

文字包摂ガイドライン

デジタル庁デジタル社会共通機能グループ地方業務システム基盤チーム
2024/5

【監修】

笹原宏之（早稲田大学 社会科学総合学術院 社会科学部 教授）

山下真里（熊本大学 大学院人文社会科学部 准教授）

目次

1. 文字包摂ガイドラインの目的.....	1
2. 用語の解説	2
3. 文字同定作業の手順・予備知識・留意点.....	7
4. 包摂可能な部首の字形について.....	9
5. 包摂可能な基準について.....	11
6. 同定先文字の優先設定について.....	25
7. 参考文献	29

1. 文字包摂ガイドラインの目的

地方公共団体においては、地方公共団体情報システムの標準化に関する法律（令和3年法律第40号。以下「標準化法」という。）に基づき、標準化対象事務に係るシステムの標準化に向けて取組を進めていただいているところである。

国では、標準化法第7条第1項に規定する共通する基準のうち、同法第5条第2項第3号イ（電磁的記録において用いられる用語及び符号の相互運用性の確保その他の地方公共団体情報システムに係る互換性の確保に係る事項）に関するものについて規定する「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」を策定している。当該仕様書においては、各標準準拠システムが保持する氏名等の文字セットは行政事務標準文字とすることとしている。

他方、標準化対象事務に係る現行システムでは、地方公共団体や事業者が作成した文字が使用されている場合も多く、標準準拠システムへの移行に当たって、行政事務標準文字に同定する必要がある。

標準化対象事務に係る現行システムで使用している文字（以下「使用文字」という。）のうち同定作業を行う必要がある文字について、同定作業の従事者が同定作業を行う際の参考として、行政事務標準文字に包摂できるかどうかを判断するためのガイドライン（以下、「本書」という。）を設ける。

具体的には、地方公共団体職員が、同定支援ツールで示されている同定結果から同定先を確定させる際に使用することを想定している。本書とは別に地方公共団体で法令等に規定されている内容に沿う形で独自の包摂基準を定めることや既に行っている同定作業を妨げるものではなく、今後、同定作業を進める際の支援策として本書を策定している点に御留意いただきたい。

なお、本書は、地方自治法（昭和22年法律第67号）第245条の4第1項の規定に基づく技術的助言であることを申し添える。

2. 用語の解説

本書で用いる用語について定義する。

2-1 字体

我々が、異なった漢字を別々の漢字であると判別できるのは、それぞれの点画（2-12 点画参照）の数や組合せなど基本となる骨組みや、漢字の音や意味が異なっていると判断するからである。

字体について、国語審議会が 1981 年に答申し内閣告示された常用漢字表の中では、「字体（文字の骨組み）」とある。

2016 年の文化庁文化審議会国語分科会による「常用漢字表の字体・字形に関する指針（報告）」では、「字体とは、同じ文字として様々に肉付けされた数多い個別の文字の形状それぞれから抽出される共通した特徴である。文字の具体的な形状を背後で支えている抽象的な概念と言うこともできる。」とされている。

これらを踏まえ、「字体」とは文字の骨組みであり、個別の文字の形状それぞれから抽出される共通した特徴と考え、本書においては「字体」を「その文字を成り立たせている抽象的な骨組み」とする。

2-2 字形

字形について、「常用漢字表の字体・字形に関する指針（報告）」では、「字体」と対比させつつ「手書き文字、印刷文字を問わず、具体的に出現した個々の文字の形状のことを『字形』と言う。（中略）字形の違いが字体の違いにまで及ばない限り、特定の字形だけが正しく、他は誤りであると判定することはできない。」としている。

これを踏まえて、字体は抽象的な概念であることと比較し、字形は、具体的に出現した形状であると考え、本書においては「字形」を「字体が実物に起こされた際に現れる、具体的な個々の文字の形状」とする。

2-3 文字包摂

本書において「文字包摂」とは、「異なる字形を区別せず、それらを同じ文字として捉えること」とする。

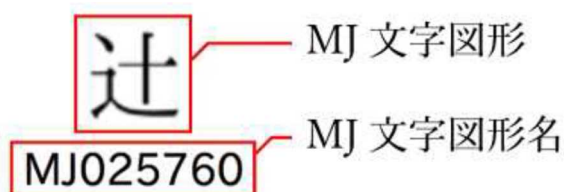
2-4 文字同定

本書において「文字同定」とは、「異なる字形同士が同じ文字かどうかを判断すること」とする。

2-5 IPA 文字情報基盤

「IPA 文字情報基盤」とは、2010 年度電子経済産業省推進費（文字情報基盤構築に関する研究開発事業）によりスタートした、行政で用いられる人名漢字等約 6 万文字の漢字を整備するプロジェクトで整備された文字セット・文字フォントである。現在独立行政法人 情報処理推進機構より信託譲渡を受けた一般社団法人 文字情報技術促進協議会において、成果物として主に「文字情報基盤 文字情報一覧表 (MJ 文字情報一覧表)」と「IPAmj 明朝フォント」を公開し、その普及促進や国際標準化の活動を行っている。成果物は必要に応じてバージョンアップされ、本作業において対象とする MJ 文字情報一覧表のバージョンは、Ver. 006. 02 となり、IPA 文字情報基盤が提供する字形は「明朝体」（書体）である。

図 1 MJ 文字図形と MJ 文字図形名について



2-5-1 MJ 文字図形

本書において、「MJ 文字図形」とは、IPA 文字情報基盤で整備された文字の字形とする（MJ 文字図形と MJ 文字図形名の関係は、「図 1 MJ 文字図形と MJ 文字図形名について」を参照）。

2-5-2 MJ 文字図形名

本書において、「MJ 文字図形名」とは、IPA 文字情報基盤の成果物である MJ 文字情報一覧表で定義される「MJ 文字図形名」とする（MJ 文字図形と MJ 文字図形名の関係は、図 1 を参照）。MJ 文字図形名を用いて MJ 文字図形を一意に特定することができる。

2-6 行政事務標準文字

行政事務標準文字は、標準準拠システムにおいて共通で使う文字・文字セットであり、文字情報基盤を拡張した文字セットである。

ただし、行政事務標準文字であっても、標準化対象事務に関する個別法令等で使用できないと規定されている文字については、その事務で使用することはできないことに留意いただきたい。

2-6-1 GJ 文字図形

本書において、「GJ 文字図形」とは、デジタル庁により行政事務標準文字（文字情報基盤文字を除く。）として整備された文字の字形とする。

2-6-2 GJ 文字図形名

本書において、「GJ 文字図形名」とは、「行政事務標準文字【1.0 版】等について(情報提供)(令和6年3月29日付事務連絡)」添付資料「GJ 文字一覧(1.0 版)」で定義される「図形名」とする。GJ 文字図形名を用いて GJ 文字図形を一意に特定することができる。

2-7 書体

字形について、「常用漢字表の字体・字形に関する指針(報告)」では、「字体を基に具現化された字形には、一定の特徴や様式が現れることがある。そのような、文字に施された一定の特徴や様式の体系を「書体」と言う。(中略)印刷文字の書体としては、明朝体、ゴシック体、教科書体などといった体系が形成されている。書体という用語は、印刷文字のデザインの体系以外に、印刷文字より古くから歴史的に形成されてきた一定の特徴や様式の体系について言う場合にも用いられる。」としている。

本書においては「書体」を「一定のデザイン上の特徴や様式を備えた文字の集合」とする。

2-7-1 手書き書体

本書において、手書き書体とは、手で書かれた書体のことを指す。

手書き書体の字形は、通常、四角形のほか、菱形(◇の形)又は三角形(△・▽の形)等を意識して構成要素を配置する。横画はやや右上がりになり、縦画は下部の線幅が細くなりがちである。

手書き書体は、書道の慣例や書き手が覚えた字体によって様々な字形が生まれる。さらに楷書体であっても「崩す」「続ける」という点画の簡略化が発生し、各手書き書体は様々な字形の幅を持つ。

一方で、コミュニケーションにおいては、郵便物の宛名から選挙の投票用紙にいたるまで、手で書かれた1点1画の有無や形態が違って「同じ文字」だと判断してコミュニケーションがなされてきた。これを漢字の「渡り」という。

2-7-2 印刷書体

本書において印刷書体とは、印刷を行うために生まれ、現代においても印刷や画面表示で利用されている明朝体やゴシック体を指す。

印刷書体は、手書き書体と比較して正方形(□)を意識して構成要素が配置される。明朝体やゴシック体などでは、横画は水平となり、縦画は垂直に下ろして線の幅は一定となる。

特に本書では、文字同定先となる MJ・GJ 文字図形や使用文字の書体である明朝体について定義する。

2-7-2-1 明朝体

図 2 MJ 文字図形による明朝体の例と対応する MJ 文字図形名 左から剛、興、語、摂 2 種



明朝体は、手書き書体と異なったデザイン規則を持つが、元となったのは手書きの楷書体である。楷書体が元になっていることを示す形の一例として、図2のように横画の終筆におかれる三角のウロコがあり、これは楷書体の収筆のおさえをデフォルメしたものである。

明朝体の漢字は水平の細い横画と垂直の太い縦画が直交する特徴を持つが、ごんべん（言）の第一画のように斜めの画は水平（ときには垂直）に再解釈され、デザインし直される（図2の「語」を参照）。

日本では19世紀末に中国・上海経由で伝えられた活字を基に金属活字製作が始まり、以降複数の金属活字鋳造所ができ、明朝体が作られるようになった。その後の明朝体は金属活字に続き、フィルムにより字形が作られる写真植字、デジタル化によるドットで表現されたビットマップフォントやベジェ曲線で表現されたベクターフォントなど、技術に合わせて表示方法を変えつつ、現代においても広く利用されている書体となっている。

同定支援ツールの実証事業では、提出いただいた多くの文字が明朝体であった。

2-8 構成要素

本書における構成要素とは、1つの漢字の字形において、複数の点画によって構成される一定のまとまりを指す。部首も構成要素に含まれる。

例えば「松」は、大きく分けると「木」と「公」に分けられる。さらに、「公」は、「ハ」（八）と「ム」（ム）に分けられる。そうすると「松」の文字構成要素は、「木」、「ハ」、「ム」と定義される。

また、「吉」は、大きく分けると「土」と「口」に分けられる。「土」と「口」は、これ以上分けられない。

※ 注意 「土」は、一般には「十」と「一」とは分離できない。「吉」の文字構成要素は、「土」と「口」と定義される。

「欠」は、これ以上分けられない。

「欠」のように、全体が1つの文字構成要素である例も数多く存在する。

2-9 部首

漢字を構成要素に基づいて分類したときに、その基準となる各部の共通する構成要素を

指す。康熙字典に定義された 214 種類の部首を指す（本来的には部首といえ、その各部の先頭の字を指す）。特に本書においては、部首として例示した構成要素を指す。

2-10 部首番号

本書における部首番号は、康熙字典の部首の通し番号に基づく 1 から 214 までの通し番号を指す。

2-11 点画

本書において点画とは、「漢字の字形を構成する最小の要素である点や線」とする。

3. 文字同定作業の手順・予備知識・留意点

作業者は、本項で示す予備知識を持ち、留意点を理解して、文字同定作業を行う。

3-1 予備知識・留意点

冪等性（べきとうせい：同じ操作を何度繰り返しても、同じ結果が得られること）を高めるために、作業者は本項の留意事項を理解して同定作業を実施する。

また、作業者は、文字の包摂基準を適用する場合に、留意する点がある。

3-1-1 使用文字（印刷書体）と手書き書体の関係性についての理解

前述のとおり、使用文字には手書き書体から明朝体デザイン化された字形が含まれている。

使用文字に手書き書体から明朝体デザイン化された字形が存在する理由として、窓口で受け付けた文字が手書きであり、楷書で丁寧にとの働きかけにもかかわらず、手書き書体毎の個性が字形に表れ、それら手書きの字形を当時の同定先となる文字集合と同定し、同定できなかった文字を、印刷書体の明朝体として再デザインし、その結果、使用文字に至ったと推測される。

3-1-2 字形から見る手書き書体と印刷書体の違い

手書き書体の字形は、菱形（◇）又は三角形（△・▽）を意識して構成要素が配置されることがあるが、印刷書体は通常、正方形（□）を意識して構成要素が配置される。

横画は手書き書体ではやや右上がりになりがちとなるが、印刷書体では水平になる。縦画は手書き書体では下部の線幅が細くなる傾向があるが、印刷書体では垂直に下ろし幅は一定である。このように、手書き書体と印刷書体では、デザインの特徴と点画をつなぐリズムのようなものが異なる。

また、明朝体デザイン化される時に参照された手書きされた字形に対する理解として、手書き書体は明朝体と異なり同一の文字を同一の書体で書き表しても、その字形は継ぎ目のない何段階もの違いが生まれる。

一方で印刷書体である明朝体は、個性を消して、どの職人が描いても同じようになるようにという工業製品の発想から、筆の痕跡をデフォルメしつつ残しながらも、水平と垂直を用いた字形となる。

その場限りで書かれた、字形に幅のある手書き書体の特徴は、外字として明朝体にデザイン化された際に表れ、手書き書体と印刷書体それぞれの特徴が混ざり合った字形となってしまうため、手書き書体と印刷書体の違いを理解することは重要と言える。

3-1-3 字形を表現する道具による同一文字の字形不一致に関する理解

字体が字形となる時の手法は、書体毎に、アナログであれば石や金属に彫る、紙に筆、万年筆、ボールペンで書く、フィルムで作る、デジタルであればドットを打つ、ベジェ曲線で作るなど、それぞれ発生的方式が異なり、それらによって筆運びの痕跡、かすれ、にじみ、データ特有の形が生まれる。

さらに、発生的方式によって、独自の特徴や習慣が発展し、結果として同じ文字であっても時代や技術に応じて、様々に字形の違いが生まれている。

さらに、字体は人それぞれが字形を見て覚えた文字のイメージであるため、すべての人が同一の、統一された書体で文字を書き表して同じ字形を表出させることはできない。これによって、同じ字体を書き表したつもりであっても、人によって書き表そうとしている字形が異なり、やはり細かな違いが生まれる。

上記のような様々な要因により、同じ字体であっても様々な字形が生まれ、さらにそれを読む人によって同じ字体であると認識されない場合が起きる。

上記の字形の違いは、手書きのみならず、デジタル上でドットやベジェ曲線などで作字を行う場合にもその道具の特性が字形に表れる。

3-2 同定支援ツール

同定支援ツールは、使用文字の字形を戸籍システムで保持している文字の字形と比較し、一致度が高い、戸籍システムで保持している文字を導き出し、その文字と行政事務標準文字の結び付けは戸籍における文字整備事業（法務省実施）の成果を活用して候補文字を表示する。また、使用文字の字形を戸籍システムで保持している文字の字形と比較し、類似の字形がない場合は、使用文字の字形と行政事務標準文字の字形を比較し、一致度が高い行政事務標準文字を候補文字として表示する。

3-3 同定手順

令和6年3月公開の同定支援ツール説明書を確認いただきたい。

4. 包摂可能な部首の字形について

部首の字形 214 種については、その字形と画数が定まっているので、文字同定作業を行う上で他の文字構成要素同士と比較より、包摂可能とする範囲を広げている。

比較対象が同じ位置であれば、部首の字形の一部の画数が見た目に少し違って見える場合でも、部首同士は画数の差異とはせず包摂可能とする。

部首の字形の包摂字例（1）※一例である。

No	部首	よみ	部首画数	包摂される部首の字形の例			
1	𠂔	すでのつくり	四画	𠂔	𠂔	𠂔	
2	片	かた	四画	片	片		
3	牙	きば	四画	牙	牙		
4	王	たま (おうへん)	五画	王	王		
5	肉	ぐうのあし	五画	肉	肉		

No	部首	よみ	部首画数	包摂される部首の字形の例			
6	舛	ます	六画				
7	韋	なめしがわ	九画				
8	麥	むぎ	十一画				
9	黽	べん	十三画				
10	鼎	かなえ	十三画				
11	龍	りゅう	十六画				
12	龜	かめ	十六画				

なお、部首の字形に「大小・高低の差」、「長短の差」が含まれる文字について、その文字の部首が変わらない場合は軽微な差として、包摂可能とする。

部首の字形の包摂字例（2）※一例である。

No	MJ文字図形名	イメージ	使用文字	部首字形
1	MJ018093		← →	さ
2	MJ016168		← →	しんによ
3	MJ025572		← →	くるま
4	MJ029813		← →	しれ (まだれ)
5	MJ007338		← →	わかんむり

5. 包摂可能な基準について

一般的に使用される各種の明朝体では、同じ文字でありながら、微細なところで形状の相違の見られるものがある。しかし、それらの相違は、いずれも書体設計上の表現の差、すなわちデザインの違いに属する事柄であり、字体の違いではないと考えられる。字形の異なりを字体の違いと考えなくともよい判断基準を「常用漢字表(平成22年内閣告示第2号)(付)字体についての解説」、総務省「市区町村が使用する外字の実態調査」報告書(2012年3月)、文化審議会国語分科会「常用漢字表の字体・字形に関する指針(報告)」を参考にした上で、本書で定義している。判断基準では、デザイン差を含む字形の差を包摂可能な範囲としている。

以下に包摂可能な基準5-1から5-20とその字例を示す。

区分	基準
5-1	大小・高低
5-2	長短
5-3	接触・非接触
5-4	接触位置
5-5	交差有無
5-6	点か線か
5-7	傾斜方向
5-8	曲げ方・折り方
5-9	曲げ方・跳ね方
5-10	止め・払い
5-11	止め・抜き
5-12	止め・跳ね
5-13	八屋根
5-14	筆おさえの有無
5-15	払い・跳ね
5-16	牙の縦横線
5-17	縦線か横線か
5-18	縦線と左払い
5-19	運筆方向
5-20	類推判断

5-1 大小・高低

文字構成要素の大きさの違い、上下左右の高さの違いについて、文字構成要素の配置の違いにならない範囲は、包摂可能な基準の範囲と定義した。

左右の字例に示した‘青い矢印’の箇所
に違いがある。
主に右の使用文字から左の文字へ包摂
可能な関係性を表している。
左の文字の下には、MJ文字図形名を
表示している。

5-1 大小・高低

文字構成要素の大きさの違い、上下左右の高さの違いについて、文字構成要素の配置の違いにならない範囲は、包摂可能な基準の範囲と定義する。

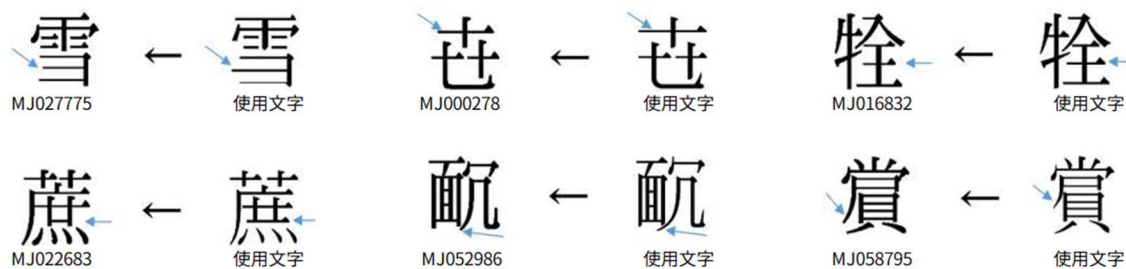
なお、「包摂可能な部首の字形について」で示しているとおおり、配置の差とも取れる範囲であっても、部首の変更を伴わない場合は、包摂可能な基準の範囲としている。

(包摂しない字例)

携 × 携 部首がてへん以外の可能性もあり、包摂不可と判断する。

5-2 長短

文字構成要素内のそれぞれの画の長さについては、包摂可能な基準の範囲と定義する。

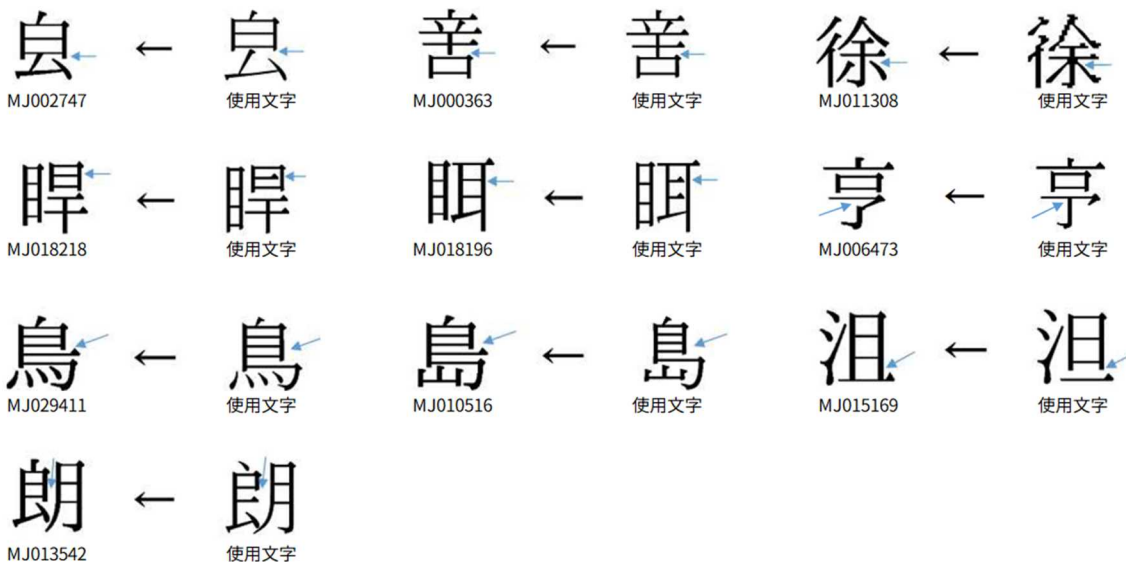


なお、「土」と「土」、「未」と「未」、「吉」と「吉」はそれぞれ単独字形に限り包摂しない文字として取扱うが、これらが別の文字構成要素となる場合は包摂可能な範囲に含む。

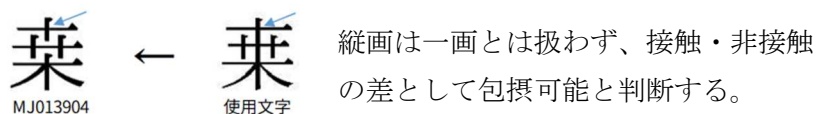


5-3 接触・非接触

文字構成要素内の点画と点画の接触及び非接触について、文字構成要素の画数の違いと異なる範囲は、包摂可能な基準の範囲と定義する。



(例外的に包摂した字例)



(包摂しない字例)



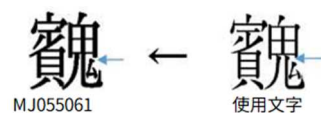
上記は、字体の違いとし、包摂しないと判断する

5-4 接触位置

文字構成要素内の画の接触位置について、文字構成要素の画数の違いとならない範囲は、包摂可能な基準の範囲と定義する。



なお、部首の字形（ここでは「鬼」、部首画数10画）はその字形と画数が定まっていることにより、画数違いとは判断せず、包摂可能な範囲に含む。



(包摂しない字例)



文字構成要素の画数の違いとなり、包摂しないと判断する。

5-5 交差有無

文字構成要素内の画と画の交差、画の突き抜けについて、文字構成要素の画数の違いとならない範囲は、包摂可能な基準の範囲と定義する。



なお、「田」と「冊」はそれぞれ単独字形に限り包摂しない文字として取扱うが、これらが別の文字構成要素となる場合は包摂可能な範囲に含む。



(包摂しない字例)



なお、「山 (やま)」と「巾 (てつ)」の違いは、文字構成要素の画数の違いとならない範囲に含めず、包摂しない。



5-6 点か線か

文字構成要素内の画の点状の表現と線状の表現差について、線状の表現が極端に長さの違いがない範囲は、包摂可能な基準の範囲と定義する。



なお、横線「月」と点「月」、「令」と「令」は、単独字形に限り包摂しない文字として扱うが、これらが別の文字構成要素となっている場合は、線の表現が極端に長さの違いがない範囲に含む。



5-7 傾斜方向

文字構成要素内の画の傾斜について、文字構成要素の画数の違いとならない範囲は、包摂可能な基準の範囲と定義する。



※於の右下がㇿの字形は旧字体とする漢和辞典もあるが、包摂可能とする。

なお、この基準の範囲の一部であるが、筆の運びが「右から左へ」に対して「左から右へ」のような逆向きとなっている場合については、別の包摂基準「5-19 運筆方向の差」と定義する。

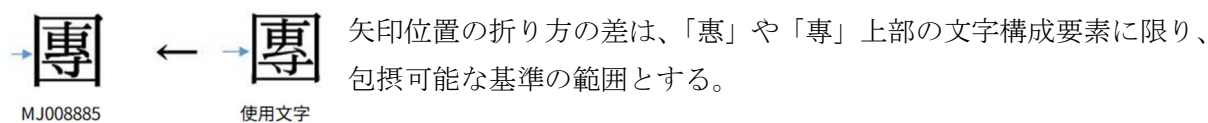
(包摂しない字例)



5-8 曲げ方・折り方

文字構成要素内のいわゆる曲げ方と折り方について、文字構成要素の画数の違いとならない範囲は、包摂可能な基準の範囲と定義する。

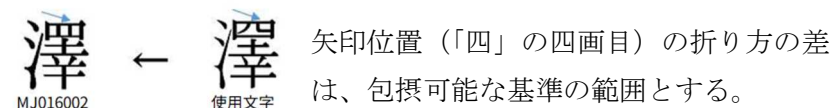
なお、折り方の違いについても画数の違いとはせず、この範囲に含む。



矢印位置の折り方の差は、「恵」や「專」上部の文字構成要素に限り、包摂可能な基準の範囲とする。

※瓜 ム型を1画少ないとみなし、旧字体とする漢和辞典もあるが、包摂可能とする。

なお、「四」や「西」のように、囲われた部分内の右側の画が縦線か、途中で曲がっているかの違いは、文字構成要素内のいわゆる曲げ方と折り方について、文字構成要素の画数の違いとならない範囲に含む。



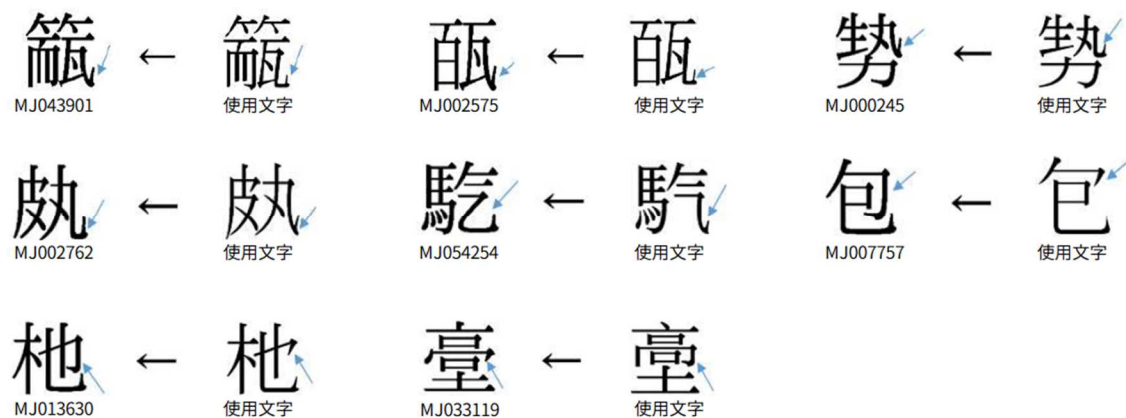
矢印位置（「四」の四画目）の折り方の差は、包摂可能な基準の範囲とする。

例外として、「すいによろ」と「とまた」、「はこがまえ」と「かくしがまえ」のように、部首の字形の違いとなる場合でも文字構成要素内のいわゆる曲げ方と折り方について、文字構成要素の画数の違いとならない範囲に含む。



5-9 曲げ方・跳ね方

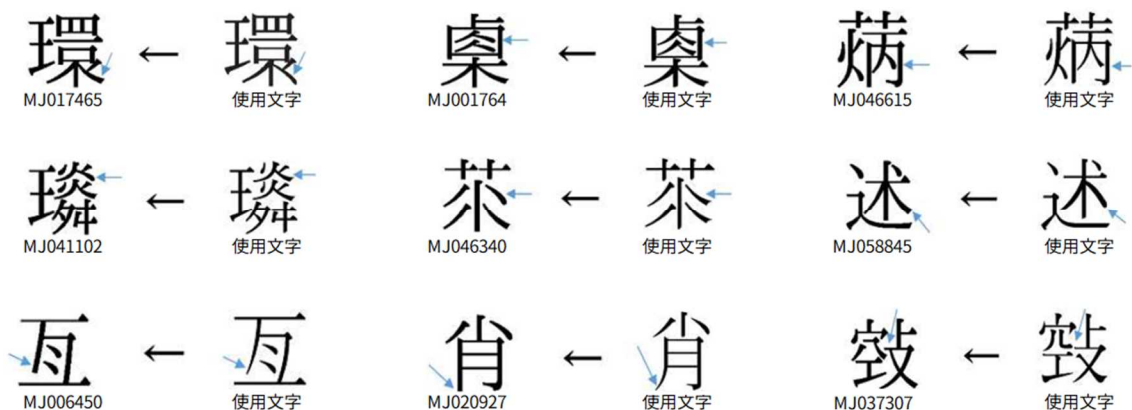
文字構成要素内の曲げ跳ねと跳ねについて、文字構成要素の画数の違いとならない範囲は、包摂可能な基準の範囲と定義する。



5-10 止め・払い

文字構成要素内の止めと払いについて、文字構成要素の画数の違いとならない範囲は、包摂可能な基準の範囲と定義する。

なお、曲げ止めと払いについてもこの範囲に含む。



5-11 止め・抜き

文字構成要素内の止めと抜きについて、文字構成要素の画数の違いとならない範囲は、包摂可能な基準の範囲と定義する。



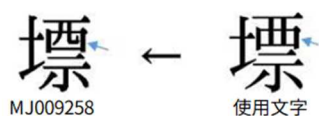
5-12 止め・跳ね

文字構成要素内の曲げ止めと曲げ跳ねについて、文字構成要素の画数の違いとならない範囲は、包摂可能な基準の範囲と定義する。

なお、止めと曲げ跳ねについても文字構成要素の画数の違いとならない範囲に含む。



なお、「四」や「西」のように、囲われた部分内の右側の画が縦線か、途中で曲がって跳ねているかの違いは、文字構成要素の画数の違いとならない範囲に含む。



(包摂しない字例)

氤 × 氤
MJ056871 包摂しない 使用文字

禎 × 禎
MJ018895 包摂しない 使用文字

积 × 积
MJ019045 包摂しない 使用文字

甄 × 甄
MJ017577 包摂しない 使用文字

柳 × 柳
MJ013797 包摂しない 使用文字

上記については、文字構成要素の画数の違いとし、包摂しない。

5-13 八屋根

文字構成要素内のいわゆる八屋根については、包摂可能な基準の範囲と定義する。

睽 ← 睽
MJ018204 使用文字

楡 ← 楡
MJ001782 使用文字

谷 ← 谷
MJ000142 使用文字

院 ← 院
MJ052375 使用文字

5-14 筆おさえの有無

文字構成要素内のいわゆる筆おさえについては、包摂可能な基準の範囲と定義する。

筆おさえの一例として、縦画の始筆にある墨だまり、横画の終筆におかれる三角のウロコなどがある。

霁 ← 霁
MJ027797 使用文字

𠂇 ← 𠂇
MJ057622 使用文字

瞞 ← 瞞
MJ013379 使用文字

堆 ← 堆
MJ052541 使用文字

𠂇 ← 𠂇
MJ000795 使用文字

颯 ← 颯
MJ005635 使用文字

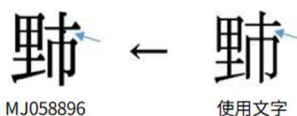
鞫 ← 鞫
MJ053144 使用文字

嵩 ← 嵩
MJ010633 使用文字

穉 ← 穉
MJ019150 使用文字

尢 ← 尢
MJ010306 使用文字

なお、「柿（かき）」と「柿（こけら）」のみ、文字構成要素内の筆おさえの対象外とするが、同様の構成要素で使用されている文字はこの範囲に含む。



なお、人屋根と入り屋根の違いについても文字構成要素内の筆おさえの範囲に含むが、入り屋根と判断できない場合は包摂しない。

(包摂する字例)



※漢和辞典では、「全」が単独の場合も構成要素の場合も、新字体・旧字体という字体差となるが、文字構成要素内の筆おさえの範囲とする。

(包摂しない字例)



なお、「几（ひとあし）」と「几（つくえ）」の違いについても文字構成要素内の筆おさえの範囲に含むが、一部の字形で別の文字と判断し、包摂しない場合がある。

(包摂する字例)



(包摂しない字例)



なお、「山（やま）」と「屮（てつ）」のどちらかが分からない場合は、文字構成要素内の筆おさえの範囲に含めず包摂しない。



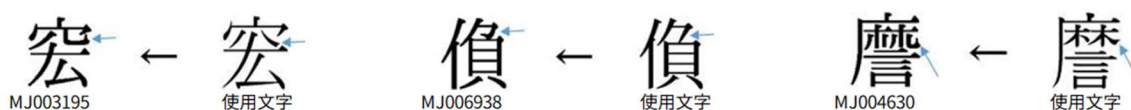
なお、三画くさかんむり、四画くさかんむりに代表される画数の違いが特徴的な文字については、文字構成要素内の筆おさえの範囲に含まない。



5-15 払い・跳ね

文字構成要素内の払いと跳ねについて、文字構成要素の画数の違いとならない範囲は、包摂可能な基準の範囲と定義する。

なお、払いと曲げ跳ねについてもこの範囲に含む。



※上記の「麻」の字形差は、漢和辞典では新字体・旧字体という字体差であるが、文字構成要素の画数の違いとならない範囲とし、包摂可能な基準の範囲とする。

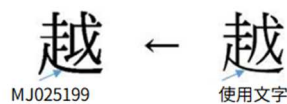
なお、「戌」と「戌」のように跳ねと払いの方向が逆となっている場合、同じ方向でも別の文字と判断する場合は、文字構成要素の違いとし包摂しない。

(包摂しない字例)



ただし、「越」に限り、別の文字の候補がないため、例外的に包摂可能としている。

(例外的に包摂する字例)



5-16 牙の縦横線

「牙」「𠂔」「𠂔」など、文字構成要素内の牙の縦横線について、文字構成要素の画数の違いとはせず、包摂可能な基準の範囲と定義する。



5-17 縦線か横線か

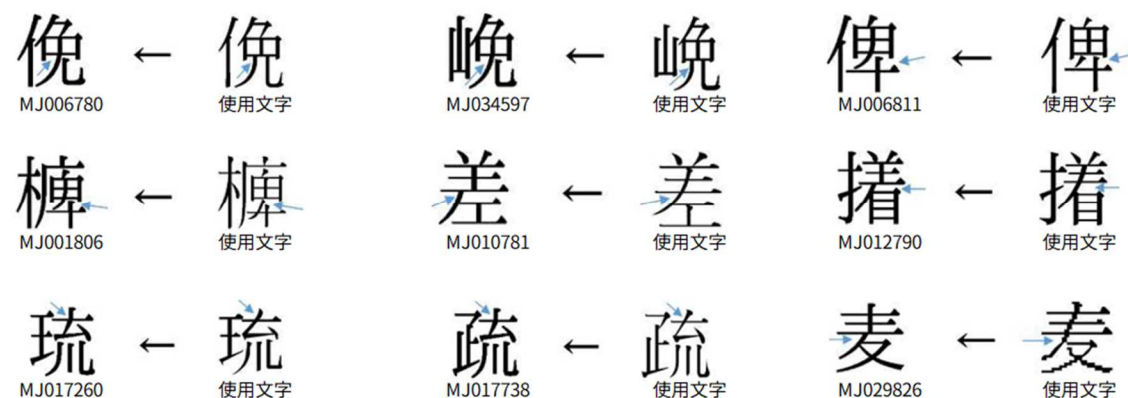
文字構成要素内の縦線か横線かの差によって、別の文字構成要素とならない場合は、包摂可能な基準の範囲と定義する。

基本的には、短い横画とその直下に位置する長い横画の組合せで、片方（主に短い方）の横線が縦線になる。



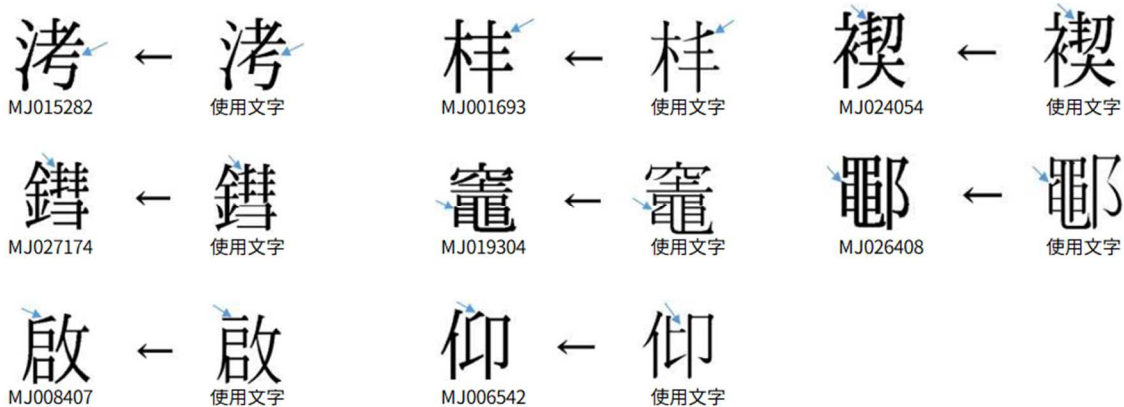
5-18 縦線と左払い

「免」「兎」「卑」「着」「流」「麦」の文字構成要素内の縦線とその下の左払いについて、文字構成要素の画数の違いはあるが、包摂可能な基準の範囲と定義する。



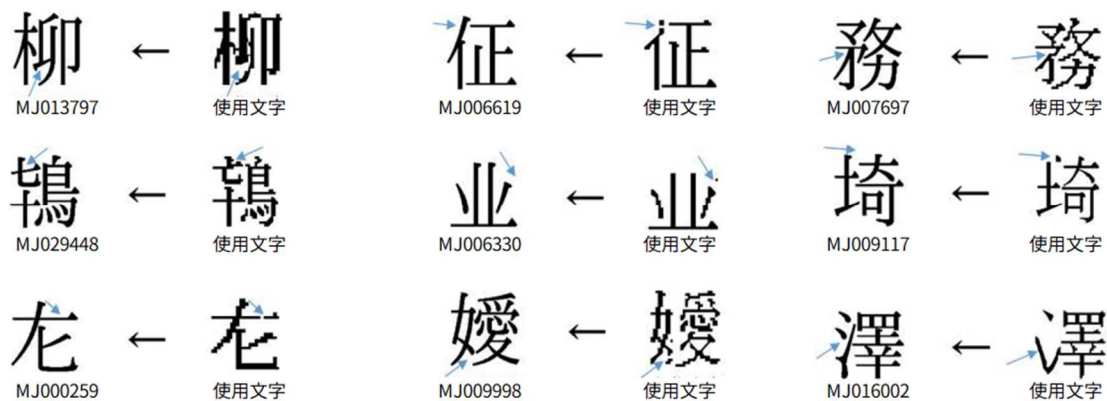
5-19 運筆方向

文字構成要素内の運筆方向の差（筆の運びが「右から左へ」に対して「左から右へ」のような逆向きとなっている場合）については、包摂可能な基準の範囲と定義する。



5-20 類推判断

文字の一部に判断のつかない点や線が含まれている、又は画線のかすれや粗さにより、他の包摂可能な基準に当てはめることはできないが、全体として文字構成要素の違いとは判断しなかった範囲を包摂可能な基準の範囲と定義する。



6. 同定先文字の優先設定について

これまでに示した包摂基準を用いて同定候補文字を洗い出した結果、複数の候補が挙げられた場合は、同定先文字を確定する必要がある。

6-1 複数の同定候補文字が挙がる要因

以下の要因により、同定候補文字が複数挙がるケースが考えられる。

6-1-1 同定先に含まれる同形文字

文字情報基盤に含まれる同形異字や戸籍統一文字に重複して収録されている文字など、同定先文字には同形文字の組合せが含まれている。

6-1-2 包摂区分の優劣

同定支援ツールの同定結果において、同定候補文字が複数挙がり、その包摂区分の組合せによっては優劣がつけられないことがある。優劣がつけられない場合は、6-2 の条件で優劣をつける。

6-2 同定先文字を確定させる条件

以下の優先順位に従い、同定先文字を確定する。

- ① 文字情報基盤文字である
- ② 戸籍統一文字である
- ③ 漢字施策（常用漢字）が定義されている
- ④ 住基ネット統一文字コードがある
- ⑤ 一意な縮退マップに縮退先が存在する
- ⑥ 実装した UCS がある
- ⑦ サロゲート範囲ではない
- ⑧ 戸籍統一文字番号の若い文字を優先する
- ⑨ GJ 文字番号が若い文字を優先する

例えば、同定の優劣が付けられない同定候補文字が2つあった場合、そのうち1つが文字情報基盤文字かつ戸籍統一文字であり、もう一方が文字情報基盤文字ではあるが、戸籍統一文字ではない場合は、前者を同定先として選ぶといったように①の条件から比較を行い、優劣がつけられるまで比較を行う。

【確認方法】

- ① 文字情報基盤文字である。



次の一覧の中にある文字かを確認する。

MJ 文字情報一覧表 Ver. 006. 02

MJ 文字情報一覧表 変体仮名編 Ver. 002. 01

([MJ 文字情報一覧 | 一般社団法人 文字情報技術促進協議会 \(moji.or.jp\)](http://moji.or.jp))

- ② 戸籍統一文字である。



①の一覧において戸籍統一文字番号欄に付番されているものが戸籍統一文字となる。

図形	MJ文字図形名	font	対応する UCS	実装した UCS	実装した Moji_Johoコ レクション	実装した SVS	戸籍統一文 字番号
二	MJ000001	々	U+3005	U+3005			
二	MJ000002	𠄎	U+3006	U+3006			001340
二	MJ000003	𠄎	U+303B	U+303B			001070

- ③ 漢字施策（常用漢字）が定義されている。



①の一覧において漢字施策欄に「常用漢字」と入力があるものが、漢字施策（常用漢字）が定義されている文字となる。

図形	MJ文字図形名	font	対応する UCS	実装した UCS	実装した Moji_Johoコ レクション	実装した SVS	戸籍統一文 字番号	住基ネット 統一文字 コード	入管正字 コード	入管外字 コード	漢字施策
二	MJ006294	一	U+4E00	U+4E00			000010	J+4E00	0x4E00		常用漢字
二	MJ006295	丁	U+4E01	U+4E01			000020	J+4E01	0x4E01		常用漢字

- ④ 住基ネット統一文字コードがある。



①の一覧において住基ネット統一文字コード欄に付番されているものが住基ネット統一文字コードがある文字となる。

図形	MJ文字図形名	font	対応する UCS	実装した UCS	実装した Moji_Johoコ レクション	実装した SVS	戸籍統一文 字番号	住基ネット 統一文字 コード	入管正字 コード	入管外字 コード	漢字施策
二	MJ006294	一	U+4E00	U+4E00			000010	J+4E00	0x4E00		常用漢字
二	MJ006295	丁	U+4E01	U+4E01			000020	J+4E01	0x4E01		常用漢字

⑤ 一意な縮退マップに縮退先が存在する。



[MJ 縮退マップ一意な変換表 | 一般社団法人 文字情報技術促進協議会 \(moji.or.jp\)](https://moji.or.jp/)より

MJSU.1.2.0.json(<https://moji.or.jp/wp-content/mojikiban/lab/xb657/MJSU.1.2.0.json>)をブラウザで開く。

変換先の JIS X 0213 が”1-01-18”となっているものは縮退先が存在しないものである。

```
[
  {
    "MJ文字図形名": "MJ000098",
    "変換先": {
      "JIS X 0213": "1-76-40",
      "UCS": "U+8CB3"
    },
    "備考": "戸籍統一文字情報 親字・正字"
  },
  {
    "MJ文字図形名": "MJ000099",
    "変換先": {
      "JIS X 0213": "1-01-18",
      "UCS": "U+FF3F"
    },
    "備考": "候補なし"
  }
]
```

MJ000098 は縮退先が存在する

MJ000099 は縮退先が存在しない

⑥ 実装した UCS がある。



①の一覧において実装した UCS 欄にコードの入力があるものが実装した UCS があるとなる。

図形	MJ文字図形名	font	対応する UCS	実装した UCS	実装した Moji_Johoコ レクション	実装した SVS	戸籍統一文 字番号	住基ネット 統一文字 コード	入管正字 コード	入管外字 コード	漢字施策
二	MJ006294	一	U+4E00	U+4E00			000010	J+4E00	0x4E00		常用漢字
二	MJ006295	丁	U+4E01	U+4E01			000020	J+4E01	0x4E01		常用漢字

⑦ サロゲート範囲ではない。



①の一覧において対応する UCS 欄に記載されている UCS 符号位置が U+0000~FFFF の範囲内であるものがサロゲート範囲内ではないとなる。

図形	MJ文字図形名	font	対応する UCS	実装した UCS	実装した Moji_Johoコ レクション	実装した SVS	戸籍統一文 字番号	住基ネット 統一文字 コード	入管正字 コード	入管外字 コード	漢字施策
二	MJ000006	七七	U+3402	U+3402			000690	J+3402	0x3402		
二	MJ000007	中	U+3404		3404_E0101		152700	J+3404			

⑧ 戸籍統一文字番号の若い文字を優先する。



①の一覧において戸籍統一文字番号欄を比較する。

図形	MJ文字図形名	font	対応する UCS	実装した UCS	実装した Moji_Johoコ レクシヨ	実装した SVS	戸籍統一文字番号	住基ネット 統一文字 コード	入管正字 コード	入管外字 コード	漢字施策
二	MJ000022	吏	U+342A	U+342A	342A_E0103		003590	J+342A			
二	MJ000023	吏	U+342A		342A_E0101		003600	J+C0B3			

⑨ GJ 文字番号が若い文字を優先する。



「行政事務標準文字【1.0版】等について(情報提供)(令和6年3月29日付事務連絡)」の添付資料「GJ文字一覧(1.0版)」において文字番号欄を比較する。

No	文字番号	図形名	コード	文字	No	文字番号	図形名	コード	文字	No	文字番号	図形名	コード	文字
1	100000	GJ000001	U+100000	𠄎	2	100001	GJ000002	U+100001	𠄏	3	100002	GJ000003	U+100002	𠄐
4	100003	GJ000004	U+100003	𠄑	5	100004	GJ000005	U+100004	𠄒	6	100005	GJ000006	U+100005	𠄓

7. 参考文献

- 『MJ 文字情報一覧』（一般社団法人文字情報技術促進協議会）
- 『文字情報基盤文字情報一覧表（MJ 文字情報一覧表）』（一般社団法人文字情報技術促進協議会）
- 『MJ 縮退マップ一意な変換表』（一般社団法人文字情報技術促進協議会）
- 『康熙字典』（1716）
- 江守賢治『漢字字体の解明』（日本習字普及協会／1966）
- 江守賢治『硬筆毛筆書写検定理論問題のすべて』（日本習字普及協会／1995）
- 江守賢治『楷書の基本 100 パターン』（日本習字普及協会／1987）
- 『常用漢字表の字体・字形に関する指針：文化審議会国語分科会報告』（文化庁／2016）
- 『人名用漢字と誤字俗字関係通達の解説』（日本加除出版／2011）
- 『汎用電子情報交換環境整備プログラム成果報告書』（日本規格協会／2004）
- 『標準校正必携第8版』（日本エディタスクール／2011）
- 江守賢治『解説字体辞典』（三省堂／1986）
- 江守賢治『楷行草筆順・字体字典』（三省堂／1983）
- 芝野耕司『JIS 漢字字典増補改訂』（日本規格協会／2002）
- 児玉幸多編『くずし字解読辞典机上版』（近藤出版社／1978）
- 文化庁文化部国語課『明朝体活字字形一覧（上・下）1820年-1946年（漢字字体関係参考資料集）』（大蔵省印刷局／1999）
- 『常用漢字表』（文化庁／2010）
- 『表外漢字字体表』（文化庁／2000）
- 『JIS X 0213:2004 包摂規準』（日本規格協会／2004）
- 『「市区町村が使用する外字の実態調査」報告書』（総務省／2012）
- 『通用規範漢字表』（中華人民共和国国務院／2013）