

# 専修西洋写本 MS 1 『ポリクロニコン』と 「単語単位」認識プログラム

松 下 知 紀・寺 沢 憲 吾

## 0. はじめに

人間には言語という優れた認知能力が備わっており、物事を伝達することができる。情報工学の文字認識の研究は目覚ましい進歩をしており、現在の活字資料であればスキャナーで高度の認識率を示す。しかし、初期印刷本や西洋手稿写本の場合、現在も認識率が低い状態が続いている。その原因として、現代の活字は基本的に1文字に対して1つの字形であり、様々な単語のどの位置に現れても同じ形状をしているので、文字単位の認識が基本的可能である。しかし、古代文字の場合、同一の文字に対して複数の字形が存在する。そのような問題に取り組む姿勢が情報工学の分野に芽生えて来た。情報工学と人文科学の学際性は現在極めて重要性を増してきた。2011年に開催された「じんもんこん 2011」(於 京都、龍谷大学)において、次の発表がされた。

- (1) 「古代木簡解読支援のための画像処理および字体検索の高度化」 末代誠仁・中川正樹他
- (2) 「文書画像から全文検索のオンラインサービス」 寺沢憲吾・川嶋稔夫
- (3) 「古代エジプト神官文字の画像を利用した字形データベースについて」 永井正勝
- (4) 「古辞書のテキストアーカイブズ構築について」 白須裕之

以上の研究は画像データを文字データに変換する試みであり、精度が高ま

れば実用化が一層可能になる。筆者も西洋写本の画像データを文字データに変換する試みをしているが、G. Chaucer の *The Canterbury Tales* の 80 写本や W. Langland の *Piers Plowman* の 70 写本の文字データを作成し、国際的水準のデータベースを構築すれば、これまで未開拓であった文献学の詳細な研究が進むことが期待できる。

西洋初期印刷本の場合、同じ文字に対して複数の木製活字が使用されたため、対応関係が多様になる。さらに、1つの単語を構成する文字フォントの組み合わせが一層多様になり、認識作業の演算が複雑になる。そのため、「文字単位」の認識では、単一の文字のもつ多様性を標準化して、1つの文字に対して1つまたは少数の標準フォントを作成し、作業を行ってきた。

さらに、西洋手稿写本の場合、写字生による筆写により作成されるため、全ての文字フォントが異なる。そのため、情報工学の文字認識では多様なフォント認識が充分になされていない。さらに、「文字単位」で行う認識作業が隣接文字との繋がり、分離文字、隣接文字への被さりなどの問題も生じている。しかし、西洋写本の文字認識を「単語単位」で行えば、これらの問題が解決することが寺沢（公立はこだて未来大学）のプログラムで示された。

初期印刷本は現存する保存状態のよいものが少なく、入手する場合も極めて高価である。さらに、西洋写本は全てが一点限りのものであり、同一のものが流布していることはない。また、グーテンベルクの印刷術が行われるまでの、中世の学問・文化を支えた写本のうち、校訂本の基本写本に選ばれた写本を除く数十万点に及ぶ現存写本がほとんど未解読であり、研究の対象にはなっていない。

本稿では専修大学社会知性開発研究センター/言語・文化研究センターにおける「文字単位」の研究をまとめ、寺沢プログラムによる専修西洋写本 MS 1 『ポリクロニコン』を「単語単位」による研究の成果と将来の可能性について述べる。

And tutoz of your treasure, and take you at nede  
 For husbandye and hit holden together.  
 Than I prayned byz fayre for hym that me made  
 n That dungeon in the dale, that dreedfull is of syght  
 What may it be to meane, madame I you besech.  
 That is the castell of care, who so commeth therein  
 May barne that he borne was, to body or to soule,  
 Therin wommeth a wyght that wronge is I hote  
 Father of falsehead, and founded it him selfe  
 Adam and Eue he egged to yll,  
 Councelled Cayne to kyl hys brother  
 Judas he laped wyth Jewes siluer  
 And stithen on an elder hanged hym after  
 He is letter of loue, and lyeth hem all  
 That truste in hys treasure betrayeth he sonest  
 That had I woder in my wyt, what womā it were  
 That suche wyse wordes of holy wyzte shewed  
 And I asked her on þ heig ht name or the thēce yede  
 What she were wysely, that wyshed me so fayre  
 Holy church I am, quod she, þ oughtest me to know  
 I vnderfeng the fyrst, and the faryn taught  
 Thou broughtest me bozowes, m<sup>y</sup> hiddings to ful-  
 And to loue me lelly, þ whyle the life dureth (fyll  
 Than I courbid on my knees, and cried her of grace  
 And prayed byz pituouly pray for my synnes  
 And also kenne me kyndely, on Chziste to beleue  
 That I myght worke hys wil þ wrought me to  
 Teache me to no treasure, but tell me thys like  
 Howe I may saue my soule, that saynt art yholden.  
 When al treasures are tried, quod she, truth is best  
 I do it on Deus charitas, to deme the sothe  
 It is as dere woorth a drury, as deare God him selfe,  
 go go

図 1. Crowley (1550) William Langland, *Piers Plowman*. I, 56-88, fol. 5v.

© Senshu University Library

## I. 初期印刷本『農夫ピアズ』（Crowley 1550）の「文字単位」による認識

情報工学における従来の文字認識の研究は「文字単位」で行われ、スキャナーと解析ソフトの精密化により、現代の印刷本についての読み取りは高精度に行われるようになった。その原因として、使用される文字フォントが画一的であり、且つ印刷状態が良く、汚れやシミなどがまったくないので、認識作業が充分に行える。

しかし、初期印刷本の場合、同じ文字でも木版活字が同一の文字に対して多数準備され、認識上同一とは認められない場合が多く発生する。さらに、古書の場合、印刷状態にムラがあったり、紙に汚れやシミができる場合もある。筆者は Crowley 編、1550 年版の William Langland, *Piers Plowman* の文字認識を行った。

この本の文字を切り出して比較すると、木版活字がそれぞれ微妙に異なり、文字認識研究の進展を阻害してきた。



図 2. 「農夫ピアズ」の文字 “n”<sup>1</sup>



図 3. 「農夫ピアズ」の文字 “A”<sup>2</sup>

*Piers Plowman* の 1550 年版の校訂本の文字フォントも文字認識上困難な問題を抱えている。文字フォントの変位の幅が見られることと印刷状態が濃い場合と薄い場合がある。図 2 と図 3 は小文字 **n** と大文字 **A** についての調査結果を示した。

1550 年当時の文字フォントは同一の書籍に多種類使用されていたことが分かる。そのため、図 4 に示すように、**t** と **e** の文字を対象に研究し、高い相関関係のものを同じ文字とし、低いものを異なる文字とした。一番目と四番目の **t** は相関関係がそれぞれ 0.917 と 0.870、二番目、三番目、五番目の **e** は相関関係がそれぞれ 0.848、0.832、0.830 であった。しかし、六番目の **gy** は **g** と相関関係が極めて低く 0.089 であり、同一文字とは認識されなかった。(松下他：2010：124)

「文字単位」の分析は、反証可能性を充たす点で科学的な方法論ではあるが、演算量が極めて大きいうえに、単語としての認識精度が低くなることが問題であった。さらに、「誤分析」による正解を導く方法も導入する必要があった。

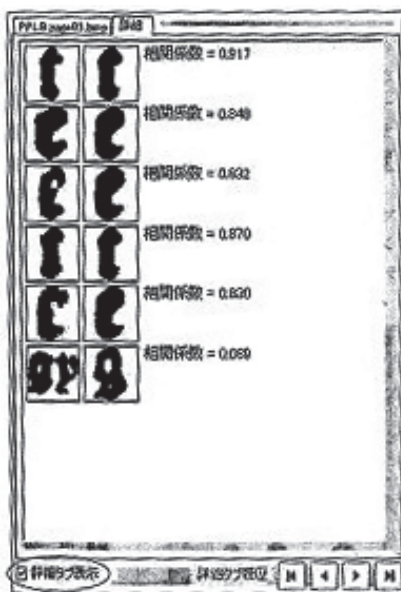


図 4. t と e の相関係数<sup>3</sup>

実験では t と e の輪郭マッチングを行い、図 5 と図 6 に示すようにそれぞれの第 1 主成分から第 5 主成分まで調査した。

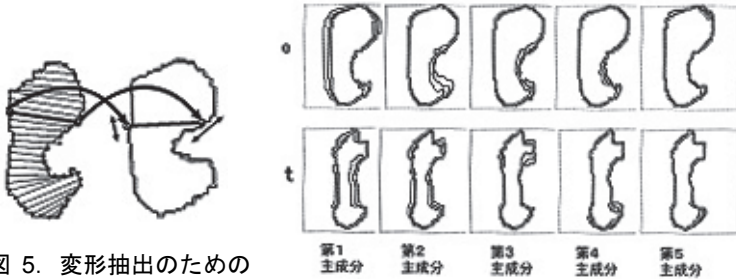


図 5. 変形抽出のための輪郭マッチング<sup>4</sup>

図 6. “e” と “t” の固有変形成分<sup>5</sup>

1550 年版『農夫ピアズ』において、合字 (図 7) と略字 (図 8) が問題となった。fa、ffe、ft など合字の例であり、be、b(a)t、w(i)t(h)、a(n)、e(n)などが略字の例である。



図 7. 合字活字の例<sup>6</sup>



図 8. 省略のための特殊な表現<sup>7</sup>

## II. 初期印刷本『農夫ピアズ』(Crowley 1550) の「単語単位」の認識

初期印刷本『農夫ピアズ』(1550) のための「単語単位認識プログラム」を開発し、文字数が 5 文字程度までの Truth などの語彙を中心にプログラムを修正すれば、次の best (1 行目), siker (2 行目), shal (2 行目) などの文字の接着、分離や被さりの問題を回避できるうえ、wende (2 行目), troweth (3

行目), hem (3 行目) などの文字間が離れた単語も 1 画像として画像特徴量をもつ単語として登録されるので正しく認識される可能性が高い。



図 9. Crowley, ed. (1550) William Langland, *Piers Plowman*

「文字単位」分析では wen\_de, tro\_weth, he\_m のように前半の文字列と後半の文字列が分離して認識される可能性がある。しかし、「単語単位」分析では容認可能な文字列のみを識別できる。1 行目の **Truth** と 3 行目の **Truth** の同定認識を行う。文字画像が同一と認識されない場合 'Truth-1' と 'Truth-2' のように複数の変異形 (variants) を認める。

従来の文字認識研究は「文字単位」で行われ、十分な成果が出なかったが、「単語単位」は文字画像の特徴量を用いて照合するので高い精度を確保できる。

初期印刷本に特徴のある文字表記を単語の画像特徴と捉えて転写を行う。例えば、次の p(a)t, pe, he(m), therew(i)t(h), w(i)t(h), tha(n), nu(m)bre, b(o)u, ma(n) には省略表記がある。



図 10. Crowley, ed. (1550) William Langland, *Piers Plowman* における省略表記

「単語単位」分析では「上付き文字・記号 (t, e, -, t, t, u, -)」の画像特徴量を把握して同定作業を行うので、複雑なプログラムを加えなくても認識が容易である。また、「母音の上の長音記号 (-)」が n を指すか m を指すかが「単語単位」分析では問題にならず、有望な分析であることが判る。従来の「文字単位」分析では分離した p と t、p と e のような単語を「文字や分離・省略単語」として認識することは困難である。

### III. 専修大学西洋写本 MS 1 *Polychronicon* の「単語単位」による認識

同一作品の写本であっても、写本ごとに綴り、形態、語順、統語法が異なる上、内容が加筆や削除、置換、誤記などにより異なっている。さらに、作者自身が写本を作成した例は少なく、そのため、現存する写本と作者の意図がどの程度離れているか問題が生じる。西洋写本の文字認識の問題点として、単語の文字の間に隙間がなく合字の状態であること、また、時代の経過によって文字が薄れていたり、欠ける場合がある。このような問題を取り組んできた研究として、蜂谷大翼他「欠損を含む文字パターンを対象とした文字認識手法の試作」(2005)と寺沢憲吾・長崎健・川嶋稔夫「文字切出しによらない毛筆手書き文字検索のための部分空間法」(2005)などがある。西洋写本の文字認識を対象にする研究には不可欠な視点である。

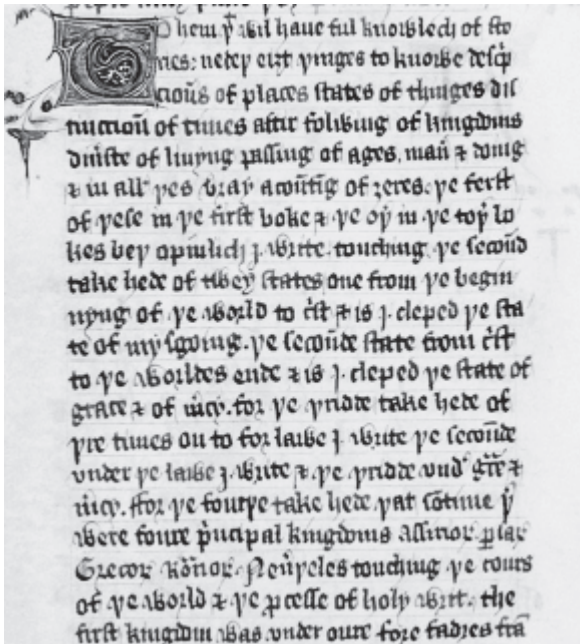


図 11. 専修大学図書館所蔵 MS 1 *Polychronicon*, fol. 16v. (部分)

© 専修大学図書館



本稿は比較的保存状態が良好で、規則的に写字された専修大学図書館写本 MS 1 *Polychronicon* を中心に研究を行った。松下が同写本の画像デジタルを作成し、寺沢が画像処理を加え、「単語単位」認識プログラムを作成して検索を行った。

専修写本 *Polychronicon* の文字フォント・セットは次のとおりである。

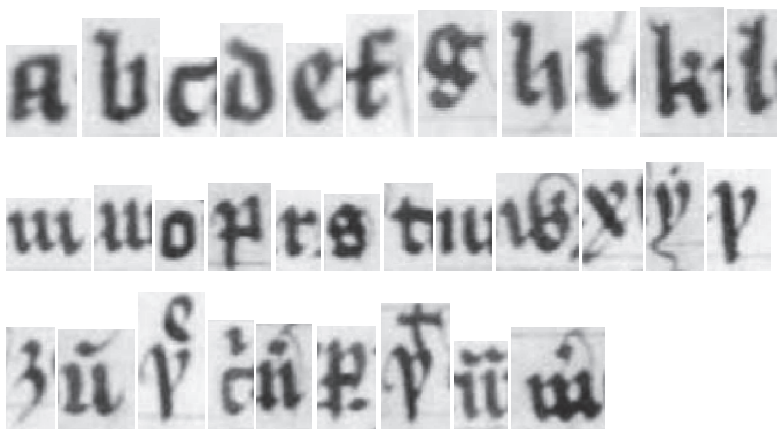


図 12. 文字フォント. Senshu University Library, MS 1. *Polychronicon*.

「単語単位」の分析は「文字単位」分析のもつ、文字接着・文字分離 (i, j, ā) ・被さり (s, f) の問題を回避できるうえ、必ず 1 画像 1 単語として登録されるので正しく認識される。専修写本『ポリクロニコン』の全 214 葉を研究対象とした。現行プログラムでは単語単位の指定が手動で行われる。プログラムを作動すると、写本検索の指示として次のように示される。

- ・画像をドラッグすると範囲選択ができ、選択状態で検索ボタンを押すと検索できます。
- ・検索文字をきれいに囲まなくても、文字のはじめと終わりが適切であれば、選択幅は文字より小さくて構いません。

次の表に示すように、alwey は初出 5 例のうち 3 例が正解で 60%、bope は初出 6 例のうち 6 例が正解で 100%、erpe は初出 20 例のうち正解が 18 例 (90%)、king は 19 例 (95%)、god は 16 例 (80%)、lawe は初出 6 例のうち 6 例が正

polychronicon

検索元画像	1: foI022v.jpg	2: foI033v.jpg	3: foI029r.jpg	4: foI041r.jpg	5: foI033v.jpg					
alwey	alwey	alwey	alwey	alwey	alwey					
検索元画像	1: foI016v.jpg	2: foI049r.jpg	3: foI038r.jpg	4: foI046r.jpg	5: foI050v.jpg	6: foI032r.jpg				
bope	bope	bope	bope	bope	bope	bope				
検索元画像	1: foI036v.jpg	2: foI017v.jpg	3: foI030r.jpg	4: foI019r.jpg	5: foI017r.jpg	6: foI025r.jpg	7: foI018v.jpg			
erpe	erpe	erpe	erpe	erpe	erpe	erpe	erpe			
検索元画像	1: foI020r.jpg	2: foI010v.jpg	3: foI026v.jpg	4: foI004r.jpg	5: foI026v.jpg	6: foI043r.jpg	7: foI001r.jpg			
king	king	king	king	king	king	king	king			
検索元画像	1: foI030r.jpg	2: foI025r.jpg	3: foI022r.jpg	4: foI033r.jpg	5: foI016v.jpg	6: foI044v.jpg	7: foI030v.jpg	8: foI007v.jpg		
god	god	god	god	god	god	god	god			
検索元画像	1: foI045r.jpg	2: foI016v.jpg	3: foI011r.jpg	4: foI024v.jpg	5: foI011r.jpg	6: foI012v.jpg				
lawe	lawe	lawe	lawe	lawe	lawe	lawe				
検索元画像	1: foI023v.jpg	2: foI023v.jpg	3: foI019r.jpg	4: foI028r.jpg	5: foI021v.jpg	6: foI001r.jpg				
lonte	lonte	lonte	lonte	lonte	lonte	lonte				
検索元画像	1: foI021v.jpg	2: foI025v.jpg	3: foI023v.jpg	4: foI010r.jpg	5: foI031r.jpg					
name	name	name	name	name	name					
検索元画像	1: foI032v.jpg	2: foI028r.jpg	3: foI028r.jpg	4: foI032r.jpg	5: foI018v.jpg	6: foI019r.jpg	7: foI020r.jpg	8: foI026v.jpg	9: foI018r.jpg	
out	out	out	out	out	out	out	out	out		
検索元画像	1: foI014r.jpg	2: foI0031v.jpg	3: foI025v.jpg	4: foI0037v.jpg	5: foI0037v.jpg	6: foI047v.jpg	7: foI011v.jpg			
put	put	put	put	put	put	put	put			
検索元画像	1: foI013v.jpg	2: foI036v.jpg	3: foI019v.jpg	4: foI002v.jpg	5: foI032r.jpg	6: foI012v.jpg				
lone	lone	lone	lone	lone	lone	lone				
検索元画像	1: foI015r.jpg	2: foI017r.jpg	3: foI050v.jpg	4: foI0031r.jpg	5: foI015r.jpg	6: foI017v.jpg	7: foI016r.jpg	8: foI051r.jpg	9: foI018r.jpg	10: foI018r.jpg
ye	ye	ye	ye	ye	ye	ye	ye	ye	ye	

図 13. 「単語単位」の認識結果 (標準表記): 専修大学図書館写本 MS 1 Polychronicon

解で 100%、londe は初出 20 例のうち 20 例が正解で 100%、name は初出 5 例のうち 5 例が正解で 100%、out は初出 20 例のうち 13 例が正解で 65%、put は 7 例のうち 6 例が正解で 86%、sone は 6 例のうち 5 例が正解で 83%、pe は 20 例のうち正解が 20 で 100%だった。さらに、of, pat は正解が 20 例で 100%を示している。そのほか、Engelond, bousand, hundred のような 7~8 の文字から成る単語も認識率が高く 100%であった。しかし、william, abrah(a)m, passiou(n)は 10%~30%の認識率に終わっている。また、ラテン語のように赤で記されている libro は 10%程度の認識率であり、書かれた文字の色も今後の課題である。

今後、最高頻度語彙 and, be, to, pat, of, for, I, in や高頻度語彙 a, he, is, it, as, his, wib, hym, hem, ne, be, alle, me, so, no3t, bei, was, by, men, quod, hir, no, my などの機能語を中心に認識精度の調査し問題点を整理する。また、「単語単位」分析では容認可能な文字列のみを識別できる特徴がある。

#### IV. 省略表記を含む単語の認識

「単語単位認識プログラム」は西洋写本に多く含まれる省略表記を認識し、高い認識率を示すことができる利点が認められる。例えば、p(ro)uendre, p(er)sonage, neu(er)e, L. no(m)(i)(n)e, p(re)st, L. sup(er), L. man(us), b(a)t, L. h(ab)eo, L. d(omi)ni, g(r)ace などは「文字単位」分析では認識不可能な場合が多かった。



図 14. 「単語単位」の認識結果（省略表記）

しかし、「単語単位」分析では、単語画像の特徴量を利用して認識を行うので、規則的に書かれた写本を容易に転写することができる。また、「単語単位」分析の果たすべき大きな普遍性はヨーロッパに多数所蔵される西洋写本の転写テキストの作成を通じて内容の実質的な研究への貢献が多いに期待される。

実験では、b(a)t, ob(er), w(i)t(h), he(m), man(er), sp(r)ing, p(ar)adis, dep(ar)ted, breb(er)en について調査した。

polychronicon



図 15. 「単語単位」の認識結果（省略表記）

b(a)t は初出 20 例のうち 20 例認識され、認識率は 100%だった。ob(er)は 20 例で 100%、w(i)t(h)は 20 例で 100%、he(m)は 11 例で 55%、man(er)は 13 例で 65%、sp(r)ing は 16 例で 80%、p(ar)adis は 20 例で 100%、dep(ar)ted は 19 例で 95%、brep(er)en は 7 例で 35%だった。省略表記を含む単語の認識率は認識空間が一般の単語と比べて広く不規則であるため一般に認識率が低下する傾向があるが、実験段階でも高い認識率を示す単語が見られ、今後の展望が開けた。

写本には文字ごとに変異差が大きい上に、文字の接着、被さりが多く、従来「文字単位」分析では演算量が膨大になり、実用性が極めて低かった。本研究では「単語単位」で画像特徴量を把握して分析を行うので、文字接着や分離と被さりの問題が生じることはなく、文字単位の分析に比べて演算量が小さくて済む。

#### V. Polychronicon 写本から The Canterbury Tales 写本群への文字認識

グーテンベルクの印刷術が発明される以前は、羊皮紙または紙に黒インクまたはイカ墨で筆写された。現存する写本だけでも数十万点に及ぶと考えられる。今後「単語単位」認識プログラムを実用化することによって、詳細な研究が可能になるだろう。The *Canterbury Tales* 写本は約 80 点存在するが、筆者は Glasgow, Alnwick, Lichfield 写本を現地に出向いて撮影しカラー・ファクシミリを作成した。Manly-Rickert によるモノクロ・マイクロフィルム (1940) は写本研究に必須の資料であるが、現代では高精度のデジタル画像が必要である。



写真 1. Lichfield Cathedral





写真 2. Alnwick Castle, Northern England



写真 3. The *Canterbury Tales* 写本の撮影風景 Alnwick Castle, Northern England

専修写本 MS 1 *Polychronicon* の「単語単位」の認識は極めて有望と考えられる。しかし、それぞれの写本の文字フォントは同一の *anglicana formata* 体で筆写されていても、写字生が異なれば同一の認識プログラムで作業が進むことは困難であろう。「単語単位」の認識を円滑に進めるために、写本ごとのテンプレートを作成し、「単語画像」を「デジタル綴り」に変換する「単語画像＝デジタル綴り辞書」を作成する。The *Canterbury Tales* であれ、*Polychronicon* であれ、ある写本のテンプレートと別写本のテンプレートの対応関係を示して置換すれば、個別に「単語画像＝デジタル綴り辞書」を作成する負担を回避することができる。

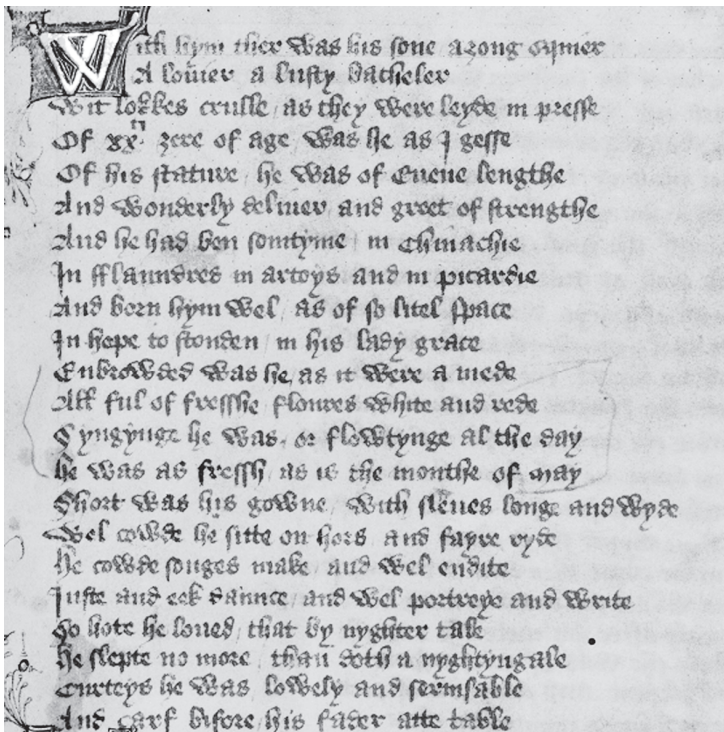


図 16. The *Canterbury Tales*, Lichfield MS 2, fol. 2r (部分)  
© Lichfield Cathedral



Lichfield の *The Canterbury Tales* 写本の文字フォント・セットは次のとおりである。

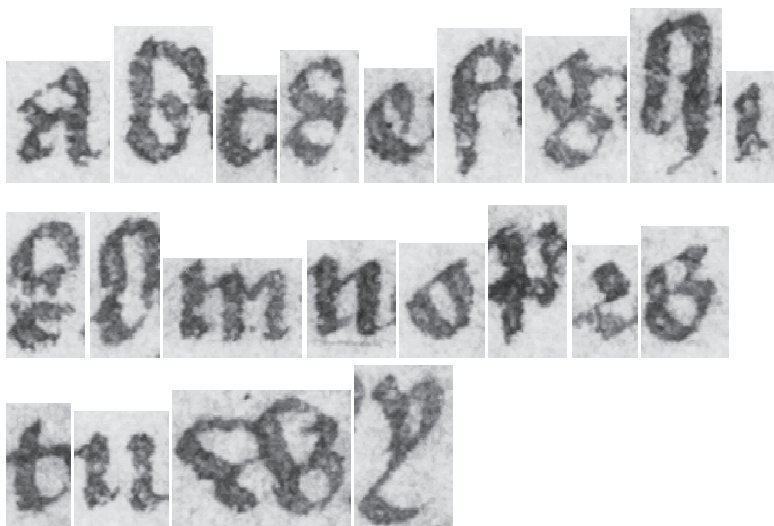


図 17. 文字フォント. *The Canterbury Tales*, Lichfield Cathedral.

## VI. 写本間の文字フォントの変換

西洋写本の「単語単位」認識を促進するために、ある写本の文字フォントを別の写本の文字フォントに変換し、「単語画像」を作成する。一写本の「単語画像」＝「文字綴り」という「辞書」を準備し自動的に変換する。例えば、専修大学図書館の *Polychronicon* 写本の「辞書」から Lichfield Cathedral の *The Canterbury Tales* 写本用の「単語画像」を生成する。

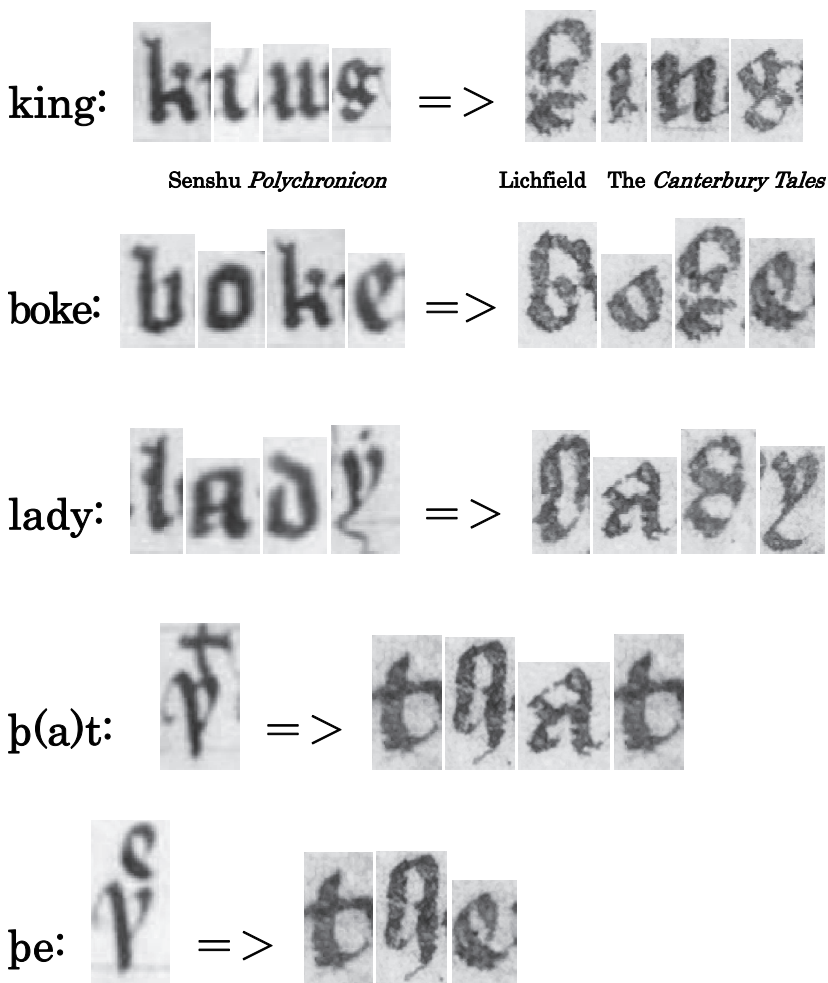


図 18. 写本間の文字フォントの変換

写本間の転写作業は各写本の文字フォントを抜き出してフォント・リストを作成し、自動で文字フォントの変換を行い、「単語画像」を生成する。例えば、専修写本 MS 1 の king, boke, lady は Lichfield 写本の文字フォントに変換される。さらに、前者の p(a)t, pe は後者の that, the として「単語画像」が作

成される。今後、The *Canterbury Tales* 80 写本や *Piers Plowman* 70 写本の転写作業に適用されれば、効率のよい成果が期待される。

## VII. 頻度表を活用した「単語単位」認識プログラム

The *Canterbury Tales* 写本は Geoffrey Chaucer によって作成された、弱強五歩格で書かれた韻文を中心として構成されている。そのため、写本間の相違が少ない作品である。本研究は今後の展望として頻度表を活用した「単語単位」認識プログラムにより、作品全体の単語数に占める認識の度合いを高める。and, that, the, of は頻度が 5,000 を超えている。また、any, time など は頻度が 200 以上である。

and	: 8414	an	: 323
that	: 5596/1	than	: 323/3
the	: 5212/3	thee	: 323/63
of	: 5157	upon	: 318/2
to	: 3825/7	greet	: 317
he	: 3541/128	thus	: 315/20
in	: 3418/8	us	: 313/6
his	: 3025	every	: 290
a	: 2933/1	yet	: 288
for	: 2921	moore	: 282/35
i	: 2766/46	out	: 269
is	: 2366/40	how	: 266/6
as	: 2210/1	o	: 256
this	: 2014/38	certes	: 255
it	: 1668/11	noon	: 254/40
was	: 1570/41	good	: 252/21
hym	: 1261	after	: 245/1
so	: 1220/59	seye	: 243/88
but	: 1215	love	: 241/33
by	: 1136/8	anon	: 238/99
with	: 1102	myn	: 238/2
al	: 1080/37	telle	: 238/50
nat	: 1059/1	seyo	: 237/35
be	: 1007/117	do	: 233/12
or	: 981	kan	: 231/51
she	: 971/79	folk	: 226
ye	: 955/43	nevere	: 223/1
my	: 919	thise	: 221
ne	: 899	many	: 218
man	: 891/106	fro	: 217/20
hir	: 882	any	: 211
they	: 846/1	tyme	: 211/13
no	: 826/1	Crist	: 210/1
been	: 782/6	nyghte	: 210/20
hire	: 778/1	therfore	: 206/7
me	: 748/156	oother	: 204/21
yow	: 706/15	heere	: 197/83

図 19. The *Canterbury Tales* の語彙の頻度表（部分）  
Oizumi ed. *A Complete Concordance to the works of Geoffrey Chaucer*, vol.3.

The *Canterbury Tales* には約 80 写本存在するが、基本的にある単語の頻度はほぼ同一であると考えられる。高頻度の認識を示す単語のうち文字数が 5 以下の単語を選び、優先的に「単語単位」の認識をすれば、効率のよい転写テキストの作成が可能である。

## VIII. まとめ

*Polychronicon* 写本の校訂本は Babington-Lumby (1865-86) 全 9 巻と Waldron (2004) Book VI のみがあり、転写テキストは Seeger (1975) による Huntington 写本 MS. 28561 だけである。本研究は「単語単位」認識プログラムを専修大学写本に適用して十分な精度が得られることを示した。さらに、プログラムの自動化と The *Canterbury Tales* 写本への応用を「単語画像」＝「スペリング」辞書により効率のよい複数の写本転写テキスト作成の可能性も示唆した。

## 注

松下知紀 専修大学文学部教授

寺沢憲吾 公立はこだて未来大学システム情報科学部准教授

1. 山口他 「中世英文学資料のデジタルテキスト化の試み」 227 頁
2. 同書、同頁
3. 松下他 「西洋初期印刷本の文字認識プログラムと手稿写本への応用」『ことばの普遍と変容』Vol. 5, 124 頁
4. 山口他 「中世英文学資料のデジタルテキスト化の試み」 235 頁
5. 同書、236 頁
6. 同書、228 頁
7. 同書、同頁

## 参考文献

- Beal, Jane. *John Trevisa and the English Polychronicon*. Temple, Arizona: Arizona Center for Medieval and Renaissance Studies & Brepols, 2012.
- Edwards, A. S. G. “John Trevisa.” In *Middle English Prose: A Critical Guide to Major Authors and Genres*, ed. idem, New Brunswick, N.J.: Rutgers University Press, 133-46, 1984.
- 蜂谷大翼他 「欠損を含む文字パターンを対象とした文字認識手法の試作」『信学技報、PRMU2004-260』2005.
- 樫村雅章 『貴重書デジタルアーカイブの実践技法』慶応義塾大学出版会、2010年.
- Manly, John M. and Edith Rickert eds. *The Text of the Canterbury Tales on the Basis of All known Manuscripts*. University of Chicago Press, 1940.
- Matsushita, Tomonori ed. *Polychronicon*. (Senshu University Manuscripts in Facsimile, v. 1), Senshu University Press, 2010.
- 松下知紀、Skander Hannachi, 飯田周作、和氣美紀夫「西洋初期印刷本の文字認識プログラムと手稿写本への応用」『ことばの普遍と変容』専修大学社会知性開発研究センター/言語・文化研究センター、Vol. 5 (2010) : 115-28.
- Seeger, Richard Arthur. ed. *The English Polychronicon: A Text of John Trevisa's Translation of Higden's Polychronicon, Based on Huntington MS. 28561*. Ann Arbor: Michigan: UMI, 1975.
- Taylor, John. *The Universal Chronicle of Ranulf Higden*. Oxford at Clarendon Press, 1966.
- 寺沢憲吾・長崎健・川嶋稔夫 「文字切出しによらない毛筆手書き文字検索のための部分空間法」『信学技報、PRMU2004-272』2005.
- 寺沢憲吾・川嶋稔夫 「文書画像からの全文検索のオンラインサービス」『デジタル・アーカイブ』再考』じんもんこん 2011、329-34.
- Trevisa, John, trans. *Polychronicon Ranulphi Higden Monachi Cestrensis*

*Together with the English Translations of John Trevisa and of an Unknown Writer of the Fifteenth Century*, ed. Churchill Babington and Rawson Lumby. 9 vols. Rolls Series. London: Her Majesty's Stationery Office, 1865-1886.

山口晃典他 「文字認識による中世英文学資料のデジタル化の試み」『ことばの普遍と変容』専修大学社会知性開発研究センター/言語・文化研究センター、Vol. 4 (2009) : 87-101.

山口晃典他 「中世英文学資料のデジタルテキスト化の試み」松下知紀・篠田勝英編『Anglo-Saxon 語の継承と変容 II: 中世英文学』専修大学出版局、2010、221-43.

Waldron, Ronald. "The Manuscripts of Trevisa's Translation of the *Polychronicon*: Towards a New Edition." *Modern Language Quarterly* 51 (1990): 281-331.

Waldron, Ronald ed. *John Trevisa's Translation of the Polychronicon of Ranulph Higden, Book VI: an Edition based on British Library MS Cotton Tiberius D. VII*. Heidelberg: Winter, 2004.