Bulletin of Toyohashi Sozo University
2010, No. 14, 49–59

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［第2報］

舟久保　　登

本紀要の前号において執筆者は表題の報告をしたが，この論文はその後に行った研究内容を述べるものである．ここでの対象であるつづき文字部分の第2文字を識別するためには，第1文字に対応した標準文字パターンの縦方向大きさを用いていて，その結果の識別率は53.85％であった．今回この縦方向大きさをいくつかの条件を考慮して改変したところ，上記識別率が78.6％にまで向上する事柄が得られた．しかし依然として識別できない場合が残っており，さらなる検討に備え，その具体的なつづき文字パターンの状態を正しく識別する状況と比較対照した形で示した．このような次第で，本課題の研究は未だ継続中であることを，最後にお断りさせていただきます．

Key Words：continuous character parts discrimination, engraved printing Edo period book, minimum distance method, standard character pattern dictionary, height length of character pattern

I　まえがき

本報告は，昨年度の豊橋創造大学紀要・第13号に執筆者が書いた表題の研究に対し，その後に行った内容を記したものである．したがって既に述べた事項については特に繰り返すことなくそれに基づいた対象として使っているので，これらはそれが載っている最後に記した文献をぜひ参照されたい．

また残念ながら，ここで扱っている研究課題はまだその結果を達成できたとは云えない．本報告を読んでいただく前に，この事柄をお断りしておく．

II　情報処理学会で発表した識別結果における状況

表記学会の第71回全国大会（平成21年3月開催）で，執筆者はつづき文字部分の第2文字につき識別率が53.85％であると発表した．ただそこでは全国大会に許された紙幅の関係上，この値が得られた諸種の条件（状況）を余り明確で詳しく記すことができなかったので，報告の最初のこの章においてこれに関し述べることにする．

ここで対象としたつづき文字部分は後に一括して載せた図1（発表別刷と第13号紀要でも同じく図1）に示したパターンで，これは本研究が取り扱うこととした江戸版本「女五経」

中の見開き2頁から，文字パターンがつづいて版刻されている部分を抜き出したものである．ただし上記の識別率はこの図1の中のつづき文字パターンを全て対象にしたものではなく，発表文にも書いているようにその目的が，分離した個別な文字状態の場合に正しく識別されることの判明している文字パターンに対して，それがつづけて存在する際にどうなるかを知りたい要求であったゆえ，そのような文字パターンから成るつづき文字部分の第2文字のみを対象としている．そして問題であるつづけのための影響を純粋に調べたいので，識別の際のつづき文字部分の先頭左上縦横座標画素およびつづき効果を直接受ける第2文字の仮想的文字パターン枠左上横座標画素位置の三つの値は，人間による設定値を採用した．つまりここでのつづきによる影響は，結局つづき文字部分内における第2文字パターンの縦方向位置に集約されていると想定した次第である．

このようにして得られた識別に関する全ての詳細な状況を，これも後に集めた表1の中に挙げた．表1は大別してグループ1・右半部，左半部とグループ2・右半部，左半部の四つから成っており，その各部の図1における対応箇所はそこから容易に判ると思う．そして各半部には縦書きの日本文に対して右端から番号付けした（縦）行が数えられていて，対象であるつづき文字部分の存在場所が明確に示されている．実際のつづき文字部分内容に続く下の三つの値はその対象部分の先頭左上横座標画素位置と縦座標画素位置，また第2文字の左上横座標画素位置で，これらは上述したように人間（執筆者）が与えた．引き続く4番目の値は，先頭文字を識別した文字名（これは常に正解）に対する標準パターンの縦方向大きさである．この値は各標準パターンについて唯一の大きさに決めており，この値に基づき対象である第2文字識別のための文字切出し（分割）が行われるわけである．

次の横欄が以上の状況下においての第2文字識別結果で，合っているときには「正解」，識別を誤ったものは何と間違ったかの文字名を記している．これに基づきその統計を各表の小計から取ると，この場合の対象文字個数は 29＋33＋22＋19＝103個，正解文字数は 12＋21＋13＋10＝56個で，よって全体の識別率としては 56÷103＝0.544＝54.4%という結果が得られるのである．（この識別率の値は，先の情報処理学会において報告した53.85%と微妙に違っている．その原因は今回表1の内容を整理した際に，単独に分離された状態でもその正しい識別に無理があると判断した一つのつづき文字部分を除いたためであり，ここでは統計的な計算の誤差範囲内として読者の許容をお願いしたいと思う．）

Ⅲ　標準文字パターン縦方向大きさの改変と識別率向上

ここで述べている文字識別には原理的で簡単な最短距離法を用いているが，その際使用した各文字についての標準パターンは図2のようである．そして本研究課題のつづき文字部分についての識別に対する中心問題とした対象の第2文字についての分割箇所は，その左上縦横座標画素の位置を人が設定しているので正しく識別されることが保証されている先頭の第1文字標準パターンの縦方向大きさにより決めているという事項は，前章で記したところであった．またこのときに用いた実際の縦方向大きさの値は，第13号の紀要に述べたよう

に，図2に示したパターンをパソコンのディスプレイ上に表示して執筆者が観察し，それから各文字毎に唯一つずつ決定したものであった．

しかしこうして決めた値と，つづき文字部分中の第2文字を正しく識別できるよう個々の第2文字に対して人が設定した分割箇所の座標値とは，各文字により相当の違いのあることが分っている．（これについての端的な概観は，前号紀要の図5を参照のこと．なお学会発表にも同様な図が存在するが，こちらにはその作成時に若干のデータのミスがあったので，紀要の図の方を見て欲しい．）

そこでつづき文字部分の第2文字識別のための分割箇所に直接影響を与える各標準文字パターンの縦方向大きさについて，その変更の検討を行うことにした．この際に考えた基本的な立脚点は，次の3点である．

① 変更する大きさの値は，これまでの決定値と正しく識別できるよう人間が設定した値との間のものとする．

② こうして新たに決める縦方向大きさ値は各標準文字パターン毎に唯一とするが，その値にはこの標準パターンに対応する第2文字の正しい識別個数がなるべく多くなるような（各文字概念毎における識別率が最大）ものにとる．

③ 複数の大きさの値に対してその識別率が同じなら，小さい値の方を採用する．この理由は文字のつづき効果により文字パターンに余計な縦部分が生じるであろう筈のため，それを後に明確に考慮する根拠を残しておきたいゆえである．

上のようにして改変した標準パターンの縦方向大きさの値は，表1の第8行（「改変縦大きさ」）に記してある．そこで「同じ」とあるのは値の変更なし，具体的な数値を示してあるものは変更した値である．ざっと見て2／3位を改変した．

さてこのときに得られたつづき文字部分中の第2文字に対する識別結果は，その次の「新識別結果」の行に載せてある．そしてこれの全体についての識別率は大分良くなって，81÷103＝0.786＝78.6%と向上した．

Ⅳ　誤識別な第2文字パターンの分割に関する検討

表1–1〜4の「新識別結果」欄が表しているように，この報告でもっぱら対象としてきたつづき文字部分内の103個の第2文字について，その分割箇所を決める標準パターンの縦方向大きさを改変（変更・調整）しても，22個の誤り識別を生じてしまうことが判った．そこでこの事実に対処する方法を今後見出さねばならないが，そのための参考資料となる具体的な分割文字パターンを図3–1に挙げた．この掲載順序は右列から「あいうえお」に従い，また1概念の第2文字について誤りが複数場合あるときは，表1の順番に依って横方向に右側より並べている．

しかしながらこの図3–1の内容を見ても，文字がつづけて版刻されたことによる誤識別の原因の様子は余りはっきり摑めないようである．そこで前の紀要で実施した第2文字を正しく識別するように人間が設定した文字枠左上縦座標位置（表1–1〜4の最下欄）を用い，正解し

ている際における第2文字パターンの分割の様子を図3–2に示した．こうすれば先の図3–1の表示結果と比較して，つづきのための識別誤りの原因が視覚的に把握でき易いと思う．

実際これを行ってみると，識別を誤ったほとんどの第2文字パターンについて，その上の部分につづきの影響による余分な範囲の存在していることが見て取れ，これが誤りの原因になっていると考えられる．ただしほんの少数であるが，標準パターンの縦方向大きさの値がいき過ぎ，第2文字パターンの上部が欠けてしまっているものも見られた．今回の報告の内容はここまでであるが，今後はこの点にも注目して研究を進めていく必要があろう．

V　まとめと今後に残された問題

この報告においては，先ず昨年の情報処理学会の大会で発表したつづき文字部分の第2文字を識別した際の諸条件（状況）を，詳細で明確に説明した．また続いてこの第2文字識別のために行うつづきに対する分割箇所の決定は，それに先行する第1文字についての標準パターンの縦方向大きさを用いて決めているので，この値の改変を実施した結果，その識別率が54.4%から78.6%に上昇することが得られた．

しかしながらこれでもまだ103個の第2文字中22個の誤りが生じている．そこでこのときの具体的な分割した文字パターンと正確に識別された際の文字パターンの様子を，比較できるよう図3–1，2に挙げた．したがって残された問題は，この状態に対処できる処理法を開発することである．この事柄は元来の科学研究費の中心主題でもあったが，これについてこれまでの研究経過から次の内容が示唆されたように思われる．

図3–2に示した正しい識別結果を得たときの状態は，人による第2文字の左上縦座標位置の設定にあった．そこでこれから分ったのは，少なくとも本研究で扱ってきた版刻した文字の場合，つづきの効果は字体の変形をもたらすほどのものではなく，その縦位置の変動を表す程度のそれであるらしいということである．この事実をきっかけとして踏まえてここでの問題を解決すべく，さらに研究を進める所存である．

謝辞

本研究がその端緒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，心からなる感謝を申し上げます．

【参考文献】

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．江戸版本におけるつづき文字部分の識別についての検討，豊橋創造大学紀要第13号，平成21年3月
3．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

図1　対象とするつづき文字部分
（上半：グループ1・右，左半部，下半：グループ2・右，左半部）

あ	さ	な	ま	ら
い	し	に	み	り
う	す	(ぬ)	む	る
お	せ	(ね)	め	れ
か	そ	の	も	ろ
き	た	は	や	わ
く	ち	ひ	(い)	を
け	つ	ふ	ゆ	ん
こ	て	へ	(え)	
ゝ	と	ほ	よ	

図2　標準文字パターン辞書

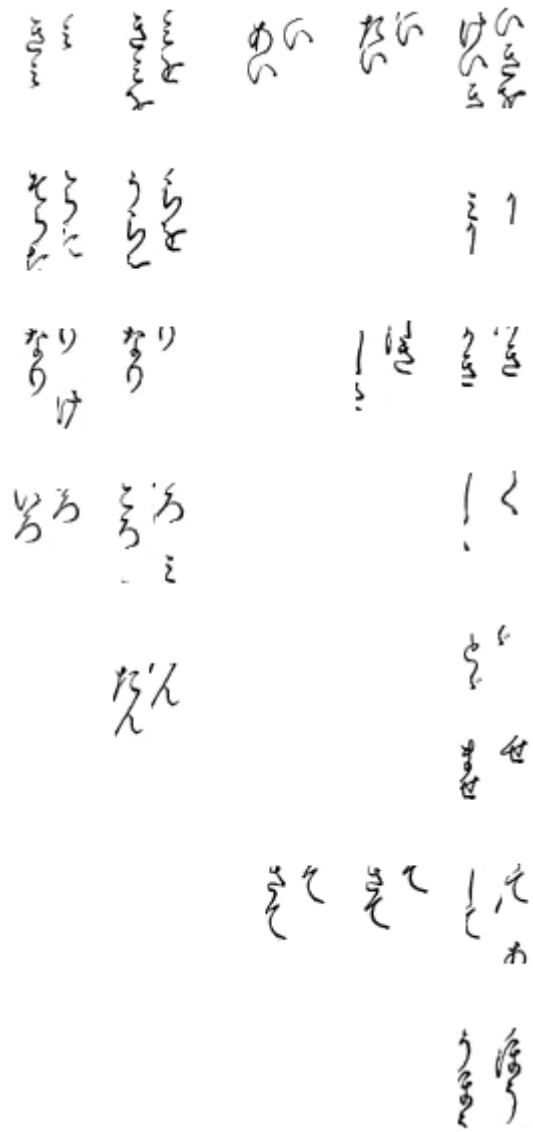

図3–1　第2文字のパターン分割状態

図3–2　正識別された第2文字のパターン分割状態

表1–1　つづき文字部分の識別結果（グループ1・右半部）

	6		5				4	3	1	右からの行目
く	あ	み	た	み	ま	か	い	や	な	つづき文字
れ	け	ち	ま	ち	へ	し	ふ	う	り	部分内容
476	476	500	498	514	514	517	516	536	572	対象部左上横
98	60	83	27	320	236	199	28	55	266	〃　左上縦
474	478	498	500	512	512	518	511	539	574	第2字左上横
17	17	14	17	14	16	10	13	13	21	標準縦大きさ
と	正解	と	と	て	正解	正解	正解	か	↘	識別結果
10	同じ	15	14	15	18	5	10	11	同じ	改変縦大きさ
正解	正解	正解	正解	正解	正解	正解	正解	正解	↘	新識別結果
ー	ー	ー	ー	ー	ー	ー	ー	ー	282	第2字左上縦

			8					7	6	
や	し	た	か	い	や	み	こ	け	し	
う	て	ん	し	ん	く	か	ろ	る	て	
430	439	434	439	450	450	458	458	455	478	
279	212	136	78	305	275	172	124	25	228	
434	433	434	442	450	452	456	454	457	472	
13	18	17	10	13	13	14	13	18	18	
か	正解	正解	↘	正解	正解	と	ん	正解	く	
11	同じ	14	5	10	11	15	同じ	20	同じ	
正解	正解	正解	正解	正解	正解	と	ん	正解	く	
ー	ー	ー	ー	ー	ー	188	140	ー	248	

小計			13	12	11				10
	こ	か	わ	さ	め	き	ま	ま	い
29	ひ	き	ず	せ	し	み	せ	し	ろ
	334	336	336	354	374	390	394	394	392
	303	252	156	230	166	266	165	131	27
	330	334	332	350	380	393	388	397	391
	13	10	17	16	14	18	16	16	13
12	正解	る	ふ	す	正解	し	ふ	正解	つ
	同じ	5	14	18	12	19	18	18	10
21	正解	て	正解	正解	正解	し	ふ	正解	ん
	ー	265	ー	ー	ー	289	184	ー	39

表1–2　つづき文字部分の識別結果（グループ1・左半部）

	5		4		3		2		1	右からの行目
ま	か	け	な	た	け	あ	さ	ひ	ま	つづき文字
へ	し	り	り	↘	れ	け	ん	し	も	部分内容
187	188	205	204	222	225	242	244	261	265	対象部左上横
68	32	96	46	226	75	304	258	312	192	〃　左上縦
182	189	208	207	229	222	243	242	266	265	第2字左上横
16	10	18	21	17	18	17	16	19	16	標準縦大きさ
つ	正解	正解	↘	か	正解	正解	正解	正解	と	識別結果
18	5	20	同じ	14	20	同じ	18	同じ	18	改変縦大きさ
正解	正解	正解	↘	正解	正解	正解	正解	正解	正解	新識別結果
ー	ー	ー	62	ー	ー	ー	ー	ー	ー	第2字左上縦

9					8		6		5	
う	き	な	た	け	お	い	ち	け	あ	
ち	ん	れ	ん	る	も	し	か	れ	り	
105	124	123	124	123	121	163	164	182	182	
174	356	292	178	80	14	134	45	294	260	
105	122	120	124	126	126	170	168	180	184	
16	18	21	17	18	16	13	17	18	17	
正解	正解	正解	正解	正解	と	正解	↘	正解	正解	
同じ	19	同じ	14	20	14	10	19	20	同じ	
正解	正解	正解	る	正解	正解	正解	正解	正解	正解	
ー	ー	ー	198	ー	ー	ー	ー	ー	ー	

12				11				10	9	
さ	う	き	た	た	つ	せ	き	け	や	
ん	ら	ん	い	る	く	ん	を	い	う	
44	66	64	61	61	80	82	85	82	100	
70	277	174	84	17	282	232	66	30	323	
41	66	61	61	64	82	84	83	82	105	
16	16	18	17	17	13	13	18	18	13	
る	し	正解	正解	と	正解	正解	く	く	正解	
18	同じ	19	14	14	同じ	同じ	19	20	11	
正解	し	正解	く	正解	正解	正解	正解	と	正解	
ー	297	ー	102	ー	ー	ー	ー	52	ー	

小計							13		12	
							こ	ろ	れ	
33							↘	う	い	
							23	40	40	
							316	290	250	
							25	44	40	
							13	17	17	
21							正解	正解	し	
							同じ	同じ	24	
28							正解	正解	正解	
							ー	ー	ー	

表1–3　つづき文字部分の識別結果（グループ2・右半部）

	6	5		4			3	2	1	右からの行目
け	い	さ	き	た	い	さ	な	な	ろ	つづき文字
う	つ	て	み	ん	ら	し	り	み	う	部分内容
461	463	485	502	504	520	521	523	546	564	対象部左上横
199	124	60	260	62	376	300	234	40	38	〃　左上縦
464	463	484	504	504	525	524	524	549	568	第2字左上横
18	13	16	18	17	13	16	21	21	17	標準縦大きさ
↘	こ	正解	て	正解	う	正解	正解	正解	正解	識別結果
20	10	18	19	14	10	18	同じ	同じ	同じ	改変縦大きさ
正解	正解	か	し	正解	正解	正解	正解	正解	正解	新識別結果
ー	ー	74	282	ー	ー	ー	ー	ー	ー	第2字左上縦

	12	11			10		9			7
け	け	う	か	そ	に	そ	す	さ	り	よ
ん	ん	へ	し	の	ほ	ら	を	か	か	の
341	341	364	381	381	380	398	400	440	443	445
42	42	164	346	256	60	336	112	352	220	124
342	342	360	383	378	380	401	400	440	443	440
18	18	16	10	15	15	15	18	16	17	18
正解	正解	正解	正解	い	は	正解	く	と	↘	正解
20	20	同じ	5	10	16	10	20	18	18	同じ
正解	正解	正解	正解	正解	正解	と	正解	正解	正解	正解
ー	ー	ー	ー	ー	ー	352	ー	ー	ー	ー

小計								13	12
								み	も
22								す	ん
								322	343
								364	138
								318	342
								14	17
13								正解	正解
								15	18
19								正解	正解
								ー	ー

表1–4　つづき文字部分の識別結果（グループ2・左半部）

				4	3		2			1	右からの行目
	あ	さ	け	を	め	な	す	う	ゝ	う	つづき文字
	い	て	る	し	し	り	か	ち	ろ	ち	部分内容
	200	203	201	200	221	240	240	261	263	264	対象部左上横
	347	144	106	72	84	256	188	328	104	34	〃　左上縦
	200	202	203	207	228	243	245	259	258	262	第2字左上横
	17	16	18	17	14	21	18	16	12	16	標準縦大きさ
	と	正解	正解	正解	正解	正解	ゝ	正解	を	正解	識別結果
	同じ	18	20	同じ	12	同じ	20	同じ	4	同じ	改変縦大きさ
	と	か	正解	正解	正解	正解	正解	正解	正解	正解	新識別結果
	367	160	ー	ー	ー	ー	ー	ー	ー	ー	第2字左上縦
小計				13	12	11	8	7	6	5	
		よ	と	ひ	も	と	し	し	う	ま	
19		く	き	と	の	ゝ	く	き	ほ	り	
		20	19	17	38	61	124	144	164	184	
全体の		330	206	50	234	140	208	272	248	32	
識別率		18	17	21	35	65	118	137	158	185	
56/103		18	14	19	17	14	18	18	16	16	
＝0.544	10	正解	よ	正解	つ	し	ん	く	ん	正解	
81/103		同じ	22	同じ	18	22	同じ	同じ	同じ	18	
＝0.786	13	正解	正解	正解	正解	し	ん	く	ん	正解	
		ー	ー	ー	ー	164	230	296	268	ー	