製表技術参考資料６

ISSN 1348－7345

市区町村コード自動格付に関するアルゴリズムの研究

N S T A C

Working Paper No.6

平成 19 年 8 月

独立行政法人 統計センター

製表技術参考資料は、独立行政法人 統計センターの職員がその業務に関連して行った製表技術に関する研究の結果を紹介するためのものである。

ただし、本資料の内容は執筆者の個人的見解を示すものであり、機関の見解を示すものではない。

目　次

市区町村コード自動格付に関するアルゴリズムの研究

磯部　祥子*，堀内　泰志*，畠山　昌子*

要　旨

本稿は、平成20年住宅･土地統計調査における、市区町村コード自動格付による審査事務効率化のためのアルゴリズム構築及び、ＯＣＲ機発注仕様書作成に関する基礎資料を得るために、平成17年導入のＯＣＲ機及び平成15年の調査データを用いて検証した結果をとりまとめたものである。

市区町村コード自動格付は、平成15年住宅･土地統計調査、平成17年国勢調査第3次試験調査において試験的に導入されたが、文字認識精度及び格付精度が不十分で本調査で採用するには至らず、新たに導入された平成17年導入機での検証が必要となった。

検証では、調査票と調査データの内容から、主に以下の3点について分析を試みた。1点目は、調査票設計に起因する誤読及び誤格付に関して、調査票記入欄とＯＣＲ機の文字認識範囲との関係について、2点目は、ＯＣＲ機文字認識の際に辞書データとして用いる住所辞書に関して、登録内容の妥当性について調べた。また3点目は、ＯＣＲ機文字認識結果の正誤読及び、自動格付結果の正誤格付の傾向分析から、自動格付アルゴリズムを構築するというものである。

正誤読及び正誤格付の傾向分析においては、文字認識及び自動格付の際にその精度を示す「認識精度向上フラグ」の信頼性を検証し、このフラグを適確に利用することで人手審査と同等の信頼性があることを確認した。また、平成20年住宅･土地統計調査での自動格付に適用するアルゴリズムを、ＯＣＲ機自動格付後に要人手審査としなければならないデータの除外を含めて構築したところ、格付率は 71.18%、正解率は 99.79%となった。さらに、ＯＣＲ機文字認識機能のみを使用し、前述のデータ除外を含む自動格付を今回構築したアルゴリズムで行うこととすると、格付率は 71.91%、正解率は 99.72%にまで上がった。

アルゴリズムの改善に加え、住所辞書及び調査票設計の改善を図ることで、さらなる格付率の向上が期待されるが、引き続き検証すべき点も明らかになった。住所辞書の改善に関しては、登録数を増やすことで認識可能な文字や住所が増える反面、類似した文字や住所も増え誤読が多くなる可能性もあることから、登録内容の検討のため、ＯＣＲ機による調査票読み取りテストを伴う検証が必要であると考えられる。また、調査票設計ではＯＣＲ機の文字認識設定領域を考慮した記入欄の設定等による改善効果について、さらなる検証が必要であると考えられる。

* 統計センター研究センター（E-mail : research@nstac.go.jp）

市区町村コード自動格付に関するアルゴリズムの研究

磯部　祥子，　堀内　泰志，　畠山　昌子

1　研究の目的

調査票に記入された市区町村名の文字をＯＣＲ機によって読み取り、市区町村コードを自動格付する方式は、平成15年住宅･土地統計調査で導入された。しかし、このときに使用したＯＣＲ機（以下、前回機と記す）は、ファイナルテスト時に行われた精度検証での文字認識精度及び格付精度が共に低かったため、自動格付された市区町村コードを人手で全数検査せざるを得ず、結果として省力化の効果は少なかった。その後、平成17年国勢調査第3次試験調査において、これとは別のＯＣＲ機を試用して精度検証を行ったが、その結果も文字認識精度が不十分で、本調査で採用するには至らなかった。

平成17年国勢調査（本調査）においては、これらとは別な新たなＯＣＲ機（以下、現有機と記す）が導入された。そこで、このＯＣＲ機による文字認識精度及び自動格付精度等について検証し、平成20年住宅･土地統計調査における自動格付による審査事務効率化の可能性について検討を行うこととなった。このため、平成15年住宅･土地統計調査の調査票について再度現有機で読み取り、これまでの精度検証で得られた前回機の文字認識状況等と比較することにより文字認識精度がどの程度向上したか把握するとともに、現有機で市区町村コードの自動格付を行う場合のアルゴリズムを構築し、同時に、今後のＯＣＲ機発注の際の仕様書作成のための基礎資料を得る。

2　研究の方法

今回の検証では、平成15年住宅･土地統計調査調査票の広島県、大分県、東京都特別区のデータを用いて検証を行った。

まず、前回機による読み取り結果と、現有機による読み取り結果について、文字認識精度及び自動格付精度の比較検証を行うこととした。また、両機には読み取りの精度の指標となる認識精度向上フラグ付与機能が設けられているが、このフラグを有効利用するための検討、さらに両機について誤認識の傾向分析を行った。

現有機読み取りデータについては、対象県の調査票を新たに読み取ることにより作成し、文字認識精度及び自動格付精度についてさらに詳細な検証を行うこととした。比較検証に用いる前回機読み取りデータについては平成15年住宅･土地統計調査で試験導入された際の精度検証結果データ広島県乙調査票を用いた。

具体的には、広島県、大分県（調査票不備による読み取り不具合のため約半数）、東京都特別区のうち人口移動の多い6区（大田区、世田谷区、杉並区、板橋区、練馬区、江戸川区）の調査票約69,000 枚のうち、「7　前住居」「30　土地の所在地（1～4区画目）」「38　農地･山林の所在地（1～4区

画目）」の 9 項目において、記入または読み取りのあったデータを抽出し、調査票に基づいて人手で記入文字の入力を行った。このうち、前回機との比較に使用した検証データは、広島県乙調査票約 15,000 枚の中から、記入又は読み取りのあったデータを抽出した。

3　本分析で用いる用語

(1)　人手により判断した状況を示すもの

ｱ 調査票の記入状況

記入あり：調査票になんらかの記入がある

記入なし：調査票に記入がない

ｲ 調査票の記入方法

正記入：統計局が想定している記入方法のとおりに記入がされている

誤記入：正記入以外

うち過剰記入：記入の必要のない町丁字等の記入がされている

(正記入の例)

	都道府県欄	市郡支庁欄	区町村欄
政令指定都市	広島(県)	広島(市)	安佐北(区)
東京都特別区	東京(都)		世田谷(区)
	東京(都)	世田谷(区)	
その他の市町村	広島(県)	呉(市)	
	広島(県)	安芸(郡)	府中(町)

(注)　( ) 内は省略可

(誤記入の例)

	都道府県欄	市郡支庁欄	区町村欄	
政令指定都市	広島(県)	広島(市)		区が未記入
	広島(県)		安佐北(区)	市が未記入
東京都特別区	東京(都)	世田谷(区)	奥沢	町丁字が過剰記入
その他の市町村	広島(県)	呉(市)	中央	町丁字が過剰記入
	広島(県)		府中(町)	郡が未記入

(注)　( ) 内は省略可

ｳ 読み取り状況

調査票を人手で確認したときに、ＯＣＲがどのように読み取りをしたかを示す

正読：調査票の記入と同じ内容で読み取りがされている
誤読：調査票の記入と異なる内容で読み取りがされている
うち不読：文字として認識されず、「？」と読み取られている

(2) ＯＣＲ機が判断した状況を示すもの

ｱ 文字認識状況

文字認識あり：ＯＣＲ機によって、文字として読み取りがされている
文字認識なし：ＯＣＲ機によって、文字として読み取りがされていない

ｲ 認識精度向上フラグ

ＯＣＲ機文字認識の際に付与される、知識処理が行われたか否か、また知識処理の結果どのように処理したかを示すフラグ

1（特定）：文字認識段階から認識した
2（類推）：文字認識段階では認識不能であったが、住所辞書を用いて知識処理を行い、住所を類推した
＠（不読）：文字認識において「？」と判断された（「不読」又は「過剰記入」）
△（ブランク）：空白として読み取られた

(3) その他

ｱ 住所辞書

文字認識及び自動格付を行う際にベースとなる住所漢字データを登録したファイルで、文字認識の際の判読不明文字の推測や、格付の際の基礎情報として使用。

統計センターより提示した市区町村一覧を基に、ＯＣＲ受注業者が作成。

ｲ 知識処理

住所辞書に登録された住所の文字を基に、判読不明文字を機械的に推測し文字認識する処理。

また、誤記入データ及び誤読データについても住所辞書に登録されている住所の文字、前後の文字、前後の記入欄等を基に市区町村コードを推測し格付を行う処理。

4　分析結果

認識された文字が調査票の記入内容と合致するかを検証するために、対象データについて人手により調査票記入欄の住所入力を行った。また、調査票記入欄は統計局で想定しているとおりの記入がされていない場合もあるため、それぞれのデータについて正しい記入方式での住所の入力も行った。

図1　検証データ作成の流れ

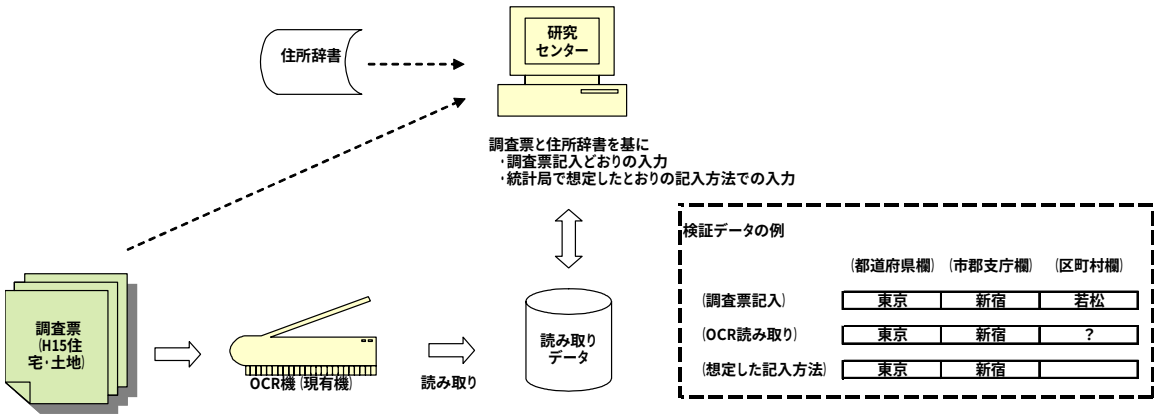


表1　検証データの基本数

	現有機	前回機
調査票枚数	約　65,000　枚（甲乙比／6：1）	約　15,000　枚（乙のみ）
データ件数（項目数）	約　139,286　件	約　135,000　件
うち記入あり件数	20,052　件	1,845　件
欄数（1件につき、都道府県欄、市郡支庁欄、区町村欄の3欄）	417,858　欄	405,000　欄
うち記入あり件数における欄数	60,256　欄	5,535　欄

(注) 項目とは「前住居」「土地の所在地」「農地・山林の所在地」で、甲調査票は1枚につき1項目、乙調査票は1枚につき9項目

なお、現有機の分析では、現有機の住所辞書に登録されていない市区町村のデータ 473 件を除外している（12 ﾍﾟｰｼﾞの4-(3)-ｲにて詳細を説明）。

(1)　ＯＣＲ機の文字認識状況及び格付状況

ア 仕様

現有機の仕様は以下のとおり。

		現有機
住所辞書		平成１７年度の市区町村一覧に基づいて作成され、漢字のみでなく、それぞれに対応したひらがな表記にも対応している
文字認識	調査票(注1)	1文字単位に枠線のある調査票及び、枠線のない調査票、共に対応(注2) 欄全体を読み取り領域とし、領域内に自由に書かれた文字について、1文字1文字の境界を判断
	文字認識と格付	住所辞書に存在する文字とその組合せを基に文字認識を行う。判読不明文字については知識処理によって可能な限り文字認識を行い、住所辞書に登録されていない文字（地域）と判断された場合は不読としている さらに、その文字を基に、知識処理を行いながら市区町村コード格付を行っている
「都道府県」「市郡支庁」「区町村」の末尾文字		末尾文字あり、なしともに正とし、末尾文字の有無に関わらず読み取りを行う 例）末尾文字あり　･･･　東京都 末尾文字なし　･･･　東京
認識精度向上フラグ		文字認識及び格付の際に、各記入欄について知識処理を行ったか否かを示すフラグ

(注1) 読み取りに使用した調査票の住所記入欄は、1文字単位に枠線のある仕様の調査票
(注2) 1文字単位に枠線のある調査票及び、枠線のない調査票は参考1を参照

イ 文字認識状況

現有機の読み取りでは、記入があって読み取りが行われたものは 94.5%、残り 5.5%は汚れや消し残りを文字として認識したものであった（表2）。

表2　正誤記入及び正誤読別、文字認識件数

	文字認識あり	文字認識なし
計	20,049 (100.0)	3
記入あり	18,949 (94.5)	3
正記入	15,098 (75.3)	2
正読	12,551 (62.6)	-
誤読	2,547 (12.7)	-
その他	-	2
誤記入	3,851 (19.2)	1
正読	232 (1.2)	-
誤読	3,619 (18.1)	-
その他	-	1
記入なし	1,100 (5.5)	-

ｳ　平成15年導入の前回機との精度比較

平成15年住宅･土地統計調査で試験導入された際の精度検証結果を用い、この前回機読み取り結果と、今回の検証データを比較検証した。なお、この比較におけるデータは、広島県乙調査票を使用している。

前回機の仕様は以下のとおり。

		前回機
住所辞書		平成15年度の市区町村一覧に基づいて作成されている
文字認識	調査票(注)	1文字単位に枠線のある調査票に対応
		1文字単位に読み取り領域を指定
	文字認識と格付	単漢字辞書にて文字認識を行った後に、読み取られた文字が住所辞書に登録されているかを判断し、市区町村コード格付を行っている
「都道府県」「市郡支庁」「区町村」の末尾文字		末尾文字なしを正とし、末尾文字があればそれを自動的に削除して読み取りを行う
認識精度向上フラグ		文字認識の際に、各記入欄について何文字の知識処理を行ったかを示すフラグ

(注)　読み取りに使用した調査票の住所記入欄は、1文字単位に枠線のある仕様の調査票

前回機と現有機との比較では、広島県乙調査票のみで行ったが、読み取りの状況は以下の各表のとおりとなった。文字認識ありデータのうち、記入なしデータは前回機では 32.6%であったが、現有機では 7.4%にとどまり、現有機では汚れや消し残りを誤読しにくくなったことがわかる（表3）。

表3　記入の有無別文字認識率（前回機と現有機の比較）

	前回機	現有機
読み取り対象	100.0	100.0
記入あり	67.4	92.6
記入なし	32.6	7.4

さらに記入あり文字認識ありデータについて見たところ、正誤読の内訳や自動格付率とその正解率において、どちらの精度が高いという特徴はなかった（表4）。しかし、認識精度向上フラグが妥当に付与されているか否かを人手にて検証したところ、現有機が 99.5～99.7%妥当に付与されていたのに比べ前回機は 81.8%にとどまり、現有機のほうが精度が高いことがわかった（表5）。

表4　正誤記入及び正誤読別、自動格付率とその正解率（前回機と現有機の比較）

	前回機			現有機		
	文字認識ありデータの割合	格付率	正解率	文字認識ありデータの割合	格付率	正解率
計	100.0	90.4	99.3	100.0	86.8	99.3
正記入	73.1	94.5	99.7	67.3	90.7	99.5
正読	67.9	100.0	100.0	56.7	98.9	100.0
誤読	5.2	22.9	81.8	10.6	46.9	93.5
誤記入	26.9	79.0	98.2	25.7	76.8	98.9
正読	6.7	99.2	100.0	2.7	98.0	100.0
誤読	20.2	72.3	97.4	23.0	74.4	98.7

(注1) 文字認識ありデータの割合は、文字認識あり件数を100として割合を算出(%)
(注2) 格付率は文字認識ありデータ数、正解率は格付数をそれぞれ100として割合を算出(%)

表5　認識精度向上フラグが妥当に付与されている割合（前回機と現有機の比較）

［前回機］　平均　81.8　%

	都道府県欄	市郡支庁欄	区町村欄
知識処理あり	93.1	87.2	65.1

［現有機］　平均　99.9　%

フラグの種類	都道府県欄	市郡支庁欄	区町村欄
1	100.0	99.9	99.8
2	100.0	98.5	100.0
@	100.0	100.0	100.0
△	100.0	100.0	100.0

(2)　調査票設計

調査票の設計に関しては、広島県乙調査票すべてについて人手で確認し、記入内容による読み取りへの影響を調べた。

ｱ 記入の影響

記入内容については、広島県乙調査票の「都道府県」等末尾文字への丸囲み、右詰記入等をすべて人手で検証した。正記入データの誤読率が全体では 15.8%であるのに対し、丸囲みは 24.7%、右詰は 29.3%と高くなっており、丸囲みや右詰記入によって正読率が低下することがわかった。これは、都道府県欄末尾文字の丸囲みの一部を読み取り、文字と認識したことが原因と考えられ、事例としては「広島市」を「東広島市」と誤読したものなどがある。

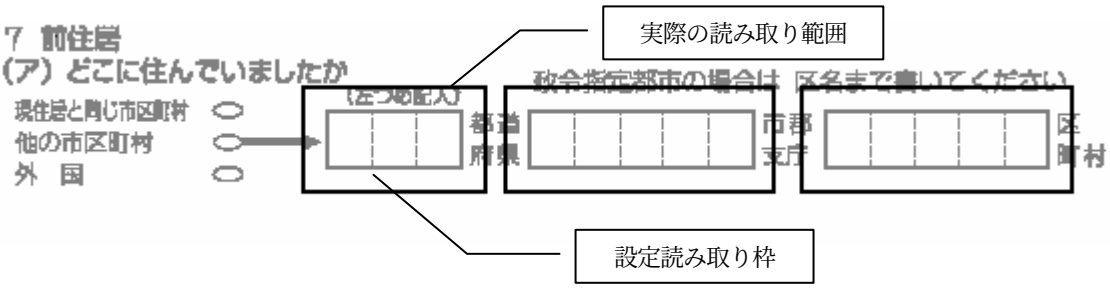
また、不読や誤読の要因として、行書での記入や、記入欄が小さいために文字潰れが生じている例がみられた。文字認識は、住所辞書に登録されている情報や文字認識の際の文字の形によって行われ、文字が潰れていても正読されるものもあるが、文字の潰れがなくなれば誤読や不読は減ると思われ、記入欄の大きさの検討や、文字間隔の設定等の検討は今後の課題であると思われる。

ｲ 調査票における読み取り枠と読み取り範囲

ＯＣＲ機で調査票読み取りを行う際に、フィードのずれが生じることがある。現有機では、このずれに対応するため、設定した枠の外側3mm までを実際の読み取り範囲として文字認識処理を行っている（受注業者に確認済み）。

調査票をみると、外側から3mm の範囲には「都道府県」「市郡支庁」「区町村」等の末尾文字が大きく重なり、誤読の原因となる可能性があることがわかる（図2）。

図2　平成15年住宅･土地統計調査調査票の設定読み取り枠と実際の読み取り範囲



(3)　住所辞書

現有機では、文字認識及び格付の際に住所辞書を使用している。そこで、住所辞書の内容が読み取りの精度に影響すると考え、現有機の住所辞書と、それを使用して読み取られたデータについて検証した。

（現有機の住所に登録されている地域数は 3,731 地域、辞書内データ数としては 15,113 件）

ア 住所辞書に登録されている地域のうち、今回の検証データにおける検証カバー率

本分析で使用した検証データは、住所辞書に登録されている地域のうち 45.2%をカバーしている。なお、1地域に対しデータ数が1件のみの地域を除くと 30%であった。

イ 住所辞書作成時期と調査実施時期

現有機の読み取りでは、平成15年の調査票を平成17年の住所辞書で読み取りを行っているために、平成15年から平成17年にかけて新設された地域については登録がされていない。このような、住所辞書に登録されていない住所については丁寧な文字で記入されていてもすべて不読となり、その件数は 473 件であった。

このことから、現有機では住所辞書の内容が読み取りに大きく影響しており、住所辞書の充実と調査時期に合った住所辞書の差し替えが重要であることがわかった。

ウ ひらがな、カタカナの文字認識状況

ひらがな表記の地域を除いた、ひらがなでの記入状況を検証したところ、記入あり欄数のうち 0.03%にとどまり、そのうちの 25%は誤読されていた。また、文字認識あり欄数のうち、ひらがなに読み取られたものは 0.09%、そのうち 73%が誤読で、その記入内容は、ひらがな以外の記入をひらがなとして読み取っていたものと記入のないものであった（表6）。また、認識精度向上フラグにおけるフラグの誤付与のうち、8%は漢字記入や空欄をひらがなに誤読したものであった（参考2参照）。

現有機の住所辞書には現在、全地域の読み仮名をひらがなで登録しているが、実際の記入にはひらがなでの記入が少なく、また読み取ったものも大部分が誤読であるため、住所辞書にはひらがなを含めないほうが効率的であると考えられる。

表6　ひらがな記入及び、ひらがな読み取りの欄数と誤読

記入あり欄数	48,127	
うち、ひらがなで記入されたもの	16	(0.03)
うち誤読	4	[25]

(注1) 元の地域名がひらがなのみの地域を除く
(注2) ( )は「記入あり欄数」に対する割合(%)
(注3) [ ]は「うち、ひらがなで記入されたもの」に対する割合

読み取りあり欄数	49,473	
うち、ひらがなに読み取られたもの	44	(0.09)
うち誤読	32	[73]
ひらがな以外の記入	24	
記入なし	8	

(注4) 元の地域名がひらがなのみの地域を除く
(注5) ( )は「読み取りあり欄数」に対する割合(%)
(注6) [ ]は「うち、ひらがなで読み取られたもの」に対する割合

参考までに、カタカナでの記入状況も調べたが、カタカナ表記の地域を除いたカタカナでの記入は、全データのうち4件で、欄は6欄であった。

ｴ 市区町村合併による影響

住所辞書に登録されていない地域については正しい読み取りが行われず、調査時期と住所辞書作成時期を合わせることが極めて重要であることは「ｲ 住所辞書作成時期と調査実施時期」で述べた。しかし、近年では市区町村合併が頻繁に行われており、その影響で記入者が旧市区町村名を記入する可能性も考えられるため、その実態について検証した。

まず、市区町村合併が増え始めた平成13年から、検証データの調査時期である平成15年までの市区町村の増減を調べ、また実際に旧市区町村がどの程度記入されたかを調べた。市区町村数の増減は表7のとおりで、81 の市区町村が廃止され、それらを記入していたデータは 114 件（記入あり件数に占める割合は 0.6%、1旧市区町村あたり約 1.4 件）の旧市区町村記入があった。同様に平成17年から平成19年の市区町村の増減を調べたところ、廃止されたのは 1,336 地域、仮に平成13年から平成15年と同じ確率で旧市区町村の記入がされるとすると、平成19年に調査した場合、単純計算して約1,880件（記入あり件数に占める割合は約9.9%）の不読が発生することが想定される。このことから、市区町村合併の多い時期に実施される調査に限って、廃止となった市区町村を住所辞書に含めるという案も考えられる。

しかし、市区町村合併によって増減される市区町村の名称は、1文字違いの似た名称や、県内同名市郡が設定されることが多いため、旧市区町村を住所辞書に含めることで誤読が増える可能性も考えられる。例えば愛知県の「西春日井郡」の2町村は「北名古屋市」となり、既に存在している「名古屋市」と似た名称となる。また、県内同名市郡は平成19年に 58 組存在するが、平成17年以降の旧市区町村を加えると 24 組増え 82 組と大きく増えることとなる。

今回の検証では、住所辞書を差し替えた読み取りテストが行えず、詳細については検証できなかったが、現時点では慎重な対応を取る必要があり、また今後も検証を行う必要があると思われる。

表7　市区町村合併による市区町村数の増減と、旧市区町村記入の状況

住所辞書			調査年度以前の旧市区町村の記入件数	
作成年度（平成）	2年間での市区町村数増減			
	増	減	調査年度	件数（率）
13年～15年	28	81	15年	114 (0.6)
17年～19年	213	1,336	19年	1,880 (9.9)

(注1) 太枠部分は、平成17年から平成19年の市区町村増減に基づいた推定件数
(注2) ( )は記入あり件数を100として割合を算出(%)

また、「前住居」「土地･農地･山林」別に記入あり･文字認識あり件数について比較したところ、記入件数では前住居が 88.5%と圧倒的に多く、またその市区町村合併前の旧市区町村名での記入も、前住居の方がわずかではあるが旧市区町村名を記入している割合が高い（表8）。前住居は5年前の居住地を記入する欄であるため、市区町村合併後の地名を知らずに記入する例は多いことが想像でき、そのため 0.11 ポイントの差が生じていると考えられる。このことを重要視すると、調査項目によって住所辞書を別に用意することで、より効率的な読み取りを行える可能性も考えられる。

この差の要因については、さらに詳細な検証が必要と思われる。

表8　前住居記入欄及び農地･山林記入欄の記入状況の比較

	総数　(率)	前住居　(率)	土地･農地･山林　(率)
記入あり･文字認識あり件数	20,525 (100.00)	18,156 (88.46)	2,369 (11.54)
旧市町村名での記入 (H15.10.1以前)	114 [0.56]	103 [0.57]	11 [0.46]

(注1) ( )は総数を100として割合を算出(%)
(注2) [ ]は記入あり･文字認識あり件数を100として割合を算出(%)

ｵ　北海道の郡と支庁

北海道の郡部は 19 件であったが、そのうち正読されたのは1件で、他の地域に比べ誤読率が著しく高い。そこで、北海道の郡部における記入の状況と住所辞書について詳しく検証したところ、現有機における北海道の住所辞書では、町村部の市郡支庁名について郡ではなく支庁で登録されているにもかかわらず、実際の記入のほとんどが郡名での記入であることがわかった（表9）。

このことから、北海道の住所辞書における町村部の市郡支庁名については、支庁名での登録を郡名での登録に差し替える必要があると考えられる。

なお、長崎県の支庁のデータは郡名記入の1件であったが、平成16年の市区町村合併にて支庁が廃止されているため考慮の必要はない。また、東京都の支庁は郡がないが、支庁が記入されているものは半数以下であった。残りは市郡支庁欄が空欄であるか、島名など誤った記入で、支庁という意識は比較的低いと思われる。

表9　市郡支庁欄における、郡名と支庁名の記入数

	郡部の記入	支庁名	郡名	誤記入
北海道（全域）	19	1	16	2
東京都（大島･八丈･三宅･小笠原）支庁　(注)	16	7	-	9
長崎県（対馬支庁）(注)	1	0	1	0

(注) 参考情報

(4)　自動格付アルゴリズム

自動格付を行うにあたっては、統計分類特有の構造を分析し、個別調査事項の回答内容を基に、適切なアルゴリズムを探求しなければならない。今回のテーマである市区町村名とＯＣＲ機文字認識との関係には様々な特性や傾向があり、これらを踏まえてよりよい自動格付を行うため、そのアルゴリズムを検討した。

ア　認識精度向上フラグ

認識精度向上フラグ（以下「フラグ」と記す）は、前回機では信頼性及び有効性が低く実用には至らなかった。現有機では、フラグが実用にどの程度耐え得るかを改めて検証するため、フラグが人手による文字認識の正誤読とどの程度整合性があるかを検証した。

フラグは、文字認識ありのデータのすべてに、都道府県欄、市郡支庁欄及び区町村欄ごとに付与される。このフラグが妥当に付与されているかを、全てのデータについて人手で検証したところ、次のようなことがわかった。

まず、記入がされていない欄のフラグについてはすべて妥当である△（ブランク）が付与され、記入がないにもかかわらずフラグが付与されるという例はなかった。

また、フラグ1は都道府県欄 82.5%、市郡支庁欄 62%、区町村欄 44.7%と、記入の多い欄ほど多く付与されているが、フラグ2は市郡支庁欄が 7.6%と高く、市郡支庁欄に知識処理が多いことがわかる。フラグ@は区町村欄で約 23.3%と他の2欄より高いが、これは不読の中でも知識処理による過剰記入が多いためである（表 10）。

表 10　3欄のいずれかに記入のあるデータのうち、記入欄、フラグの種類別欄数（割合）

記入	フラグ	計	都道府県欄	市郡支庁欄	区町村欄
計		60,156 (100.0)	20,052 (100.0)	20,052 (100.0)	20,052 (100.0)
あり	1	37,933 (63.1)	16,545 (82.5)	12,425 (62.0)	8,963 (44.7)
	2	2,943 (4.9)	957 (4.8)	1,530 (7.6)	456 (2.3)
	@	8,597 (14.3)	1,817 (9.1)	2,107 (10.5)	4,673 (23.3)
なし	△	10,683 (17.8)	733 (3.7)	3,990 (19.9)	5,960 (29.7)

(注1) フラグ
1：特定　･･･　文字認識段階から認識した
2：類推　･･･　文字認識段階では認識不能であったが、住所辞書を用いて知識処理を行い、住所を類推した
@：不読　･･･　文字認識において「?」と判断された（「不読」又は「過剰記入」）
△：ブランク　･･･　空白として読み取られた
(注2) 各記入欄の計を100として割合を算出(%)

次に、フラグが妥当に付与されているものの割合をみると、フラグ1で正しく付与された正読のものはフラグ1の付与された総件数を 100 すると、99.82%（表 11 の注1の軽微な誤読を除くと 99.73%）であり、信頼性が高いと思われる（表 11）。

表 11　フラグ1が妥当に付与されているものの件数(割合)

	総数	正しく付与	正読	軽微な誤読(注1)	誤って付与（「2」が付与されるべき）
件数	37,933	37,863	37,832	31	70
割合(%)	100.00	99.82	99.73	0.08	0.18

(注1)「髙」を「高」と読み取る、あるいは「さいたま」を「さいたまし」と読み取るなど、「誤読とするには至らないと思われる軽微な読み取りの差異
(注2) フラグ1が付与された総件数を100として割合を算出(%)

フラグ2について見ると、99.46%は正しく付与されていた（表 12）。

しかし、フラグ2が付与された読み取りは知識処理にて類推された結果であるため、正しく付与されたもののうちの正読は 89.98%（表 12 の注1の軽微な誤読を除くと 88.35%）であった。この正読率からもわかるとおり、フラグ2を利用する際には、なんらかの工夫が必要である。

そこで、フラグ2のフラグ誤りは、読み取った内容がすべて「？」であることを考慮し、「？」及び「？」を含む誤読はすべて審査することにし、この 16 件及び 250 件を除外すると、98.32%という比較的高い信頼性となった（表 13）。

表 12　フラグ2が妥当に付与されているものの件数(割合) と、正誤読の内訳

	総数	正しく付与	正読	軽微な誤読(注1)	誤読	明らかな誤読	「？」を含む誤読	誤って付与（「@」が付与されるべき）	「？」に誤読
件数	2,943	2,927	2648	48	281	29	250	16	16
割合(%)	100.00	99.46	89.98	1.63	9.55	0.99	8.49	0.54	0.54

(注1)「髙」を「高」と読み取る、あるいは「さいたま」を「さいたまし」と読み取るなど、「誤読とするには至らないと思われる軽微な読み取りの差異
(注2) フラグ2が付与された総件数を100として割合を算出(%)

表 13　フラグ2のうち「？」を含む誤読を要審査とし、信頼性算出の対象外とした場合の件数（割合）

	総数	正しく付与	正読	軽微な誤読(注1)	誤読	明らかな誤読	要審査（「？」を含む誤読）	正しく付与	誤って付与（「@」が付与されるべき）
件数	2,943	2,677	2,632	48	31	29	266	250	16
割合(%)	-	100.00	98.32	1.79	1.16	1.08	-	-	-

(注1)「髙」を「高」と読み取る、あるいは「さいたま」を「さいたまし」と読み取るなど、「誤読とするには至らないと思われる軽微な読み取りの差異

(注2) フラグ2で正しく付与されたもののうち、「？」を含む誤読を除外した件数を100として割合を算出(%)

また読み取りは、都道府県、市郡支庁、区町村の順に読み取りが行われ、住所辞書に登録された地域数も欄ごとにそれぞれ異なっている。都道府県名は 47 種類の名称しかないが、区町村名は 3,000 種類（平成17年時点）を越える。そのため、表 14 のように、欄ごとに信頼性に違いが出る。フラグ2を利用するときには、読み取られた欄の「？」の有無や、他の2欄のフラグも考慮に入れる必要がある。

なお、フラグが誤って付与されている例を表15に示す。

表 14　フラグ2のうち、誤読の記入欄別件数（割合）

	計	都道府県	市郡支庁	区町村
住所辞書登録数（末尾文字、ひらがなを含む）	13,447 (100.00)	188 (1.40)	2,074 (15.42)	11,185 (83.18)
件数	2,943 (100.00)	957 (32.52)	1,530 (51.99)	456 (15.49)
うち誤読	315 (100.00)	58 (18.41)	81 (25.71)	176 (55.87)
明らかな誤読	29 (100.00)	0 (0.00)	25 (86.21)	4 (13.79)
「？」を含む誤読	270 (100.00)	57 (21.11)	56 (20.74)	157 (58.15)
「？」に誤読	16 (100.00)	1 (6.25)	0 (0.00)	15 (93.75)

(注) それぞれの計を100として割合を算出(%)

表 15　フラグが誤って付与されている例

記入内容				読み取り内容			フラグ						
都道府県欄	市郡支庁欄	区町村欄		都道府県欄	市郡支庁欄	区町村欄	現有機				正しいもの		
京都	京都	下京	→	?	京？都	?	@	2	@	→	@	@	@
岡山	邑久	牛窓	→	岡山	邑久	牛窓	1	1	1	→	1	2	2
広島	佐伯	湯来	→	広島	かも	?	2	1	@	→	2	2	@
愛媛	越智	岡村	→	愛媛	越智	関前	1	1	1	→	1	1	2

(注) 網掛け部分は誤りの箇所

京都の例：　読み取りできない文字があることから、フラグ@が妥当と思われる。

岡山の例：　市郡支庁欄について、調査票には「邑久」とされているが、「邑」が読み取り領域からはみ出してしまったため、「品」に近い文字として読み取られている。しかし、読み取りは正しく「邑久」とされているため、フラグ2が妥当と思われる。

広島の例：　記入内容と異なるものが読み取られていることから、フラグ2が妥当と思われる。
なお、これは「佐伯」を「かも」と読み取っていることからフラグ1の誤読である。

愛媛の例：　結果的に処理はあっているものの、知識処理されていることからフラグ2が妥当と思われる。

検証の結果、フラグが1であるものについて高い信頼性があることがわかった。また、フラグが2であるものについても、機械的に判別が可能である明らかな誤読を除外することにより、フラグ1に比べて精度は落ちるが、充分信頼できるものであると考えられる。

コンピュータで自動格付を行う場合、正読及び正記入を、機械的にどう判断させるかが重要となる。これまでの検証結果から、住所辞書に存在するデータを正記入と判断し、また、信頼性が高く誤読もほぼ発生しないフラグ1及び2と、空白を示すフラグ△のみ付与されているデータを正読と判断することが可能と思われる。これを認識された文字の信頼できる基準とすることで、記入あり・文字認識ありの件数に占める正記入正読の割合は 64.6%となり、現在の市区町村コード自動格付アルゴリズムにおいて自動格付によって処理できるといえる。(表 16)。

なお、3欄中にフラグ@が存在する場合については、フラグ@となった欄の読み取りが不読であるため、自動格付をそのまま信頼することはできないが、後述するアルゴリズムの構築により、人手審査不要の自動格付の割合を高めることが可能である。

表 16　記入あり・文字認識ありデータのフラグ状況別内訳に関する自動格付数 (割合) 及びその正解数 (割合)

	記入あり・文字認識あり件数	計		3欄中にフラグ1、2、△のみ		3欄中にフラグ@が存在	
		自動格付数	正解数	自動格付数	正解数	自動格付数	正解数
計	18,949 (100.0)	15,899 (83.9)	12,270 (64.8)	12,482 (65.9)	12,270 (64.8)	3,417 (18.0)	0 (0.0)
正記入	15,098 (79.7)	13,085 (69.1)	12,235 (64.6)	12,345 (65.1)	12,235 (64.6)	740 (3.9)	0 (0.0)
正読	12,551 (66.2)	12,234 (64.6)	12,234 (64.6)	12,234 (64.6)	12,234 (64.6)	0 (0.0)	0 (0.0)
誤記入	3,851 (20.3)	2,814 (14.9)	93 (0.5)	137 (0.7)	93 (0.5)	2,677 (14.1)	0 (0.0)
正読	232 (1.2)	36 (0.2)	1 (0.0)	36 (0.2)	1 (0.0)	0 (0.0)	0 (0.0)

(注)　記入あり・文字認識あり件数計を100として割合を算出(%)

イ これまで用いられていた格付のチェック審査基準

平成15年住宅･土地統計調査の審査事務に用いた製表事務手続には、所在地の記入状況から市区町村コードを格付する際の付与方法が明記されている（表17）。この付与方法を利用することにより、大部分の文字認識と格付の関係を定義できる。

表 17　平成15年住宅･土地統計調査でのチェック審査の基準（製表事務手続より）

前住居、所在地の記入状況				付与する市区町村コード		
	都道府県名	市郡支庁名	区町村名	都道府県コード	市区町村コード	注意事項
市区町村番号なし						
都道府県のみ	○	×	－（×）	該当コード	VVV	
市区町村番号100番台						
正記入	○	○	○	該当コード	該当コード	
政令市名（特別区を除く）	○	○	×	該当コード	政令指定都市コード	
都道府県名、区名	○	×	○	該当コード	該当コード	
政令市名、区名	×	○	○	該当コード	該当コード	
政令市名（特別区を除く）	×	○	×	該当コード	政令指定都市コード	
区名のみ	×	×	○	該当コード	該当コード	全国で他に同一区名が存在しない
市区町村番号200番台						
正記入	○	○	－	該当コード	該当コード	
市名のみ	×	○	－	該当コード	該当コード	全国で他に同一市名が存在しない
市区町村番号300番台～						
正記入	○	○	○	該当コード	該当コード	
都道府県名、区名郡支庁名	○	○	×	該当コード	該当コード	郡支庁内に町村が1つしかない
都道府県名、町村名	○	×	○	該当コード	該当コード	県内で他に同一町村名が存在しない
郡支庁名、町村名	×	○	○	該当コード	該当コード	
郡支庁名のみ	×	○	×	該当コード	該当コード	郡支庁内に町村が1つしかない
町村名のみ	×	×	○	該当コード	該当コード	全国で他に同一町村名が存在しない

(注) '○' … 記入がある、 '×' … 記入がない、 '－' … 記入の必要がない

ｳ 読み取りの傾向からみた格付アルゴリズム

(ｱ) 県内同名市郡

現有機では、県内に同名の市と郡が存在する場合に、区町村欄が空欄で市として格付されるべきであったものは298 件あったが（参考3参照）、そのうち 297 件が格付されていなかった（格付がされた1件は、「宇佐市」と末尾文字の「市」も記入がされていた）。また、区町村欄が不読である場合には、ほとんどが市として格付されていたが、その中には郡部の町村名が記入され不読となっていたものが2割以上含まれていた。

これに関しては、ＯＣＲ機格付機能の不具合としてシステム修正を行う、もしくはＯＣＲ機格付後の後処理でのフォローが必要である。

例）　市部：　埼玉県　入間市
郡部：　埼玉県　入間郡　三芳町

(ｲ) 県外同名市

県内同名市郡と同様に考えると、都道府県欄が不読であるのに自動格付を行う場合には、県外で同名の市について人手審査を行う必要がある。

例）　東京都　府中市
広島県　府中市

(ｳ) 1郡1町村

1つの郡に町村が1つしか存在しない地域は、現有機に用いた住所辞書に 57 町村存在する。これらに該当するデータは 194 件あったが（参考4参照）、そのうち誤記入または誤読であったものの大部分は格付がされていなかった（表 18）。

また、市郡支庁欄または区町村欄のどちらかが不読となった場合に、もう一方の欄から地域を特定し自動格付を行うと、正解率は 88.57%となった。このようなデータは、自動格付を行い、人手審査を行うアルゴリズムとする。

例）　広島県　深安郡　神辺町

表 18　1郡1町村データの格付状況

		件数	未格付		データの例		
			都道府県コード	市区町村コード	都道府県	市郡支庁	区町村
1郡1町村に該当するデータ		194	2	46	-	-	-
うち誤記入または誤読		59	2	45	-	-	-
市郡支庁欄	記入なし	0	0	0	広島	△	神辺
	不読	3	0	0	広島	?	神辺
区町村欄	記入なし	24	1	24	広島	深安	△
	不読	32	1	21	広島	深安	?

(注) 都道府県コード未格付の2件は、すべて市区町村コードも未格付

(ｴ)　区町村欄に記入のない郡部

都道府県欄が正記入、市郡支庁欄が郡部の正記入で、区町村欄が未記入であった場合、都道府県コードは該当都道府県コードを、市区町村コード（3桁）は不詳（「VVV」）を自動格付することとする。

なお、現有機における格付では、都道府県コード及び市区町村コード（3桁）を特定できない場合に「@」をコードの桁数分付与しているが、対象データの市区町村コード（3桁）は全て「@@@」と格付されていた。

例）　長野県　北佐久郡　軽井沢（20321）　→　長野県　北佐久郡　△（20VVV）

(オ)　記入欄末尾文字丸囲みの影響による誤読

記入欄末尾文字の丸囲みが誤読に関係していることは「(2)　調査票設計」にて述べた。これについて詳細に分析を行ったところ、影響の大きな原因が2点みつかった。

1点目は、「広島市」を「東広島市」と誤読したように、丸囲みを何らかの文字として読み先頭に1文字加えた地域として読まれたもので、2点目は、「三原市」を「庄原市」と誤読したように、先頭の文字に丸囲みが接近して他の文字として読まれたものであった（表 19）。

表 19　丸囲みの影響で誤読された市区町村の、記入と誤読結果の内訳

【広島市を東広島市と誤読（広島県）】

	件数　(割合)
記入が「広島」	2,835 (100.00)
「広島」	2,486 (87.69)
「東広島」	98 (3.46)
「安芸」	1 (0.04)
「こし」	1 (0.04)
「なよろ」	1 (0.04)
「とよた」	1 (0.04)
不読	246 (8.68)
読み取りなし	1 (0.04)

(注1) ほか2地域（うち誤読はそれぞれ1件ずつ）

【三原市を庄原市と誤読（広島県）】

	件数　(割合)
記入が「三原」	98 (100.00)
「三原」	56 (57.14)
「庄原」	22 (22.45)
不読	20 (20.41)

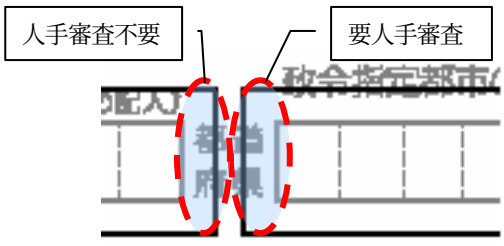
(注2) ほか11地域（うち誤読はそれぞれ1件ずつ）

これら特定の誤読のある市区町村は、自動格付を行った場合の正解率が低い。同様に丸囲みの影響が考えられる市区町村は全国に多数存在し、このような市区町村は自動格付後にすべて人手審査をするアルゴリズムとする。

ただし、次の3点については人手審査を行わないこととする。

1点目は、都道府県欄の末尾文字が左側に位置する「都」と「府」に該当するもので、これらは、図3のとおり丸囲みの影響を受けにくい。このことから、「東京都」「京都府」「大阪府」について人手審査は行わないこととする。

図3　市郡支庁欄について、丸囲みの影響が考えられる末尾文字



2点目は、記入方法の異なるもの同士の誤読で、これらは、読み取った内容を住所辞書とマッチングした際にアンマッチとなり、人手審査が必要であると機械的に判断が可能である。例えば、図4のように市が他の市に誤読された場合には、3欄とも住所辞書と一致するため誤読であるのに正読と判断され問題があるが、図5の政令指定都市が市に誤読された場合、また図6のように郡が他の郡に誤読された場合を考えると、区町村欄でアンマッチとなり、機械的に誤読を見逃すことはない。このような、内容と住所辞書がアンマッチとなる組合せについては、人手審査は行わないこととする。

図4　誤読されたデータの住所辞書マッチング例　その1　（要人手審査）

[市が他の市に誤読された場合]

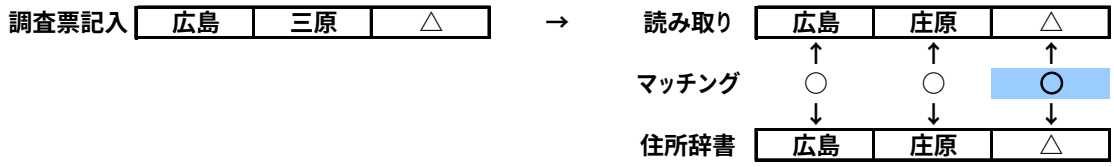


図5　誤読されたデータの住所辞書マッチング例　その2　（人手審査不要）

[政令指定都市が市に誤読された場合]

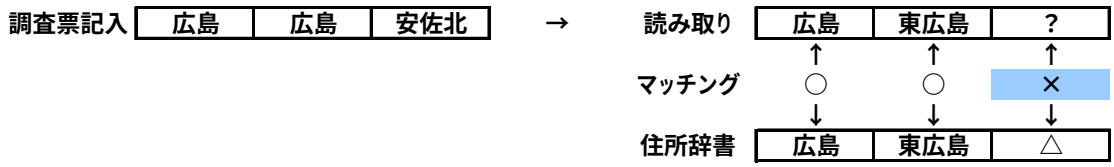
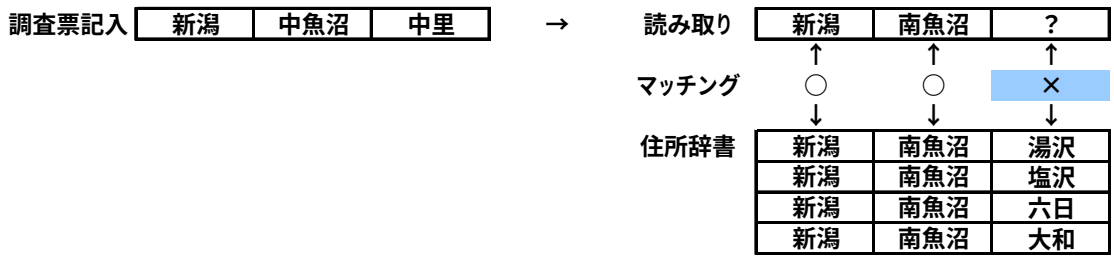


図6　誤読されたデータの住所辞書マッチング例　その3　（人手審査不要）

[郡が他の郡に誤読された場合]



3点目は、都道府県欄が正読の場合の県をまたいだ組合せである。都道府県が正読の場合、その都道府県内の市郡支庁を基準に市郡支庁欄の文字認識を行うため、他県の市郡支庁が読み取られることは稀である。したがって、都道府県欄が正読の場合、県をまたいだ組合せに該当するものは人手審査を行わないこととする。

以上のことより、以下の表 20 の組合せに関してのみ、人手審査を行うこととする。

なお、平成17年住所辞書より作成した、特定の誤読により要人手審査とする市区町村の組合せ一覧を参考5に示す。

表 20　丸囲みの影響で誤読された場合に人手審査を要するものの一覧

			県内		県外		根拠
			市郡支庁名の先頭に1文字追加すると、県内の他の市郡支庁名となる地域 例） 広島県広島市 ↓ 広島県東広島市	市郡支庁名の先頭が1文字異なると、県内の他の市郡支庁名となる地域 例） 広島県三原市 ↓ 広島県庄原市	市郡支庁名の先頭に1文字追加すると、県外の他の市郡支庁名となる地域 例） 愛媛県松山市 ↓ 埼玉県東松山市	市郡支庁名の先頭が1文字異なると、県外の他の市郡支庁名となる地域 例） 青森県八戸市 ↓ 岩手県二戸市	
対象都道府県		都	－	－	－	－	
		道	要	要	要	要	末尾文字が市郡支庁欄寄りにあるため、誤読の影響が強い
		府	－	－	－	－	
		県	要	要	要	要	末尾文字が市郡支庁欄寄りにあるため、誤読の影響が強い
要審査地域と理由となる地域の組合せ	要審査地域	理由となる地域					
	政令指定都市	政令指定都市	－	－	－	－	住所辞書マッチングにて、区町村欄がアンマッチ
		市部	－	－	－	－	住所辞書マッチングにて、区町村欄がアンマッチ
		郡部	－	－	－	－	住所辞書マッチングにて、区町村欄がアンマッチ
	市部	政令指定都市	要（過剰記入データのみ）		要（都道府県欄が不読で、かつ他の2欄が正記入のデータのみ）		住所辞書マッチングにて、区町村欄がアンマッチ
		市部	要	要	要	要	
		郡部	要（区町村欄が未記入の同名市データのみ）		要（都道府県欄が不読、市郡支庁欄が正記入、区町村欄が未記入の同名市データのみ）		住所辞書マッチングにて、区町村欄がアンマッチ
	郡部	政令指定都市	要（区町村欄が不読の1郡1町村データ）		－	－	住所辞書マッチングにて、区町村欄がアンマッチ
		市部	－	－	－	－	住所辞書マッチングにて、区町村欄がアンマッチ
		郡部	要（区町村欄が不読の1郡1町村データ）		－	－	住所辞書マッチングにて、区町村欄がアンマッチ

(注) '要' … 要人手審査、 '－' … 人手審査不要

(カ)　過剰記入として読み取りされた市部

過剰記入として読み取りがされるのは、市部の区町村欄と、東京都特別区の市郡支庁欄または区町村欄のみである。このうち東京都特別区については、正しく過剰記入として読み取られ、なおかつ正しく格付されていた。しかし、市部については、郡部を市と格付しているものが多いことがわかった。

郡部を市部として自動格付したものは、ほとんどが県内同名市郡のもので、県内同名市郡のうちの誤格付の割合は 88.37%に及んでいる。また、県内同名市郡以外は、丸囲みや住所辞書の影響を除くと、郡部を市部と誤格付したものは 0.33%であった（表 21）。

これらのことから、県内同名市郡については人手審査を行わざるを得ず、またそれ以外の誤読についても、丸囲みの影響のある市区町村など原因のわかっているものについては人手審査を行う必要がある。したがって、これらを除いたものを正しく過剰記入と読み取られたものとして、自動格付を行うアルゴリズムとする。

表 21　過剰記入処理されたデータの誤読率とその内訳

過剰記入欄のあるデータの内訳	件数　（割合）
過剰記入として読み取られたもの	2,688　（　100.00　）
県内同名市郡のもの	86　（　3.20　）
都道府県欄、市区町村欄が正読	10　［　11.63　］
都道府県欄、市区町村欄が誤読	76　［　88.37　］
以外	2,481　（　92.30　）
都道府県欄、市区町村欄が正読	2,472　（　91.96　）
都道府県欄、市区町村欄が誤読（注3）	9　（　0.33　）

(注1) [ ]は過剰記入処理されたもののうち同名市郡のもの計を100として割合を算出(%)
(注2) ( )は過剰記入処理されたものの計を100として割合を算出(%)
(注3) 記入欄末尾文字丸囲みの影響及び、住所辞書の影響による誤読を除く

(ｷ) 不読文字「？」を取り除くと正読となる欄

不読には、「？」「？？？」などのように欄全体が不読となったものの他に、「東京？」「さいた？」のように欄の一部が不読となったものがある（表22）。このうち、「東京？」のように正しい地域名の末尾に不読文字「？」が付与されたものについて、「？」を取り除いて検証したところ、都道府県欄は30件すべてが正記入正読であった。また、市郡支庁欄、区町村欄は、他の2欄が正読のものに限るとすべて（市郡支庁欄40件中18件、区町村欄11件中4件）正記入正読であった。

これらのデータは、不読文字「？」を取り除き、正読として扱うアルゴリズムとする。

表22　末尾が不読文字「？」と読まれた例

読み取り			調査票記入			正記入（住所辞書）		
都道府県	市郡支庁	区町村	都道府県	市郡支庁	区町村	都道府県	市郡支庁	区町村
東京？		目黒	東京		目黒	東京		目黒
埼玉？	比企	嵐山	埼玉	比企	嵐山	埼玉	比企	嵐山
大阪？	豊中		大阪	豊中		大阪	豊中	
東京	あきる野？		東京	あきる野		東京	あきる野	
兵庫	神戸？	西	兵庫	神戸	西	兵庫	神戸	西
神奈川	横浜	港北？	神奈川	横浜	港北	神奈川	横浜	港北

(ｸ) 都道府県のみ不読

都道府県欄が不読の場合、他の2欄が正記入であれば自動格付が可能である。現有機では、該当データの大部分が正しく格付されていたが、うち7.3%は格付がされていなかった。

これらのデータは自動格付を行うアルゴリズムとする。ただし、県外同名市及び、丸囲みの影響のある市区町村は要人手審査とする。

(ｹ) 都道府県のみ正読

市区町村コードの5桁すべてを自動格付することができないもののうち、都道府県欄が正読であれば都道府県コードの自動格付が可能である。現有機では、該当データの大部分が正しく格付されていたが、都道府県欄が正読となったうちの 1.8%は未格付となっていた。

これらのデータは都道府県コードについて全て自動格付を行い、市区町村コード（3桁）のみ人手で格付を行うアルゴリズムとする。

㈡　自動格付アルゴリズム

これまでの分析結果について、改めて自動格付アルゴリズムとしてまとめると次のとおりとなる。

[前提条件]

① 住所辞書とマッチングし、3欄とも完全一致したものを「正記入」とする

② 認識精度向上フラグが「1」または「2」である欄を「正読」とする（15 ページ 4-(4)-ｱ）

③ 上記により「正記入」かつ「正読」と判断できた場合は自動格付を行い、これを「人手審査不要」とする

[アルゴリズム詳細]

① 都道府県欄について、正しい地域名の末尾に不読文字「？」が付与されている場合は「正読」と同様に扱う（26 ページ 4-(4)-ｳ-(ｷ)）

② 市郡支庁欄及び区町村欄について、正しい地域名の末尾に不読文字「？」が付与されている場合、他の2欄が正読であれば「正読」と同様に扱う（26 ページ 4-(4)-ｳ-(ｷ)）

③ 記入欄末尾文字丸囲みの影響の考えられる地域については、予めﾘｽﾄを作成し、該当するデータを「要人手審査」とする（22 ページ 4-(4)-ｳ-(ｵ)）

④ 「製表事務手続」で示された表に基づき、自動格付を行う（19 ページ 4-(4)-ｲ）

⑤ 1郡1町村である地域は、都道府県欄が正記入の場合に、市郡支庁欄、区町村欄のどちらかが不読であっても、もう一方の欄から地域を特定し自動格付を行う。ただし、要人手審査とする（20 ページ 4-(4)-ｳ-(ｳ)）

⑥ 都道府県欄、市郡支庁欄が正記入で、区町村欄に記入のない郡部は、都道府県コードを該当都道府県コードに、市区町村コード（3桁）を不詳「VVV」に自動格付を行う。ただし、記入欄末尾丸囲みの影響の考えられる地域は「要人手審査」とする。（21 ページ 4-(4)-ｳ-(ｴ)）

⑦ 過剰記入として読み取りされたもののうち、都道府県欄及び市郡支庁欄が正記入の場合に、市郡支庁欄が市部であれば該当市に自動格付を行う。ただし、県内同名市郡、記入欄末尾丸囲みの影響の考えられる地域及び、市郡支庁欄のフラグが「2」であるものについては「要人手審査」とする（25 ページ 4-(4)-ｳ-(ｶ)）

⑧ 都道府県欄が不読で、他の2欄が正記入である場合は、自動格付を行う。ただし、県外同名市、記入欄末尾丸囲みの影響の考えられる地域は「要人手審査」とする（26 ページ4-(4)-ｳ-(ｸ)）

⑨ 要人手審査としたもののうち、都道府県欄のみ正記入正読である場合は、都道府県コードのみ自動格付を行う(26 ページ4-(4)-ｳ-(ｹ))。

ｴ 格付率及び正解率

自動格付アルゴリズムに従って検証データを自動格付した場合、格付に偏りが見られる「三原･庄原」のような例など正解率の低い事例を要人手審査とすると、人手審査不要データの格付率は 73.53%、その正解率は 99.91%となった（表 23）。

また、自動格付は行うが人手による審査が必要なデータは 8.69%、人手による格付が必要なデータは 17.78%であった。さらに、人手による格付が必要なデータのうちの 7.71%は、都道府県コードのみ自動格付が可能で、その正解率も 100%であった。

表 23 アルゴリズムによる検証データの自動格付結果

アルゴリズムによる自動格付の内訳	格付	（率）	正解	（率）
全データ	20,525	-	-	-
記入あり･文字認識ありデータ	20,049	100.00	-	-
人手審査不要	**14,742**	**73.53**	**14,729**	**99.91**
除外ケース1（ひらがなに読まれたもの）	*43*	*0.21*	*10*	*23.26*
製表事務手続に基づくものすべて				
製表事務手続に基づくもの	12,288	61.29	12,278	99.92
除外ケース2（広島･東広島（県内）などの正データ）	*102*	*0.51*	*101*	*99.02*
除外ケース3（三原･庄原（県内）などの正データ）	*523*	*2.61*	*508*	*97.13*
過剰記入された市部すべて				
過剰記入された市部	1,839	9.17	1,837	99.89
除外ケース4（広島･東広島（県内）など）	*228*	*1.14*	*131*	*57.46*
除外ケース5（三原･庄原（県内）など）	*395*	*1.97*	*374*	*94.68*
除外ケース6（同名市郡（県内））	*88*	*0.44*	*69*	*78.41*
除外ケース7（市郡支庁欄の読み取りフラグが「2」）	*169*	*0.84*	*156*	*92.31*
地域＋？	39	0.19	39	100.00
都道府県？	25	0.12	25	100.00
市郡支庁？	11	0.05	11	100.00
区町村？	3	0.01	3	100.00
不詳に格付してよいもの（製表事務手続で対応済を除く）すべて				
不詳に格付してよいもの（製表事務手続で対応済を除く）	56	0.28	56	100.00
除外ケース8（広島･東広島（県内）など）	*0*	*0.00*	*0*	*0.00*
除外ケース9（三原･庄原（県内）など）	*0*	*0.00*	*0*	*0.00*
都道府県欄が不読でその他の欄が正記入のものすべて				
都道府県欄が不読でその他の欄が正記入のもの	520	2.59	519	99.81
除外ケース10広島･東広島（県内･県外）など）	*9*	*0.04*	*6*	*66.67*
除外ケース11三原･庄原（県内･県外）など）	*0*	*0.00*	*0*	*0.00*
除外ケース12（同名市（県外））	*1*	*0.00*	-	-
要人手審査	**1,743**	**8.69**	**1,481**	**84.97**
1郡1町村（市郡支庁欄、区町村欄のどちらかが不読である場合）	35	0.17	31	88.57
除外ケース2（広島･東広島（県内）などの正データ）	330	1.65	232	70.30
除外ケース3（三原･庄原（県内）などの正データ）	820	4.09	807	98.41
除外ケース4（広島･東広島（県内）など）	228	1.14	131	57.46
除外ケース5（三原･庄原（県内）など）	64	0.32	49	76.56
除外ケース6（同名市郡（県内））	88	0.44	69	78.41
除外ケース7（市郡支庁欄の読み取りフラグが「2」）	169	0.84	156	92.31
除外ケース10（広島･東広島（県内･県外）など）	9	0.04	6	66.67
除外ケース12（同名市（県外））	（除外ケース10に含まれている）			
人手格付	**3,564**	**17.78**	**1,545**	**43.35**
市区町村コードのみ人手格付（都道府県コードのみ格付）	1,545	7.71	1,545	100.00
都道府県･市区町村とも人手格付すべて				
都道府県･市区町村とも人手格付	2,019	10.07	-	-
除外ケース1（ひらがなに読まれたもの）	*43*	*0.21*	*12*	*27.91*

（注1）不読の欄以外の読み取りフラグは全て「1」まはた「2」であること
（注2）人手審査不要のうち斜体文字部分は、要人手審査とするもの
（注3）除外ケース2に関して正解率が高いのは、当研究で用いた検証データに地域の偏りがあるため

(5)　特異データ

検証データの中には、記入があるにもかかわらず、記入があることを認識できていないデータが3件（全体の 0.01%）存在した。

この3件について調査票記入欄左側「所在地について」の択一マーク欄（図6）の記入を確認すると、うち2件については「他の市区町村」にマーク記入があるにもかかわらず、データ上には記入欄同様読み取られていなかった（表 24）。

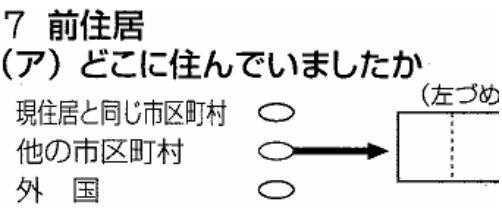
文字の記入、マーク記入ともに記入があることを認識されていないデータについては、「前住居」に関してはデータチェックの際に発見できる可能性があるが、「土地の所在地」及び「農地・山林の所在地」については、人手審査を行う以外の方法でこれらのデータを救うことは困難だと思われる。

表 24　記入はあるにもかかわらず、記入があることを認識できていないデータ

OCR読み取り			調査票記入			正記入			択一マーク	
都道府県	市郡支庁	区町村	都道府県	市郡支庁	区町村	都道府県	市郡支庁	区町村	読み取りの有無	調査票へのマーク記入の有無
			ー						「現住居と同じ市区町村」にマークあり	「現住居と同じ市区町村」にマークあり
			広島	山県	豊平	広島	山県	豊平	マークなし	「他の市区町村」にマークあり
			広島	広島	西	広島	広島	西	マークなし	「他の市区町村」にマークあり

(注) 択一マーク：所在地について「現住居と同じ市区町村」「他の市区町村」のマーク欄

図6　「所在地について」の択一マーク欄（前住居）



5　提言

(1)　調査票設計

ア ＯＣＲ機の読み取り設定領域を考慮し、記入欄の間隔を空けること

イ 末尾文字の丸囲みはしないよう指導を行い、情報として必要であれば「マークにする」、丸囲みをする場合に記入欄へのはみ出しがないよう「囲み誘導線を引く」等の工夫をすること

ウ 調査票上に、楷書での記入を促すこと

エ 漢字記入欄の左詰め記入の指導を強化すること

(2)　住所辞書

ア 可能な限り調査時期に合わせた市区町村一覧より作成する

イ 正式名称以外のひらがな、カタカナ及び、調査時点以外の地域名は含めない
ただし、末尾文字の「都道府県」「市郡」「区町村」が付与された名称は含む
例:「東京」「新宿区」「ひたちなか」「ニセコ」 ･･･ ○
「ﾄｳｷｮｳ」「しんじゅくく」 ･･･ ×

ウ 北海道の支庁は辞書から除き、郡のみとする

エ 末尾文字丸囲みの影響による誤読に関する要審査ﾘｽﾄ作成
下記についてそれぞれﾘｽﾄを作成する
・市郡支庁名の先頭に1文字追加すると県内の他の市郡支庁名となる地域
例 ：「広島県広島市」と「広島県東広島市」
・市郡支庁名の先頭に1文字追加すると県外の他の市郡支庁名となる地域
例 ：「愛媛県松山市」と「埼玉県東松山市」
・市郡支庁名の先頭が1文字異なると県内の他の市郡支庁名となる地域
例 ：「広島県三原市」と「広島県庄原市」
・市郡支庁名の先頭が1文字異なると県外の他の市郡支庁名となる地域
例 ：「青森県八戸市」と「岩手県二戸市」

(3)　現有機にて読み取られたデータを自動格付するためのアルゴリズム

現有機を利用して自動格付を行う場合、現時点ではその格付結果をすべてそのまま利用することはできない。そこで、当研究で得られたアルゴリズムを用いて後処理を行い、利用が可能な水準まで正解率を上げるための方法として、下記2案の自動格付アルゴリズムを提案する。

ア 現有機による自動格付を活かしたアルゴリズム

現有機の文字認識機能及び自動格付機能を利用し、自動格付が行われたデータに対して後処理にて正解率の低い要人手審査データを抽出するアルゴリズム。

(ア) 自動格付結果より要人手審査データを抽出するアルゴリズム

下記「前提条件」に該当するデータを自動格付済みデータとする。

ただし、下記「アルゴリズム詳細」に示す条件に該当するデータは「要人手審査」とする。

[前提条件]

① ＯＣＲ機読み取り時に、都道府県コード及び市区町村コードが格付されたもの

② 都道府県コードが格付され、市区町村コードが「@@@」と格付されたもの

[アルゴリズム詳細]

① 記入欄末尾文字丸囲みの影響の考えられる地域のリストを予め作成し、該当する市区町村に自動格付されたものについては「要人手審査」とする（22 ページ 4-(4)-ウ-(オ)）

② 過剰記入として読み取りされ、市として自動格付されたもののうち、県内同名市郡、記入欄末尾丸囲みの影響の考えられる地域及び、市郡支庁欄の認識精度向上フラグが「2」であるものについては「要人手審査」とする。また、市以外に自動格付されたものはすべて「要人手審査」とする（25 ページ 4-(4)-ウ-(カ)）

③ 市区町村コードが「@@@」に自動格付されたもののうち、以下の2点に該当するもの以外は「要人手審査」とする。1点目は、都道府県欄が正読で他の2欄に記入のないもの、2点目は、都道府県欄、市郡支庁欄が正読で区町村欄に記入のないもの。ただし、記入欄末尾文字丸囲みの影響の考えられる地域については、すべて「要人手審査」とする（20 ページ 4-(4)-ウ-(ア)）。

④ 都道府県が不読かつ他の2欄が正記入で自動格付されたもののうち、県外同名市及び、記入欄末尾文字丸囲みの影響の考えられる地域については「要人手審査」とする（26 ページ 4-(4)-ウ-(ク)）

⑤ 都道府県欄が不読で自動格付されたもののうち、市郡支庁欄か区町村欄のいずれか1つが不読のものについては「要人手審査」とする（26 ページ 4-(4)-ウ-(ク)）

⑥ 1郡1町村の町村に自動格付されたもののうち、3欄中いずれかの欄に不読のあるものについては「要人手審査」とする（20 ページ 4-(4)-ウ-(ウ)）

(ｲ) 格付率及び正解率

現有機によって自動格付された結果から、格付に偏りが見られる事例及び正解率の低い事例を人手審査不要データから除外し、要人手審査とすると、表 25 のとおりとなった。

人手審査不要データの格付率は 72.19%、その正解率は 99.68%で、後述のＯＣＲ文字認識機能のみを利用した自動格付結果（格付率及び正解率は 28 ページ表 23）と比較すると、ともに低い結果となった。

表 25　アルゴリズムによる検証データの自動格付結果（現有機自動格付後）

アルゴリズムによる自動格付の内訳	格付	（率）	正解	（率）
全データ	20,525	(-)	-	(-)
記入あり･文字認識ありデータ	20,049	(100.00)	-	(-)
現有機による自動格付	**18,176**	**(90.66)**	**15,900**	**(87.48)**
人手審査不要	**14,473**	**(72.19)**	**14,426**	**(99.68)**
除外ケース1（ひらがなに読まれたもの）	*29*	*(0.14)*	*10*	*(34.48)*
過剰記入され格付されたもの				
過剰記入された市部	1,844	(9.20)	1,837	(99.62)
除外ケース4（広島･東広島（県内）など）	*226*	*(1.13)*	*131*	*(57.96)*
除外ケース5（三原･庄原（県内）など）	*387*	*(1.93)*	*374*	*(96.64)*
除外ケース6（同名市郡（県内））	*77*	*(0.38)*	*64*	*(83.12)*
除外ケース7（市郡支庁欄の読み取りフラグが「2」）	*264*	*(1.32)*	*154*	*(58.33)*
市区町村コードが不詳に格付されたもの				
都道府県欄が正記入、他の２欄が記入なし	66	(0.33)	66	(100.00)
都道府県･市郡支庁欄が正記入、区町村欄が記入なし	43	(0.21)	43	(100.00)
除外ケース8（広島･東広島（県内）など）	*4*	*(0.02)*	*0*	*(0.00)*
除外ケース9（三原･庄原（県内･県外）など）	*9*	*(0.04)*	*0*	*(0.00)*
その他				
除外ケース10（その他）	*1,934*	*9.65*	*283*	*(14.63)*
都道府県欄が不読で格付されたもの				
うち、その他の欄が正記入のもの	482	(2.40)	479	(99.38)
除外ケース11（広島･東広島（県内･県外）など）	*7*	*(0.03)*	*4*	*(57.14)*
除外ケース12（三原･庄原（県内･県外）など）	*0*	*(0.00)*	*0*	*(0.00)*
除外ケース13（同名市（県外））	*3*	*(0.01)*	*0*	*(0.00)*
うち、他の１欄が不読のもの				
除外ケース14（都道府県欄と他１欄が不読）	*221*	*(1.10)*	*206*	*(93.21)*
その他	12,038	(60.04)	12,001	(99.69)
除外ケース2（広島･東広島（県内）など）	*102*	*(0.51)*	*101*	*(99.02)*
除外ケース3（三原･庄原（県内）など）	*582*	*(2.90)*	*570*	*(97.94)*
除外ケース15（1郡1町村（いずれかの欄に不読がある場合））	*34*	*(0.17)*	*13*	*(38.24)*
要人手審査	**3,671**	**(18.31)**	**1,533**	**(41.76)**
除外ケース2（広島･東広島（県内）などの正データ）	102	(0.51)	101	(99.02)
除外ケース3（三原･庄原（県内）などの正データ）	582	(2.44)	570	(97.55)
除外ケース4（広島･東広島（県内）など）	226	(1.13)	131	(57.96)
除外ケース5（三原･庄原（県内）など）	387	(1.94)	374	(96.39)
除外ケース6（同名市郡（県内））	77	(0.37)	64	(85.33)
除外ケース7（市郡支庁欄の読み取りフラグが「2」）	264	(0.65)	154	(96.15)
除外ケース10（その他）	1,934	(9.65)	283	(14.63)
除外ケース11（広島･東広島（県内･県外）など）	7	(0.03)	4	(57.14)
除外ケース13（同名市（県外））	3	(0.01)	0	(0.00)
除外ケース14（都道府県欄と他１欄が不読）	221	(1.10)	206	(93.21)
除外ケース15（1郡1町村（いずれかの欄に不読がある場合））	34	(0.17)	13	(38.24)
人手格付	**1,905**	**(9.50)**	-	(-)
人手格付のものすべて				
除外ケース1（ひらがなに読まれたもの）	*29*	*(0.14)*	*10*	*(34.48)*

(注1) 人手審査不要のうち斜体文字部分は、要人手審査とするもの
(注2) 除外ケース2に関して正解率が高いのは、当研究で用いた検証データに地域の偏りがあるため
(注3) 複数の除外ケースに重複したデータが存在する

ｲ 認識された文字から自動格付を行うｱﾙｺﾞﾘｽﾞﾑ

現有機の文字認識機能のみを利用し、自動格付を後処理にて行う場合のｱﾙｺﾞﾘｽﾞﾑ。

詳細は 27 ページ 4-(4)-ｳ-(ｺ)「自動格付ｱﾙｺﾞﾘｽﾞﾑ」及び、28 ﾍﾟｰｼﾞ 4-(4)-ｴ「格付率及び正解率」を参照。

認識された文字から自動格付を行うｱﾙｺﾞﾘｽﾞﾑでの格付率は 73.53%、正解率は 99.91%で、ＯＣＲ機自動格付を利用した場合より、ともに高い結果となった。

ＯＣＲ自動格付を利用した場合のｱﾙｺﾞﾘｽﾞﾑでは、現有機の自動格付機能不具合に関するフォローも同時に行う必要があり、当ｱﾙｺﾞﾘｽﾞﾑに比べ処理が複雑となる。また、ＯＣＲ自動格付を利用した場合の要人手審査「除外ケース 10（その他）」に関しては、自動格付は行われているものの格付がされているのは都道府県ｺｰﾄﾞのみで、実際には人手格付を要するのと同等の内容となっている。

これらのことから、現有機による自動格付では、認識された文字から自動格付を行うｱﾙｺﾞﾘｽﾞﾑの利用を推奨する。

⑷　その他

これまで述べた事柄以外に、今後のＯＣＲ機発注基本仕様作成に資すると考えられる事柄についてまとめた。

ｱ 文字認識及び自動格付時の付与フラグ

現有機の認識精度向上フラグ付与機能をさらに有効利用するため、以下のようなフラグ仕様を提案する。

① 文字認識フラグ

知識処理の有無を付与。

知識処理を行った
知識処理を行わなかった

② 文字認識知識処理フラグ

文字認識の際に知識処理を行ったものについて、知識処理の内容別に付与。

フラグの種類	内容	読み取り結果
類推	知識処理によって文字を読み取った	読み取った文字を出力
不読	知識処理によって文字として読み取ることができなかった	「?」を出力
過剰記入	知識処理によって過剰記入処理がされた	読み取り文字を出力 読み取れなかった場合は「?」を出力
空白	記入欄がなかった	文字出力なし

③ 類推フラグ

文字認識知識処理フラグに類推が付与されたものについて、その根拠を付与。

類推の種類	内容
文字	欄内の他の文字より類推
欄	他の欄より類推

④ 類推文字の特定情報

どの文字が類推されたのか判別できるよう、先頭から何文字目かの情報を付与

例：　１文字目なら「１」、３文字目なら「３」、など

⑤ 格付知識処理フラグ

格付の際に知識処理を行ったものについて、都道府県コード、市区町村コード（３桁）別、知識処理の内容別に付与。

フラグの種類		内容	読み取り結果
都道府県コード	特定	都道府県コードの格付を行った	該当コードを格付
	類推	知識処理によって都道府県コードの格付を行った	該当コードを格付
	格付不能	知識処理を用いても、格付することができなかった	？？を格付
	未格付	読み取りがなかった	@@を格付
市区町村コード	特定	市区町村コードの格付を行った	該当コードを格付
	類推	知識処理によって市区町村コードの格付を行った	該当コードを格付
	格付不能	知識処理を用いても、格付することができなかった	？？？を格付
	空白	読み取りがなかった	@@@を格付

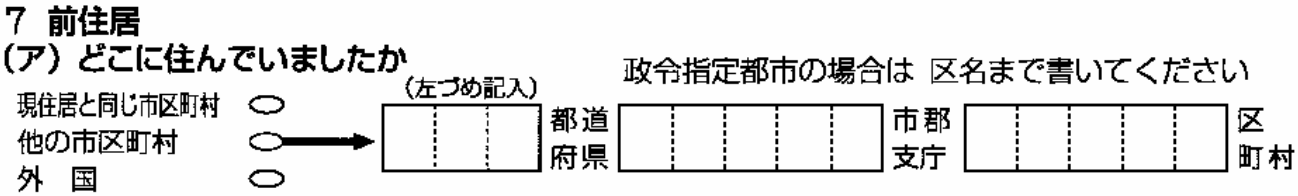
ｲ 文字認識例、格付例の提示

文字認識及び自動格付の方法を明確にするために、正記入誤読の場合の文字認識結果及び格付結果と、誤記入の場合の文字認識結果及び格付結果について、それぞれ具体的な例示をＯＣＲ機受注業者に示す必要があると思われる。

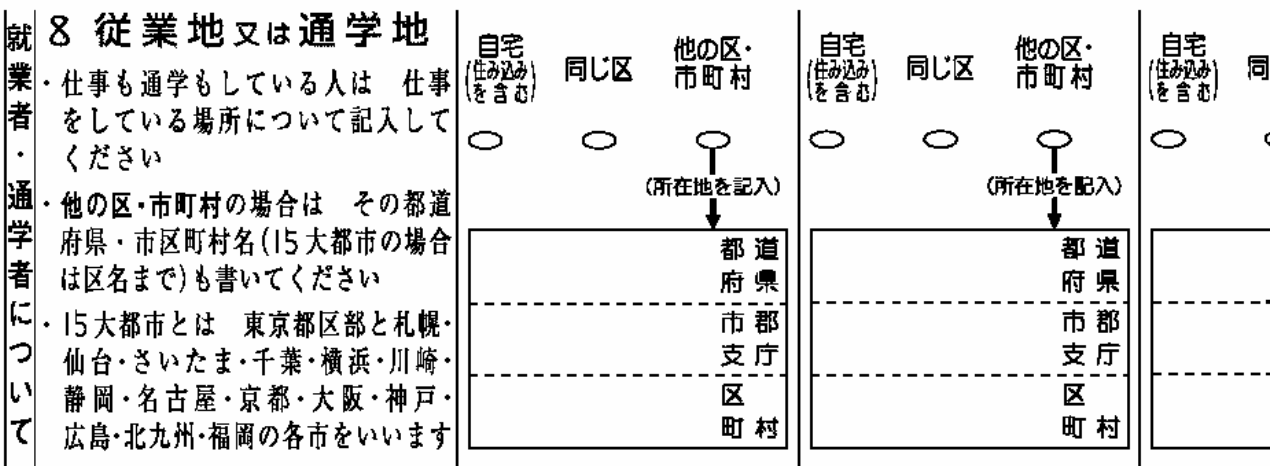
なお、現有機での文字認識結果及び格付結果については、参考6に示す。

参考 1

【平成15年住宅･土地統計調査調査票　市区町村名記入欄(抜粋)】



【平成17年国勢調査調査票　市区町村名記入欄(抜粋)】



参考　2

【ひらがなで記入されているもの一覧】

調査票記入	OCR読み取り	正記入	フラグ
さくら	さくら	さくら	1
あいち	あいち	あいち	1
こうなん	こうなん	こうなん	1
のばく	？		@
あきるの	あきるの	あきるの	1
さくら	さくら	さくら	1
とち木	？	栃木	@
ちがさき	ちがさき	ちがさき	2
つるがしま	つるがしま	つるがしま	1
つるがしま	つるがしま	つるがしま	1
わらび	わらび	わらび	1
ふかや	？？？？	ふかや	@
すずか	すずか	すずか	1
ひろしま	ひろしま	ひろしま	1
たかはし	？	たかはし	@
ふくやま	ふくやま	ふくやま	1

(注) もとがひらがな表記の地域を除く

【ひらがなで読み取られているもの一覧】

・正読

調査票	読み取り	正記入	フラグ
さくら	さくら	さくら	1
あいち	あいち	あいち	1
こうなん	こうなん	こうなん	1
あきるの	あきるの	あきるの	1
さくら	さくら	さくら	1
ちがさき	ちがさき	ちがさき	2
つるがしま	つるがしま	つるがしま	1
つるがしま	つるがしま	つるがしま	1
わらび	わらび	わらび	1
すずか	すずか	すずか	1
ひろしま	ひろしま	ひろしま	1
ふくやま	ふくやま	ふくやま	1

・誤読

調査票	読み取り	正記入	フラグ
長野	なら	長野	2
	かこ		2
	つ		1
入間	にいざ	入間	2
岡山	？いみ	岡山	2
新吉町	あおば	港北	2
	い？い		2
新潟	あらい	新潟	2
市川	ゆり	市川	2
神戸	すもと	神戸	2
宇部	やない	宇部	1
	っ		1
徳山	みね	徳山	2
甲田	しか	甲田	2
広島	こし	広島	2
	すき		2
	こし		2
広島	なよろ	広島	2
佐伯	かも	佐伯	1
	もじ		2
広島	とよた	広島	2
鳴門	いこま	鳴門	2
広島	なら	広島	2
	にし		1
吉浦	おお？		2
佐伯	かも	佐伯	1
安浦	うちこ	安浦	2
木江	おき？	木江	2
玖珠	なす	玖珠	1
宇佐	はやみ	宇佐	2
大分	うさし	大分	2
北九州	みい	北九州	1

参考　3

【県内同名市郡一覧（平成１７年住所辞書)】

都道府県		市郡支庁	
ほっかい	道	あばしり	市
ほっかい	道	あばしり	郡
ほっかい	道	いしかり	市
ほっかい	道	いしかり	郡
ほっかい	道	くしろ	市
ほっかい	道	くしろ	郡
ほっかい	道	ねむろ	市
ほっかい	道	ねむろ	郡
ほっかい	道	るもい	市
ほっかい	道	るもい	郡
北海	道	釧路	市
北海	道	釧路	郡
北海	道	根室	市
北海	道	根室	郡
北海	道	石狩	市
北海	道	石狩	郡
北海	道	網走	市
北海	道	網走	郡
北海	道	留萌	市
北海	道	留萌	郡
いわて	県	にのへ	市
いわて	県	にのへ	郡
岩手	県	二戸	市
岩手	県	二戸	郡
ふくしま	県	そうま	市
ふくしま	県	そうま	郡
福島	県	相馬	市
福島	県	相馬	郡
いばらき	県	かしま	市
いばらき	県	かしま	郡
茨城	県	つくば	市
いばらき	県	つくば	市
いばらき	県	つくば	郡
いばらき	県	ゆうき	市
いばらき	県	ゆうき	郡
茨城	県	結城	市
茨城	県	結城	郡
さいたま	県	いるま	市
さいたま	県	いるま	郡
さいたま	県	ちちぶ	市
さいたま	県	ちちぶ	郡
埼玉	県	秩父	市
埼玉	県	秩父	郡
埼玉	県	入間	市
埼玉	県	入間	郡
かながわ	県	みうら	市
かながわ	県	みうら	郡
神奈川	県	三浦	市
神奈川	県	三浦	郡
石川	県	かほく	市
いしかわ	県	かほく	市
いしかわ	県	かほく	郡
いしかわ	県	すず	市
いしかわ	県	すず	郡

都道府県		市郡支庁	
いしかわ	県	はくい	市
いしかわ	県	はくい	郡
石川	県	羽咋	市
石川	県	羽咋	郡
石川	県	珠洲	市
石川	県	珠洲	郡
福井	県	あわら	市
ふくい	県	あわら	市
ふくい	県	おおの	市
ふくい	県	おおの	郡
福井	県	大野	市
福井	県	大野	郡
ながの	県	すわ	市
ながの	県	すわ	郡
長野	県	諏訪	市
長野	県	諏訪	郡
ぎふ	県	えな	市
ぎふ	県	えな	郡
ぎふ	県	かに	市
ぎふ	県	かに	郡
ぎふ	県	とき	市
ぎふ	県	とき	郡
ぎふ	県	はしま	市
ぎふ	県	はしま	郡
ぎふ	県	もとす	市
ぎふ	県	もとす	郡
岐阜	県	羽島	市
岐阜	県	羽島	郡
岐阜	県	可児	市
岐阜	県	可児	郡
岐阜	県	恵那	市
岐阜	県	恵那	郡
岐阜	県	土岐	市
岐阜	県	土岐	郡
岐阜	県	本巣	市
岐阜	県	本巣	郡
しずおか	県	いわた	市
しずおか	県	いわた	郡
しずおか	県	ふじ	市
しずおか	県	ふじ	郡
静岡	県	磐田	市
静岡	県	磐田	郡
静岡	県	富士	市
静岡	県	富士	郡
あいち	県	ちた	市
あいち	県	ちた	郡
愛知	県	知多	市
愛知	県	知多	郡
三重	県	いなべ	市
みえ	県	いなべ	市
みえ	県	いなべ	郡
みえ	県	くわな	市
みえ	県	くわな	郡
みえ	県	すずか	市
みえ	県	すずか	郡

都道府県		市郡支庁	
三重	県	桑名	市
三重	県	桑名	郡
三重	県	鈴鹿	市
三重	県	鈴鹿	郡
おおさか	府	せんなん	市
おおさか	府	せんなん	郡
大阪	府	泉南	市
大阪	府	泉南	郡
ひょうご	県	あこう	市
ひょうご	県	あこう	郡
兵庫	県	赤穂	市
兵庫	県	赤穂	郡
なら	県	いこま	市
なら	県	いこま	郡
奈良	県	生駒	市
奈良	県	生駒	郡
わかやま	県	ありだ	市
わかやま	県	ありだ	郡
和歌山	県	有田	市
和歌山	県	有田	郡
やまぐち	県	みね	市
やまぐち	県	みね	郡
山口	県	美祢	市
山口	県	美祢	郡
えひめ	県	いよ	市
えひめ	県	いよ	郡
愛媛	県	伊予	市
愛媛	県	伊予	郡
こうち	県	あき	市
こうち	県	あき	郡
こうち	県	とさ	市
こうち	県	とさ	郡
高知	県	安芸	市
高知	県	安芸	郡
高知	県	土佐	市
高知	県	土佐	郡
ふくおか	県	たがわ	市
ふくおか	県	たがわ	郡
ふくおか	県	むなかた	市
ふくおか	県	むなかた	郡
ふくおか	県	やめ	市
ふくおか	県	やめ	郡
福岡	県	宗像	市
福岡	県	宗像	郡
福岡	県	田川	市
福岡	県	田川	郡
福岡	県	八女	市
福岡	県	八女	郡
さが	県	さが	市
さが	県	さが	郡
佐賀	県	佐賀	市
佐賀	県	佐賀	郡
くまもと	県	うと	市
くまもと	県	うと	郡

都道府県		市郡支庁	
くまもと	県	きくち	市
くまもと	県	きくち	郡
くまもと	県	たまな	市
くまもと	県	たまな	郡
くまもと	県	やつしろ	市
くまもと	県	やつしろ	郡
熊本	県	宇土	市
熊本	県	宇土	郡
熊本	県	菊池	市
熊本	県	菊池	郡
熊本	県	玉名	市
熊本	県	玉名	郡
熊本	県	八代	市
熊本	県	八代	郡
おおいた	県	うさ	市
おおいた	県	うさ	郡
おおいた	県	おおいた	市
おおいた	県	おおいた	郡
おおいた	県	ひた	市
おおいた	県	ひた	郡
大分	県	宇佐	市
大分	県	宇佐	郡
大分	県	大分	市
大分	県	大分	郡
大分	県	日田	市
大分	県	日田	郡
みやざき	県	みやざき	市
みやざき	県	みやざき	郡
宮崎	県	宮崎	市
宮崎	県	宮崎	郡
かごしま	県	いずみ	市
かごしま	県	いずみ	郡
かごしま	県	いぶすき	市
かごしま	県	いぶすき	郡
かごしま	県	かごしま	市
かごしま	県	かごしま	郡
鹿児島	県	鹿児島	市
鹿児島	県	鹿児島	郡
鹿児島	県	出水	市
鹿児島	県	出水	郡

参考　4

【1郡1町村一覧（平成１７年住所辞書)】

都道府県		市郡支庁		区町村	
岩手	県	気仙	郡	住田	町
宮城	県	伊具	郡	丸森	町
秋田	県	鹿角	郡	小坂	町
山形	県	北村山	郡	大石田	町
山形	県	西田川	郡	温海	町
福島	県	北会津	郡	北会津	村
茨城	県	多賀	郡	十王	町
群馬	県	碓氷	郡	松井田	町
群馬	県	山田	郡	大間々	町
千葉	県	東葛飾	郡	沼南	町
東京	都	小笠原支	庁	小笠原	村
神奈川	県	三浦	郡	葉山	町
神奈川	県	高座	郡	寒川	町
新潟	県	古志	郡	山古志	村
石川	県	江沼	郡	山中	町
石川	県	珠洲	郡	内浦	町
福井	県	足羽	郡	美山	町
福井	県	大野	郡	和泉	村
長野	県	更級	郡	大岡	村
長野	県	埴科	郡	坂城	町
岐阜	県	本巣	郡	北方	町
岐阜	県	土岐	郡	笠原	町
静岡	県	富士	郡	芝川	町
愛知	県	葉栗	郡	木曽川	町
愛知	県	渥美	郡	渥美	町
三重	県	員弁	郡	東員	町
三重	県	鈴鹿	郡	関	町
三重	県	名賀	郡	青山	町
滋賀	県	滋賀	郡	志賀	町
京都	府	乙訓	郡	大山崎	町
京都	府	久世	郡	久御山	町
京都	府	加佐	郡	大江	町
大阪	府	三島	郡	島本	町
大阪	府	泉北	郡	忠岡	町
兵庫	県	川辺	郡	猪名川	町
兵庫	県	美嚢	郡	吉川	町
兵庫	県	赤穂	郡	上郡	町
奈良	県	添上	郡	月ヶ瀬	村
岡山	県	児島	郡	灘崎	町
岡山	県	後月	郡	芳井	町
岡山	県	吉備	郡	真備	町
岡山	県	上房	郡	北房	町
岡山	県	加賀	郡	吉備中央	町
広島	県	世羅	郡	世羅	町
広島	県	沼隈	郡	沼隈	町
広島	県	深安	郡	神辺	町
広島	県	甲奴	郡	総領	町
山口	県	大島	郡	周防大島	町
山口	県	佐波	郡	徳地	町
徳島	県	名東	郡	佐那河内	村
愛媛	県	温泉	郡	中島	町
愛媛	県	南宇和	郡	愛南	町
福岡	県	筑紫	郡	那珂川	町

都道府県		市郡支庁		区町村	
福岡	県	三池	郡	高田	町
長崎	県	南松浦	郡	新上五島	町
大分	県	北海部	郡	佐賀関	町
鹿児島	県	伊佐	郡	菱刈	町

参考　5

【特定の誤読により要人手審査とする市区町村の組合せ一覧（平成17年住所辞書）】

・市郡支庁名の先頭に1文字追加すると県内の他の市郡支庁名となる地域

①政令指定都市を市部に誤読

市部（要審査地域）				政令指定都市（理由となる地域）			
都道府県		市郡支庁		都道府県		市郡支庁	
広島	県	東広島	市	広島	県	広島	市

②市部を他の市部に誤読

市部（要審査地域）	市部（理由となる地域）
該当なし	

③郡部を市部に誤読

市部（要審査地域）				郡部（理由となる地域）			
都道府県		市郡支庁		都道府県		市郡支庁	
熊本	県	上天草	市	熊本	県	天草	郡

・市郡支庁名の先頭に1文字追加すると県外の他の市郡支庁名となる地域

①政令指定都市を市部に誤読

市部（要審査地域）				政令指定都市（理由となる地域）			
都道府県		市郡支庁		都道府県		市郡支庁	
鳥取	県	境港	市	東京	都	港	区
静岡	県	浜北	市	東京	都	北	区

②市部を他の市部に誤読

市部（要審査地域）				市部（理由となる地域）			
都道府県		市郡支庁		都道府県		市郡支庁	
北海	道	苫小牧	市	愛知	県	小牧	市
北海	道	北広島	市	広島	県	広島	市
岩手	県	一関	市	岐阜	県	関	市
茨城	県	高萩	市	山口	県	萩	市
栃木	県	日光	市	山口	県	光	市
栃木	県	大田原	市	愛知	県	田原	市
埼玉	県	東松山	市	愛媛	県	松山	市
埼玉	県	和光	市	山口	県	光	市
埼玉	県	上福岡	市	福岡	県	福岡	市
千葉	県	君津	市	三重	県	津	市
千葉	県	富津	市	三重	県	津	市
神奈川	県	小田原	市	愛知	県	田原	市
新潟	県	新津	市	三重	県	津	市
富山	県	魚津	市	三重	県	津	市
静岡	県	沼津	市	三重	県	津	市
静岡	県	焼津	市	三重	県	津	市
滋賀	県	大津	市	三重	県	津	市
滋賀	県	草津	市	三重	県	津	市
兵庫	県	加古川	市	宮城	県	古川	市
島根	県	江津	市	三重	県	津	市
山口	県	下関	市	岐阜	県	関	市
山口	県	小野田	市	千葉	県	野田	市
佐賀	県	唐津	市	三重	県	津	市
大分	県	中津	市	三重	県	津	市
群馬	県	安中	市	神奈川	県	中	郡
鳥取	県	境港	市	東京	都	港	区
静岡	県	浜北	市	東京	都	北	区
広島	県	府中	市	神奈川	県	中	郡
鹿児島	県	鹿児島	市	岡山	県	児島	郡

③郡部を市部に誤読

市部（要審査地域）				郡部（理由となる地域）			
都道府県		市郡支庁		都道府県		市郡支庁	
群馬	県	安中	市	神奈川	県	中	郡
広島	県	府中	市	神奈川	県	中	郡
鹿児島	県	鹿児島	市	岡山	県	児島	郡

・市郡支庁名の先頭が1文字異なると県内の他の市郡支庁名となる地域

①政令指定都市を市部に誤読

市部				政令指定都市			
都道府県		市郡支庁		都道府県		市郡支庁	
広島	県	因島	市	広島	県	広島	市

②市部を他の市部に誤読

市部				市部			
都道府県		市郡支庁		都道府県		市郡支庁	
北海	道	芦別	市	北海	道	江別	市
北海	道	芦別	市	北海	道	士別	市
北海	道	芦別	市	北海	道	登別	市
北海	道	芦別	市	北海	道	紋別	市
北海	道	江別	市	北海	道	芦別	市
北海	道	江別	市	北海	道	士別	市
北海	道	江別	市	北海	道	登別	市
北海	道	江別	市	北海	道	紋別	市
北海	道	士別	市	北海	道	芦別	市
北海	道	士別	市	北海	道	江別	市
北海	道	士別	市	北海	道	登別	市
北海	道	士別	市	北海	道	紋別	市
北海	道	登別	市	北海	道	芦別	市
北海	道	登別	市	北海	道	江別	市
北海	道	登別	市	北海	道	士別	市
北海	道	登別	市	北海	道	紋別	市
北海	道	紋別	市	北海	道	芦別	市
北海	道	紋別	市	北海	道	江別	市
北海	道	紋別	市	北海	道	士別	市
北海	道	紋別	市	北海	道	登別	市
北海	道	旭川	市	北海	道	砂川	市
北海	道	旭川	市	北海	道	深川	市
北海	道	旭川	市	北海	道	滝川	市
北海	道	砂川	市	北海	道	旭川	市
北海	道	砂川	市	北海	道	深川	市
北海	道	砂川	市	北海	道	滝川	市
北海	道	深川	市	北海	道	旭川	市
北海	道	深川	市	北海	道	砂川	市
北海	道	深川	市	北海	道	滝川	市
北海	道	滝川	市	北海	道	旭川	市
北海	道	滝川	市	北海	道	砂川	市
北海	道	滝川	市	北海	道	深川	市
山形	県	上山	市	山形	県	村山	市
山形	県	村山	市	山形	県	上山	市
群馬	県	沼田	市	群馬	県	太田	市
群馬	県	太田	市	群馬	県	沼田	市
群馬	県	藤岡	市	群馬	県	富岡	市
群馬	県	富岡	市	群馬	県	藤岡	市
群馬	県	沼田	市	群馬	県	太田	市
群馬	県	太田	市	群馬	県	沼田	市
埼玉	県	越谷	市	埼玉	県	熊谷	市
埼玉	県	越谷	市	埼玉	県	深谷	市
埼玉	県	熊谷	市	埼玉	県	越谷	市
埼玉	県	熊谷	市	埼玉	県	深谷	市
埼玉	県	深谷	市	埼玉	県	越谷	市
埼玉	県	深谷	市	埼玉	県	熊谷	市

（続き）

市部				市部			
都道府県		市郡支庁		都道府県		市郡支庁	
埼玉	県	戸田	市	埼玉	県	行田	市
埼玉	県	戸田	市	埼玉	県	蓮田	市
埼玉	県	行田	市	埼玉	県	戸田	市
埼玉	県	行田	市	埼玉	県	蓮田	市
埼玉	県	蓮田	市	埼玉	県	戸田	市
埼玉	県	蓮田	市	埼玉	県	行田	市
埼玉	県	桶川	市	埼玉	県	吉川	市
埼玉	県	吉川	市	埼玉	県	桶川	市
千葉	県	成田	市	千葉	県	野田	市
千葉	県	野田	市	千葉	県	成田	市
千葉	県	佐原	市	千葉	県	市原	市
千葉	県	佐原	市	千葉	県	茂原	市
千葉	県	市原	市	千葉	県	佐原	市
千葉	県	市原	市	千葉	県	茂原	市
千葉	県	茂原	市	千葉	県	佐原	市
千葉	県	茂原	市	千葉	県	市原	市
千葉	県	館山	市	千葉	県	流山	市
千葉	県	流山	市	千葉	県	館山	市
千葉	県	鴨川	市	千葉	県	市川	市
千葉	県	市川	市	千葉	県	鴨川	市
千葉	県	君津	市	千葉	県	富津	市
千葉	県	富津	市	千葉	県	君津	市
長野	県	上田	市	長野	県	飯田	市
長野	県	飯田	市	長野	県	上田	市
長野	県	茅野	市	長野	県	中野	市
長野	県	茅野	市	長野	県	長野	市
長野	県	中野	市	長野	県	茅野	市
長野	県	中野	市	長野	県	長野	市
長野	県	長野	市	長野	県	茅野	市
長野	県	長野	市	長野	県	中野	市
静岡	県	沼津	市	静岡	県	焼津	市
静岡	県	焼津	市	静岡	県	沼津	市
静岡	県	下田	市	静岡	県	島田	市
静岡	県	下田	市	静岡	県	磐田	市
静岡	県	島田	市	静岡	県	下田	市
静岡	県	島田	市	静岡	県	磐田	市
静岡	県	磐田	市	静岡	県	下田	市
静岡	県	磐田	市	静岡	県	島田	市
愛知	県	江南	市	愛知	県	碧南	市
愛知	県	碧南	市	愛知	県	江南	市
愛知	県	安城	市	愛知	県	新城	市
愛知	県	新城	市	愛知	県	安城	市
愛知	県	半田	市	愛知	県	豊田	市
愛知	県	豊田	市	愛知	県	半田	市
三重	県	熊野	市	三重	県	上野	市
三重	県	上野	市	三重	県	熊野	市
滋賀	県	草津	市	滋賀	県	大津	市
滋賀	県	大津	市	滋賀	県	草津	市
兵庫	県	小野	市	兵庫	県	龍野	市
兵庫	県	龍野	市	兵庫	県	小野	市
兵庫	県	加西	市	兵庫	県	川西	市
兵庫	県	川西	市	兵庫	県	加西	市

（続き）

市部				市部			
都道府県		市郡支庁		都道府県		市郡支庁	
島根	県	益田	市	島根	県	大田	市
島根	県	益田	市	島根	県	浜田	市
島根	県	益田	市	島根	県	平田	市
島根	県	大田	市	島根	県	益田	市
島根	県	大田	市	島根	県	浜田	市
島根	県	大田	市	島根	県	平田	市
島根	県	浜田	市	島根	県	益田	市
島根	県	浜田	市	島根	県	大田	市
島根	県	浜田	市	島根	県	平田	市
島根	県	平田	市	島根	県	益田	市
島根	県	平田	市	島根	県	大田	市
島根	県	平田	市	島根	県	浜田	市
岡山	県	岡山	市	岡山	県	津山	市
岡山	県	津山	市	岡山	県	岡山	市
広島	県	三原	市	広島	県	庄原	市
広島	県	三原	市	広島	県	竹原	市
広島	県	庄原	市	広島	県	三原	市
広島	県	庄原	市	広島	県	竹原	市
広島	県	竹原	市	広島	県	三原	市
広島	県	竹原	市	広島	県	庄原	市
愛媛	県	西条	市	愛媛	県	北条	市
愛媛	県	北条	市	愛媛	県	西条	市
愛媛	県	伊予	市	愛媛	県	西予	市
愛媛	県	伊予	市	愛媛	県	東予	市
愛媛	県	西予	市	愛媛	県	伊予	市
愛媛	県	西予	市	愛媛	県	東予	市
愛媛	県	東予	市	愛媛	県	伊予	市
愛媛	県	東予	市	愛媛	県	西予	市
福岡	県	大川	市	福岡	県	田川	市
福岡	県	大川	市	福岡	県	柳川	市
福岡	県	田川	市	福岡	県	大川	市
福岡	県	田川	市	福岡	県	柳川	市
福岡	県	柳川	市	福岡	県	大川	市
福岡	県	柳川	市	福岡	県	田川	市
大分	県	竹田	市	大分	県	日田	市
大分	県	日田	市	大分	県	竹田	市
鹿児島	県	出水	市	鹿児島	県	垂水	市
鹿児島	県	垂水	市	鹿児島	県	出水	市

③郡部を市部に誤読

市部				郡部			
都道府県		市郡支庁		都道府県		市郡支庁	
北海	道	旭川	市	北海	道	上川	郡
北海	道	砂川	市	北海	道	上川	郡
北海	道	深川	市	北海	道	上川	郡
北海	道	滝川	市	北海	道	上川	郡
青森	県	八戸	市	青森	県	三戸	郡
岩手	県	水沢	市	岩手	県	胆沢	郡
岩手	県	二戸	市	岩手	県	九戸	郡
宮城	県	角田	市	宮城	県	遠田	郡
宮城	県	角田	市	宮城	県	刈田	郡
宮城	県	角田	市	宮城	県	志田	郡
宮城	県	角田	市	宮城	県	柴田	郡
宮城	県	古川	市	宮城	県	黒川	郡
秋田	県	男鹿	市	秋田	県	平鹿	郡
茨城	県	北茨城	市	茨城	県	西茨城	郡
茨城	県	北茨城	市	茨城	県	東茨城	郡
群馬	県	沼田	市	群馬	県	山田	郡
群馬	県	沼田	市	群馬	県	新田	郡
群馬	県	太田	市	群馬	県	山田	郡
群馬	県	太田	市	群馬	県	新田	郡
石川	県	輪島	市	石川	県	鹿島	郡
福井	県	福井	市	福井	県	坂井	郡
福井	県	武生	市	福井	県	丹生	郡
静岡	県	伊東	市	静岡	県	駿東	郡
愛知	県	津島	市	愛知	県	中島	郡
愛知	県	半田	市	愛知	県	額田	郡
愛知	県	豊田	市	愛知	県	額田	郡
三重	県	亀山	市	三重	県	阿山	郡
滋賀	県	甲賀	市	滋賀	県	滋賀	郡
兵庫	県	尼崎	市	兵庫	県	城崎	郡
兵庫	県	尼崎	市	兵庫	県	神崎	郡
兵庫	県	明石	市	兵庫	県	出石	郡
奈良	県	葛城	市	奈良	県	磯城	郡
福岡	県	古賀	市	福岡	県	遠賀	郡
佐賀	県	鹿島	市	佐賀	県	杵島	郡
佐賀	県	唐津	市	佐賀	県	藤津	郡
熊本	県	熊本	市	熊本	県	鹿本	郡
熊本	県	玉名	市	熊本	県	玉名	郡
鹿児島	県	指宿	市	鹿児島	県	揖宿	郡

・市郡支庁名の先頭が1文字異なると県外の他の市郡支庁名となる地域

①政令指定都市を市部に誤読

市部				政令指定都市			
都道府県		市郡支庁		都道府県		市郡支庁	
北海	道	旭川	市	東京	都	品川	区
北海	道	旭川	市	東京	都	荒川	区
北海	道	滝川	市	東京	都	品川	区
北海	道	滝川	市	東京	都	荒川	区
北海	道	砂川	市	東京	都	品川	区
北海	道	砂川	市	東京	都	荒川	区
北海	道	深川	市	東京	都	品川	区
北海	道	深川	市	東京	都	荒川	区
青森	県	八戸	市	兵庫	県	神戸	区
岩手	県	盛岡	市	福岡	県	福岡	区
岩手	県	遠野	市	東京	都	中野	区
岩手	県	二戸	市	兵庫	県	神戸	区
宮城	県	古川	市	東京	都	品川	区
宮城	県	古川	市	東京	都	荒川	区
山形	県	鶴岡	市	福岡	県	福岡	区
福島	県	福島	市	広島	県	広島	区
福島	県	福島	市	東京	都	豊島	区
福島	県	相馬	市	東京	都	練馬	区
茨城	県	水戸	市	兵庫	県	神戸	区
茨城	県	日立	市	東京	都	足立	区
茨城	県	石岡	市	福岡	県	福岡	区
茨城	県	守谷	市	東京	都	渋谷	区
栃木	県	佐野	市	東京	都	中野	区
栃木	県	真岡	市	福岡	県	福岡	区
群馬	県	前橋	市	東京	都	板橋	区
群馬	県	高崎	市	神奈川	県	川崎	区
群馬	県	渋川	市	東京	都	品川	区
群馬	県	渋川	市	東京	都	荒川	区
群馬	県	藤岡	市	福岡	県	福岡	区
群馬	県	富岡	市	福岡	県	福岡	区
埼玉	県	熊谷	市	東京	都	渋谷	区
埼玉	県	深谷	市	東京	都	渋谷	区
埼玉	県	越谷	市	東京	都	渋谷	区
埼玉	県	戸田	市	東京	都	墨田	区
埼玉	県	戸田	市	東京	都	大田	区
埼玉	県	桶川	市	東京	都	品川	区
埼玉	県	桶川	市	東京	都	荒川	区
埼玉	県	蓮田	市	東京	都	墨田	区
埼玉	県	蓮田	市	東京	都	大田	区
埼玉	県	坂戸	市	兵庫	県	神戸	区
埼玉	県	吉川	市	東京	都	品川	区
埼玉	県	吉川	市	東京	都	荒川	区
千葉	県	市川	市	東京	都	品川	区
千葉	県	市川	市	東京	都	荒川	区
千葉	県	船橋	市	東京	都	板橋	区
千葉	県	松戸	市	兵庫	県	神戸	区

他99組

②市部を他の市部に誤読

市部				市部			
都道府県		市郡支庁		都道府県		市郡支庁	
北海	道	函館	市	茨城	県	下館	市
北海	道	函館	市	秋田	県	大館	市
北海	道	旭川	市	愛知	県	豊川	市
北海	道	旭川	市	沖縄	県	石川	市
北海	道	旭川	市	宮城	県	古川	市
北海	道	旭川	市	群馬	県	渋川	市
北海	道	旭川	市	埼玉	県	吉川	市
北海	道	旭川	市	埼玉	県	桶川	市
北海	道	旭川	市	静岡	県	掛川	市
北海	道	旭川	市	千葉	県	市川	市
北海	道	旭川	市	千葉	県	鴨川	市
北海	道	旭川	市	東京	都	立川	市
北海	道	旭川	市	富山	県	滑川	市
北海	道	旭川	市	福岡	県	大川	市
北海	道	旭川	市	福岡	県	田川	市
北海	道	旭川	市	福岡	県	柳川	市
北海	道	旭川	市	北海	道	滝川	市
北海	道	旭川	市	北海	道	砂川	市
北海	道	旭川	市	北海	道	深川	市
北海	道	釧路	市	兵庫	県	姫路	市
北海	道	北見	市	岡山	県	新見	市
北海	道	北見	市	富山	県	氷見	市
北海	道	夕張	市	三重	県	名張	市
北海	道	稚内	市	鹿児島	県	川内	市
北海	道	赤平	市	東京	都	小平	市
北海	道	滝川	市	愛知	県	豊川	市
北海	道	滝川	市	沖縄	県	石川	市
北海	道	滝川	市	宮城	県	古川	市
北海	道	滝川	市	群馬	県	渋川	市
北海	道	滝川	市	埼玉	県	吉川	市
北海	道	滝川	市	埼玉	県	桶川	市
北海	道	滝川	市	静岡	県	掛川	市
北海	道	滝川	市	千葉	県	市川	市
北海	道	滝川	市	千葉	県	鴨川	市
北海	道	滝川	市	東京	都	立川	市
北海	道	滝川	市	富山	県	滑川	市
北海	道	滝川	市	福岡	県	大川	市
北海	道	滝川	市	福岡	県	田川	市
北海	道	滝川	市	福岡	県	柳川	市
北海	道	滝川	市	北海	道	旭川	市
北海	道	滝川	市	北海	道	砂川	市
北海	道	滝川	市	北海	道	深川	市

他2,490組

③郡部を市部に誤読

市部				郡部			
都道府県		市郡支庁		都道府県		市郡支庁	
北海	道	旭川	市	宮城	県	黒川	郡
北海	道	旭川	市	香川	県	香川	郡
北海	道	旭川	市	高知	県	吾川	郡
北海	道	旭川	市	石川	県	石川	郡
北海	道	旭川	市	島根	県	簸川	郡
北海	道	旭川	市	福岡	県	田川	郡
北海	道	旭川	市	福島	県	石川	郡
北海	道	旭川	市	北海	道	上川	郡
北海	道	北見	市	大分	県	速見	郡
北海	道	稚内	市	栃木	県	河内	郡
北海	道	三笠	市	静岡	県	小笠	郡
北海	道	滝川	市	宮城	県	黒川	郡
北海	道	滝川	市	香川	県	香川	郡
北海	道	滝川	市	高知	県	吾川	郡
北海	道	滝川	市	石川	県	石川	郡
北海	道	滝川	市	島根	県	簸川	郡
北海	道	滝川	市	福岡	県	田川	郡
北海	道	滝川	市	福島	県	石川	郡
北海	道	滝川	市	北海	道	上川	郡
北海	道	砂川	市	宮城	県	黒川	郡
北海	道	砂川	市	香川	県	香川	郡
北海	道	砂川	市	高知	県	吾川	郡
北海	道	砂川	市	石川	県	石川	郡
北海	道	砂川	市	島根	県	簸川	郡
北海	道	砂川	市	福岡	県	田川	郡
北海	道	砂川	市	福島	県	石川	郡
北海	道	砂川	市	北海	道	上川	郡
北海	道	深川	市	宮城	県	黒川	郡
北海	道	深川	市	香川	県	香川	郡
北海	道	深川	市	高知	県	吾川	郡
北海	道	深川	市	石川	県	石川	郡
北海	道	深川	市	島根	県	簸川	郡
北海	道	深川	市	福岡	県	田川	郡
北海	道	深川	市	福島	県	石川	郡
北海	道	深川	市	北海	道	上川	郡
北海	道	恵庭	市	岡山	県	真庭	郡
北海	道	伊達	市	福島	県	安達	郡
青森	県	八戸	市	岩手	県	九戸	郡
青森	県	八戸	市	岩手	県	二戸	郡
青森	県	八戸	市	青森	県	三戸	郡
青森	県	黒石	市	広島	県	神石	郡
青森	県	黒石	市	島根	県	飯石	郡
青森	県	黒石	市	兵庫	県	出石	郡
青森	県	三沢	市	岩手	県	胆沢	郡
岩手	県	盛岡	市	高知	県	高岡	郡
岩手	県	盛岡	市	高知	県	長岡	郡
岩手	県	宮古	市	兵庫	県	加古	郡
岩手	県	水沢	市	岩手	県	胆沢	郡
岩手	県	北上	市	山形	県	最上	郡
岩手	県	北上	市	滋賀	県	犬上	郡
岩手	県	北上	市	千葉	県	海上	郡
岩手	県	北上	市	奈良	県	添上	郡
岩手	県	北上	市	福岡	県	築上	郡
岩手	県	北上	市	兵庫	県	氷上	郡

他1,151組

参考　6

【正記入誤読の場合の文字認識結果及び格付結果（現有機）】

調査票記入状況			読み取り結果				正解か	備　考
都道府県	市郡支庁	区町村	都道府県	市郡支庁	区町村	自動格付		
（政令指定都市）								
広島	広島	安佐北	広島	広島	？	34@@@	×	34110と格付されるべき
						@@@@@	×	
広島	広島	安佐北	広島	？	安佐北	34106	○	
広島	広島	安佐北	？	広島	安佐北	34106	○	
広島	広島	安佐北	広島	？	？	34@@@	○	
広島	広島	安佐北	？	？	安佐北	34106	○	県外同名区が存在する場合は未格付とすべき
広島	広島	中	？	？	中	@@@@@	○	
広島	広島	安佐北	？	広島	？	34@@@	○	
広島	広島	安佐北	？	？	？	@@@@@	○	
（東京都特別区）								
東京		新宿	東京		？	13@@@	○	
東京		新宿	？		新宿	13104	○	
東京		新宿	？		？	@@@@@	○	
東京	新宿		東京	？		13@@@	○	
東京	新宿		？	新宿		13104	○	
東京			？	？		@@@@@	○	
（市部）								
広島	福山		広島	？		34@@@	○	
広島	福山		？	福山		34304	○	
広島	福山		？	？		@@@@@	○	
大分	宇佐		大分	宇佐		44@@@	×	県内同名市郡の場合、区町村欄が空欄であれば市として格付されるべき なお、「市」まで記入してある場合は正しく格付されている
大分	宇佐市		大分	宇佐市		44211	○	
（郡部）								
広島	安芸	海田	広島	安芸	？	34@@@	○	
広島	安芸	海田	広島	？	海田	34304	○	県内同名町村が存在する場合は34@@@とすべき
広島	安芸	海田	？	安芸	海田	34304	○	
広島	安芸	海田	広島	？	？	34@@@	○	
広島	安芸	海田	？	安芸	？	@@@@@	○	同名郡が存在する場合は@@@@@とすべき
広島	沼隈	沼隈	？	沼隈	？	34@@@	○	
広島	安芸	海田	？	？	海田	34304	○	同名町村が存在する場合は34@@@とすべき
広島	安芸	海田	？	？	？	@@@@@	○	

【誤記入の場合の文字認識結果及び格付結果（現有機）】

調査票記入状況			読み取り結果				正解か	備　考
都道府県	市郡支庁	区町村	都道府県	市郡支庁	区町村	自動格付		
(政令指定都市)								
北海道	札幌		北海道	札幌		01@@@	○	01110と格付すべき
北海道	札幌	発寒	北海道	札幌	?	01@@@	○	
北海道	さいたま	西	北海道	?	西	01107	×	県内同名区町村が存在する場合は01@@@とすべき
			北海道	?	?	01@@@	○	
北海道	千葉	中央	?	千葉	中央	12101	△	
北海道		札幌	北海道		?	01@@@	○	
北海道		西	北海道		西	01107	○	
北海道	西		－	－	－	－	－	該当データなし
	札幌	西		札幌	西	01107	○	
	札幌		－	－	－	－	－	該当データなし
	札幌	発寒	－	－	－	－	－	該当データなし
(東京都特別区)								
東京	新宿	若松町	東京	新宿	?	13104	○	
東京		若松町	東京		?	13@@@	○	
東京	新宿	品川	東京	新宿	?	13104	○	
			東京	?	品川	13109	×	
東京		新宿区若松町	東京		?	13@@@	○	
東京	新宿区若松町	19－1	－	－	－	－	－	該当データなし
東京	西東京(旧保谷市)		東京	西東京市		13229	○	
	新宿			新宿		13104	○	
				?		@@@@@	○	
		新宿			新宿	13104	○	
					?	@@@@@	○	
(市部)								
大阪	新宿		大阪	?		27@@@	△	
北海道	さいたま	阿蘇町	北海道	?	?	01@@@	△	
			?	さいたま	?	11@@@	△	
北海道	さいたま	阿蘇町	?	?	阿蘇町	43422	○	同名区町村が存在する場合は未格付とすべき
静岡		伊豆高原	静岡		?	22@@@	○	
神奈川	厚木	旭	神奈川	厚木	?	14212	○	
神奈川	厚木	旭	神奈川	?	旭	14112	×	同名区町村が存在する場合は14@@@とすべき
神奈川		厚木	神奈川		?		○	
	厚木		－	－	－	－	－	該当データなし
		厚木	－	－	－	－	－	該当データなし
	千葉県	八千代		?	?	@@@@@	○	
	千葉	八千代		千葉	?	12@@@	○	
(郡部)								
広島	賀茂		広島	賀茂		34@@@	○	
広島	賀茂	春日野	広島	賀茂	?	34@@@	○	
広島	さいたま	黒瀬	広島	?	黒瀬	34402	○	県内同名区町村が存在する場合は34@@@とすべき
			広島	?	黒瀬	34@@@	×	
			広島	?	?	34@@@	○	
広島	千葉	中央	?	千葉	中央	12101	△	
広島		賀茂	広島		?	34@@@	○	
広島		黒瀬	広島		黒瀬	34402	○	県内同名区町村が存在する場合は34@@@とすべき
広島	黒瀬		－	－	－	－	－	該当データなし
	賀茂	黒瀬	－	－	－	－	－	該当データなし
	賀茂		－	－	－	－	－	該当データなし
	賀茂	春日野	－	－	－	－	－	該当データなし
茨城	日立	旭	茨城	?	旭	08401		同名区町村が存在する場合は14@@@とすべき

(注) 正解か否か：'○' … 正解、'×' … 不正解、'△' … 人手審査なしでは判断不能

製 表 技 術 参 考 資 料 6

平成 19 年 8 月発行

編集・発行　　独立行政法人　統計センター

〒162－8668
東京都新宿区若松町 19－1
電　話　代表　03 (5273) 1200

研究センター　TEL : 03－5273－1286
E-mail : research@nstac.go.jp