

JGPとしての検討課題への対応

■本資料の目的

JGP会合(10)資料5で示した今後の検討課題について調査・検討を進め、CJKおよびIPと調整する際のJGP方針案としてまとめたので、合意を得たい。

なお、本資料の方針案については、CGPおよびKGPのLGR-1が確定した後に影響があると判断した場合は再度相談する。

■検討課題への対応方針案

- 凡例 ★：JGP内で相談・検討し、方針を定める
☆：JGPのこれまでの方針に従いCJK調整担当が主に検討・調整を進め、必要に応じJGPに相談する
－：JGPのこれまでの方針に従いCJK調整担当が検討・調整を進める

A. JGPが独自にIPと進めるもの

(A1) 明らかな禁止事項等のルールをWLEで定義するか？(★)

JGPでの仮合意は、オブジェクションのプロセスで排除できるので日本語LGRでは特別なWLEの定義は必要ないとしている。
日本語JPドメイン名での使用状況を確認し、不自然な使われ方がないもしくは極端に少なければWLEを定義しないという方針を維持するのが妥当であると考えられる。

2014年10月時点での日本語JPドメイン名登録(125628ドメイン名)における、不自然と考えられる長音および踊り字の登録状況(数字はドメイン名数)を示す。

文字列の先頭に長音：	0
漢字の後に長音：	4
文字列の先頭に平仮名踊り字：	0
片仮名・漢字の後に平仮名踊り字：	0
文字列の先頭に片仮名踊り字：	0
平仮名・漢字の後に片仮名踊り字：	0
文字列の先頭に漢字踊り字：	0
平仮名・片仮名の後に漢字踊り字：	0

(参考)

平仮名の後に長音：	642
文字列の先頭に小平仮名：	6
文字列の先頭に小片仮名：	0

★JGPとしての対応方針(案)

不自然と思われる文字列は「漢字の後ろに長音」の4例しか存在せず、これは全体の0.003%と無視できる数であるため、従来のWLEを定義しないという方針を維持する。

(A2) 多すぎるallocatableを減らす方法は？(★)

JGPでの仮合意は、C-LGR-1の影響は非常に限定的であり統合アルゴリズム案によって異体字とされる文字が登録可能(allocatable)の異体字タイプになるのであれば許容できるとしている。

統合アルゴリズム案はこの考え方に基づいて設計されているが、IPはこのアルゴリズムではJ-LGR-2のWLE適用後に多数の登録可能(allocatable)文字列(ラベル)ができることを懸念し、改善を要求している。

IPからの要求に対してJGPが取り得る対応は以下の3つが考えられる。

- (1) RootLGRのプロセス外に、申請者が登録可能ラベルの中から登録するラベルを選定するプロセスを作るようICANN/IPに要求する
- (2) CGPIに異体字の数を減らすよう要求する
- (3) J-LGR-2に基づいて受身的異体字定義および異体字サブタイプ定義をした第2世代のJ-LGR-1を作成する

資料4で示された今後の検討課題にあるように、ICANN/IPは事前の自動判断を強く求めており、(1)は合意に至る可能性が低い。また、CGPIはCDNCでの15年におよぶ異体字運用の経験に基づきかつ可能な限り最小限の異体字定義をしており、そこからさらに異体字を減らせと要求する(2)も合意に至る可能性は低い。

したがって下記(B3)の分析を行った後、(3)を検討するのが妥当であると考えられる。この場合、コミュニティへはJ-LGR-1の作成は他GPやIPからのフィードバックを得て複数世代に渡って行われたものであり、基本概念は第1世代J-LGR-1を作成したときのままであることを十分に説明する必要がある(それに矛盾のない第n世代J-LGR-1を作成する必要がある)。

★JGPとしての対応方針(案)

5/16-17のCJK調整委員会でのCJK各LGR-1を元に生成したJ-LGR-2に基づき、2014年10月時点での日本語JPドメイン名登録(125628ドメイン名)を対象にどれほどの異体字ラベルが生成され得るか、実際にその中で登録されているドメイン名数がどれほどあるかを調査した。その結果、1つのドメイン名から生成される異体字ラベル数の最大は20736通りであること、および全体の80%のドメイン名は生成される異体字ラベル数が6通り以内であることがわかった。また、実際に登録されている125628ドメイン名の中で異体字ラベル同士のもは210組(2ラベルが207組、3ラベルが1組、4ラベルが2組の計425ドメイン名)あり、全体の0.3%であった。なお、これら210組に共通のルール(文字列中のすべての新字体が旧字体に置き換わっているなどの規則性)を見出すことはできなかった。したがって、第2世代のJ-LGR-1で異体字サブタイプを定めて登録可能(allocatable)ラベルの数を減らすことは困難であるが、申請者が異体字ラベルを複数選ぶ可能性は極めて低いと考えられるため、従来他LGR-1で定義された異体字の異体字タイプはallocatableとする方針を維持する。

B. CJKが一体となって検討するもの

(B1) 用語及びその定義の合意 (一)

2015/05のCJK調整委員会での議論を踏まえ、米谷が修正を反映し、CJK GP MLおよびIPで共有する。ICANNブエノスアイレスまでに用語及びその定義の合意を得る。

(B2) アルゴリズムのスペック(入出力仕様とアルゴリズム)の完成 (☆)

2015/05のCJK調整委員会での議論を踏まえ、米谷がアルゴリズムのスペックをドラフトし、CJK GP MLで共有する。ICANNブエノスアイレスまでにスペックの合意を得る。

(B3) 各言語LGR-1の提案が互いに与える影響を分析し、問題ある場合は必要に応じLGR-1を修正する (★)

JGPでの仮合意は、C-LGR-1の影響は非常に限定的であり統合アルゴリズム案によって異体字とされる文字が登録可能な異体字タイプになるのであれば許容できるとしている。

121:
122: 2015/05のCJK調整委員会用に提出されたCJK各LGR-1(*1)を統合アルゴリズム
123: 案で統合した結果導き出されるJ-LGR-2の異体字について、再度日本語JPド
124: メイン名登録における異体字とみなされる登録の状況を調査し、影響がない
125: ことを確認することが妥当と考えられる。
126:
127: その上で、上記(A2)の問題を解決するための検討と併せてJ-LGR-2からの
128: フィードバックとして新たにJ-LGR-1を再定義する検討(J-LGR-2の異体字を
129: 使い異体字サブタイプを定義するか、等)を実施するのが妥当であると考え
130: られる。
131:
132: (*1) C-LGR-1
133:
134: <http://mm.icann.org/pipermail/japanesegp/attachments/20150430/4fd4c763/mss-var-
135: 20150430-0001.xml>
136: J-LGR-1
137:
138: <http://mm.icann.org/pipermail/japanesegp/attachments/20150427/3bbfc797/J-LGR-1-
139: 0001.xml>
140: K-LGR-1
141:
142: <http://mm.icann.org/pipermail/japanesegp/attachments/20150515/643c2fac/K2-KLGR-
143: 1-0001.xlsx>
144:
145: ★JGPとしての対応方針(案)
146: 調査結果はA2で述べたとおりであり、C-LGR-1の影響は限定的であるとい
147: えるため、従来の他LGR-1の異体字定義を受け入れる方針を維持する。
148:
149: C. その他
150: (C1) 現レパートリ案から不適切な文字を削除 (★)
151:
152: JGPでの仮合意は、JIS X 0208第1水準・第2水準の平仮名・片仮名・漢字お
153: よびそれらに準ずる文字を文字範囲(レパートリー)とし、それに対する文字
154: の追加・削除の合理的な理由はないとしている。
155: MSRのように、ICANN/IPが定める外的要因がない限り、この方針は変更しな
156: いのが妥当であると考えられる。JGPチェアからは「各言語GPIはその言語お
157: よびScriptの範囲を超えた判断はすべきではなく、その判断はGPの外部で行
158: うべき」というアピールをICANNに対して実施しており、CGPチェアからも
159: ICANNスタッフがいる場で賛同を得ている。
160:
161: 日本語LGRに対しICANNが実施するパブリックコメントで指摘が行われる可能
162: 性があるので、回答は用意しておく必要がある。
163:
164: なお、2014年10月時点での日本語JPドメイン名登録(125628ドメイン名)中、
165: 以下の文字の登録状況(数字はドメイン名数)を示す。
166:
167: 卍 : 2
168: あ : 19
169: い : 67
170: う : 6
171: え : 51
172: お : 61
173: ん : 1
174: ゑ : 3
175: せ : 3
176: ㄣ : 8
177:
178: ★JGPとしての対応方針(案)
179: 上記文字を排除する合理的な理由は存在しないため従来の範囲を維持する。
180:

175: 以上