

国土地理院の地名情報で使用する文字の標準化について

Standardization of Characters used in Geographical Name Information

基本図情報部 倉田憲¹・五島健太
National Mapping Department KURATA Tadashi and GOTO Kenta

要 旨

国土地理院が整備、管理している地名情報データベースで使用する文字について、以下のとおりとしたので報告する。

- ・日本語文字集合の範囲 JIS X 0213 : 2012 (第1～第4水準漢字)
- ・文字コード JIS X 0221 : 2020
- ・文字の符号化形式 UTF-8

1. はじめに

国土地理院は、日本の国家地図作成機関として我が国の国土に関する情報を更に価値ある使いやすいものとなるよう改善を図りつつ整備・提供を進めており、日本国内の地名に関して、約42万件の地名を地名情報データベース（以下「地名DB」という。）として整備し、管理している。地名DBで取り扱うデータは地名とともにその読みや位置等の属性情報を保有しており、国土地理院の地図に記載される地名の基データの位置づけである。

日本では、地名を漢字や平仮名、片仮名のいずれかの文字又はこれらの組み合わせによって表記しているが、一部の漢字については、文字フォントや文字コードが提供されておらず、これまでシステム環境によっては正しく表示されなかった。この正しく表示されない漢字について、国土地理院の地図では、利便性を考慮しこれまで一般的なパソコンなどで取り扱える代替の漢字又は平仮名に置き換えて表記してきたが、令和5年6月9日に「デジタル社会の実

現に向けた重点計画」が閣議決定されたことを受け、その重点的な取組の一つであるデータ連携推進の一環として、国土地理院では、誰もが一意の文字情報を得られるようにデジタル庁が推奨する文字仕様を導入し、地名DBで使用する文字の標準化・共通化に取り組んだ。

2. 日本において地名等に使用される漢字の現状

日本において、一般の社会生活で使用する常用の漢字の目安として、平成22年内閣告示の常用漢字表に約2千文字が定められている。これに対し、地名や氏名など行政業務に使用される漢字の数は6万文字以上あると言われており、僅かなデザインの違いであっても異なる漢字として使用される例（図1）もあることから、地名や氏名に使用する漢字は膨大な数となり、多くの行政機関が独自に文字情報を作成し、行政業務に使用してきた。この結果、文字情報交換の際に不整合が生じ、一意の文字情報が得られない等の問題が生じていた。これを解決するため、デジタル庁は文字情報を標準化・共通化した文字情報基盤を導入し、全ての漢字について文字フォントの整備と文字コードの国際規格化を完了させた。

現在、デジタル庁は原則として全ての地方公共団体に対し、2025年度までに基幹業務システムにおける文字コードの統一、国際規格化への対応を進めている。

接触非接触		接触位置		交差有無		点か棒か		曲げ方折り方	
眞	真	家	家	非	非	帰	帰	頑	頑
運筆方向		止め跳ね		八屋根		曲げ跳ね		筆おさえ	
考	考	配	配	八	八	執	執	更	更

図-1 僅かなデザインの差であっても異なる漢字として使用されている例

3. 地名 DB の漢字に関する課題

国土地理院では、地元で呼び習わされている地名については地方公共団体に確認し、また収集した告示情報に基づいて、可能な限り正確な文字を地図に表記してきた。これまで地名 DB で使用する文字は、日本産業規格 JIS X 0221 の基本日本文字集合（第 1 及び第 2 水準漢字）の例示字形（JIS の規格票に示されている字形）に該当する文字を使用していた。JIS によって規格化されていない漢字、あるいはシステム環境によって扱えない漢字については、平仮名又

は検索の利便性を考慮して代替文字を使用することとし、本来の漢字との対応を示す付属資料（図 2）を提供してきた。また、漢字には文字の意味と読みが同じであっても字体が異なる異体字（図 3）も多く、これらの漢字についても付属資料の提供で対応してきた。

しかし、地図利用者が正しい地名の文字を代替した漢字や平仮名であると誤解する、あるいは行政機関が文字の使用に当たり正しい文字と代替した文字とを混同する、などの問題が生じていた。

仮コード	文字	代替文字	Unicode（参考）
E004	𪗇	「うつき」	
E007	𪗈	「なる」	
E008	𪗉	「れい」	U+3691 E0102
E009	𪗊	「ぼき」	

図-2 システム環境によっては扱えない漢字とその代替文字

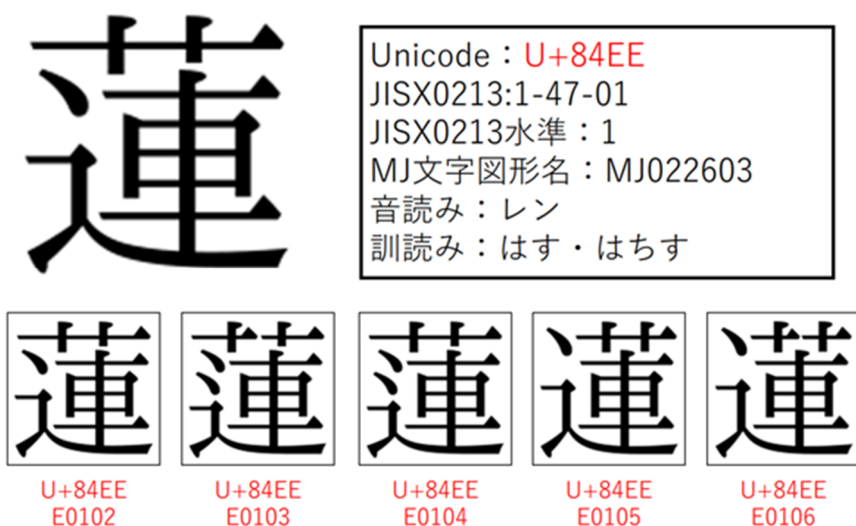


図-3 文字の意味と読みが同じでも字体が異なる漢字（異体字）の例

4. 文字の標準化の取組と効果

前述したとおり、漢字に関する国内状況や地名を地図に表記する際の課題等を踏まえ、地名 DB で使用する文字について、文字情報基盤の活用を前提に、国際標準に準じた仕様への変更を検討した。

コンピュータで文字を扱うためには、「取り扱う日本語文字集合の範囲」を定めるだけでなく、文字をコンピュータが理解できる形式に符号化する必要がある。符号化に当たっては、文字に与える番号である「文字コード」と、この番号を文書ファイルの保存や交換の際にどのように表現するかを指定する「符号化形式」も必要である。

地名や氏名における文字は、アイデンティティを確立する上で非常に重要な要素であるが、文字に含まれる漢字の正確性を重視した場合、取り扱う漢字の数を増加させ、複雑化につながる。そのため地名 DB で取り扱う文字の仕様については、文字コード及び文字の符号化形式は旧来どおりそれぞれ「JIS X 0221」及び「UTF-8」とし、日本語文字集合の範囲を一般的なパソコンやスマートフォンに標準的に搭載されている「JIS X 0213(第1～第4水準漢字)」とした。この結果、「取り扱う日本語文字集合の範囲」の拡張により代替文字又は平仮名へ代替する必要がある漢字(図4)は数十程度となり、異体字等の外字の使用も不要となった。

文字	代替文字
𣎵	𣎵
𣎵	「なる」
𣎵	𣎵
𣎵	「ぼき」

図-4 代替文字又は平仮名への代替例

なお、今回の文字仕様の標準化にあたり、日本語文字集合の範囲外の漢字については引き続き、代替した漢字や平仮名を使用する。また、異体字については Unicode の基底文字へ統一し、文字仕様の標準化を行ってもなお一部残るシステム環境によっては扱えない漢字についても、代替する漢字又は平仮名とした。

5. 地名情報に係る測量成果への反映

地名 DB で取り扱う文字の仕様変更を受けて、地名情報に係る測量成果について次のとおり対応を行った。

○数値地図(国土基本情報)地名情報

- ・ ファイル仕様及び数値地図(国土基本情報)等閲覧ツールの改良を行い、国土地理院ホームページから公開した。
- ・ 変更後のファイル仕様に適合する形で「数値地図(国土基本情報)地名情報」のデータ更新を行った。

○電子国土基本図(地名情報)住居表示住所

- ・ ファイル仕様及び変更対象の地名を含むデータについて、地方公共団体単位で更新を行い、ホームページから公開した。

6. おわりに

文字の標準化はデジタル庁において推進されており、相互運用性を確保した機械判読可能なデータは、電子テキストや音声認識において正しい情報を伝えることを可能とし、システム環境によって適切に表示されない文字の回避に繋がる。今後、地図に表記する地名だけでなく、各行政機関においてもこの文字の標準化が広がることで、更なる利便性の向上、行政サービスのデジタル化の推進が期待できる。

国土地理院では地名 DB の効率的かつ適正な整備、利用者にとって有益でわかりやすい地名情報の提供について引き続き取り組んで行く。

(公開日：令和6年10月4日)

参 考 文 献

- 文化審議会国語分科会 (2016) : 常用漢字表の字体・字形に関する指針 (報告),
https://www.bunka.go.jp/seisaku/bunkashingikai/kokugo/hokoku/pdf/92550601_01.pdf (accessed 24 June 2024)
- デジタル庁 (2023) : 地方公共団体情報システムデータ要件・連携要件標準仕様書【第 2.0 版】「文字要件説明資料」, https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/4748d2a4-45c7-45de-a55f-dfac4cb04a70/0e0c8a2c/20230609_local_governments_character_specification_materials_01.pdf (accessed 24 June 2024)
- デジタル庁 (2023) : デジタル社会推進実践ガイドブック DS-461「文字環境導入実践ガイドブック」,
https://github.com/JDA-DM/GIF/blob/main/460_%E5%AE%9F%E8%B7%B5%E3%82%AC%E3%82%A4%E3%83%89%E3%83%96%E3%83%83%E3%82%AF/docx/461_%E6%96%87%E5%AD%97%E7%92%B0%E5%A2%83%E5%B0%8E%E5%85%A5%E5%AE%9F%E8%B7%B5%E3%82%AC%E3%82%A4%E3%83%89%E3%83%96%E3%83%83%E3%82%AF.docx
(accessed 2 Oct. 2024)
- デジタル庁 (2023) : デジタル社会の実現に向けた重点計画,
https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/5ecac8cc-50f1-4168-b989-2bcaabffe870/b24ac613/20230609_policies_priority_outline_05.pdf (accessed 24 June 2024)
- 文字情報技術促進協議会 : 文字情報基盤整備事業について, <https://moji.or.jp/mojikiban/> (accessed 24 June 2024)
- 総務省 (2012) : 「市区町村が使用する外字の実態調査」報告書,
https://www.soumu.go.jp/main_content/000157022.pdf (accessed 24 June 2024)