

中世英語写本文字認識の問題点について

松下 知紀

専修大学文学部

連絡先：〒214-8580 神奈川県川崎市多摩区東三田2-1-1 社会知性開発研究センター

Tel: 044-911-1349

千葉 淳一

専修大学文学研究科

Fax: 044-911-1349

植竹 朋文

専修大学経営学部

iruk@nifty.com

専修大学『Anglo-Saxon語の継承と変容』プロジェクトが2005年から研究している、中世英語写本文字認識プログラムの開発に関して、現時点までに明らかになった問題点と、その対策について述べる。中世英語写本の文字認識で重要なことは、第一に、文字の研究やテンプレート収集のコストと、文字認識プログラムによる自動認識で節約されるコストのバランスである。文字研究やプログラム開発に時間がかかり過ぎては、自動化の目的が達成されない。第二に、判定が難しい文字を決定するために、当時の英語の知識も不可欠である。これらの問題点を述べながら、文字認識プログラムの現実的な用途について述べる。

Tomonori Matsushita

School of Literature,

Senshu University

Junichi Chiba

School of Humanities,

Senshu University

Tomofumi Uetake

School of Business Administration,

Senshu University

Problems on Developing Character Recognition Program for Medieval English Manuscripts

Since 2005 when Senshu University launched a project "The Development of Anglo-Saxon Language and Linguistic Universals", we have been attempting to devise a software which optically recognizes characters in medieval English manuscripts. In the course of two years, it can be shown that there are two major technical problems. One problem is unpredictability of time needed for preparing effective templates from a manuscript. In order to take good advantage of the automation, the time must be shorter than that of transcribing the target manuscript manually. The other problem is the need of ability for deciphering ambiguous characters, based on the knowledge of medieval manuscripts. Considering these problems, we propose to make realistic employment of character recognition.

1. まえがき

平成17年度に文部科学省私立大学学術高度化推進事業オープン・リサーチ・センター整備事業に選定された専修大学『Anglo-Saxon語の継承と変容』プロジェクトは、その主要な目的の1つとして、写本文字を自動認識するソフトウェア「写本文字認識プログラム」の開発を試みてきた[1][2]。本プロジェクトが発足する前年(平成16年)に文部科学省に提出した構想調書に記されている開発の目的を以下に記す。

日本国内で初めて貴重な手稿写本(主要備品)を購入し、中世英文学頭韻詩と脚韻詩の研究

体制を確立し、資料研究のレベルを国際的レベルに向上させる。また、写本文字を自動認識する、世界初の「写本文字認識プログラム」をアドバック社と産学協同により開発して、写本を高速に転写するシステムを確立を目指す(ただし、初期段階においてインキュナビラ等の初期印刷本の文字認識から開始する)こうして、ヨーロッパ諸国の国立図書館・大学図書館・古文書館に所蔵されている多くの写本に本プログラムを適用すれば、転写テキストを作成することが可能になり、これまで組織的な研究から免れていた膨大な資料が一斉に研究対象となるので、国際的に大きな貢献を果たすことになりうる。(構想調書)

この構想を本プロジェクトの目標とし、2005年度から文字認識プログラムの開発を行なっている。

中世の写本を対象とした文字認識プロジェクトの多くは、文字を書き写した写字生を特定するために、統計的な手法で文字特徴を抽出し、比較するという研究を行なっている。それに対し本プロジェクトは、文字認識技術により写本画像からテキストを起こすことを趣旨としている。文字認識プログラムで作成する転写テキストは、作品研究や言語研究に利用する計画であり、それゆえ高い認識率が求められる。

文字認識の分類としては、オフラインの、手書きの続き欧文文字(筆記体に類似)の認識である。オフライン手書き文字認識は、最も難しい部類のものである。また、続き文字であるため、前後の文字と分離しにくい。さらに、一般に中世の写本は、綴り字の確立に貢献した印刷本が普及する以前の資料であるため、綴り字が一定していない問題がある。現代の多くの文字認識ソフトウェアは、実用的な認識率を達成するために、辞書的な知識により補正しているが、綴り字の安定していない資料では、単純な辞書による補正は行なうことができない。写字生による誤りや、意図を持った綴り字の変更も、作品研究で重要となる場合もあり、辞書的な知識による自動補正が適切でないことが多い。

本プロジェクトが中世英文学を軸に研究を進めていることから、この時代に転写された比較的有名な写本である「エレズミア写本」[3]および、専修大学所蔵「ポリクロニコン」[4]を認識対象として選択した。前者は、ほかの写本と比べて書体が安定しているので、文字認識で問題点が少なそうであったこと、また、広く研究され、参考となる転写テキストがすでに存在するため、認識の判断がしやすいことから、選択した。転写テキストの存在するものを文字認識する必要性は無いが、認識結果を比較できることから、テスト用として選択した。

後者は、本プロジェクトが2005年にノルウェー・オスロのスコーエン・コレクションから購入した写本である。この写本そのものの転写テキストはないため、文字認識プログラムの出力結果が、このプログラムの最初の成果となる見込みである。文字認識結果の正確性の判断については、実際に任意のページを人間が見て判定する他

に、同作品の別の写本に基づいた既存のエディションを参考にして、大きなずれがないかどうか検証を行なうことができる。

本プロジェクトで研究してきた認識アルゴリズムは、ネオコグニトロンなど神経回路シミュレーション、統計論的枠組みのSVM(サポートベクターマシン)、パターンマッチングといった複数の手法である。プログラムはアドバック社との共同開発し、認識アルゴリズムのテストを重ねてきた。これらの比較と考察については、[1]を参照されたい。2006年度は主にパターンマッチングで実験を行なったが、アルゴリズム固有の特徴や、研究成果等については[2]に譲ることとする。

本稿では、これら2つの写本の文字特徴について、開発過程で明らかになったテンプレート作成時の問題点と文字認識の問題点、および今後の対策について述べたい。

2 テンプレート作成時の問題点

文字認識プログラムが参照する文字のテンプレート収集過程で生じた以下の問題点について述べる。

- ・行・文字の切り出しについて
- ・語中形・語尾形の扱いについて
- ・省略形の扱いについて
- ・合字の扱いについて
- ・文字分類の予測不能性について
- ・文字画像の解像度について
- ・テンプレート作成時の問題について

なお、表中の単語は特に触れない限り、エレズミア写本 fol. 11rからの引用であり、Lはこのページの行番号を示す。

2.1 行・文字の切り出し

写本によって文字の形状が異なることから、テンプレートの登録は、手動で行なわなければならない。なお、ここでは写本文字の特徴について述べるが、テンプレートの具体的な画像ファイルや、支援ソフトについての議論は行なわない。

テンプレート収集は、任意の数ページに渡り、大文字・小文字それぞれ25文字前後のアルファベットの各文字を収集することから始まる。小文字に関しては、エレズミア写本の例では1~2ページのサンプル収集で概ねカバーできる。しかし大

文字や、出現確率の低い文字(zなど)を拾うためには、十ページ以上見る必要が生じることもある。また、これはすべての文字が同じ形状と仮定し、1文字につき1つのサンプルを拾う場合のことであり、実際には以下に述べるように、注意すべき点がある。

小文字oは、表1のように概ね変化が少なく、四隅が丸いため、判読が容易である。また、隣接の文字との接触が比較的少なく、切り出しがしやすい。

一方、小文字wは写本中で最も装飾性が高い文字である。wには上の2本の腕があるが、表2 L7 whan、L2 with、L5 sorweを比較すると、太さ、方向、高さが文字ごとに異なっていることがわかる。大きさも変化しやすく、テンプレートを決定しにくい文字である。また、L5 sorweのように、上の行の文字と接触する場合がある。

アセンダやディセンダのある文字は、一部が上方向や下方向に長く伸びていることが多い。小文字hがその例であり、表2 L2 thatのように、上下とも他の文字のベースラインとアセンダラインから大きくはみ出す。L7 whanは、下の行のhisに触れてはいないものの、下の行のアセンダラインを侵している。写本の最上行のhやkは、L1 makethのように柱部分が非常に長くなる。他に、下に延びて次の行の文字と接触する例がL4 wrecchesに見られる。

hの形状についてだが、hの左の部分は、表3 L6 gentil、L10 whilomのlに似ているため、判別しにくい。

L4	on	on
L4	sou	som
L5	oupe	oure
L5	doun	doun
L9	so	so

表1 oの例(変化の少ない文字)

L2	that	that
(L6) L7 (L8)	whan whan his he	(doun f...) whan (his he...)
L1 (L2)	maketh with th	maketh (with th...)
L4 (L5)	wrecches sorwe sy	wrecches (sorwe sy...)

表2 hの例

L6	gentil	gentil
L10	whilom	whilom

表3 hとlの区別

2.2 語中形・語尾形

語中形、語尾形の異なる文字もある。表4 L13 knyght、L40 kyngのkは、下の撥ねの部分に微かな差異がある。

小文字sは、表5 L4 vsの語尾、L5 synkenの語頭のような通常のsに対して、L2 respit、L3 pitouslyの語中のsのようなlong-sとがある。long-sは、表6 L3 grufのf、L3 fillenのfやl、L8 woldeのlなどに似ている。表5 L3 pitouslyのように、long-sとlが連続する例もあり、形状がほとんど同じで、位置が異なることがわかる。また、L10 estaatのsは、次に来るtの横棒がかかり、fと誤解しやすい。

小文字rにも、表7 L2 word、L3 cridenのように、2つの形が見られるが、単語内の位置による書き分けではないようである。また、long-sとlong-rは、上下方向の位置に違いはあるものの、形状自体は類似している。

表4 L13 knyghtのhのように、ghの綴りでのみキャップライン付近に長い横棒が書かれる文字もある。また、本来はディセンダの無い文字に、装飾的なディセンダが付けられることがある。表8のmynotaurのmを見ると、表2 L1 maketh、表3

L10 whilomには見られないディセンダが付いている。

L13		knyght
L40		kyng
L8		breke

表4 語頭形・語中形のある文字の例

L4		vs
L5		synken
L2		respit
L3		pitously
L1		despit
L10		estaat

表5 通常のsと語中形のs

L3		gruf
L3		fillen
L8		wolde

表6 long-sと類似した文字の例

L2		word
L3		criden

表7 short-rとlong-r

L34		mynotaur
-----	--	----------

表8 mの例

2.3 省略形

古英語・中英語などの写本には、数多くの省略形が用いられている。アルファベットのテンプレートと平行して、こうした省略形の書式も収集しなければならない。

エズミア写本に頻繁に見られるのは、表9 L2 wt (withの省略形)、L8 bt (batの省略形)に見られ

るような、右肩付きのtである。これらは頻度が高く、字形の変化が少ないため、テンプレートの扱いによく似ている。

この他に、文字の上にcやsのような曲線を加える省略形もある。L4 m(er)cyは、mの上に、erの省略形を表すsのような曲線を追加することで、merを表し、その後cyが続いている。

鼻音n, mの省略形は、L4 wo(m)menのように、前の母音の上に横棒を引くことで表す例が見られる。

L2		wt
L8		bt
L4		m(er)cy
L4		wo(m)men
L5		in

表9 省略形の例

2.4 合字

複数の文字を詰めて1文字のように書いた合字は、エズミア写本には見られないが、ポリクロニコンには数多く見受けられる。第1巻の始まりであるfol.15 left columnからいくつか合字を表10に例示する。

L6		kindelich
L41		dedes
L5		aboute
L41		poetis

表10 合字の例

L6 kindelichやL41 dedesにあるようなdeの合字は、10行に2~3例ずつ見られ、とても高い頻度である。それより低頻度だが、L5 abouteに見られるようなboの合字、L41 poetisに見られるようなpoの合字なども散見される。

合字のうち、頻出するものは通常のアルファベットのようテンプレート的に扱うことで対処

するのが単純である。認識時に、それに似た形状の文字を該当の2文字に展開するようなルールとすれば良いからである。しかし、テンプレート登録を逃がれた合字の対処は難しい。どのような組合せの合字があるのかを予測し、適切な方法で分解しなければ、認識が不可能だからである。

2.5 文字分類の予測不能性について

h, k, s, rなど揺らぎの多い文字についてテンプレートを収集する際は、単語内の出現位置や、形状の違いに留意して、選出しなければならない。特に、hの異形など、特定の環境でしか起きないものは見落としがちである。また、テンプレート登録の問題だけでなく、プログラム側で対策を取る必要のある部分もある。例えば、hの異形に見られる横棒は、次の文字の上にかかるほど長いため、ノイズとならないよう適切な切り出しを行わなければ、次の文字の一部として認識され、認識率を低下させる要因となるかもしれない。

しかし、手書きであり、単語内の位置やその他の条件により形状が異なる文字が多いため、どの程度の違いを異形と判断するかはテンプレート作成者に委ねるしかないのが実情である。また、写本の後半までも一貫した書体である保証がないため、どの範囲で何文字のサンプルを取ると過不足ないのかを事前に予測できないという問題もある。

このように、どこまで詳細に文字分類の研究をすべきか、事前に予測を立てにくいのは問題である。もちろん、文字の揺らぎ、異形、省略形などに関する調査は、できる限り詳細に行なわれることが望ましい。しかし、それにかかる時間的なコストも同時に考慮しなければならない。文字認識で節約される時間的なコストは、ソフトウェア開発にかかるコストも当然のことながら、字形研究にかけるコストを十分に低くしなければ、本プロジェクトの構想調書に掲げる目的を達成することはできない。ちなみに、筆者(松下)はエズミア写本のテンプレート作成に10時間程度を要したが、一部の文字(tやyなどの変形)を捨てるには足りなかった。同じ時間で、[5]などの転写テキストを参考にしながら転写テキストを作成するほうが、コストが安い場合もあり得る。

しかし、合字についても問題が発生する。ここでもまた、テンプレート登録にかかるコストが問題となる。写字生によって、または書き写された

時期によって、異なる字形の合字が存在する可能性が高く、出現頻度が極端に低い合字の存在も否定できない。どのページのどの文字をテンプレート登録すれば、十分なサンプルに達するのを確認する手段がない。

この問題を解決するためには、合字と考えられる画像を、適切に2字に分解する方法の開発が重要であると考えられる。ただし、分割位置の正確な予測に加え、重なり合う線分の補完という工程が必要であると思われ、それらを誤認識のないよう実装する方法を考えなければならない。また、ひとつの写本について対策ができたとしても、それをそのまま別の写本で使えるという保証はない。

また、何と何の合字か判断が難しい例も存在する。例えば、bとoの合字か、lとoの合字かを判断できない例も実際にある。このような場合は、2で述べたnとuの区別のように、文脈や共起確率を参考にすることも視野に入れて初めて、正確な判断が期待できると考えられる。

2.6 解像度について

2005年度からアドバック社と共同開発している認識ソフトウェアでは、認識速度の都合上、テンプレートサイズの上限を19×19ピクセルとしていた。しかし、この仕様では、2.1で触れたwのような大きなサイズの文字や、ページ上部のhやkのように極端に長いアセンダを持つ文字は、認識が不可能であることが分かった。動作速度や、単純な字形のみの認識率の見積りをもとに、プログラムの仕様を決めたからであるが、本来は写本文字の研究を詳細に行ない、目的も明らかにしてから、プログラムの仕様などを具体化していくべきである。

2.7 テンプレート作成者について

テンプレート作成者は、上で述べたような中世英語写本のアルファベットの字形や、省略形の用法に、十分に慣れていることが求められる。アルファベットの形状だけで判読ができない単語も多数あり、その場合は語彙・文法・形態論の知識をもとに判断してアルファベットを特定できるからである。同時に、プログラムの仕様を良く理解し、適切なテンプレートを提案しなくてはならない。特に、転写テキストや、同作品の別エディションのテキストが無い、いわゆる「未知」の写

本の場合(これこそが本プログラム開発の対象であるが)、中世の写本文字や文学作品によく通じている必要がある。頼りになるテキスト無しでも、適切なテンプレートの提案と、必要に応じてプログラム修正の提案をしなければならないからである。また、認識結果の正確性の判定も、こうした中世英語写本の専門的な知識に委ねられる。

3 文字認識の問題点

この章では、文字を認識する過程において生じた以下の問題について述べる。

- ・形態による判定について
- ・辞書による判定について
- ・語頭・語尾の形態による判定
- ・省略形と復元形の問題

この章では、[7]でオンライン公開されているエレミア写本の転写テキストを、辞書として参考にした。文中で述べている回数は、テキストに登場する該当単語の登場回数であり、大文字・小文字の別がある。

3.1 単語の形態による判定について

表 2 L7 whan の n に注目すると、whau という綴りにも見える。なぜここで whan が選択されるかという点、英語の語尾の形態に注目して、-au という綴りより、-an という綴りのほうが遥かに妥当であろうと判断できるからである。

表 11 L1 But や L3 gruf は、母音を欠く Bnt や Grnf は不適切であると判断できるし、L2 And は、出現確率から考えて Aud よりも普通であろうという判断基準を立てることができる。

このように、一部の単語は、辞書による判定以前に、英語らしい綴りかどうかによって、ある程度の判断が可能である。

L1		houudes
L1		But
L2		And
L3		gruf
L4		m(er)cy
L6		doun
L9		pitous
L11		armes
L15		tiraunt

表 11 u と n が紛らわしい単語の例

3.2 辞書による判定について

表 11 L9 pitous についても、テンプレートを参考にするだけでは、u と n は形状がほとんど同一であるため、pitous にも pitons にも見ることができ。実際、Wright は中英語の書体では、文脈や語源の知識を援用しない限り n と u の区別ができないと述べている [6]。辞書を見ると、実際は、pitous という単語は存在し、pitons は存在しないため、ここでは pitous に決定することができる。しかし、高頻度の単語でないため、当時の単語を辞書的に知らなければ、どちらか一方を安易に棄却する判断ができない。

試しに、L11 armes の m について、類似の形状の綴りの単語を検索した結果が表 12 である。正しい armes 以外は、出現回数が極端に少ない出現回数を基準とすると、かなり正確に armes と決定することができる。しかし、この方法で決定できるのは、同時代、類似作品の辞書が利用可能な場合で、かつ、類似の綴りの単語の頻度と大きな差がある場合である。

実際に、辞書を参考にしても判断の難しい単語がある。小文字 c と o の類似性に注目して、語頭に生じる co と oo を比較した結果、oo は 193 回、co は 2004 回登場した。co のほうが頻度が高いが、oo も無視できない頻度である。それぞれ最多の単語は、cold(29 回)、oold(7 回)であった。両者とも形容詞であるため、そこに入るべき品詞

が予測できたとしても、決定できない。そうなる
と、作品の文脈を追うなど、高度な文脈情報に頼
らなければ、判別が難しいと考えられる。

上記2語は、他に old(73回)、olde(98回)、
colde(5回)の綴りも見られ、都合5種類の綴りが
あった。このように、頻出する単語においても、
綴りが安定していないことが分かる。写本内の単
語の綴りは、一般に安定していないため、例え辞
書が提供されたとしても、類似の綴りを許容する
アルゴリズムでなければ、適切な判定は難しい。

arines	0
ariues	0
arnes	0
arn	3 (be 動詞 3 人称複数形)
aruies	0
arnies	0
arnes	91
arme	1 (be 動詞 命令形)

表 12 arnes に似た綴りの出現回数

3.3 語頭・語尾の形態による判定

表 11 L15 tiraunt について、後半の un の判読
は、単独では難しく、周囲の文字の連鎖、コンテ
キストを勘案して u か n かを決定しなければなら
ない。L9 pitous は、pitons という綴りも有り得
なくはないが、on で終わる名詞が少ないという
ことも、判断基準となるかもしれない。フランス
語借用語では他にも ons で終わる単語が多く、
65 例が観察できた。

類似の綴りとは別の例であるが、この写本では
L1 Houndes に見られるような es が名詞の複数
形語尾として主に使われている。実際、es は
1045 語、s は 2351 語、es の語尾から e を削除
して検索し当たったのは 111 語であった。しかし、
同じイングランドの同年代の写本でも、北方と南
方で綴りが異なる場合があり、一定の形の複数形
語尾を、全部の地域に当てはめることはできない。
結局、写本の単語の書かれ方について詳細な研究
や、その写本についての辞書がなければ、単語の
判定基準が決められない、というジレンマがある。

このように、エレズミア写本を読み解く過程か
らわかるように、n と u など形状が類似した文字
の判定は、字形以外の知識が不可欠であるので、
光学文字認識の手法のみでは正確な出力結果を得
られないと考えられる。

3.4 省略形と復元形の問題

中世の写本には、2.3で紹介したような、肩付
き文字による省略形や、横棒による省略形が多く
用いられている。また、画像は例示していない
が、preなどの接頭辞の省略形として、pのディ
センダに=を付したような文字が使われることも
ある。これらの正確な認識は、対応が非常に困難な
問題である。

例えば、i (iの上に横線)は、この後に鼻子音(n
またはm)が続くことを意味している。当然、iだ
けでなく、他の母音字に付される可能性もある。
どの文字に付いても、ノイズとして見落すことな
く、認識しなければならない。

また、こうした省略記号を、どのように展開し
て表示するかも、議論の余地がある。iが単独で
用いられる場合は、前置詞のin以外に該当する単
語が存在しないため、容易にinと決定することが
できる。しかし、iの後にpなどの破裂音が来る場
合は判断が難しい。例を挙げると、中英語の時代
はimpossibleとimpossibleの2つの語形が共存す
るが、その写字生が省略せずに用いるのがin-な
のかim-なのかを調査し、判断するのが通例であ
る。しかしながら、Benskin and Laing [8]が指
摘しているように、写字生はテキストを忠実に写
す訳でもなく、また、自分自身の言語に合わせて
書き写しているわけでもないので、iとpの間の鼻
子音としてnを選択するべきかmを選択するべき
かは、幅広い写本研究の背景知識が必要である。

他の例では、vertuという単語が、vtuという綴
りで、vの上に装飾的な曲線を加えることでその
付近の文字を省略したことを示している場合があ
る。単語や文脈の知識を用いて判断すると、確か
にvに後続するerが省略されていることがわか
る。転写ではvertuと復元した(expanded)形で示
すのが慣例であるため、それに従えば、vertuと
展開しなければならない。

復元形は光学文字認識とは別の問題として考
え、独自の省略記号を定めて、i-possibleやv-tu
などの表記を定めるとする。しかし、それでも大
きな課題に直面する。上に例示したような省略記
号は、ラテン語の省略記号から派生したものであ
るが、現代英語のアポストロフィーのように補助
記号として確立していたわけではない。標準語さ
え確立していない中英語の時代に、省略記号の使
用は写字生によって揺れが観察されることは想像
に難くない。省略記号を見落とさないためには、

微細な特徴の洗い出しとテンプレート化、および処理方法のソフトウェアへの実装に大変な労力がかかることが予想される。

4. まとめ

写本の文字認識は、2で述べたようにテンプレート収集の問題や、3で述べたような認識時の問題点など、解決すべき課題は多い。

字形の類似性についての対策は、写本文字の詳細な研究を行なうのが最も有効であると考えられる。そのために、gameraプロジェクトの文字認識補助アプリケーション[9]など、文字の分類をある程度自動化するツールの使用は有効である。しかし、文字の研究ができるほど写本文字に慣れた研究者が、本当に文字認識プログラムを必要としているのか、という疑問も起きる。プログラム開発のコストを無視しても、テンプレートを検討するために、何十、何百時間も費やすより、手動でテキストを起こしたほうが早い場合もあるからである。

認識プログラム側での文字判定においても、中世写本の英語に見られる数多くの例外に、くまなく対処することは、あまり現実的ではない。判定の難しい文字の多くは、作品の内容や時代背景をも含んだ知識によって、総合的に判断しなければ、決定できないからである。よって、コンピュータに全面的に依存した認識プログラムを求めるのは現実的でなく、辞書や語尾などのコーパスを参考にしながら候補を提示するといった限定的な動作により、研究者の負担を軽減するための補助ツールとしてのプログラムを提案したほうが実用的であるかもしれない。最終的には、専門家の解釈に頼らざるを得ないからである。誰が何のために使うツールにするか検討し、その人の用途に即したプログラムに仕上げることが有効であるように思える。

本論では、このようにして、中世英語写本の文字認識について、研究の過程で挙げられた問題点を列挙した。困難極まりない分野であるものの、文字研究にかかるコストや、研究者からの需要も考慮しつつ、具体的な解決策を模索し、プログラムの改良を進めていく予定である。

参考文献

- [1] 千葉 他: 写本資料のデジタルテキスト化に向けて, 人文科学とコンピュータシンポジウム論文集, pp.295-301, 2006.
- [2] 千葉 他: 写本文字認識エンジンの性能評価, Anglo-Saxon語の継承と変容, 叢書2, pp.107-117, 2007.
- [3] San Marino, Huntington Library MS El. 26.C.9
- [4] Senshu University Library Manuscript of the Polychronicon (Formerly Oslo/London, Schøyen Collection, MS 194)
- [5] Frederick J. Furnivall: Chaucer's Canterbury Tales in Parallel Columns, 1868-79.
- [6] Wright, C. E.: English Vernacular Hands from the Twelfth to the Fifteenth Century, Oxford, 1960.
- [7] http://en.wikisource.org/wiki/The_Canterbury_Tales
- [8] Benskin, M. and M. Laing: Translations and Mischsprachen in Middle English manuscripts, Michael Benskin, M. L. Samuels (eds.), pp.55-106, 1981.
- [9] <http://ldp.library.jhu.edu/projects/gamera/>

謝辞

専修大学社会知性開発研究センター任期制助手市川誠氏には、本論文の執筆の際に中世写本についての貴重な助言を頂いた。ここに記して謝意を申し上げる。