

地名辞書ファイル仕様

1. ファイル形式

- ファイル名 : 地名辞書.csv
- ファイル形式 : CSV (カンマ区切りテキストファイル)
- データ項目間の区切 : “,” (カンマ)
- レコード間の区切 : CR+LF
- 文字コード : シフト JIS エンコード

2. レコード構成

各レコードは、以下の(1)~(38)の 38 のデータフィールドを持ち、各レコードの先頭から順に格納されている。

(1) 住所

- 定義 : 日本国内の地名 (住所)。
市区町村内の町域名 (字名) 相当までを対象とし、都道府県名、市区町村名、町域名を半角空白で区切る。
- 文字種 : 全角(2byte)の漢字、かな、カナ、記号および半角(1byte)空白
- 例 : 東京都 千代田区 一ツ橋

(2) 住所よみ

- 定義 : (1)住所の読み仮名を示す。
都道府県名、市区町村名、町域名を半角空白で区切る。
- 文字種 : 全角(2byte)の平仮名
- 例 : とうきょうと ちよだく ひとつばし

(3) 住所ローマ字

- 定義 : (1)住所のヘボン式ローマ字表記を示す。英字は全て大文字とする。
市区町村名、町域名を半角空白で区切る。
- 文字種 : 半角(1byte)英大文字、ハイフン(“-”)、空白
- 例 : TOKYO-TO CHIYODA-KU HITOTSUBASHI

(4) ~ (35) 慣用的表記揺れ 1 ~ 3 2

- 定義 : (1)住所の慣用的な表記揺れ表現を示す。
- 文字種 : 全角(2byte)の漢字、かな、カナ、数字、記号および半角(1byte)空白。市区町村名、町域名を半角空白で区切る。
想定しうる表記揺れを、慣用的表記揺れ 1 フィールドから順にならべ、存在しない場合は、以降を null 値とする。

例 1 : 東京都 千代田区 一つ橋

例 2 : 東京都 千代田区 一ッ橋

(36) 緯度

定義 : (1)住所の緯度(北緯)を世界測地系で示す。

文字種 : 半角(1byte)数字、"."(小数点)。
整数部 2 桁+小数点+小数点以下 1 3 桁。
小数点以下の末尾桁の"0"は、省略する。

例 : 35.6930145199516

(37) 経度

定義 : (1)住所の経度(東経)を世界測地系で示す。

文字種 : 半角(1byte)数字、"."(小数点)。
整数部 3 桁+小数点+小数点以下 1 2 桁。
小数点以下の末尾桁の"0"は、省略する。

例 : 139.740571962963

(38) フラグ

定義 : (36)緯度、(37)経度を付与する際に、住所のどのレベルまでを参照したかを示す。

値 : 以下の 3 値のうちの一つ。
3 : 町域名まで参照
2 : 市区町村名まで参照
1 : 都道府県名まで参照

(39) 緯度経度付与対象住所

定義 : (36)緯度、(37)経度を付与する際に、実際の対象とした住所を示す。
(38)フラグの値に応じて、(38)フラグが"3"の場合は、住所の町域名までが、"2"の場合は市区町村名までが、"1"の場合は都道府県名が格納される。

文字種 : 全角(2byte)の漢字、かな、カナ、記号。

3. 収録内容

- ・ レコード件数 : 117,075 件
- ・ 収録基準、その他
 - 住所は、郵便事業株式会社発行の郵便番号簿（平成 20(2008)年 1 月末現在）から抽出した。
 - 慣用的表記揺れは、住所表記に特有の表記揺れパターンをもとに、生成した。
◇ ツ → つ、ッ （例：一ッ橋→一つ橋、一ッ橋）

- ◇ ケ → が、ヶ (例:千駄ヶ谷→千駄が谷、千駄ヶ谷)
- ◇ 旧字体 → 新字体 (例:龍ヶ崎→竜ヶ崎)
- ◇ 平仮名 → 漢字 (例:ふ頭→埠頭)
など。

➤ 緯度、経度のデータは、国土交通省の街区レベル位置参照情報を基に、以下のロジックで算出した。

- ◇ 当該住所（町域名レベル）が、街区レベル位置参照情報に存在する場合は、その全てのデータの平均値を当該住所の緯度、経度とした。
この場合、(38)フラグには、値”3”を入れる。
- ◇ 当該住所（町域名レベル）が、街区レベル位置参照情報に存在しない場合は、街区レベル位置参照情報の当該市区町村の平均値を採用した。
この場合、(38)フラグには、値”2”を入れる。
市区町村の平均値であるため、緯度、経度が指し示す地点が、当該住所が示す地域内に入るとは限らない点に注意を要する。
- ◇ 街区レベル位置参照情報に、市区町村名も存在しない場合は、都道府県の平均値を採用した。
この場合、(38)フラグには、値”1”を入れる。
都道府県の平均値であるため、緯度、経度が指し示す地点が、当該住所が示す地域内に入るとは限らない点に注意を要する。

4. 表記についての留意事項

4. 1 住所表記

(1)原則

日本郵便のホームページからダウンロードした郵便番号データの表記に従う。

(2) 区切り

都道府県名と郡市町村名との間及び郡市町村名と地名との間は半角スペースで区切る。

例:「沖縄県 八重山郡与那国町 与那国」

(3) 展開

郵便番号データの表記にカッコ書きが付加されている場合、その内容が地名と見做されればカッコの前にある地名に付記する。

例：「北海道 稚内市 声問村（曙）」
→「北海道 稚内市 声問村曙」

(4)例外

・京都市の「通」

地名の後に付く「～通～上る」「～通～下る」「～通～西入」「～通～東入」は削除する。

例：「京都府 京都市上京区 今出川町今出川通烏丸西入」
→「京都府 京都市上京区 今出川町」

4. 2 住所読み

(1)原則

日本郵便のホームページからダウンロードした郵便番号データの表記に従う。

(2)区切り

都道府県名と郡市町村名との間及び郡市町村名と地名との間は半角スペースで区切る。

例：「おきなわけん やえやまぐんよなぐにちょう よなぐに」

4. 3 住所ローマ字

(1)原則

国土地理院の次のページに準拠し、全て大文字にて綴る。

<http://www.gsi.go.jp/REPORT/JIHO/vol108/8.pdf>

へボン式に準じているが、「い」の長音は「II」となる。

(2)区切り

都道府県名と郡市名との間、郡名と町村名との間及び市町村名と地名との間は半角スペースで区切る。また、都道府県及び郡市町村の接尾辞は半角マイナス「-」でつなぐ。

例： 「OKINAWA-KEN YAEYAMA-GUN YONAGUNI-CHO YONAGUNI」

4. 4 慣用的表記揺れ

(1) 読みに伴う「が」「の」「つ」

読みに伴いつなぎで「が」「の」「つ」が入るものは①～③の通りとする。

① 「が」

「が」「ガ」「ケ」「ヶ」を採用する。

例：「岐阜県 各務原市」
「ぎふけん かかみがはらし」
→「岐阜県 各務ヶ原市」 etc

ただし、読みによっては「が」「ガ」が「か」「カ」となるものもある。

例：「宮城県 刈田郡七ヶ宿町」
「みやぎけん かったぐんしちかしゆくまち」
→「宮城県 刈田郡七カ宿町」 etc

② 「の」

「の」「ノ」「之」を採用する。

例：「大阪府 大阪市住之江区」
「おおさかふ おおさかしすみのえく」
→「大阪府 大阪市住ノ江区」 etc

③ 「つ」

「つ」「ツ」「ッ」を採用する。

例：「東京都 新宿区 四谷」
「とうきょうと しんじゅくく よつや」
→「東京都 新宿区 四ッ谷」 etc

(2) 異体字

使用頻度の高い異体字・類字による表記を採用する。原則として旧字→新字。

ただし、JIS78 と JIS83 で字体が入れ替わった 22 組 44 文字を含むものも採用する。

別表 1 に旧字体と新字体との対応、別表 2 に JIS78 と JIS83 とで字体が入れ替わった文字

を示す。

例：「宮城県 塩竈市」
→「宮城県 塩釜市」「宮城県 塩竈市」

「茨城県 龍ヶ崎市」
→「茨城県 竜ヶ崎市」 etc

「東京都 渋谷区 鶯谷町」
→「東京都 渋谷区 鶯谷町」

(3)ひらがな

ひらがな地名は、元の漢字が判明する場合はその漢字による表記を採用する。ただし、郡市町村名のみ。

例：「茨城県 つくば市」
→「茨城県 筑波市」

(4)数字

①「条」「線」「区」「号」

北海道で、算用数字及び漢数字と「条」「線」「区」「号」の接尾辞との組み合わせからなる地名は、漢数字及び算用数字との組み合わせを採用する。

例：「北海道 札幌市北区 北四十条西」
→「北海道 札幌市北区 北40条西」

②「番町」

漢数字と「番町」の接尾辞との組み合わせからなる地名は、算用数字との組み合わせを採用する。

例：「東京都 千代田区 二番町」
→「東京都 千代田区 2番町」

(5)カタカナ

①「エ」

つなぎで「エ」が入っているものは「エ」による表記揺れを採用する。

例：「岩手県 紫波郡矢巾町 又兵エ新田」
→「岩手県 紫波郡矢巾町 又兵エ新田」

②「エ」

「エ」による表記揺れを採用する。

例：「北海道 礼文郡礼文町 船泊村ウエンナイ」
→「北海道 礼文郡礼文町 船泊村ウエンナイ」

(6)長音記号

カタカナ地名の長音記号の有無による表記揺れを採用する。

例：「青森県 八戸市 卸センター」
→「青森県 八戸市 卸センタ」

(7)送りがな

しばしば送りがなが付加・省略されるものは採用する。

例：「千葉県 浦安市 鉄鋼通り」
→「千葉県 浦安市 鉄鋼通」

(8)かな・漢字

しばしばひらがなと漢字の両方が用いられるものは採用する。

例：「神奈川県 横浜市中区 本牧ふ頭」
→「神奈川県 横浜市中区 本牧埠頭」

・別表1 旧字体と新字体との対応

別表1. 1 旧字体

亞,圍,壹,榮,驛,應,櫻,假,會,懷,覺,樂,陷,歡,氣,戲,據,挾,區,徑,
溪,輕,藝,儉,圈,權,嚴,恆,國,齋,雜,蠶,殘,兒,實,釋,從,縱,敘,燒,
條,剩,壤,釀,眞,盡,醉,髓,聲,竊,淺,錢,禪,爭,插,騷,屬,對,滯,擇,
單,斷,癡,鑄,勅,鐵,傳,黨,鬪,屆,腦,廢,發,蠻,拂,邊,瓣,寶,沒,滿,
藥,餘,樣,亂,兩,禮,靈,爐,灣,惡,醫,飲,營,圓,歐,奧,價,繪,擴,學,
罐,勸,觀,歸,犧,舉,狹,驅,莖,經,繼,缺,劍,檢,顯,廣,鑛,碎,劑,參,
慘,絲,辭,舍,壽,澁,肅,將,證,乘,疊,孃,觸,寢,圖,穗,樞,齊,攝,戰,
潛,雙,莊,裝,藏,續,體,臺,澤,膽,彈,蟲,廳,鎮,點,燈,盜,獨,貳,霸,
賣,髮,祕,佛,變,辯,豐,翻,默,與,譽,謠,覽,獵,勵,齡,勞,壓,爲,隱,
衛,鹽,毆,穩,畫,壞,殼,嶽,卷,關,顏,偽,舊,峽,曉,勳,惠,螢,鷄,縣,
險,獻,驗,効,號,濟,冊,棧,贊,齒,濕,寫,收,獸,處,稱,獎,淨,繩,讓,
囑,慎,粹,隨,數,靜,專,踐,織,壯,搜,總,臟,墮,帶,瀧,擔,團,遲,晝,
聽,遞,轉,當,稻,讀,惱,拜,麥,拔,濱,竝,辨,舖,褻,萬,譯,豫,搖,來,
龍,壘,隸,戀,樓

別表1. 2 新字体

亜,围,壹,榮,驛,应,樱,假,会,懷,覺,樂,陷,歡,氣,戲,挾,挟,区,径,
溪,輕,芸,儉,圈,權,嚴,恆,国,齋,雜,蚕,殘,兒,実,釈,從,縱,叙,燒,
条,剩,壤,釀,眞,尽,醉,髓,声,窃,淺,錢,禪,爭,插,騷,属,对,滯,択,
单,断,痴,鑄,勅,鉄,伝,党,鬪,届,腦,廢,発,蛮,払,辺,弁,宝,沒,滿,
藥,余,樣,乱,兩,礼,靈,炉,灣,惡,医,飲,營,円,欧,奥,価,繪,拡,学,
缶,勸,觀,歸,犧,举,狹,驅,莖,經,繼,欠,劍,檢,顯,広,鉦,碎,劑,參,
慘,糸,辭,舍,寿,渋,肅,将,証,乘,疊,孃,触,寢,囧,穗,枢,齐,摂,戰,
潛,双,莊,裝,藏,続,体,台,沢,胆,彈,虫,庁,鎮,点,灯,盜,独,貳,霸,
売,髮,秘,仏,変,弁,豊,翻,黙,与,譽,謠,覽,獵,勵,齡,勞,圧,為,隱,
衛,塩,毆,穩,画,壞,殼,岳,卷,関,顏,偽,旧,峽,曉,勳,惠,螢,鷄,県,
險,献,驗,効,号,濟,冊,棧,贊,齒,湿,写,収,獸,処,称,獎,淨,繩,讓,
囑,慎,粹,隨,數,静,專,踐,織,壯,搜,総,臟,墮,帶,滝,担,団,遅,晝,
聴,遞,転,当,稻,読,惱,拜,麦,拔,浜,並,弁,舖,褻,万,訳,予,揺,来,
竜,壘,隸,恋,楼

・別表2 JIS78 と JIS83 とで字体が入れ替わった文字

別表2. 1 JIS78 第1水準→JIS83 第2水準

鯨,鶯,蠣,攪,竈,灌,諫,頸,礪,藥,靱,賤,壺,礪,檣,濤,邇,蠅,檜,儘,
藪,籠

別表2. 2 JIS78 第2水準→JIS83 第1水準

鯨,鶯,蛎,攪,竈,灌,諫,頸,礪,蕊,靱,賤,壺,礪,檣,濤,迓,蠅,桧,俣,
藪,籠