使用する日本語と中国語においても、日本語では全く別の字である「機」と「机」が中国語では互いに異体字として扱われているというように、異体字のとらえ方は異なっている。しかし、IETF における標準は世界中どこでも共通して使えるプロトコルである必要があるため、中国語の異体字問題をプロトコル内で解決することは、世界共通に作られたプロトコルを地域化(localization)することになり、インターネットの分断を招く恐れがあるとされ、IDN プロトコルの検討対象外とされた。

中国語の異体字問題の解決方法については、漢字を使う言語文化圏の中国・日本・韓国・台湾の NIC(Network Information Center)が中心となって結成した JET(Joint Engineering Team)で検討が行われた。その結果、ある IDN を登録する時に、その IDN 文字列が異体字を持つ文字を 1 つ以上含む場合は、その文字列において異体字を持つ文字それぞれを異体字に置き換えてできる全ての文字列を不可分なまとまり(パッケージ)とし、そのパッケージ全体を同一登録者に紐付けるという運用方式が合意された。この方式は JET Guidelines としてまとめられ、2004 年に RFC 3743 として発行された。また、この考え方は ICANN が策定した IDN Implementation Guidelines にも取り入れられ、IDN の登録サ

¹ 2010 年に、RFC 5890、5891、5892、5893、5894 で更新された。これらの標準を区別するため、2003 年版を IDNA2003、2010 年版を(方式が決定した年から)IDNA2008 と呼んでいる。

² 一般には、同音同義の漢字であって、字体だけが異なり、どの文脈でも交換可能である漢字と定義される。本書では、ある言語において同音同義の漢字であって、Unicode のコードポイントは異なるが同一の文字として取り扱われるべき漢字と定義する。

ービスを行うレジストリは、IDN として使用可能なすべての文字とそれら個々の文字の異体字を定義したテーブルを IANA に登録することが推奨されている。

1.2 JET Guidelines の目的

JET Guidelines は、前述のように、IDN の登録において、その IDN が異体字を持つ文字を 1 つ以上含む場合は、その文字列において異体字を持つ文字それぞれを異体字に置き換えてできる全ての IDN 文字列を不可分なまとまり(パッケージ)とし、そのパッケージ全体を同一登録者に紐付けるという運用方式を推奨するものである。具体的には、JET Guidelines では、登録可能文字(ドメイン名において使用可能な文字)とその文字の異体字を結び付ける IDN テーブルのフォーマットと、パッケージを生成するアルゴリズムを規定している。JET Guidelines で規定する IDN テーブルの構成を以下に示す。

➤ 構成要素

- ✓ その IDN テーブルが適用される言語
- ✓ その IDN テーブルに含まれる文字の出典（出典が複数ある場合はそのリスト）（個々の出典には参照番号を付与）
- ✓ その IDN テーブルのバージョン番号
- ✓ 3 カラムからなる行(各行は登録可能文字 1 文字に対応する)を並べた IDN テーブル本体

➤ IDN テーブル本体の各カラムの意味

第 1 カラム：登録可能文字(Unicode のコードポイント)と、出典の参照番号。

第 2 カラム：第 1 カラムの文字に異体字がある場合は、全異体字のうち IDN での使用が推奨される異体字(群)と、その出典の参照番号。第 1 カラムの文字に異体字がない場合は、第 1 カラムと同じ文字。

第 3 カラム：第 1 カラムの文字の異体字がある場合は、全異体字のうち IDN では使用されず予約される異体字(群)と、その出典の参照番号。第 1 カラムの文字に異体字がない場合は、空。

各カラムはセミコロン (;) で区切られる。第 2 カラムと第 3 カラムにおいては、カラム内に文字が複数ある場合は、各文字はカンマ (,) で区切られる。以下に、例を示す(JET Guidelines の第 4 章 b から引用)。

Language Variant Table for zh-tw # 言語は台湾の中国語

Reference 1 CP950 (commonly known as BIG5) # 出典の参照番号 1

Reference 2 zVariant, zTradVariant, zSimpVariant in Unihan.txt

Reference 3 List of Simplified Character Table (Traditional column)

Reference 4 zTradVariant in Unihan.txt

Version 1 20020701 # July 2002 # バージョン番号

以下、IDN テーブル本体

#登録可能文字;推奨される異体字;予約される異体字

5718(1);5718(4);56E2(2),56E3(2) # 團;團;团,団

60F3(1);60F3(1); # 想;想;

6559(1);6559(1);654E(2) # 教;教;教;

6E05(1);6E05(1);6DF8(2) # 清;清;清

771F(1);771F(1);771E(2) # 真;真;眞

806F(1);806F(3);8054(2),8068(2) # 聯;聯;联,聯

96C6(1);96C6(1); # 集;集;

66

67 たとえば、この例では、登録可能な IDN の中で使用可能な文字「團」(Unicode で 5718
68 というコードポイントを持ちその出典は CP950)の異体字は、「團」「团」「団」の 3 つであ
69 り、そのうち「團」だけが IDN で使用可能であり、「团」と「団」は使用されず予約される
70 文字であることを示している。

71 また、この例で、「清真教」(それぞれの文字のコードポイントは、清(U+6E05³) 真
72 (U+771F) 教(U+6559))という文字列の IDN が登録申請されたとすると、登録者(ここでは
73 登録者 A とする)に紐付けられる IDN のパッケージは以下となる。

74

³ U+16 進数は、Unicode のコードポイントを示す。

登録文字列：清(U+6E05) 真(U+771F) 教(U+6559)
予約文字列：清(U+6E05) 眞(U+771E) 教(U+6559)
清(U+6E05) 眞(U+771E) 教(U+654E)
清(U+6E05) 真(U+771F) 教(U+654E)
清(U+6DF8) 真(U+771F) 教(U+6559)
清(U+6DF8) 眞(U+771E) 教(U+6559)
清(U+6DF8) 眞(U+771E) 教(U+654E)
清(U+6DF8) 真(U+771F) 教(U+654E)

すなわち、「清」「真」「教」それぞれが 2 つの文字を異体字として持つため、文字列「清真教」が属するパッケージには、 $2 \times 2 \times 2$ の 8 種の文字列が含まれる。「清」「真」「教」はすべてドメイン名において使用可能な文字であるため、「清真教」は IDN として登録できる文字列である。そして、その登録と同時に、他の 7 種の文字列は登録者 A に紐付けられた予約文字列となる。

また、この例では、「清」「真」「教」のみが登録可能文字であるため、8 種の文字列のうち、登録された「清真教」のみが IDN として登録可能であり、残りの 7 種は登録不可能である。その結果、登録者 A は、7 種の予約文字列に相当する IDN を登録することはできない。また、別の登録者(ここでは登録者 B とする)が後から 7 種の予約文字列のうちのいずれかを登録申請しても、それは A に紐付けられたパッケージに含まれる予約文字列であるため、登録できない。

このように、異体字を持つ字を含む文字列を IDN として登録申請した場合、

- ・文字列中に登録可能文字以外の文字が含まれる場合
 - ・他者が登録済みの IDN のパッケージに属する場合
- のいずれかのとき、その IDN は登録できない。

1.3 各 TLD の内部における異体字の定義

いくつかのレジストリは、そのレジストリが運用する TLD 配下で(主には第 2 レベルドメイン名ラベルとして)IDN を使用可能としている。

1.3.1 中国語ドメイン名において

前述のように、中国語では、繁体字と簡体字が異体字の関係にあり、ドメイン名においても、異体字同士を同一文字とみなすこととしている。.CN、.TW を中心とする中国語ドメイン名を登録可能としているレジストリは、協力して中国語ドメイン名用の IDN テーブルを一つ定義し、そこから各レジストリが自レジストリで使う部分を切り出すことにより自レ

ジストリ用の IDN テーブルを定義し、それをドメイン名登録・使用時のルールとしており、すでに 15 年近くの運用実績を持っている。

1.3.2 日本語ドメイン名において

日本語で使われる漢字においても、旧字体と新字体等、異体字の関係にあるととらえられるものが存在する。たとえば、「國」と「国」は異体字ととらえられることができる。

JP ドメイン名においては、2001 年にサービス開始した汎用 JP ドメイン名において、日本語 JP ドメイン名を使用可能とした。

この日本語 JP ドメイン名のルールは、2000 年 9 月より JPNIC が中心となり、ドメイン名の専門家だけでなく商標専門家や文字の専門家等、多様な専門家により議論し、一般利用者からの意見も取り入れつつ、作ったものである。そのルールの主な特徴は、使用可能文字を JIS 第 1 水準と第 2 水準とし、異体字は設けないというものである。異体字を設けない理由は、日本では異なる文字コードを持つ文字は異なる文字として扱うことが適切であると判断したためである。たとえば、「國」と「国」は意識的に異なる文字として使われることが多いという判断である。

そのルールに基づき JPRS が 2001 年からサービスを提供しており、すでに 15 年近くの運用実績を持っている。

1.4 TLD 創設時における異体字問題の回避策

TLD 内(たとえば汎用 JP ドメイン名のような第 2 レベルドメイン)でのドメイン名登録においては、登録希望者がオンラインで登録申請し、希望したドメイン名が申請時点で使用可能であれば即時に登録されその申請者が使用開始できるというサービスを行っているレジストリが多い。これは、たとえば何十万、何百万といったドメイン名が存在する空間に新しくドメイン名を追加する処理を安価かつスピーディーに行えるようにし、インターネットの価値を向上させることに役立っている。

この場合、もし登録者 A がドメイン名 X を登録し使用することに問題があるときは、そのドメイン名の登録後に、他者 B がその登録・使用に問題がある旨を申し立て、A および B からは中立な第三者が問題の有無および解決策を決定している。その典型的なものが、他人の商標もしくはそれに類似したドメイン名を悪意を持って使用するケース等を個々のケース毎に解決するための仕組みである DRP(紛争解決ポリシー:JP ドメイン名の場合は JP-DRP)であり、多くのレジストリで使われている。

中国語の場合は、そもそも対応する簡体字と繁体字は、見た目や Unicode 上のコードが違えど同じ文字(異体字)であるという定義であるため、簡体字と繁体字を入れ替えた文字列

同士は、個々に判断するまでもなく、見た目やコードは違えど同じ文字列(ここでは便宜上「異体文字列」と呼ぶ)であるということになる。このため、第三者が他意なくもしくは悪意を持ってすでに登録された文字列の異体文字列を登録しないように、各文字の異体字を前もって定義し、登録された文字列の異体文字列を自動計算(JET Guidelines におけるパッケージ化に相当)して第三者からの登録を申請時点でブロックできるようにするという処理をしている。

一方、新しい TLD の申請と創設は、2012 年 1 月 12 日に申請受付が開始された新 gTLD プログラムが代表的なものであるが、これには、1930 件の申請が行われ、その中で 75 件が漢字を含む IDN であった。このプログラムでは、個々の TLD 申請には 185,000 ドルの申請料とともに、使用目的や技術的背景などを記した申請書提出が必要とされた。申請された TLD が創設承認されるまでの間には、その文字列や申請者が誰であるかを含めた申請書が公開され、文字列自体の無問題性などに対して中立的なパネルによる評価や第三者からの異議申し立てとその評価が行われ、問題があれば創設は承認されないという仕組みが組み込まれた。この時点では、どの文字にも異体字は定義されておらず、異体字相当のものに起因する問題があれば、それらは中立的なパネルによる評価や第三者からの異議申し立てにより解決されている。

このように、TLD 内のドメイン名登録と比較して、TLD そのものの創設は、申請にかかるコスト、申請の手軽さや登録されるドメイン数、創設までの審査の慎重さが大きく異なっている。

ただし、申請者にとっては、文字列が既存の TLD の異体文字列である等の理由により新 gTLD 創設が拒否される等の問題が存在することが先にわかっていることが望ましく、また、すでに TLD を創設した者にとっても、異体文字列であることを理由に異議申し立てをするという作業がなくなることは望ましいことであろう。

なお、次回の新 gTLD 申請ラウンドは早ければ 2016 年に開始すると見られており、異体文字列を事前排除することをプログラムに組み入れるのであれば、2015 年半ばまでに異体文字列を自動的に知ることができる仕組みを作っておくことは有用である。

2 RootLGR と言語 LGR

2.1 RootLGR とは

RootLGR(Root Label Generation Rules)とは、DNS のルートゾーンに登録されるラベル、すなわち、TLD の文字列に関するルールである。

TLD に ASCII 以外の文字列が使われる場合、その TLD 文字列で使える文字の集合や異

体字を RootLGR として定義し、それにしたがって TLD 文字列をルートゾーンに登録し、必要な場合 TLD 文字列を予約し、運用することが必要になる。TLD 文字列で使われる言語は複数あり、いずれの TLD 文字列もルートゾーンという一つの空間に共存することになるため、RootLGR は次の手順で作られることが ICANN で合意されている。

Step1 TLD で使う可能性のある文字の集合を定義する(MSR : Maximal Starting Repertoire と呼ぶ)

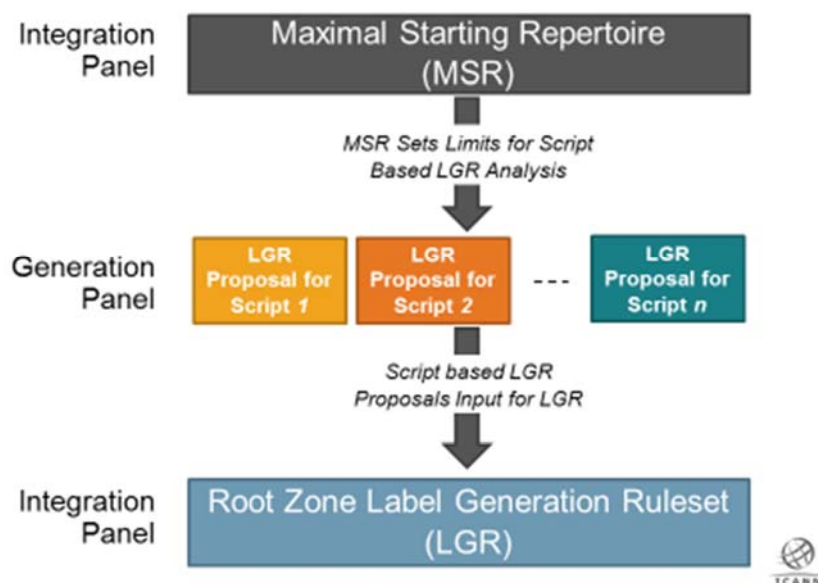
Step2 TLD で使う言語毎に、使える文字の集合や異体字を言語 LGR として定義する(ここで定義される文字集合は Step1 で定義された MSR の部分集合である)

Step3 Step2 で定義された言語 LGR を全言語にわたって整合性を保ちつつ統合し、1 つの RootLGR を定義する

TLD 登録申請の処理において使われる各言語用のルール(言語 LGR)は、Step3 で定義された RootLGR を構成するルール集合から各言語に適用される部分集合を切り出すことにより完成すると考えることができる。換言すれば、TLD 登録申請者は、登録しようとする文字列が属する言語を指定し、各登録申請は、指定された言語用の RootLGR 部分を適用して処理される。

上記 3 つのステップのうち Step1 と Step3 は 1 つの統合パネル(IP : Integration Panel)が担い、Step2 は言語毎に設置された言語生成パネル(GP : Generation Panel)が担う(次の図参照)。IP、GP とも、最終的には ICANN により承認されたメンバーにより作業を実施し、結果を ICANN に提出する。2015 年 1 月 6 日時点で、IP および一部の言語に対する GP が活動している。Step1 はすでに完了しており、現在、種々の言語に対して、GP が Step2 の活動を行っている状況である。日本語に対しては、2014 年 8 月より日本語生成パネル(JGP : Japanese Generation Panel)が有志により活動を開始している。2015 年 2 月頃に JGP 設立を正式に ICANN に対し提案する予定である。

LGR Development Process



2.2 言語 LGR

各 GP が作成する言語 LGR は JET Guidelines の IDN テーブルとほぼ同様のものであり、以下の情報から構成される。

➤ 申請可能文字

MSR⁴の中から申請するドメイン名において使用可能な文字の範囲を各 GP が選択する。選択できる文字は、GP がその言語で使用・すると宣言した Script⁵の範囲に限られる。

➤ 異体字リスト

個々の申請可能文字について、その文字が持つ異体字を一覧(リスト)にしたもの。異体字リストに含まれる文字は、その言語における申請可能文字でなければならない。

➤ 各異体字の属性

- ・割り当て可能(その文字を含む文字列をドメイン名として Root ゾーンに登録することが可能)

- ・ブロック(その文字を含む文字列は Root ゾーンに登録することが不可能)

のいずれであるかが、異体字リストに含まれる個々の文字について定義される。

⁵ Script は平仮名、片仮名、漢字などある言語を書き表すために使用される★同種の★「同種の」の意味がわからない★文字(用字)のこと。

- 219 ➤ 文字列評価ルール
220 申請された文字列全体に対して適用されるルール。たとえば、「簡体字と繁体字が混在
221 した文字列は許容しない(ブロックする)」などの定義を行う。

222
223 JET Guidelines の IDN テーブルと RootLGR における言語 LGR の違いは以下の通りで
224 ある。

- 225
226 ➤ JET Guidelines では、予約される異体字は登録可能文字の一覧に現れなくともよいが、
227 言語 LGR では異体字リストに含まれるすべての文字は申請可能文字でなければならない
228 い。
229 ➤ JET Guidelines ではドメイン名登録を行うレジストリが自身が管理するドメイン名空
230 間に適用する独自の IDN テーブルを定義することとしているが、ルートゾーンに適用
231 するための各言語 LGR の定義では次の 2 点を考慮して GP が定義する必要がある。
232 ・各言語の言語コミュニティ(当該言語を使用している国・地域などの言語文化圏)の合
233 意に基づくこと
234 ・IP が一つの RootLGR に適切に統合できるよう、文字を共有する他 GP と協調する
235 こと
236 ➤ JET Guidelines の IDN テーブルはそれ自身が単独で適用されるルールであるが、言
237 語 LGR は統合され RootLGR となるためのインプットであり単独で適用されることは
238 ない。

240 2.3 RootLGR

241
242 IDN TLD として申請された文字列に対して適用されるルールは、GP が作成した各言語
243 個別の言語 LGR ではなく、すべての GP が作成した言語 LGR を IP が統合した統合ルール
244 (RootLGR)である。これは、TLD 文字列はさまざまな言語の文字列が混在するため、特定
245 の言語 LGR のみを適用できないためである。

246 各言語 LGR の統合の結果できる RootLGR は、概念的には、各言語の言語 LGR の和集
247 合と考えることができる。すなわち、Script を共有しない異なる言語同士は、それぞれの言
248 語 LGR がそのまま RootLGR 内に並存する。また、異なる言語であっても同じ Script を共
249 有するもの(中国語と日本語の漢字など)であれば、同じ文字(たとえば、日本語で使う「愛」
250 も中国語で使う「愛」も Unicode において同じコードポイント U+611B を持つため同じ文
251 字とみなされている)に対して定義される異体字リストは、それぞれの言語で定義された異
252 体字リストの和集合となる。なお、その和集合としての異体字リストにおいて、どの文字が
253 割り当て可能か、どの文字がブロックされるか、さらにどういう文字列評価ルールが適用さ
254 れるかは、言語ごとに異なる。

したがって、RootLGR に基づいてドメイン名を登録申請する場合は、そのドメイン名が属する言語を指定する必要がある、申請の処理には、RootLGR におけるその言語に適用されるルール部分が当てはめられることになる。

例えば、

- ・中国語 LGR では「愛(U+611B)」と「爱(U+7231)」は互いに他を異体字として持ち、両方とも割り当て可能

- ・日本語 LGR では「愛(U+611B)」は異体字なしの登録可能文字と定義され、「爱(U+7231)」は登録可能文字ではなく日本語 LGR に現れないと定義している場合、RootLGR では、以下のようなになる。

- ・中国語の「愛(U+611B)」は「爱(U+7231)」を異体字として持ち、「愛(U+611B)」が入った中国語 TLD の登録申請者は、その TLD が登録可能であると同時に、「愛(U+611B)」を「爱(U+7231)」に入れ替えた TLD も登録可能(他の登録申請者は登録不可能)

- ・中国語の「爱(U+7231)」は「愛(U+611B)」を異体字として持ち、「爱(U+7231)」が入った中国語 TLD の登録申請者は、その TLD が登録可能であると同時に、「爱(U+7231)」を「愛(U+611B)」に入れ替えた TLD も登録可能(他の登録申請者は登録不可能)

- ・日本語の「愛(U+611B)」は「爱(U+7231)」を異体字として持ち、「愛(U+611B)」が入った日本語ドメイン名の登録申請者は、そのドメイン名は登録可能であるが、「愛(U+611B)」を「爱(U+7231)」に入れ替えたドメイン名は登録不可能(他の登録申請者も登録不可能)

3 文字を共有する言語 LGR 間の干渉

3.1 中国語 LGR の異体字が日本語コミュニティに与える影響

中国語 LGR で定義された異体字が RootLGR に採用された場合、日本語であると宣言して TLD 文字列の登録申請を行っても、漢字の異体字はすべての言語で共通に異体字とみなされるため、中国語の異体字も日本語の異体字として扱われる。そのため、次のような状況が発生し得る。

たとえば、「機(U+6A5F)」と「机(U+673A)」は日本語 LGR では異体字同士ではないが、中国語 LGR では異体字であった場合を想定する。このとき、中国語 LGR と日本語 LGR が統合された結果の RootLGR では、「機」と「机」は異体字となる。この状況で、登録者 A が日本語として TLD 「機上」を登録した場合、A の意図にかかわらず、TLD 「机上」も A に紐付けられる。そして、TLD 「机上」がブロックされて誰も登録できなくなる。

このように、日本語のコミュニティが作成したルールは中国語のコミュニティが作成したルールと統合されて TLD に適用されるため、日本語 JP ドメイン名のルールで得た利用

者の経験は、TLD 文字列用のルールである RootLGR にて利用者が得る経験とは異なるものとなる。したがって、日本語 LGR が異体字を定義する・しないに関わらず、中国語 LGR の異体字定義は日本語 TLD に影響を与えることを一般利用者が理解する必要があり、その事実を十分に TLD 登録申請を考える者や一般のインターネットユーザーに周知する必要がある。

3.2 日本語 LGR の異体字が中国語コミュニティに与える影響

日本語 LGR で異体字を定義した場合で、かつ、その定義が中国語 LGR の異体字定義と異なる場合、上記「3.1 中国語 LGR の異体字が日本語コミュニティに与える影響」で述べたことと同様の状況が、中国語コミュニティに発生し得る。

4 JGP の具体的課題定義(案)

4.1 登録可能文字の範囲の決定

日本語 LGR では、日本語の TLD として登録可能とする文字の範囲(集合)を合理的な理由に基づき MSR の中から選ぶことになる。日本語は複数の Script で構成されるため、登録可能な Script の範囲として、たとえば以下のものが考えられる。(アラビア数字は MSR に含まれておらず TLD に使うことはできない)

案 1 平仮名・片仮名・漢字

案 2 英字・平仮名・片仮名・漢字

さらに、漢字については、たとえば以下のような範囲が考えられる。

案 a 常用漢字⁶の範囲(のうち MSR に含まれるもの)

案 b 常用漢字+人名用漢字⁷の範囲(のうち MSR に含まれるもの)

案 c JIS X 0208:2012 の第一水準・第二水準漢字の範囲 (これは MSR の部分集合)

案 d JIS X 0213:2012 の第一水準・第二水準・第三水準・第四水準の範囲(のうち MSR に含まれるもの)

⁶ 常用漢字は法令、公用文書、新聞、雑誌、放送など、一般の社会生活において、現代の国語を書き表す場合の漢字使用の目安で、2136 文字が定義されている(2014 年 10 月末現在)。

⁷ 人名用漢字は日本における戸籍に子の名として記載できる漢字のうち、常用漢字に含まれないもので、861 文字が定義されている(2014 年 10 月末現在)。

案 e IICORE⁸の範囲(これは MSR の部分集合)

案 f JGP が独自に設定した範囲

など

これらのうち、どの案を選択するかは、次のような観点で判断する必要がある。

- ・ 日本語の TLD 文字列として最低限必要と思われるものが入っている
- ・ 日本語コミュニティにとって識別や理解が困難なものは入っていない
- ・ インターネットにつながる端末や OS、アプリケーションで問題なく入力や表示ができる
- ・ .JP 等の TLD 内での約 15 年の経験において、日本語ドメイン名を構成する文字としての適切さが実証されている
- ・ .JP 等の日本語ドメイン名のルールとの差異が利用者に混乱を与えない
- ・ 日本語コミュニティに説明できる(権威ある)根拠がある

4.2 日本語 LGR での異体字の定義有無の決定

中国語 GP は、既に異体字を多く定義した中国語 LGR 案を作成中であり、これまでの.CN や.TW の SLD での経験からも、中国語 LGR で多くの漢字に対する異体字が定義されることは確実である。この状況も踏まえつつ、日本語 LGR において異体字を定義するか否かを決める必要がある。

案 1 日本語コミュニティとして積極的に異体字定義を行う(積極的異体字定義)

案 2 中国語 LGR と共存するのに適した異体字定義を行う(受身的異体字定義)

案 3 異体字を定義しない(つまり、個々の文字は互いに独立した文字であるとする)

どちらの案を選択するかは、次のような観点で判断する必要がある。

- ・ TLD 登録者は、異体字を同一視したいという積極的要求が強いかな
- ・ TLD 登録者は、異体文字列を自分のためにブロックしたいという要求が強いかな
- ・ TLD 登録申請者が、他登録者が自動的に予約する文字列が多くなり登録可能文字列が減ることを大きな問題ととらえるかな
- ・ 日本語コミュニティに属するインターネットユーザにとって、異体字が定義されている方が相手を的確に識別できるか
- ・ 中国語コミュニティと適切な共存ができるか

また、日本語 LGR において異体字を定義する場合には、どのような基準で個々の異体字を

⁸ ISO/IEC 10646:2003/ Amendment 1:2005 の一部として発行された国際標準で、漢字圏で共通に使える日常生活の用を満たす漢字集合。

356 定義するかという課題がある。これは、どのような観点により異体字を定義するかによって
357 異なる。

358 まず、案 1(日本語コミュニティとして積極的に異体字定義を行う)の場合、異体字の定義
359 根拠としては次のようなものがある。

360 案 a 新字と旧字

361 案 b JIS 参照文字

362 案 c その他(たとえば JGP が独自に定義)

363

364 さらに、それぞれの異体字に対し、次のような案がある。

365 案 x すべての異体字を登録可能とする

366 案 y 個々の異体字の使用状況を考慮する

367 案 z 申請された文字以外の異体字をすべて登録不可能とする

368

369

370 案の選択に当たっては、異体字の存否判断と同様に、次のような観点で判断する必要があ
371 る。

372 ・ TLD 登録申請者にとって、どういう異体字を同一視したいという積極的要求が強いか

373 ・ 日本語コミュニティに属するインターネットユーザにとって、どういう異体字が定義さ
374 れている方が相手を的確に識別できるか

375 ・ 中国語コミュニティと適切な共存ができるか

376 ・ 日本語コミュニティのコンセンサスであるという権威がどれくらいあるか

377 ・ 残存する異体字問題が、TLD 創設申請時の異議申し立てにより効果的に解決可能であ
378 るか

379

380 次に、案 2(中国語 LGR と共存するのに適した異体字定義を行う)の場合、日本語 LGR と
381 しては、中国語 LGR の異体字パッケージをそのまま受け入れることとなるが、そのパッケ
382 ージの中で、どの文字列を登録可能とするか、ブロックするかに関して次の案がある。

383 案 a 日本語で登録可能な各文字に対し、その文字の異体字リストの中で、日本語 LGR 上
384 登録可能な文字はすべて割り当て可能とし、登録不可能な文字はブロックする

385 案 b 日本語で登録可能な各文字に対し、その文字だけを登録可能とし、その文字の異体
386 字リストの中の他の文字はすべてブロックする

387

388 案の選択に当たっては、次のような観点で判断する必要がある。

389 ・ TLD 登録申請者の要求が最大限満たされるか

390 ・ 日本語文字列なのに「中国語 TLD として登録申請した方がよい」と登録申請者は判断
391 してしまわないか

- ・日本語コミュニティに属するインターネットユーザにとって違和感がないか

4.3 中国語 LGR の異体字を容認することの是非

CGP(中国語生成パネル)作成の中国語 LGR で定義される異体字を、日本語コミュニティとして容認するか、反対するか、を決める必要がある。

案 1 異体字定義すべてを容認する

案 2 個々の文字について個々の異体字を容認するか反対するかを判断し調整する

案 3 異体字を定義すること自体に反対する

また、案 2(個々の文字について個々の異体字を容認するか反対するかを判断し調整する)の場合、判断と調整の方針を決める必要がある。例えば、次のような案がある。

案 a 日本語 LGR の登録可能文字の範囲に入っている文字はすべて異体字から除外してもらう

案 b 日本語では異体字の関係にないと考えられる文字は異体字から除外してもらう(たとえば、新字と旧字の関係や JIS 参照文字の関係にある文字同士だけを異体字として容認する)

どの案を選択するかは、次のような観点で評価する必要がある。

- ・中国語コミュニティにとって適切なもしくは問題ない案であるという説得ができるか
- ・日本語コミュニティにとって適切な案であるという説得ができるか
- ・有限時間(たとえば 3 ヶ月)で CGP と調整ができるか

4.4 文字列評価ルールの定義

現在、JP ドメイン名では、RootLGR でも考慮すべきと考えられる特別な文字列評価ルールは定義しておらず、これまでそれに起因する問題も特に生じていない。

今回作成する日本語 LGR においては、次の案が存在する。

案 1 文字列評価ルールは設けない

案 2 新たに評価ルールを設ける(例: 長音(「ー」)や繰り返し文字(「々」)が文字列の先頭に来ない)

案の選択に当たっては、次の観点での評価が必要である。

- ・設定した文字列評価ルールが、問題を内在する TLD 文字列の申請を削減する度合い
- ・設定した文字列評価ルールが適切であることの(権威ある)根拠

428

429 以上

430

431