Bulletin of Toyohashi Sozo University
2009, No. 13, 59–72

江戸版本におけるつづき文字部分の識別についての検討

舟久保　　登

Key Words：continuous character parts discrimination, engraved printing Edo period book, minimum distance method, characters segmentation, standard character pattern dictionary

1．まえがき

執筆者は平成17～19年度の3年間，同僚の教員である島田大助，三好哲也，三輪多恵子の3氏と共に，「江戸版本の読解を支援する運筆特徴を考慮したつづき文字の認識に関する研究」なる題目の科学研究費補助金による研究を実施した．そしてこの研究期間が昨年3月末に終了したので，その内容は報告書としてまとめを行っている［1］．しかしここで取り上げた研究内容は相当に難しいもので，題目の示す研究の最終目標はまだ完成されたとはいえない．

この紀要においての報告は，上記研究について舟久保が分担して実施した分野の続きを述べるものである．その前提となる出発点は，人手により分離した状態の文字パターンとしては正しく識別できることが既に判明しているものから成るつづき文字部分を対象に，その個別文字認識を試みることである．実際に対象とするこのつづき文字部分を，図1に具体的に挙げる．また既に分離した文字の識別に対して使用し，今回の研究においても使う標準文字パターン辞書を図2に示す．

なおこれらの図の内容がどのようにして準備されたかの経過の詳細は，文献1にある科研費報告書における舟久保執筆箇所を参照されたい．

2．分割箇所を指定したつづき文字部分の識別結果

ここでは最初に，ある意味では理想的に達成可能な識別結果を与えるともいうべき，人手により分割箇所を指定した場合における状況について述べる．その例として，これを行った状態の一部を図3に示した．この対象つづき文字部分は図1に示したグループ1の右半部にあたる．

さてこの結果を得るための分割箇所は図中に横線を引き，そこに括弧内の二つの数値をもって左から切り出し枠の左上隅の画素横座標位置，縦座標位置の順に記している．そして

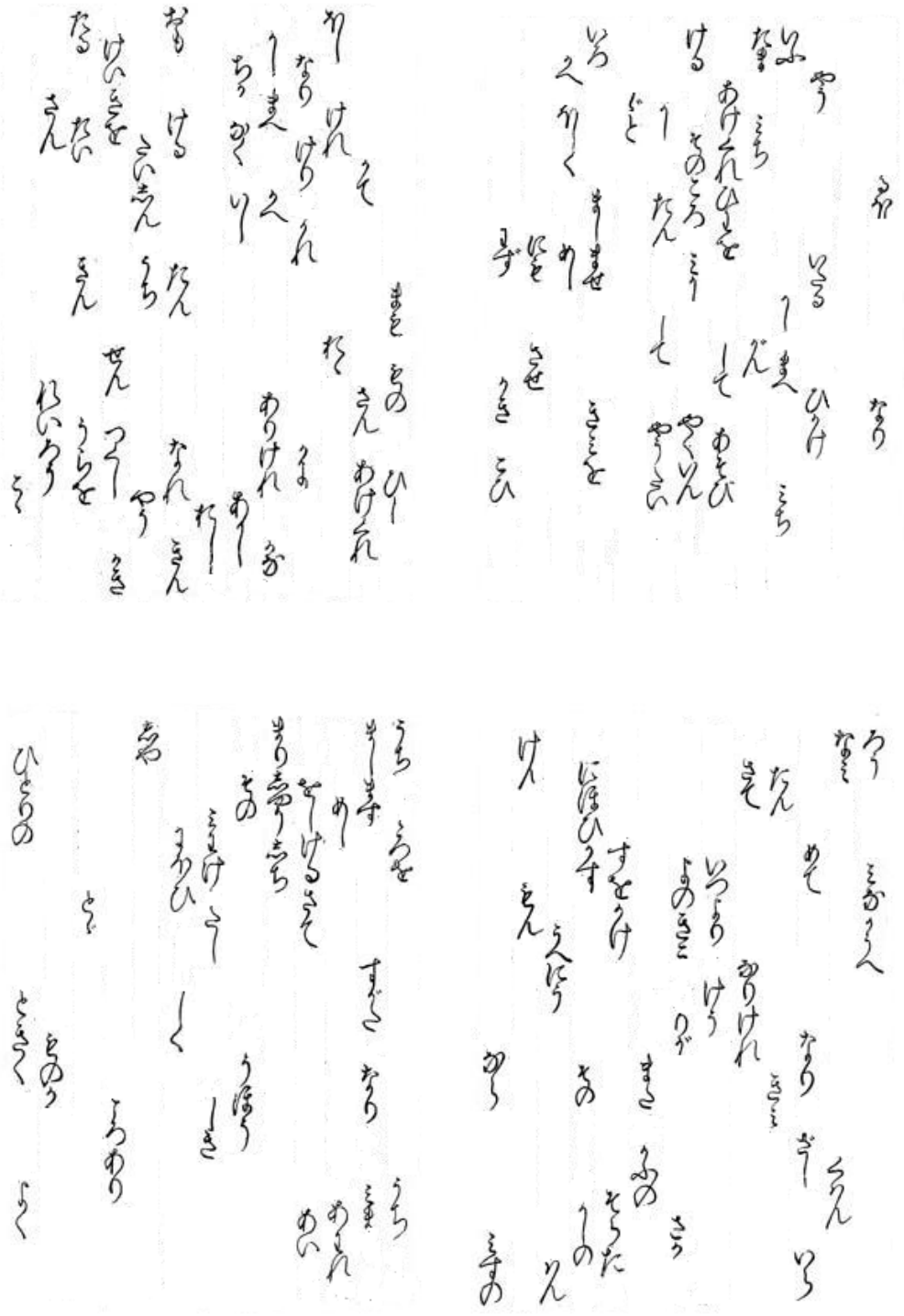

図1　対象とするつづき文字部分（上半：グループ1，下半：グループ2）

図2　標準文字パターン辞書

この分割位置は執筆者自身により，こうして枠指定された対象文字パターンができるだけ正しく識別されるように，試行錯誤的に決定したものである．表1にはこうして得られた個別文字識別の結果を，整理しまとめた形で挙げた．ここで対象とした図1に存在する文字個数は，全部で342個であった．

表1の示す内容を説明していこう．この場合正しく識別できた文字数は293個であった．したがって残りは49文字であるが，実はこの集合は2種類に分かれる．その第1のものは，上の図2の標準文字パターン辞書にない別の形状を持つ文字パターンに該当する範疇（典型例は「し」に対する「志」．その一覧は文献1中の表3.1を参照.）で，28文字あり，本来ここでの識別対象からは除外すべきパターンであった．そこでこれを引いた21個が純粋な識別誤り文字個数となり，したがって最終的な正答率は93.3％（＝293／(342－28)×100)である．

ところでここで誤った文字パターンを調べたところ「か」が12個で半分強を占め，他に比べ圧倒的に多かった（次は「そ」の2個)．またこの中でつづき文字の先頭にある「か」が12個中の8個もあって，この事実は執筆者による試行錯誤的な分割操作の不徹底を意味しているのかもしれない．さらに加えて記せば，21個の誤り識別文字パターンの内で同じくつづき文字の先頭に位置していたものは13文字となっていた．

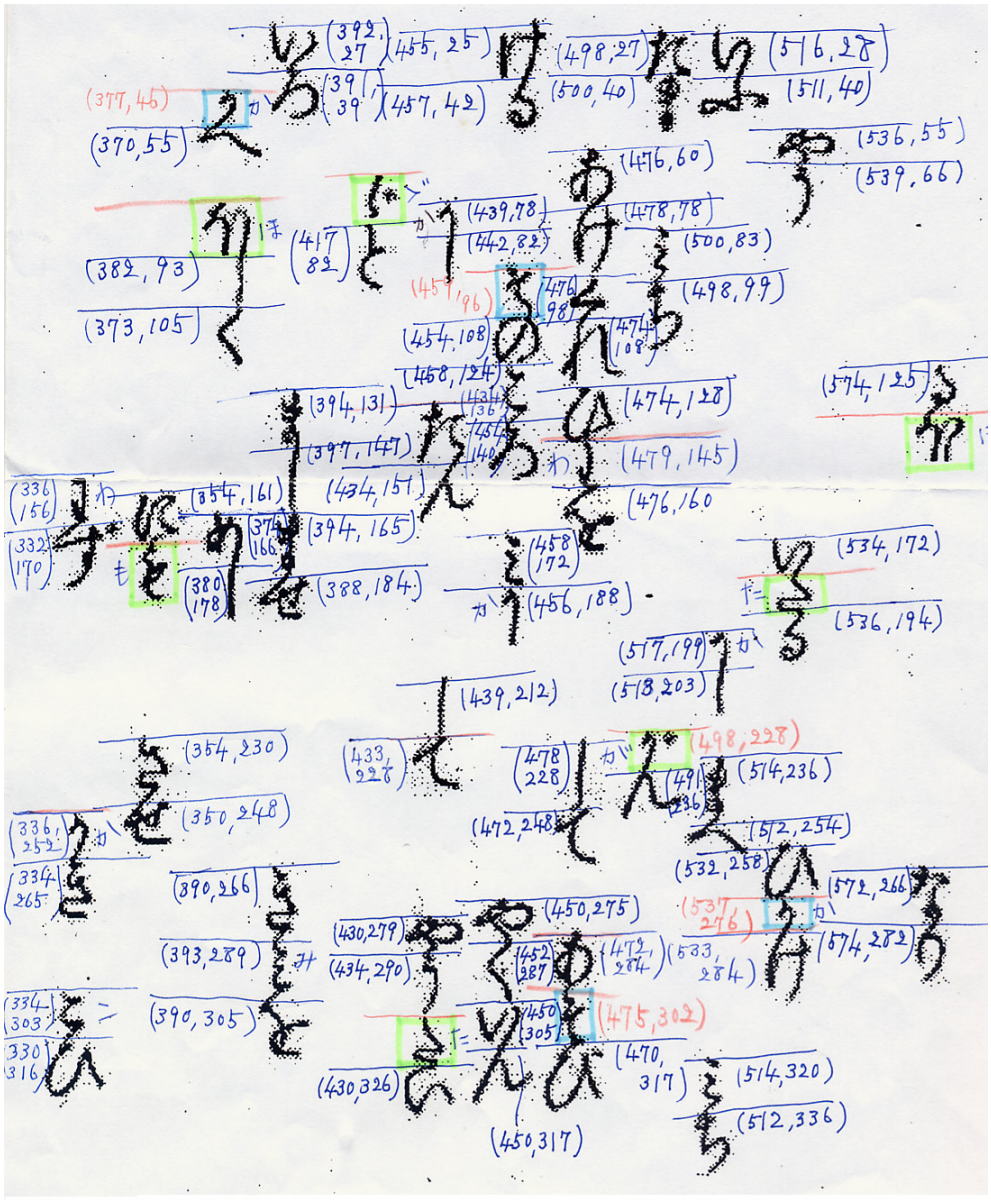

図3　分割箇所を指定したつづき文字部分の識別結果（一部）

表1　分割指定したつづき文字部分の個別文字パターン識別結果

対象文字総数	正識別文字数	辞書外文字数	誤識別文字数	
			先頭文字	非先頭文字
342個	293個	28個 （「志」など）	13個 （「か」×8 「そ」×2 「ま」、「み」、「め」）	8個 （「か」×4 「こ」,「そ」, 「ま」, 「わ」）

3. つづき文字部分の識別に関する諸特性の検討

3.1　標準文字パターンの縦と横の大きさ

文字を対象とするパターン識別，特にここで使用した最短距離法においては，文字のつづいた状態に対する分割も含めてその識別操作を行うために，標準パターンが中心的な基準としての役割を果たす．そこで個別文字パターンへの分割に際して重要な基礎情報を与える標準文字の縦と横の大きさ（画素数で表す）を，執筆者がパソコンのディスプレイ上に表示したものから読み取り，記録した．図4はその状況の結果の一部で，先の図2に示した上1／3位の部分にあたる．ここに示されているように，該当する文字パターンの右側に縦方向に対する大きさ，下側には横方向の大きさと括弧内にその半分の値（4捨5入して整数化したもの），また4角括弧中には小数点の付いたまゝの数値を記載している．なお図はメモ書きで余り見易くないが，この後に行う特性の検討の際の表にそれらの数値は明記してあるので，具体的にはそちらを見られたい．

3.2　文字パターン縦方向の大きさ特性に関する検討

つづき文字部分に存在する個別文字パターンを識別しようとする際に必要な前提条件は個々の文字パターンが縦方向どの位置範囲にあるかを決めるために，仮想的にでもせよ分割することである．そしてこのために使える基礎的な情報は，識別の基準根拠である標準文字パターンの縦方向大きさ（長さ）であろう．この値は前節において執筆者により具体的に読み取り，記録しておいた次第であったが，本当にこの測定した値が実際の正しい識別にどの

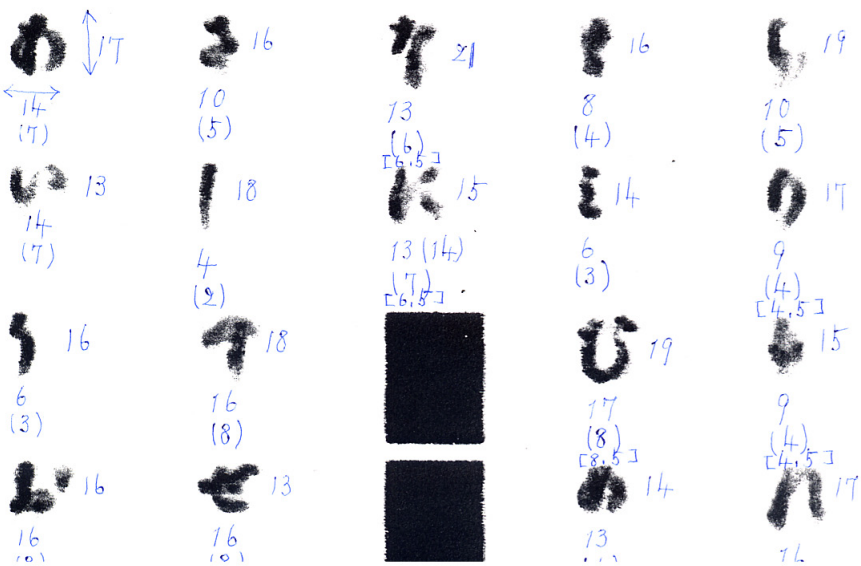

図4　標準文字パターンの大きさ測定の一部例

くらい適用できるものなのかは疑問が残る．幸いここでの報告では前の第2章において，正識別を達成する分割箇所はどの位置に設定すれば良いかのデータも人手により得ているので，本節ではこれら両者の値の間の関係を比較・検討してみることにする．

次頁から始まる表2はこれについての内容を示したものであり，まず左方の2列により各標準文字パターンについての縦の長さ（大きさ）を記してある．それには二つあり，各々「その1」，「その2」と名付けているが，「その1」は以前の文献1にある表3.2（p.36）から持ってきたもので，この数値の決定法の細かいところは失念してしまった．他方「その2」が今回の前節において測った値であり，この縦方向長さはそこでも書いたように，執筆者がディスプレイ上の標準パターンを目視して設定した．さてそこで両者は本来当然一致すべき値であるが，その理由は前者の決定法を忘れたため判然としない点があるけれどもずれている部分が存在し，その程度を一目で知ることができるよう実線（「その1」）と点線（「その2」）の折れ線グラフを用いて，図5に表示をしてみた．このグラフを観察すると，確かに数箇所で縦方向長さの大きな違いが見られる（「な」や「き」などがひどい）．

以上に対し表2の右半にある部分は，第2章に述べたできるだけ正答を得るときの各文字パターンに対する縦方向分割間長さである．この値については，得られた実際の全ての数値とその平均値を記してある．それは全てということで，例えば「い」の文字パターンについては8個，一番多い「け」には12個もある．そして今後の研究で行う予定の詳細な検討のために，これらの数値集合は；（セミコロン）により区切られており，この意味は図1の対象つづき文字部分のグループ1右半部，左半部，グループ2の右半部，左半部のどこから各数値がもたらされているかを，左からの順の；区切りにより示している次第である．

ところで識別の観点からは，この正しい識別結果の得られた実際の縦方向長さと，現実の具体的な識別過程ではそれだけしか使えない標準文字パターンの縦方向長さがどれ位違ってしまっているかが大いに気になるところである．これについて表2から個々の値は分るが全体的な様子は判断しにくいので，やはり先に挙げた図5中にグラフを用いて示すことを行った．すなわちここにある縦棒グラフが各文字において複数個ある正答となる縦方向長さ値の存在範囲であり，またそこに描かれている横棒は平均値の位置を示している．概略に見たところ半分以上は標準文字パターンの縦方向長さと重なっていると見られるが，極端な場合「れ」のように完全に離れてしまっているものもある．そこでこのような場合については，標準パターンの縦方向長さを頼りにつづき文字部分から個別文字を識別するために切り出しをしても不適当な状況となり，誤った識別を結果してしまう事態の生ずることが大層懸念される．

3.3　文字パターン横方向の大きさ特性に関する検討

前節に引き続き同様な方法と手順に従って，今度は文字パターンの横方向の大きさ（幅）について検討する．ただしつづき文字の識別に対処する分割のための情報を与える文字パターンの縦方向大きさと異なり，横方向にはこのような直接の関連はない筈である．といっても文字パターンの横方向大きさは文字毎にずい分違いがあり，また図2から推察されるようにここでの標準文字パターンは全て縦40×横50画素の文字枠の左上隅みに寄せた状態で

表２　文字パターン縦方向に対するいくつかの長さ

文字番号	文字名	標準文字パターン		つづき文字部分の分割箇所より	
		縦の長さ（その1）	縦の長さ（その2）	1文字毎の縦の長さ	平均長さ
0	あ	16	17	18;18,18;;20,18	18.4
1	い	13	13	12,12,12;10,14,16;8,10	11.8
2	う	18	16	;15,20;18,16;16,16,20	17.3
3	お	16	16	;14	14
4	か	11	10	4,4,13;6;6,4	6.2
5	き	22	18	23;18,20,22;22,22;22	21.3
6	く	19	17	10,18;12,10	12.5
7	け	18	18	20,13;20,20,18,18,20,22;17,21,18;20	18.9
8	こ	15	13	16,13;14	14.3
9	ゝ	13	12	;;;4	4
10	さ	16	16	18;14,20;16,14,18;16	16.6
11	し	18	18	20,16,18,12;;21;18,20,24,22	19
12	す	18	18	;;20,16;20	18.7
13	せ	13	13	;12	12
14	そ	17	15	;;16,10	13
15	た	16	17	13,15;12,20,14,18;18	15.7
16	ち	20	17	;19	19
17	つ	13	13	;14;14	14
18	て	19	19	-	
19	と	17	14	;;;24,16,22	20.7
20	な	16	21	16;16,20;20,22;20	19
21	に	16	15	;;16,12	14
22	の	16	15	16;;16;18	16.7
23	は	18	19	-	
24	ひ	19	19	17;20;;18	18.3
25	ふ	12	13	;;14	14
26	へ	11	10	;;10	10
27	ほ	19	19	;;18;22	20
28	ま	16	16	18,16,19;18,18;;15	17.3
29	み	15	14	16,16,16,16;;14	15.6
30	む	19	19	-	
31	め	13	14	12;;;12	12
32	も	18	17	;;18;18	18
33	や	13	13	11,12,11;13	11.8
34	ゆ	21	21	-	
35	よ	16	18	;;14,16;18	16
36	ら	20	19	;20;22	21
37	り	15	17	;16;20,18;16	17.5
38	る	17	15	;;;18	18
39	れ	17	17	20;24	22
40	ろ	18	17	;17;16;16,16	16.3
41	わ	17	17	15,14	14.5
42	を	17	17	;;;14	14
43	ん	16	16	-	

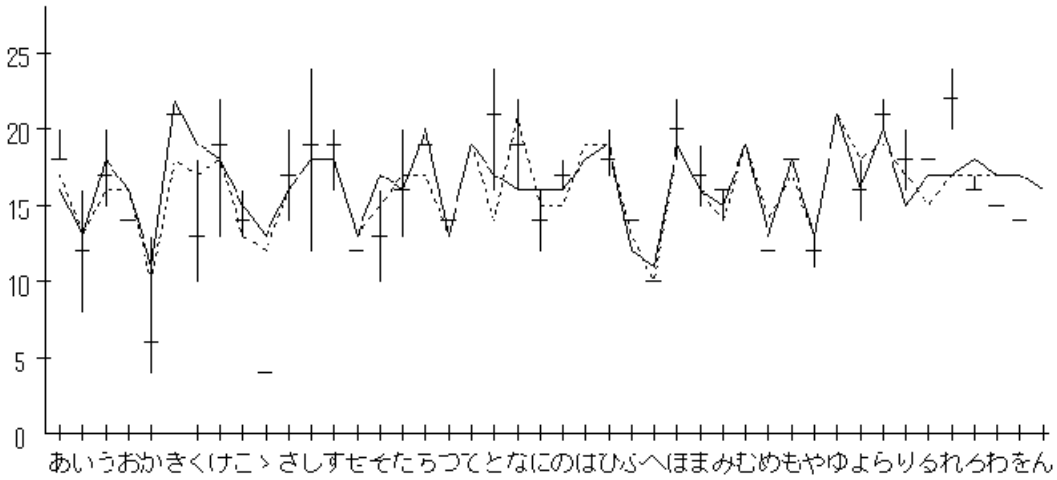

図5　文字パターン縦方向長さのグラフによる表示

用意されているので，つづき文字中の個別文字識別の際にはやはり適切なその横方向位置付けが必要になるわけである．

さてつづき文字パターン中に存在しているある一つの文字の横方向位置は，それに先立つ前の文字に影響されて決まってくると考えられる．そこでつづいて書かれているというこの状況について，図4に例示したような標準文字パターンの大きさ測定により得られたデータに基づくものと，図3で示した正しい識別のために指定した設定値によるそれとの，比較・検討を行うことにした．表3にこの結果を挙げる．この表には最上欄と最左欄に共に文字番号と文字名が記されているが，前者の最上欄の方がつづきの中で先行する文字，後者である最左欄列はこの文字につづいて次に存在している文字を表している．念のため例により述べると，最初の一番左上にある「0[;;;0]」なる箇所のデータは「あ」に続いてつづき文字「い」が存在しているという状態に対するものである．次にここにあるデータ自体の内容について説明すると，最初にある1個の数値は先行する文字パターンに対し後続の文字パターンが右横方向にどれ位ずれているかを表す量である．上の例ではこの値が0となっているので，先行の文字パターン[あ]に対し後続の文字パターン「い」は横方向位置ずれなし（0）ということである．またこのずれ量には当然先行する文字に対し右向きだけでなく左向きの場合もありうるので，＋の数値により右向き，－の数値で左向きを示すようにしている．ところで遅ればせながら，これらの数値は前の3.1節で執筆者が標準文字パターンから読み取り・測定した横方向大きさから算出したものであることに注意をしておく（横方向の大きさそれ自体の数値は，3列目に載せてある）．またこの表における先行文字－後続文字の組み合わせ列と行は，紙幅の節約上次に述べる[ ]内データが存在する際についてのみ記しているのを，お断りする．

表3　文字パターン横方向大きさに対するいくつかの比較した結果

文字番号	文字名	文字横大きさ	0 あ	1 い	2 う	3 お	4 か	5 き	6 く
0	あ	14							
1	い	14	0 [;;0]		-4 [-4]				-3 [-2]
2	う	6					0 [;;-4]		
4	か	6	. .						
5	き	10		2 [3]			-2 [-2]		
6	く	9						1 [;;0]	
7	け	14	0 [2;1]						
8	こ	8						1 [;2]	
9	ふ	5							
10	さ	10							
11	し	4	5 [;7]	5 [7]			1 [1;3;1;2]		3 [;6]
12	す	16							
13	せ	16							
15	た	13							
16	ち	11			-3 [0;-2-2]				
17	つ	15		-1 [;;0]	-5 [;-3]				
18	て	10							
19	と	7							
21	に	13							
24	の	15							
26	ひ	15	-1 [-2]						
27	ふ	18		-2 [-5]					
28	へ	15			-5 [;;-3;-4]				
29	ほ	14			-4 [;;-5]				
30	ま	8							
31	み	6						2 [3;;2]	
34	も	8				4 [;5]			
35	や	17							
39	よ	8							
40	ら	10		2 [;;5]	-2 [;0]				
41	り	9	3 [2;;3]						
42	る	9		3 [2]					
43	れ	16	-1 [;;-1]						-4 [-2;-2]
44	ろ	15		-1 [-1;0]					
45	わ	7							
46	を	15						-3 [;-3]	
47	ん	16		-1 [0;0]				-3 [;-2,-3]	-4 [;-3]

文字番号	文字名	7 け	8 こ	10 さ	11 し
0	あ				
1	い	0 [;0]			
2	う	4 [;;3]			
4	か			2 [;0]	
5	き				-3 [;;-7]
6	く	3[-2;-1]			-3 [-3;;-3]
7	け				-5 [;;-5]
8	こ		2 [;2]		
9	ふ				
10	さ			3 [;3]	
11	し				
12	す			-3 [-4]	
13	せ				
15	た				
16	ち				
17	つ			0 [;;-1;-1]	-3 [-5;-5]
18	て				
19	と				
21	に				-5 [;;-5]
24	の		-4 [-4]		
26	ひ				
27	ふ				
28	へ				-2 [-3;;-3]
29	ほ				
30	ま				
31	み				
34	も				
35	や				
39	よ				
40	ら				
41	り	3 [;3]			
42	る	3 [2;;3;2]			
43	れ	-1 [-3;-2;-2]			
44	ろ		-4 [-4]		
45	わ				
46	を				
47	ん	-1 [;;1]		-3 [-2;-3]	

文字番号	文字名	12 す	13 せ	14 そ	15 た
0	あ				
1	い				-1 [;0]
2	う				
4	か	5 [;;5]			
5	き				
6	く				
7	け				
8	こ				
9	ふ				
10	さ				
11	し				4 [;7]
12	す				
13	せ				
15	た				
16	ち				
17	つ				
18	て				
19	と				
21	に				
24	の				
26	ひ				
27	ふ				
28	へ				
29	ほ				
30	ま				
31	み				
34	も				
35	や				
39	よ				
40	ら				
41	り	1 [;;0]		-3 [;-3]	
42	る				
43	れ				
44	ろ				3 [2]
45	わ				
46	を			0 [;3]	
47	ん				2 [3]
48	ゑ				
49	ゐ				
50	ゆ				
51	ゝ	1 [;;0]			
52	ゟ		0 [2]		-2 [0;0;0]

文字番号	文字名	16 ち	17 つ	19 と	20 な	21 に	24 の	26 ひ	27 ふ	28 へ
0	あ									
1	い									
2	う					4 [;;4]	5[;;4]			
4	か	3 [4]								
5	き			-2 [;;-2]			3 [;2]			
6	く		3 [2]							
7	け							1 [2]		
8	こ						4 [4]			
9	ふ									
10	さ			1 [;;4]						
11	し									
12	す							5 [;-5]		
13	せ							-1 [;;1]		
15	た									
16	ち									
17	つ									
18	て									
19	と									
21	に									
24	の									
26	ひ							4 [;;4]		
27	ふ									1 [;;0]
28	へ								2 [;;0]	
29	ほ									
30	ま									
31	み									
34	も									
35	や									
39	よ									
40	ら					-1 [;;0]				
41	り									
42	る				4 [;;3]					
43	れ									
44	ろ									
45	わ		4 [;;3]							
46	を									
47	ん			-1 [;;0]	2[2;3;1;3;3]					
48	ゑ									
49	ゐ				-2 [;-3]					
50	ゆ									
51	ゝ									
52	ゟ			-4 [;;-2]						
53	江							4 [5]		
54	者									
55	川									
56	遍									

29	30	31	33	34	35	39	40	41
ほ	ま	み	め	も	や	よ	ら	り
4 [;;6]					6 [3,4;5;4]			
		0 [-2;;2]						2 [;;0]
					4 [2]	-1 [;;;-2]		
								-3 [;-2;-3]
	2 [3]		5 [5;;;7]					
	-4 [;;-3]	-5 [;;-4]						
	-4 [-6]							
							-2 [;;-6]	
		-3 [-2;-2]						
				-4 [;0]		4 [;;-5]		-3 [;;-5]
-1 [;;0]								
	-4 [-2]							
	0 [0]							
								-4 [;;-4]
	-1 [;;;1]					-1 [;;-1]		
		-5 [-3]					-3 [;-2]	
	-4 [-5]							

42	43	44	45	46
ろ	れ	ろ	わ	を
		1 [;;-1]		
	1 [;0]			
		5 [4,4]		
			-4 [;;-2]	1 [;;2]
		5 [;;-5]		
-1 [;;0]				
				8 [;;;7]
			-5 [-4]	
				0 [;;-3]
	1 [0]			
		0 [;;2]	4 [-3]	
				-1 [;;-1]

それでは続いて[ ]内にある数値データの方の説明であるが，これは図3における指定した設定値から求めたものである．先般の例ではこの箇所のデータは「;;;0」となっているが，この中の「;;;」の意味は前節と同じでこのデータがグループ2の左半部に属する場合の数値であるのを示しており，その「あ」−「い」の実際の図3での連続では横方向位置ずれが0（ずれはなし）であったことを記しているわけである．この元になった値はできるだけ正答が得られるよう人間が指定したものであったから，標準文字パターンに基づく現実の識別が成功する前提としては，この数値は[ ]の前にある値と一致していることが望ましい．そこでこの両者の値が±2以上異なっていることは正しい識別の達成上不利であると予想される次第であり，この場合についてはその箇所の欄の背景を灰色にして目立つようにした．そこでこの色付けした結果を本表で観察すると半分弱はその種の箇所が存在するようであり，この事実から単に標準パターンに基づいて識別の対象文字パターンについて横方向位置設定をしたのでは，余り良い識別結果の得られないことが予測される．さらにこの状況からは，少なくともある程度の横方向正規化処理操作の必須な事柄が考えられる．

4．つづき文字部分の領域抽出に関する検討

これまでの章・節においてはもっぱら，各つづき文字パターン部分中に存在する個別文字の識別に関するいくつかのデータを検討してきた．しかしこの識別を実際に行う際には，その前に1頁の紙面にある文全体から各々のつづき文字部分領域を抽出しておかなければならない．科研費報告書である文献1の35頁に載せた図3.10に，通常広く使われる文字線につ

いての縦および横方向ヒストグラムの閾値化に基づくこの領域抽出結果が示されているので，これを用いて前章まで行ったのと同様な趣旨の検討を実施する．

頁の下にある図6は，上のようにして領域抽出されたつづき文字部分の先頭である左上隅位置の画素座標を，人間（執筆者）がやはり視察により読み取ったものの1部（グループ1右半部）である．ただしこのときその読み取りが容易なように抽出領域画像をディスプレイ上で2倍に拡大して測定をしたので，画素の座標値も現実の2倍となっている．

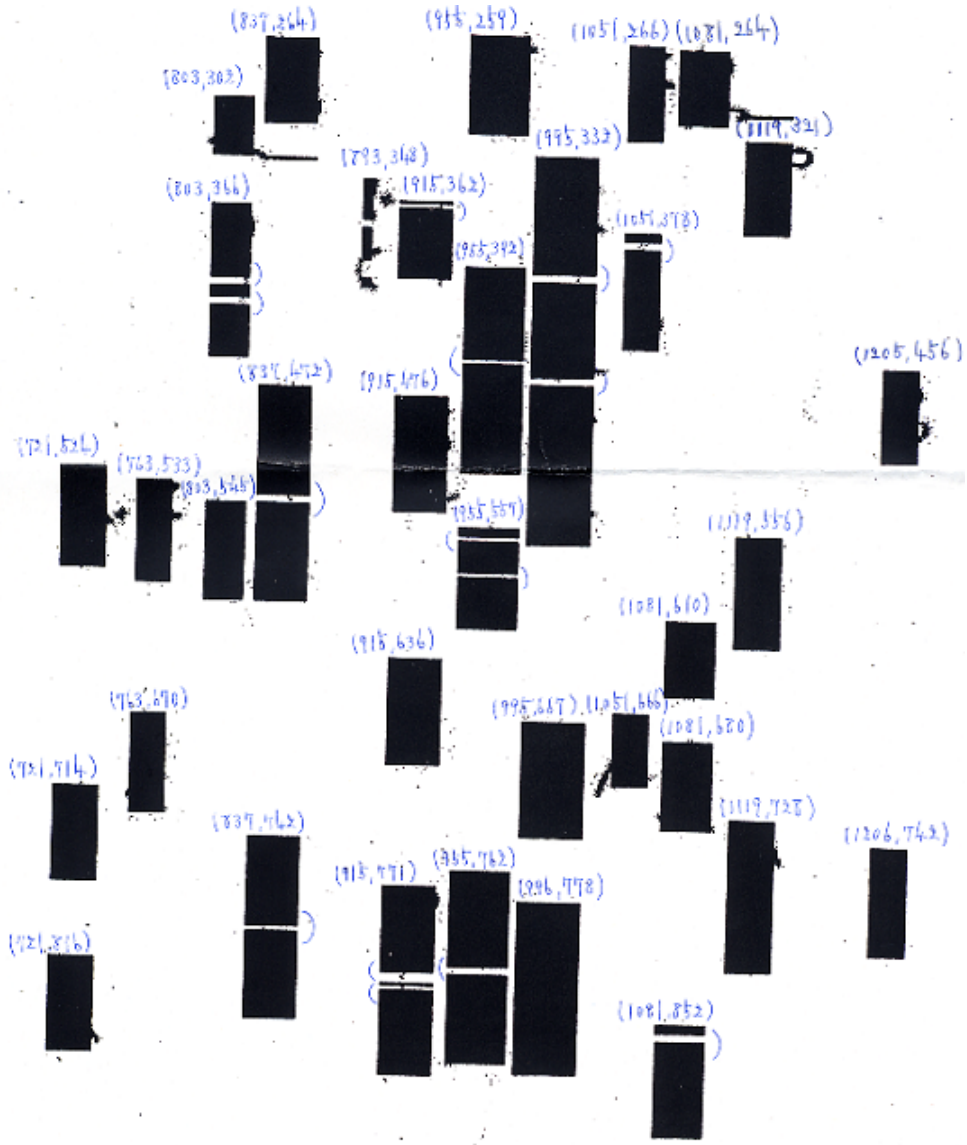

図6　つづき文字部分の抽出領域の測定結果（一部）

表4　つづき文字部分抽出領域の測定と指定値の比較
（上からグループ1右半部と左半部，グループ2右半部と左半部）

13	12	11	10	9	8	7	6	5	4	3	2	1	列番号
21 [20]	20 [20]	17 [18]	28 [29]	11 [13]	20 [20]	20 [20]	28 [21]	15 [20]	19 [21]	44 [40]			文字列の間隔量
157 [156] 252 [252] 303 [303]	162 [161] 230 [230]	46 [(46)] 78 [(-)] 168 [166]	27 [27] 131 [131] 266 [266]	69 [(-)]	76 [78] 133 [136] 213 [212] 281 [279]	25 [27] 91 [(96)] 174 [172] 276 [275]	61 [60] 229 [228] 284 [284]	28 [27] 84 [83] 228 [(228)]	27 [28] 200 [199] 235 [236] 321 [320]	56 [55] 173 [172] 259 [258]	- [-]	123 [125] 266 [266]	つづき文字部分の先頭画素位置
14 [17]	22 [21]	22 [19]	18 [20]	21 [20]	21	21 [42]	19 [18]	23 [22]	18 [20]	19 [18]	20 [21]		文字列の間隔量
315 [316]	68 [70] 254 [250]	17 [17] 85 [84] 175 [174] 278 [277]	32 [30] 231 [232] 284 [282] 365 [(366)]	97 [(-)] 174 [174] 325 [322]	14 [14] 80 [80] 178 [178] 289 [292] 355 [356]	336 [(-)]	45 [45] 89 [(-)] 134 [134] 325 [326]	30 [32] 67 [68] 132 [(132)] 262 [260] 357 [(356)]	45 [46] 97 [96] 151 [(152)] 297 [(294)]	16 [(-)] 75 [75] 224 [226]	112 [(112)] 259 [258] 305 [304]	191 [192] 244 [240] 312 [312]	つづき文字部分の先頭画素位置

13	12	11	10	9	8	7	6	5	4	3	2	1	列番号
25 [23]	17 [17]	18 [19]	19 [18]	21 [21]	24 [24]	22 [21]	21 [20]	21 [21]	16 [18]	21 [20]	23 [21]		文字列の間隔量
247 [(-)] 364 [364]	41 [42] 139 [138]	163 [164] 381 [(-)]	61 [60] 255 [256] 345 [346]	104 [112] 336 [336]	249 [250] 308 [(308)]	124 [124] 223 [220] 353 [352]	122 [124] 199 [199]	59 [60] 189 [190]	64 [62] 261 [260]	114 [(114)] 233 [234] 299 [300] 375 [376]	41 [40] 315 [314]	41 [38] 124 [126]	つづき文字部分の先頭画素位置

22 [19]	26 [26]	14 [14]	23 [20]	17 [22]	20 [20]	20 [21]	22 [23]	19 [19]	21 [18]	20 [22]	20 [18]		文字列の間隔量
51 [50] 208 [206] 331 [330]	235 [234]	144 [140]	278 [276]	35 [(-)]	103 [(92)] 210 [208]	91 [(-)] 154 [(-)] 278 [272]	67 [(70)] 249 [248]	32 [32]	72 [72] 348 [347]	84 [84] 343 [342]	35 [(34)] 187 [188] 256 [256] 334 [332]	35 [34] 99 [104] 328 [328]	つづき文字部分の先頭画素位置

さてそこでこうして求めた画素座標値を，先にできるだけ識別の正しく得られるよう人間が指定したつづき文字部分に対する分割箇所を表示した図3と比較・対照して，検討する．その結果が前頁の表4に挙げたものである．この内容には2種類あり，第1は各縦方向文字列がどれ位適切に抜き出されているか，次に第2としてそれらの文字列中で複数ある個々のつづき文字部分が如何に正確に取り出されているかである．この表において第1の問題については，対象である紙面の絶対的な位置付けを決める原点が比較している2者（図6と図3）で異なっている心配があったので，その検討は座標値自体でなく，値の差である文字列の間隔量を用い示してある．そしてその欄にある裸の数値は実際の識別時に使える図6から読み取った値に基づく量,[ ]内のそれは理想的な場合であるとして人間が設定した値から計算した量である．また加えて比較するこの両方の量が3（画素）以上違っている欄については，その背景に灰色を塗って目立つようにしている．

次にこの表4の残りの大きな部分は，各縦文字列内にある個々のつづき文字部分の左上端に対する画素位置の縦座標値についての検討結果である（この横座標値は上記の文字列抽出の際に，既に列毎に共通して決められている）．そしてこの欄の内容の意味するところはほとんど上に述べたものと同じであるが，そこにある（ ）で囲まれた縦座標値はつづき文字部分の先頭文字パターンが第2章の表1における辞書外や誤識別先頭文字であるため，その値は信頼できないゆえ対象から外すのが妥当と考えられるものである．またこの場合についても両者の値が3以上違った欄は，灰色背景にしている．

以上のような検討をしてみた結果，前章の個別文字識別の場合に比べて，今回は割合に問題を起こす灰色背景箇所の少ないことが判明した．つまり識別処理におけるこのつづき文字領域抽出部分は最終的なつづき文字の個別識別に対して，余り悪い影響をもたらさないであろうことが期待できた次第である．

5. あとがきと今後の課題

この紀要における報告では，つづき文字パターンに対する実際的な現実の識別結果の実現とその評価というよりは，それを始めるに当たってその問題点と限界を予想させるいくつかの予測的な調査型の研究を実施した．

そのため先ず対象とするつづき文字パターンについて，それができるだけ正しく識別されるようにする縦と横方向の分割位置を，人間（執筆者）により設定する試みを行った．この結果はいわば理想的な識別結果とそれをもたらす条件を調べたことになる．ところでこのようなつづき文字パターンについての識別方法は，基本的にそのための標準文字パターン自体と，それが持つ縦および横方向の大きさ情報に基づいて実現されるのが通常である．そこで各標準文字パターンについてその縦と横方向の大きさを人間が測定し，こうして得られた値を用いてつづき文字パターンの分割に必要な位置情報がいくつに設定されるかを算出した．こうして出された結果の量が先の人間が設定したものと一致すれば，正しい識別が達成できる筈である．したがって最後にこれら両方の値を比較・検討し，それほど簡単に識別を実現

できる処理操作はないらしいという予想を結論した．

以上のような過程を経れば，それでは実際の識別結果はどのような具合になるかに当然関心が向く．この実行については，本報告のまとめに続いて至急行う計画である．そしてその結果はもっとも近く開催される情報処理学会全国大会［2］で発表する積りであるから，ぜひそちらを参照してくださるようお願いする．

謝辞

本研究がその一環であった科学研究費を一緒に実施され，その中でデータの提供，いろいろな機会における議論などをして頂いた，同僚の教員である島田大助教授，三好哲也教授，三輪多恵子准教授に，心から感謝を申し上げます．

参照文献

1．江戸版本の読解を支援する運筆特徴を考慮したつづき文字の認識に関する研究，科学研究費（課題番号17500165）報告書，平成20年3月
2．江戸版本のつづき文字部分に対する識別の試み，情報処理学会第71回全国大会，平成21年3月12日，発表予定