

## Guidelines for Transcription of the Manuscripts of The Wife of Bath's Prologue

### Introduction

*Peter Robinson and Elizabeth Solopova*

This account of the principles we have established so far in our transcription of the manuscripts of The Wife of Bath's Prologue is not intended as a final statement of transcription policy even for this part of the *Canterbury Tales*. Rather, it is a discussion document, partly that we may explain to ourselves and to others what we are doing, and partly that the act of explanation may lead to debate about and refinement of our transcription of the manuscripts.<sup>1</sup>

In the course of our work we have come to realize that no transcription of these manuscripts into computer-readable form can ever be considered "final" or "definitive." Transcription for the computer is a fundamentally interpretative activity, composed of a series of acts of translation from one system of signs (that of the manuscript) to another (that of the computer). Accordingly, our transcripts are best judged on how useful they will be for others, rather than as an attempt to achieve a definitive transcription of these manuscripts. Will the distinctions we make in these transcripts and the information we record provide a base for work by other scholars? How might our transcripts be improved, to meet the needs of scholars now and to come? At the same time, we ask scholars to consider that decisions which may seem somewhat arbitrary might have a long history of argument and counter-argument behind them.

These guidelines are based on our experience of transcription of the fifty-eight surviving manuscripts and pre-1500 printed editions of The Wife of Bath's Prologue.<sup>2</sup> The first transcription of these was done partly by us, partly by other transcribers.<sup>3</sup> There were many inconsistencies from manuscript to manuscript, and indeed within manuscripts, in these first transcripts. We realized that consistency would only be possible if we established guidelines, to be applied to all new manuscripts transcribed thereafter and in the three checks to be made of each transcript. In the course of a first check of these transcripts, carried out entirely by the authors, we set ourselves the task of developing guidelines which could be so applied. This document is the first statement of these guidelines. We expect that the revised guidelines which will issue from consideration of this document will serve as a base for completion of the transcription of all the witnesses of The Wife of Bath's Prologue, and for the greater task of transcription of all the text in all the manuscripts and pre-1500 printed editions of the *Canterbury Tales*.<sup>4</sup>

These guidelines are not proposed as any sort of standard system for transcription of medieval English manuscripts. Our task is the transcription of manuscripts of the *Canterbury Tales* and these guidelines have been devised for that end. Thus, we pay particular attention to transcription of characters at the ends of words, because of the bearing this may have on final -e and hence on Chaucer's metre. Transcription of texts in non-syllabic metre or prose texts, where this is not of such importance, may be based on different principles. Thus, these guidelines may need modification when we come to transcribe the prose portions of the *Canterbury Tales*. For the sake of consistency within this Project, this modification should be slight and confined only to definition of new characters to cope with a possibly different range of abbreviation signs to those found in the manuscripts so far transcribed.

### 1. The theory of this transcription

#### 1.1 The interpretative character of transcription of manuscripts for the computer

It is useful here to review briefly the ways in which transcription of manuscripts and other primary textual sources into computer-readable form differs from the making of other electronic texts. Michael Neuman (1991, 368) has identified three "waves" in the history of the making of machine-readable texts. The first was the conversion of copyrighted text for private use; the second the conversion of public texts for public use; the third the making of electronic editions. All three presume that, for the purposes of electronic manipulation, the electronic text and the printed text are equivalent and interchangeable forms of the one text. This one text might be as well read in its electronic version as in its printed manifestation. One could read Chaucer in many different editions, printed or electronic. The choice between reading this edition or that, printed or electronic, might be made on grounds of familiarity and convenience rather than on the intrinsic value of this or that realization of the text.

This fourth wave—transcription of primary texts into computer-readable form—differs from the first three waves. It differs in that while there might be many different realizations of the one secondary or tertiary text, as different editions printed or electronic, a primary text exists in one and only one form. There are many bibles, but only one Codex Sinaiticus; many Chaucers but only one Hengwrt manuscript. Certain printed texts (the first folio of Shakespeare; Blake's printings of his own work) may have the same primary status of unique witness.

An electronic version of (for example) a reference text can reasonably claim to be as good a version of the text as the printed version. Both the electronic and the printed version are simply alternative expressions of the one reference work. The electronic version may substitute completely for the printed version, so much so that one can foresee that certain reference works (e.g. the *Oxford English Dictionary*; the British Library catalogue) might cease to exist in printed form at all.

However, no electronic version of a primary textual source can conceivably substitute for that source. The clay tablet, the manuscript, the rare first edition, all are the thing itself. They are not accidental representations of another object; they themselves are the object of interest.

This distinction, between electronic versions of primary textual sources and of non-primary textual sources, has important implications. One may assert that for a reference work (e.g. an encyclopaedia) the text has been completely captured when one has recorded every word of the text and every aspect of its structure (e.g. its headings, subheadings, and divisions into sections and subsections), regardless of exactly how that text appears in any of its printed forms. The accuracy of the

text capture can be measured objectively: if the printed text can be generated precisely from the electronic text (as happens routinely in typesetting) then the electronic text is the perfect equivalent of the printed text and can substitute for it completely.

But for a primary textual source one cannot assert that the text is the “words plus structure.” A primary text is the actual signs made upon the physical medium. For a manuscript, this will be not only the letter forms made by the scribe but their disposition upon the page: the use of colour, as emphatic or structural or decorative device; the layout of scribal signs upon the page; a hierarchy of scripts within the inscribed text; indications of correction, annotation, or deletion; the physical characteristics of the manuscript itself. These cannot be detached from the text and treated as aspects of presentation or structure. Any or all of these might have to be transcribed for satisfactory expression of that source.<sup>5</sup>

Any primary textual source then has its own semiotic system within it. As an embodiment of an aspect of a living natural language, it has its own complexities and ambiguities. The computer system with which one seeks to represent this text constitutes a different semiotic system, of electronic signs and distinct logical structure. The two semiotic systems are materially distinct, in that text written by hand is not the same as the text on the computer screen. They are formally distinct, in that a manuscript may contain an unlimited variety of letter forms but a computer font ordinarily will not. They are logically distinct, in that the computer transcription will attempt to resolve ambiguities present in the natural language of the primary source (e.g. the same graph being used for distinct letters; cf. the discussion of minims below): if the transcription does not do this, it will betray its principal aim of decoding of the primary source. Transcription is both decoding and encoding; the text in the computer system will not be the same as the text of the primary source.

Accordingly, transcription of a primary textual source cannot be regarded as an act of substitution, but as a series of acts of translation from one semiotic system (that of the primary source) to another semiotic system (that of the computer). Like all acts of translation, it must be seen as fundamentally incomplete and fundamentally interpretative.

## 1.2 The choice of level of transcription

### 1.2.1 The choices available

Up to the advent of computers, scholars transcribed primary sources for appearance in printed form either directly (as diplomatic editions) or indirectly (as absorbed in critical apparatus). Transcription for printing immediately bounds the act: there is no point transcribing what the printer cannot print. For transcription of printed texts, there is a further bound: one can identify the particular characters used in the printer's fount and simply find equivalents for each character in the transcription.

There are no such bounds in transcription of manuscripts for the computer. In theory, one can represent anything in the computer. At one extreme one could make a “graphic” representation, in which the limitless repertoire of marks in a manuscript is matched by a limitless repertoire of computer signs. At the other extreme one could make a “regularized” representation, in which the manuscript is transcribed as if for a printed edition, limiting the signs used and regularizing the spelling. One may categorize the possible levels of transcription of a manuscript of *The Wife of Bath's Prologue* as follows:

- “Graphic”: every mark in the manuscript, every space, is represented in the transcription, even to the point of decomposition of letter forms into discrete marks (as: each “i” is made up of a vertical stroke of particular breadth, length, and weight, and a dot of particular size, shape and weight in a particular position relative to the stroke). The transcription of the corpus of Norwegian runes based in the Norwegian Centre for Humanities Computing, Bergen, is an instance of a transcription on “graphic” principles.
- “Graphetic”: every distinct letter-type is distinguished (as: r “short” is transcribed apart from r “round” and r “long descender”, etc.) The transcription of Old Norse manuscripts by Hans Fix aims at a graphetic transcription (Fix 1984) and the advocacy by McIntosh (1974; 1975) and Benskin (1990) of “scribal profiles” implies graphetic transcription of Middle English manuscripts.
- “Graphemic”: every manuscript spelling is preserved (as: “she”, “sche”) without distinction of separate letter forms as in a graphetic transcription. Diplomatic transcripts, for example those of Ruggiers for the Hengwrt manuscript and Furnivall for the Chaucer Society, are centred on a graphemic reproduction.
- “Regularized”: all manuscript spellings are regularized to a particular norm, perhaps the spelling of a manuscript considered authoritative. The many editions of Chaucer which approximate the spelling of the Ellesmere or Hengwrt manuscripts are examples of this, as are the variants reported in the collations of Manly and Rickert.

In practice, most transcriptions cannot be completely defined as belonging to one or other of these categories. Typically, a transcription (or an edition) may in most respects conform to one category but include features belonging to another. For example, many editions present a regularized spelling but scrupulously follow a particular source in separating u/v and i/j, distinctions proper to a graphemic or even a graphetic representation rather than a regularized one. Diplomatic transcripts of Middle English manuscripts commonly preserve all graphemic distinctions except for abbreviations, which are expanded into standard form as in a regularized edition. Thus, the first questions we faced as we sought to define our transcription practices were these:

- At which one of these four levels should we aim in our transcription?
- Should we aim for stringent conformancy to this one level; and if we do not intend stringent conformancy, what mixture of different features from different levels might we permit?

### 1.2.2 The rejection of “regularized” or “graphic” transcription

We decided, very early, that we would not regularize the transcripts as we did them. This may appear surprising as the major immediate aim of this Project is to recover the history of the development of the tradition of *The Wife of Bath's Prologue* by analysis of the agreements and disagreements of the manuscripts. For

all manuscripts to a large extent, and for all but a few entirely, this analysis will rest upon “significant variation”, that is variation in substantive readings. The text of all manuscripts will have to be regularized before collation to yield variation in substantive readings alone. Why not then regularize as we transcribe?

We chose not to regularize as we transcribed for the following reasons:

- The computer collation program we are using (*Collate*) permits regularization as part of the collation process. This has the great advantage of allowing deferral of regularization until all the evidence of all the spellings in all the manuscripts at any one point is available. It also permits a complete record to be made of all regularizations done during the collation. *Collate* can also generate regularized-spelling versions of each file from the regularization process. Thus, there is no need then to regularize during transcription as regularized versions of each file will be made as a matter of course during collation.
- Transcription would actually be very much slower if the transcriber had to pause at almost every word and decide what the regularized spelling was. It would also be extremely difficult (if not impossible) to enforce consistency across all the transcripts, and decisions about the right regularization for particular words (especially verb forms) would be difficult to make on a manuscript-by-manuscript basis.
- Although for most manuscripts collation of the regularized text will produce sufficient information to place those manuscripts in genetic relation to one another, we were convinced that for certain important manuscripts one would need more information concerning their relationships than could be derived from a regularized collation. We had in mind the failure of Manly and Rickert to clarify the relationships of the small group of manuscripts critical to establishment of what Chaucer actually wrote: at least Hg, El, Cp, Ha<sup>4</sup>, Gg, and Dd. We believe that collation of unregularized transcripts (including graphemic information and perhaps more) of these crucial manuscripts will yield vital information which will enable us to advance beyond Manly and Rickert, who had to rely upon information from a regularized collation. Experiments with the collation of unregularized transcripts of these manuscripts suggest that this belief is justified.<sup>6</sup>

At the other extreme of this typology of transcriptions, a “graphic” transcription of the manuscripts of *The Wife of Bath’s Prologue* seemed far beyond our capacities, and also of dubious benefit at this point in the history of *Canterbury Tales* textual scholarship. The graphic transcription of the corpus of Norwegian runes has to deal with a total of only some thirty thousand characters. There are more characters than this in a single manuscript just of *The Wife of Bath’s Prologue*. Also, there are special difficulties in runic materials which make a graphic transcription desirable, notably the problem of determining just what constitutes a single letter. This is not normally a problem in the Chaucer manuscripts. Further, provision of manuscript images beside the transcripts in our electronic publication would supply the benefits of a graphic transcription regarded simply as a visual representation of the manuscript. Finally, the precision required of a graphic transcription would necessitate that they be done from the manuscripts themselves and we do not have the resources for this.

### 1.3 The choice of graphemic transcription

Our choice, then, lay between a graphemic transcription, aiming to preserve all information about distinct spellings in the manuscripts, and a graphetic transcription, aiming to preserve all information about distinct letter forms in the manuscripts. There were powerful arguments in favour of both. For a graphemic transcription:

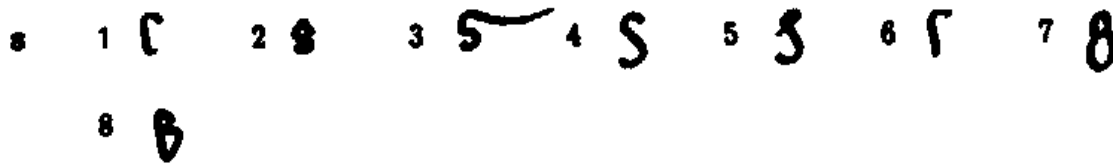
- This would give us access to a much greater volume of information about the relations between manuscripts, making it possible to refine analysis based on agreement in substantive readings with knowledge of the flow of spellings from manuscript to manuscript through the tradition.
- It would be possible, on the basis of the complete record of spellings in each manuscript and from the databases correlating all spellings in all manuscripts with their regularized forms provided by *Collate*, to make linguistic profiles for manuscripts individually and in groups, ranged both across time and place of copying. Use of this information alongside the linguistic profiles in *LALME* (McIntosh *et al.* 1986) might yield a rich harvest concerning layers of scribal copying (as suggested by Smith 1988 and Samuels 1988) and in turn extend the information in *LALME* itself.
- The record of spelling changes implicit in these spelling databases might be used to extend our understanding of the development of the language over the century of the production of these manuscripts.<sup>7</sup>
- A graphemic transcription, involving relatively straightforward literatim transcription without translation into a regularized form, seemed likely to be easier, more accurate, and more consistent in performance than a regularized transcription.

For these reasons, we determined that our transcription should be at least graphemic. The question then was, should it go further than this and be graphetic, preserving not only the distinct spellings but also the distinct letter forms which make up those spellings. We had in mind the arguments of McIntosh and Benskin in favour of “scribal profiles”, analogous to the “linguistic profiles” enabled by graphemic transcription. It seemed too that once one had done a graphetic transcription, it would be possible to generate a graphemic transcription from this simply by levelling all the different graphemes to the appropriate grapheme: all the different types of “s” to “s”, of “r” to “r”, etc. Thus one could have all the benefits of both a graphemic transcription and a graphetic transcription.

Therefore, in the first round of transcription of the manuscripts we experimented with discrimination of some graphetic forms. The purpose was to investigate the practicality and the benefits of graphetic transcription. We discriminated only the following characters: “short” r, “long-descender” r, “round” r; “short” s and “long” s; anglicana w and secretary w. Concerning the practicality of graphetic transcription: we found that while there seemed no cost in time in distinguishing these letter-forms in this first transcription, there was a marked cost in accuracy. It appeared that the concentration by transcribers on distinguishing these few characters meant that gross errors elsewhere in the transcription went undetected. Of course these could be repaired later, but the cost in time would be considerable and it was likely that this effect of distraction would persist, so that sufficient such errors would survive to damage severely the utility of the work. In itself, this difficulty in distinguishing just a few graphemes must give pause. Achievement of a transcription which is graphetic to some degree would, at the least, take rather longer to achieve the same level of accuracy as a graphemic transcription, if indeed it could achieve this same level. However, if this experiment

showed sufficient benefits the effort might be worth making.

As we carried forth this experiment, we became aware of other factors that appeared to negate the benefits of distinction of just these few graphetes. Consider the forms of s: we distinguished just two, s and "long s". But we recognized very quickly the existence of at least two other forms of s, so-called "kidney-shaped" s, and sigma s. It did not appear possible to draw any conclusions about the distribution of any form of s unless we considered these other forms, also. But how many other forms might we need to consider? Benskin's table (1990, 193) gives eight forms of s:



The closer we looked at the manuscripts, the more different types of s we saw; the more different types of s we saw, the less confident we became that we could identify with any consistency these different forms across all the manuscripts. One can see these difficulties even from Benskin's table: one can imagine infinite extension of the types of s that could be added to this table, and imagine too the difficulty (for instance) of consistent discrimination of types 7 and 8 from each other—or are types of s which appear to be almost but not quite 7 or 8 yet further separate categories? Once we have begun this exercise of categorization, where do we stop?

Another difficulty we discovered was the overlap of graphetes with graphemes. It is assumed by both McIntosh and Benskin that the relationship of graphemes to graphetes is hierarchical: so many graphetes of s; so many sub-types of each graphete; even sub-sub-types, and so on. But scrutiny of the manuscripts showed many instances where this hierarchy appears to be disturbed by graphically identical letter forms standing for quite different graphemes. In many manuscripts, for example, a form of long s is identical with a form of f; forms of o and e, of c and t, may also be identical. How are these to be treated? If we level them, into a single graphete which may have different graphemic values, we have lost immediately the advantage of automatic generation of graphemic transcription from the graphetic transcription. If we do not level them, we have obscured the scheme of graphic distinction and agreement which is the sole justification for graphetic transcription.

At the heart of these difficulties lay our sense that the novelty of graphetic transcription would involve us in problems that we could not anticipate. By contrast, graphemic transcription (the model of centuries of diplomatic transcription) is tried and well understood, not least by the scholars who would be using our transcripts. It is, for example, the basis for the transcripts contained in Parkes 1969 and Ruggiers 1979. We thought at first that graphetic transcription would be problematic simply because the microfilms from which we do so much of our work lack the detail to allow us to make accurate distinctions between letters. But when we came to work from the manuscripts themselves we realized that the difficulty lay in the act of distinction itself, and in the many problems it raises. We do not believe that these problems are insoluble, and the work of McIntosh and Benskin (as of their initial model, Bliss's 1951 study of the Auchinleck manuscript) shows the possible benefits of scribal profiles. The way to graphetic transcription lies through refinement of scribal profiles based on selected features of individual manuscripts (e.g. Svinhufvud 1978), and through progressive test transcriptions in controlled circumstances (e.g. of parts of different manuscripts written by one scribe; of a single shorter work in relatively few manuscripts).<sup>8</sup> Through this work, a methodology may develop that would permit graphetic transcription of the type not attempted here.

Accordingly, we have abandoned our experiments at a limited graphetic transcription, and no longer distinguish the different forms of r, s, and w discriminated in the first round of transcription. At some time, a graphetic transcription of at least some of the manuscripts of the *Canterbury Tales* will be done. The best basis for such a graphetic transcription will be an accurate graphemic transcription. We determined that our task is to provide this accurate graphemic transcription.

#### 1.4 Incorporation of graphic elements in the transcription

The question remains, how stringently should our transcription adhere to the graphemic model. The question was answered for us by the nature of the information in the manuscripts. For almost all the letter-forms in the manuscripts, there is a single, unambiguous and commonly accepted graphemic representation: thus for all the forms of a, or f, or s, etc. But for some marks in the manuscripts (notably marks of abbreviation, tails) there is considerable doubt as to their correct graphemic equivalent. Often, it is quite uncertain whether a particular mark has any graphemic meaning at all: thus the many signs that might or might not represent final -e. One of the reasons for attempting transcription is to provide a basis for evaluation of just these questions, with resolution of questions concerning final -e of particular importance to studies of Chaucer's metre. Therefore, the presence of these marks must be noted in the transcription. Observe that we do not try and record every mark, only those marks considered of likely significance: thus we do not transcribe flourishes judged as purely ornamental, as all those after final vowels.

These possibly significant marks are categorized, to some extent, by their graphic realization in the manuscripts. Thus, we encode as one character the flourish common on both final r (where it may represent e) and final u (where it may represent n), because in many manuscripts these signs appear graphically identical and it is impossible to distinguish them on the basis of their uncertain graphemic value. On the other hand, we distinguish this flourish character from the macron common over final u/n (where it may represent n), because in almost all manuscripts these signs appear graphically distinct, and we cannot identify them on the basis of their uncertain graphemic value. Consider these three words, taken respectively from Hg WBP 209, Hg WBP 106, and EI WBP 105:



We transcribe these as *hir*, *deuociou* and *pfecciou*. The final marks may be all purely ornamental; or they may represent three quite different abbreviations (for final e, u, or n) depending on whether we interpret the final two minims of *deuociou* and *pfecciou* as u or n; or some combination between. We do not know which of these is correct. But we are relatively certain that the scribes are using two different signs, that we can distinguish these signs consistently and usefully across all the manuscripts, and that is what our transcription reflects.

**1.5 The choice of signs to distinguish**

In the preceding section, we state that we now use the same sign in our transcription for the upward flourish mark over final r in *hir* and the upward flourish mark over final u in *pfecciou*. In the first cycle of transcription, we distinguished the sign used in *hir* from that used in *pfecciou*. This was because our first decisions about what we should and should not distinguish were based on scrutiny of the early manuscripts, particularly Hg and El. In both these manuscripts, the “lower-case hook” sign of abbreviation of final e after r, as in *hir*, is distinct in use and appearance both from the upward flourish (as in *pfecciou*) and from the “upper-case hook” sign used for -er abbreviation (as in *er<sup>2</sup>ne*).

However, as we then sought to apply this distinction over the whole range of manuscripts, we discovered that in the majority of manuscripts either this lower-case hook could not be distinguished in use or appearance from the upper-case hook, or it could not be distinguished from the flourish. In a few manuscripts, as

in Fi, one might not be able to distinguish any of the three signs from one another. From Fi, compare ou<sup>2</sup> WBP 634, where the context demands

“over”, with Vp on WBP 691, and both with forber WBP 643.

It became clear that we could not maintain the three-fold distinction of lower-case hook/upper-case hook/flourish in the face of the inconsistencies of scribal practices across all the manuscripts. We considered levelling all three signs to one. However, the clear graphemic status throughout the manuscript tradition of the

hook standing for *er/re* in such contexts as er<sup>2</sup>y El WBP 81, as against the uncertain graphemic status of the other scribal marks here discussed, persuaded us that a consistent and useful distinction could be made between the hook character (as in *er<sup>2</sup>ry*) and the flourish (as in *Vpoñ*). Thus, whereas in the last section we are swayed by the graphic realization of the macron and the flourish to distinguish characters of uncertain graphemic value, we are here swayed by the perceived graphemic value of a character, as shown by its context, to distinguish it from other characters which might have identical graphic realization.

Even though the marks over the final letters in Fi in WBP 634 and WBP 691 are identical, from the context we interpret the first as the hook character abbreviating “er”, thus ou<sup>2</sup>, and the second as an otiose flourish, thus Vpoñ. Observe that this decision is the result of our need for a predictable transcription across all the manuscripts. Were we only transcribing El and Hg, we would choose to preserve the threefold distinction.

The discussion in this and the last section illustrates that where the graphemic value of a sign is uncertain, the interplay between graphic and graphemic factors in our transcription cannot be simply formulated. The choice of a particular character to represent a particular mark in the manuscript may be determined less by the appearance of that mark on the page, and more by our perception of the sense of the passage, of the scribe’s practice throughout the manuscript, and of what is practical over a single transcription of so many manuscripts. In the detailed discussion below we outline our practice in particular cases, so that scholars might be aware of the range of choices available. In summary: our transcription aims to be graphemic at every point where it can reasonably be so. Where graphemic transcription is not possible, graphic factors are weighed in the decision of what signs to use.

**2. The practice of this transcription**

A specially designed computer screen font was used for the transcription. Standard alphabetic forms in this font were:

a b c d e f g h i j k l m n o p q r s t u v w x y  
 A B C D E F G H I J K L M N O P Q R S T U V W X Y

These were supplemented by the following Middle English characters:

þ ȝ ȝ

The following signs of abbreviation were available:

p p P P<sup>2</sup> 9 9 ƒ q ]

These characters are available as superscripts:

a e i r t u

Characters usually occurring at word ends are:

ʒ h h d

Marks of punctuation are:<sup>9</sup>

¶ , ; : . ( )

## 2.1 Characters not transcribed

Only signs held to have potentially graphemically distinctive value were transcribed. Thus:

- Dot over y or i was universally treated as part of the letter and was not separately transcribed.
- In some manuscripts capital I (the personal pronoun) appears with punctus immediately before and after. These punctus were treated as part of a letter and were not transcribed.
- Distinct forms of the tironian note ʒ were not transcribed.
- No flourish after a final vowel was transcribed.

## 3. Characters difficult to distinguish

### 3.1 y/þ


In a number of manuscripts y and thorn are so similar (even, identical) in shape as to be difficult or impossible to distinguish on their graphic representation alone, for example Lc, Ld<sup>1</sup>, Ld<sup>2</sup>, Mm, Sl<sup>1</sup>.<sup>10</sup> Often this similarity is not a problem for transcription because the context enables determination of which letter was used. In other cases, such as the personal pronouns you and þou, discrimination is more difficult. Sometimes the verb form may suggest that þou was intended, e. g. Ld<sup>2</sup> WBP 17 þo<sup>u</sup> hast. The use of such verbal forms and also the use of the possessive pronoun-þine rather than your -may indicate that þou is a regular form for a second person singular pronoun in a given manuscript. In such manuscripts it was possible therefore to determine the regular practice of the scribe and to transcribe in dubious cases the pronoun as þou rather than you. This was done in Ld<sup>1</sup> and Ld<sup>2</sup>. However, this practice does not work in all cases. Thus the scribe of Sl<sup>1</sup> does not distinguish clearly between y and þ, and there is also no uniformity in this manuscript in the use of 2nd prs sg pronoun. Sometimes the form of a verb can indicate that þou was intended: þou seyste WBP 278. However, in both Sl<sup>1</sup> and Sl<sup>2</sup> you was also used as a 2nd prs sg: Take youre disport WBP 319. When it was impossible to distinguish between þ and y by letter form, context, or scribal practice we made an arbitrary decision.


In some of these manuscripts, one may use differences in shape (albeit slight) between þ and y to help discrimination. Thus in Sl<sup>2</sup> y usually has a more curved descender (cf. 184) whereas in þ it is straighter: WBP 176 þ<sup>u</sup>g. However, it is difficult to be certain in every case in this and other manuscripts and discrimination by shape is an insecure guide.


### 3.2 Minims

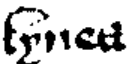
In many manuscripts minim letters pose problems for interpretation. Thus n and u are difficult to distinguish because they are both represented by two minims rather than by different letter forms. Determining whether it is u or n is possible only by context. In words like diuine we were not so much concerned with the problem of similarity between u and n because the context unambiguously suggested this reading. We transcribe three minims as in in were in this world Hg WBP 2 because this is the only reading that makes sense. The same contextual approach works in the case of four minims in Iovinian Hg WBP 653 and enuene E1 WBP 474 or six minims in diuine Ha<sup>4</sup> WBP 26.

However, there are cases where it was more difficult to decide how to transcribe what appears like two or more minims in the manuscript. Thus the name Lyuia

721 was interpreted by most transcribers in the initial transcription as Lyma from which it is indeed indistinguishable: thus in Gg  WBP

721; compare in as written a few lines later in Gg WBP 759:  We decided to transcribe it as Lyuia always when we find just three minims and it is

impossible to know what exactly was intended. However, there is evidence that some of the scribes thought the name to be Lyma. It is probably 

lima in La because the scribe generally draws strokes above each i and there is only one stroke in this word. Conversely, in Li the spelling  Lynea WBP 724 shows that some other scribes knew the name correctly.

Similarly, it can be uncertain whether a manuscript has nyl or uyl (e.g. Ph<sup>2</sup> WBP 307, WBP 319) especially because double negation was possible in Middle English. In Ph<sup>2</sup> the usual form of "will" is wil and therefore we preferred the reading nyl.

Another ambiguous case is that of the spellings -oun and -aun-. Here as in other cases where we have difficulties in distinguishing between u and n transcription is interpretative and has to depend on the context and on linguistic knowledge. We transcribe the last four minims in diffynycioun Hg WBP 25 as un and not, for example, nn because we know that the digraph "ou" derived from French was used in Middle English in this context. We transcribe a followed by four minims in comaundement and similar cases as -aun- and not -ann- for the same etymological reasons. This practice was adopted for all occasions in manuscripts where u and n can not be reliably distinguished by their graphic representation alone even though we do not exclude the possibility that unetymological spellings -ann- and -onn could have sometimes occurred. See further the discussion of flourishes and macrons over final u/n below.

### 3.3 Flourishes and tails

There is usually no difference in graphic form between a flourish over a final u (an abbreviation for n) and a flourish after the final r (sometimes an abbreviation for a final -e). As remarked above, the uncertain graphemic meaning of these signs made it impossible to distinguish these by context. They were rendered by the same character in the transcription: ū, r̄. The superscript hook <sup>2</sup>, typically a mark of abbreviation for -er-, -re can also look indistinguishable from flourishes over u and r. However, as explained above the graphemic meaning of <sup>2</sup> is not usually in doubt and this character is distinguished from the flourish character in the

transcription: thus  eu<sup>2</sup>y El WBP 81.

Often tails at the ends of words can be confused with punctuation marks. See further the discussion in the section on flourishes below, p. 34.

### 3.4 o/e

In certain hands o and e can be similar in shape and distinction between them must be by context. Examples from the manuscripts are: Tc<sup>2</sup>wel WBP 118, sent WBP 150, where WBP 237; leves Mg WBP 764.

### 3.5 c/o

In some manuscripts c and o are similar in shape and distinction between them must be by context. Examples from the manuscripts are: Ma locke WBP 317, schaf WBP 332, reckith WBP 327.

## 4. Abbreviations

With certain prescribed and rare exceptions, signs of abbreviation in the manuscripts were not expanded, but marked by special computer characters resembling those used by the scribes.<sup>11</sup> This treatment of brevisgraphs was adopted because the ambiguities and inconsistencies of scribal usage seen just in the comparatively brief section of The Wife of Bath's Prologue transcribed forbade certain assignment of any one phonetic value to any one sign. Across the forty-eight manuscripts, it was found that in different manuscripts the one brevisgraph could have different phonetic values and could even have more than one phonetic value in the same manuscript.

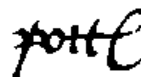
In most cases it is clear that the brevisgraph represents an abbreviation, though precisely what is abbreviated varies both within and between manuscripts. There are also cases when it is not even certain that a sign in a manuscript represents an abbreviation: it might represent an abbreviation, but it might also be simply decorative or conventional and thus have no phonetic value.

Because of all this it was decided to avoid expanding abbreviations as far as possible. The following characters were used for transcription of abbreviations.

- <sup>2</sup>-superscript hook was employed in most of the manuscripts to represent both -er- and -re- : er<sup>2</sup>ne, exp<sup>2</sup>sse. This usage is found in Bo<sup>2</sup> Bw Ch El En<sup>1</sup> En<sup>2</sup> Gg Ha<sup>4</sup> Ha<sup>5</sup> Hg Lc etc. However, in some manuscripts, for example Dd En<sup>3</sup> Ha<sup>2</sup> Hk Ne, this abbreviation was employed for -er- only. In some manuscripts (e. g. Ra<sup>3</sup>) it does not occur at all.
- p-p with a loop regularly stands for -pro- in most of the manuscripts including Hg and El. However, it was used for both -pro- and -per- in En<sup>1</sup> Bo<sup>2</sup>, and once for -pre- in Gg p̄fere WBP 96. In Bo<sup>2</sup> it is used for -pro-, -per- and -par- p̄prely WBP 224; p̄fecioun WBP 105; p̄de WBP 200.
- p-crossed p was employed to represent -per- (p̄fccion) in some of the manuscripts, for example, Bo<sup>2</sup> Ha<sup>5</sup> Hg Hk Ne Ra<sup>1</sup>. It was used for both -per- and -par- in Lc p̄ile WBP 89, page WBP 250; El p̄il WBP 89, p̄doner WBP 185, p̄dee WBP 200 and also Bw Dd Ch Cn En<sup>1</sup> En<sup>2</sup> En<sup>3</sup> Gg Ha<sup>4</sup> Ht, etc. It was almost never employed to represent -pre-, which was rendered by p followed by er-abbreviation : p<sup>2</sup>. An exception to this usage appears to be pambel Ha<sup>4</sup> WBP 805; this may be simply an erratic spelling.



- <sup>9</sup>-Abbreviation for -es (-is), or sometimes -us, is quite frequent in some of the manuscripts: husbond<sup>9</sup> Ra<sup>1</sup> WBP 6 (where -es is the usual plural form); ven<sup>9</sup> Ht WBP 594, WBP 682; lacym<sup>9</sup> Ne WBP 731.
- In Gg it seems to be used as in Latin manuscripts for -us only: þ<sup>9</sup> WBP 160, where the context determines "thus" not "this".



- v-Abbreviation for -es (-is) is comparatively rare in the manuscripts, occurring in Fi pottv WBP 287; also lordyngv WBP 112,

thyng<sup>x</sup> WBP 121, etc. It is also found in Gl, and does not appear to be used for -us.

- $\text{f}$ -Abbreviation for ser- is found in Bo<sup>2</sup> En<sup>1</sup> Hg Hk Ht Ra<sup>1</sup> in the word seruyce (101). In Ra<sup>1</sup> and Ht it also sometimes stands for sir(e): Ra<sup>1</sup>, Ht WBP 193; Ht WBP 355. In Ht it is used in prefue WBP 148.
- $\text{q}$ -q with a loop is used in some of the manuscripts, for example Gg, as an abbreviation for quod, so spelt out in other manuscripts.

#### 4.1 Macrons

The macron is used in abbreviated spellings of personal names and Latin words, most commonly in the abbreviation for “Jesus”, transcribed as Ihus Hg WBP 15, Ihu in Hg WBP 146, also ihu, ifc or Ihū, ihū, ihū depending on the position of the horizontal stroke. It appears in the abbreviation for “omnipotent”, where its position can vary: omīpotent Cn WBP 423, omīpotent Ch WBP 423.

The most frequent use of the macron is over a vowel, as an abbreviation for n or m. It is used very consistently in Gg and can occur over any vowel. A very similar usage of a bar as an abbreviation for n or m is found in Ha<sup>4</sup>: comaundemēt WBP 67, wommā WBP 87; noubre En<sup>3</sup> WBP 25, tormētrye WBP 251.

A bar over a vowel is used in spellings where o is followed by m in Bw El Hk Ht Lc probably as an abbreviation for a second m: cōmaundement Lc WBP 67, sōme Lc WBP 104. A similar usage is found in Bo<sup>2</sup> En<sup>1</sup> Ra<sup>1</sup> Ra<sup>3</sup> but it is more deliberate: the bar occurs as a rule where one would expect a double m (cōmandement).

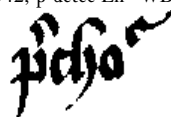
A bar over vowels followed by m or n is found in other cases than o though less commonly: En<sup>1</sup> wōmans WBP 371; sūme Dd WBP 101. This usage is common in Ch.



When a macron appears over more than one letter it was transcribed over only one: wōman Cp WBP 249. This is often a problem in Mm where macrons usually extend over more than one character and in the case of the character h over the entire word (thus too in Gl and Ry<sup>1</sup>).

#### 4.2 Superscript characters

- <sup>a</sup>-superscript a is employed in Dd and Ra<sup>1</sup>, and elsewhere, as an abbreviation for -ra- (p<sup>a</sup>y Ra<sup>1</sup> WBP 61). Sometimes, it appears in abbreviations for Abraham: abrah<sup>a</sup>m En<sup>3</sup> WBP 55.
- <sup>i</sup>-superscript i appears mainly as an abbreviation for -ri-: p<sup>i</sup>uytee Bw WBP 542, p<sup>i</sup>uetee En<sup>1</sup> WBP 531, p<sup>i</sup>vily En<sup>3</sup> WBP 718, c<sup>i</sup>st Ra<sup>1</sup> WBP 10.



- <sup>r</sup>-superscript r occurs in different abbreviations involving r (-ur-, etc): Ra<sup>1</sup> p<sup>cho</sup>r WBP 165, oth<sup>r</sup> WBP 181.
- <sup>e</sup>-superscript e occurs in several manuscripts: p<sup>e</sup> Pw WBP 729; y<sup>e</sup> Ch WBP 802-1, iiiij<sup>e</sup> Fi WBP 452.
- <sup>t</sup>-superscript t is used in abbreviations for that p<sup>t</sup> and with w<sup>t</sup> in Dd En<sup>1</sup> Gg etc.
- <sup>u</sup>-superscript u occurs in abbreviation for thou p<sup>u</sup> and you (various spellings) in Dd En<sup>1</sup> Gg etc.
- 

#### 4.3 Characters occurring at word ends

We encountered particular and recurrent problems in transcription of characters at word ends. The fundamental difficulty was the uncertain boundary between marks of decoration and marks of abbreviation. On the one hand, scribes felt a need to mark the last character, often with a flourish, a tail, or simply an extended final stroke, or perhaps (though this is not reflected in our transcription) a special letter form typically reserved for final position. On the other hand, the decay of the system of abbreviation in relation to the weakening of the inflectional system meant that signs with clear phonetic value in some contexts in some manuscripts were used with no such precision in other manuscripts, or even in other contexts in the same manuscripts. In this transcription, particular attention had to be paid to these issues because of their bearing on the presence or absence of final -e, and hence on Chaucer's metre.

##### 4.3.1 $\text{ʒ}$ (crossed double l)

A crossed double ll at the ends of words is very common in a number of manuscripts, e. g. in Ra<sup>1</sup> apostell WBP 49 ; counsell WBP 82 ; pill WBP 89. In some of them (e.g. Ht, Ra<sup>1</sup>) it alternates with a single l: Ra<sup>1</sup> shaff WBP 45; shal WBP 171; in others (e.g. Hk) with -lle. In Ch and En<sup>1</sup> it is used with great consistency and seems to be the only possible form of a final l. Single crossed l is rare but also occurs sometimes. Thus it occasionally appears in Ht in the word apostle (e.g. WBP 79). In the transcription it was represented by l with a macron.

This variety of usage makes impossible any uniform treatment of crossed l characters, beyond simple registration of their occurrence in the manuscripts. The assignment of any constant phonetic value to them would be artificial and arbitrary. A further study based on extensive manuscript material is required before any firm conclusions can be made concerning their meaning.<sup>12</sup>

##### 4.3.2 $\text{h}$ (crossed h)



Crossed h appears in En<sup>3</sup> and Dd at the ends of words or in the group ht when this occurs at the ends of words: in En<sup>3</sup> myht WBP 23; ech WBP 43; thouh WBP 53. It is also common in personal names: Iohn Dd WBP 164. In En<sup>1</sup> it appears only at the ends of words. The scribe of Gg occasionally uses crossed h in the group ht but not in other cases: WBP 77 wyght. In the majority of manuscripts this character is employed in one or both of these contexts—as a final letter of a word or in a combination with t. However, its use is often inconsistent: it freely alternates with the ordinary h.

#### 4.4 Flourishes and tails after other consonants

The downward stroke after final consonants, represented as  $\gamma$  in the The Wife of Bath's Prologue transcripts and present in many manuscripts, poses particularly difficult problems. It might represent an abbreviation, or it might be simply decorative or conventional:

- It might be an abbreviation: in a few instances, a word ending in a consonant with a stroke rhymes with a word ending in -e: þing $\gamma$ -grucchinge Ne WBP 405-WBP 6; Theofaste-fast $\gamma$ ; Dd WBP 649-WBP 50.
- It might be a decorative or conventional flourish after particular final letters, regardless of phonetic, grammatical, or metrical context. In some couplets, only one of the rhyming words ends in a consonant with a stroke: appetit $\gamma$ -whi3t Dd WBP 604-5-WBP 604-6, leef-deef $\gamma$ ; Ch WBP 613-WBP 4. That it might often be decorative is also suggested by the tendency of scribes to use it more or less often in company with certain preceding letters, rather than (for example) as a mark of abbreviation capable of occurring in any context. Thus, in Bo<sup>2</sup> it occasionally appears after the final f; in Dd after the final t. In Hk and Cn it occurs after f and g. In Ra<sup>1</sup> it is common after f and g, but is also used after t and ll. In Ch and Ne it appears after f, g, k, t. A similar usage is found in En<sup>1</sup> but in this manuscript the flourish is on the whole uncommon. In Ra<sup>3</sup> the flourish is a decorative feature and occurs after any consonant at the end of a word.
- There are many flourish marks in the manuscripts which cannot be anything other than purely decorative, e.g. those after final -e (Hg WBP 254 haue). Dd often has a upward flourish after -e, -s at line endings, and usually extends the final stroke of the -t in the same position. These appear to be orthographic decoration, and were not transcribed.
- In Cp sometimes the flourishes are so slight that one is often uncertain of their presence even when working from the manuscript itself.

This character might also be a mark of punctuation. In certain manuscripts there is constant difficulty in distinguishing between flourishes at the ends of words and virgules either on the grounds of shape or context. This is especially true of Hg and El. In these manuscripts a diagonal-like character is often linked to the final letter of a word which can be any vowel or consonant, or more rarely to the first letter of a following word. This can be a virgule joined to the previous or the next letter in a hasty and casual writing. The scribe sometimes connects the final letter of a word to the first letter of a following word (another tonne El WBP 170). A diagonal joined to a preceding letter is especially common with letters having horizontal strokes, such as f, t, g. In these cases it is likely to be a tail. However, the diagonal is not always joined to these letters and can follow them after an interval of space (cf. El WBP 73). In some cases a diagonal stroke looks more like a virgule, in others more like a tail. In a large number of cases it is impossible to decide between these two alternatives. Distinguishing by shape is possible in some rare cases though not in Hg and El. In Ch virgules usually appear as bolder strokes than flourishes. Both characters can occur together (WBP 157).

The position of punctuation marks often coincides in Hg and El but the agreement is by no means constant so that comparison of the two manuscripts will not be very helpful for deciding about the nature of diagonals. The eclectic character of punctuation and its use for both syntactic and metrical purposes makes it very difficult to argue that a punctuation mark is necessary or on the contrary unnecessary in a certain place, permitting interpretation of a diagonal stroke as a flourish or vice versa. The line of argument presuming that a diagonal stroke represents a tail and not a punctuation mark because punctuation should not be there is bound to be defective. Comparison between other manuscripts may not be helpful. Consider the line

*Byside a well, Ihe god and man*

Byside a well, Ihe god and man Ra<sup>1</sup> WBP 15. Here, comparison of manuscripts does not help to decide what the diagonal stroke after well represents. In other manuscripts we find a virgule, a final -e, or the absence of any sign after well. In Hg and El well is written with a final -e. In Hg it is also followed by a virgule.

In manuscripts like Ha<sup>5</sup> and Ra<sup>1</sup> where diagonals appear only after consonants with horizontal strokes (f, t, g) and not after other letters they can be more safely interpreted as tails. Their appearance is still ambiguous and it is their distribution rather than shape that permits such interpretation.

The varieties of circumstances in which diagonal-like characters may appear and the uncertainty of their application, outlined above, led us to devise the following rules for their transcription. A diagonal-like character is represented as a tail when it is:

- joined to a preceding letter;

*saist pat*

- looks like a tail, that is drawn like an angle  $\gamma$ : saist $\gamma$  þat $\gamma$ ; Ha<sup>4</sup> WBP 278 and not like a diagonal joined to a letter. Thus in Hg WBP

455 we have a diagonal linked to a preceding g, but in this case it is clearly written as a virgule and not as a tail:

*yongt*

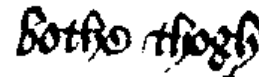
yong /;

- when it can possibly be a tail from what we know about the scribal practices in Middle English, i. e. after consonants but never after vowels, e. g. seyde Hg WBP 825, ensample El WBP 12, nombre El WBP 32, body El WBP 159.

When the character is attached to a final vowel or is separated from the preceding letter by an interval of space it was represented as a virgule, e.g.



we / Hg WBP 521. It was also transcribed as a virgule when attached to a following letter, e. g.



/ thogh Hg

WBP 313.

Our practice is a convention adopted for the sake of utility and does not claim to be a final interpretation of the meaning of these characters.

#### 4.5 $\dot{d}$ (d with a tail)

A special case of use of a downward stroke after a final letter of a word is its common occurrence after d. It usually appears different from tails after consonants with horizontal strokes like f or t. However, like other tails it can be confused with punctuation, its precise meaning is unclear and there is always a possibility of it being purely decorative. The morphological significance of final -d made us look at the cases of its use with a tail with particular attention. Because final -d is so important as a verbal ending we thought we might encounter greater deliberation in its use. To simplify statistical studies and taking into account its appearance as usually (though not always) different from tails after t, f, g, k a special character was used for d with a tail:  $\dot{d}$ .

Concerning its occurrence in particular manuscripts:

- In El it seems to stand regularly for -de (cf. in his houshold $\dot{d}$ -al of gold $\dot{d}$  WBP 99-100; in myn hond $\dot{d}$  WBP 211 and on honde WBP 226).
- The same practice is found in Ra<sup>3</sup> where in WBP 479-80 husbond $\dot{d}$  rhymes with fonde. However, it is difficult to be sure because in this manuscript almost every final consonant appears with a downward stroke.
- In Bo<sup>2</sup> in WBP 143-4 seed $\dot{d}$  rhymes with brede. Similar rhymes occur elsewhere in this manuscript.
- In Ne the use of  $\dot{d}$  is quite common and rather irregular. This can be seen in rhymes like in his houshold $\dot{d}$ -of gold $\dot{d}$  WBP 99-100, or befed $\dot{d}$ -bred WBP 143-4. However, it is also common that both rhyming words end in  $\dot{d}$  (e.g. WBP ">231-2). The flourish is more usual in rhyme but also appears inside the line (kaynard $\dot{d}$  WBP 235). Rhymes like word $\dot{d}$ -borde WBP 421-2 are also possible.
- In Ch  $\dot{d}$  is quite common and appears in nouns in the nominative and in the oblique cases and much more rarely in adjectives. It almost never appears in the word god (the only exception is in line WBP 44). In verbs and participles its use is rare. In rhymes in most cases both rhyming words end in  $\dot{d}$ . The tail often looks similar to a virgule but there are cases where both a virgule and  $\dot{d}$  are used together (WBP 451, WBP 574-5, WBP 574-6).
- In Cn  $\dot{d}$  is always used in the past forms of verbs and in past participles. It does not occur as the ending of an infinitive (WBP 85, WBP 568 wed). It is not used in the forms had WBP 17, did, WBP 384 and in the adjectives good WBP 87, glad WBP 391, wood WBP 642. The only exception is the adjective "dead" (deed $\dot{d}$  in his chest WBP 502). It never occurs in the word god. In other nouns it can be present or absent: in his bed $\dot{d}$  WBP 88, namely a bed had they myschaunce WBP 407; haue bred $\dot{d}$  of pure $\dot{d}$  whete seed WBP 143; olde Dotard $\dot{d}$  WBP 291; olde dotarde WBP 331. The flourish never occurs in the conjunction and. It appears in afturward $\dot{d}$  WBP 610 and in bakward $\dot{d}$  WBP 767.
- In En<sup>1</sup> (which is closely related to Cn)  $\dot{d}$  is also very common. It always appears in preterite forms of verbs and in past participles. Exceptions to this are very rare—only three in The Wife of Bath's Prologue. However,  $\dot{d}$  never occurs in the form had. It always appears in the infinitives (commaund $\dot{d}$  WBP 73) and in adjectives (old $\dot{d}$  WBP 242, blynd $\dot{d}$  WBP 634). It is usually (though not always) found in nouns. No distinction seems to be made in nouns between the cases that require the final -e and those where it is not expected.  $\dot{d}$  is not used in the word god except in two cases: lines WBP 5 and WBP 671). It is never found in the preposition and but appears in afterward $\dot{d}$ . There are two cases where  $\dot{d}$  rhymes with -de: seed $\dot{d}$ -brede WBP 143-4; in honde $\dot{d}$ -vnderstonde $\dot{d}$  WBP 327-8
- In Fi the tail after d is very slightly drawn and it can be difficult to decide whether it is simple d or  $\dot{d}$ : e. g. had $\dot{d}$  WBP 195, hold $\dot{d}$  WBP 198.

#### 4.6 Flourishes and macrons over u and n

Final u and n often occur accompanied by flourishes and macrons in the manuscripts. In certain manuscripts the flourish or the macron occurs in cases where no abbreviation can be expected: they appear in words like ma $\dot{n}$ , certey $\dot{n}$  or in spellings like dou $\dot{n}$  where -oun is already spelled out. In these cases we rely on the context to determine that the final letter is -n and not -u and transcribe the flourish or the macron as it appears in the manuscripts.

Particular difficulty is caused by the endings -oun, -on. These endings often appear as o followed by two minims with or without a bar over the minims. Should this be transcribed as -o $\bar{u}$  or -o $\bar{n}$ ? The question is difficult because the use of the macron is inconsistent and because in most manuscripts the two minims might equally stand for u or n. Though it is likely to be an abbreviation for n it also often occurs in the contexts where no abbreviation can be expected: thus spellings like deffini $\bar{c}$ io $\bar{u}$ n or ma $\bar{n}$  are not uncommon. It could be a diacritic mark used for distinguishing n from u.

This uncertainty led us to the following rules for treatment of -oun, -on endings with and without marks of abbreviation over the minims:

- where there is no mark of abbreviation, we interpret the minims as n;
- where there is a mark of abbreviation, we interpret the minims as u, with the mark representing abbreviation of the final n.

We adopt this policy because, in the first place, final -oun might be simplified to -on (e.g. Hg WBP 107 pfeccion, Hg WBP 156 tribulacion), but hardly to -ou; in the second place, it is usual to represent a nasal by a macron over a vowel but not to represent a vowel by a macron over n or m.

There is still a possibility of the bar being a diacritic sign over n but it is difficult to be sure that such usage existed because the practices of manuscripts are inconsistent. The notion that a bar over a vowel is an abbreviation for n has a historical basis and there are many unambiguous instances of this use in these manuscripts (see above). We believe that the use of a macron or flourish over n in words like "man" is a result of the decay of the system of abbreviations.

Our decision was made partly for the reasons for convenience and in order to achieve consistent transcription practice. We feel that following this rule leaves less scope for interpretation and decision-making by every transcriber in each individual case. We follow this practice even in a manuscript such as En<sup>1</sup> where a flourish occurs after almost every final n and abstract nouns are spelt with o followed by two minims with a flourish. There is a rhyme diffiniciou–douñ in WBP 25-6, where the flourish over the final minims of “doun” must be ornamental and it is very likely that in “diffinicion” a simple -on spelling with a flourish was intended. However, because the spelling practices are so inconsistent and difficult to interpret and we are dealing with a large number of manuscripts we felt that we had to adopt a practical rule to make transcription more uniform and predictable.

It is very uncommon for the scribes to distinguish between u and n in their handwriting. A rare instance where this distinction is made with certain consistency is Gg. The practice of the Gg scribe supports our policy concerning the -oun ending: he regularly spells -oun as -ou. Our policy is also supported by the early printed editions (Caxton, Pynson, and de Worde) which all distinguish u and n and regularly spell -oun and -aun- in diffinicioun and commaundement, etc.

A rather difficult case is represented by words such as London, bacon or lion, often spelt with o followed by two minims with a bar. In these words the spelling -oun could have been used by analogy with -oun as a suffix of abstract nouns. But it could also be simply n with a macron because in almost all the manuscripts the flourish or a macron can occur in contexts where no mark of abbreviation can be expected. The question is whether, for example, o followed by two minims with a macron in London Hg El WBP 550, should be transcribed as ou or oñ. Our decision was that in cases where spellings influenced by analogy are more likely to occur—that is in some personal and place names and words of Romance origin or borrowed through French—we would use the spelling -ou.<sup>13</sup>

#### 4.7 Flourishes and macrons

We assume that in most and perhaps all cases the difference between the macron and the flourish—an upward and backward stroke from the final minim—is purely formal: they represent the same sign but the flourish is more common in hasty and casual handwriting.

Although in very many cases the difference between a flourish and a bar is only in appearance and sometimes is even difficult to make it was decided to keep this distinction in the transcription. First of all there is some possibility that in some cases the meaning of these signs could be different. The flourish is very similar in shape with an abbreviation for a final -e used after r. That these two characters could overlap is suggested by rhymes like these in Fi: wyne–swyñ WBP 459-WBP 60; tyme–envenyñ WBP 473-4. Also sometimes an upward flourish after n or u is not high enough to make it absolutely clear that a special character was intended. Sometimes it looks as if it is just a prolonged final stroke of a letter. This can be a common problem in some of the manuscripts. Changing a flourish into a bar for the purposes of regularization during transcription in cases where one cannot be sure that a character equivalent to the bar was intended does not seem appropriate.

It was also decided that an upward flourish after the final u or n will be transcribed only when the stroke rises above the top of the final minim : e. g.



douñ in Ha<sup>4</sup> WBP 119; but not



purgacioun WBP 120.

The following subsections discuss the use of these characters in particular manuscripts.

##### 4.7.1 Manuscripts using a flourish

Flourish over n is employed consistently in El and Hg in words where the spelling -oun is expected (conclusiōu El WBP 115). Where -oun is spelled out the flourish is absent (cf. WBP 25-6 in El where diffiniciōu rhymes with doun).

In some other manuscripts, for example Ha<sup>5</sup> or Bo<sup>2</sup>, it is used both where -oun or a simple -n are expected (Ha<sup>5</sup> pfeccioū WBP 105, sayñ WBP 175, oon WBP 209; Bo<sup>2</sup> diffinicoū WBP 25, mañ WBP 18, douñ WBP 26). The same word can be spelt with or without a flourish: Bo<sup>2</sup> kañ WBP 56; kan WBP 59. In some manuscripts like Bo<sup>2</sup> and Dd the flourish appears over a final m: beem WBP 496. The scribe of Cp and Ha<sup>4</sup> occasionally uses the flourish over the final n where -oun is spelled out: douñ WBP 119.

In Cn it occurs over every final n and m including cases where -oun is spelled out (conclusioun WBP 115).

In En<sup>1</sup> the flourish is consistently used after any final n. However, it hardly ever appears in certain words such as when, than, and also man, womman, lemman (WBP 696). Abstract nouns of Romance origin are usually spelled in this manuscript with o followed by two minims with a flourish. Very often the flourish appears after the final m.

There are also cases in the same manuscripts which show that the flourish can unambiguously represent an n, e. g. wommā Bo<sup>2</sup> WBP 66; or an m, hy Bo<sup>2</sup> WBP 567.

##### 4.7.2 Manuscripts using a macron

Other manuscripts use a macron over the final n. In Hk it usually occurs in words spelled with -oun: conclusioun, generacioun WBP 115-6. Once it is found in a preposition on WBP 226.

In Ra<sup>1</sup> it occurs over final minims where abbreviation of n can be expected: diffiniciou, dou WBP 25-6, but is sometimes absent: conclusyon, gen<sup>3</sup>acyon WBP

115-6. In Ha<sup>4</sup> it is used where an abbreviation is expected: conclusioun-generaciou WBP 115-6. In Lc it is found both where -oun and simple -n occur: diffiniciou WBP 25, certayn WBP 19. In Ra<sup>3</sup> the bar appears almost over every final -n, but exceptions are possible: Whan WBP 47, can WBP 56. It can also occur over the final -m: Som̄ WBP 104.

#### 4.7.3 Manuscripts using both a macron and a flourish

In some manuscripts, for example Ht and Ne, both a macron and a flourish are common. They occur both where -oun and a simple n can be expected: Ht diffiniciou WBP 25, cristen̄ mañ WBP 48; Ne perfecciouñ-deuociou WBP 105-6.

In Bw and Dd a flourish and a bar are used in the same way but much less often. In Dd rhymes like doun-suspeciou WBP 305-6 are possible, and also like doun-purgacione WBP 119-20.

In Ch the flourish and the bar are very rare and occur in words where abbreviation can not be expected. On the other hand -oun when it is required regularly occurs unaccompanied by a flourish.

In En<sup>3</sup> a bar or a flourish occur mainly where abbreviation can be expected, but also in a few rare cases in words like mañ WBP 15, on̄ WBP 25. Sometimes it also occurs over the final m, usually in the word Som̄ (e.g. WBP 48). On the whole this manuscript shows some consistency of use of an abbreviation for a nasal consonant: the bar over a vowel is very regularly used as an abbreviation for m or n also inside the word.

#### 4.7.4 p̄ (bar over a final p)

A bar over a final p is found in El Bo<sup>2</sup> Cn Dd En<sup>1</sup> Ra<sup>3</sup> and other manuscripts and was represented as p̄. Its precise meaning is unclear. Ruggiers expands it to -pe in his transcript of Hengwrt. This practice was opposed by Burnley (1982, 177) who considers the macron to be a "graphemically meaningless" orthographic convention and not an abbreviation for a final -e.

#### 4.7.5 r̄ (final r with a flourish)

Final r is often followed by a flourish: thēr Lc WBP 210; El hir̄ WBP 130, wher̄ WBP 50, sauour̄ WBP 171. This flourish can be an equivalent of a final -e in some cases though in others it is probably meaningless.

In some manuscripts the flourish is rare or does not appear at all: thus in Dd it occurs once in a marginal note (WBP 807). Its use is often irregular. In En<sup>1</sup> it is very common but is used inconsistently: it can occur in any word ending in r, but is often absent (pardonēr WBP 163, pdoner WBP 185; both are in the nominative). The same inconsistency is characteristic of its use in possessive pronouns her, hir, our. No distinction is made between plural and singular forms. It can be difficult, in some scripts, to distinguish this flourish from a final e: endure, othere El WBP 364, WBP 368.

### 4.8 Capitalization

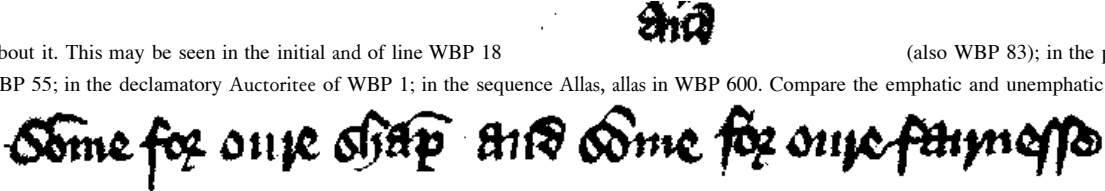
The system of emphasis used in the manuscripts is very different from any modern system of capitalization. The modern system consists of two elements: lower-case and upper-case letters. In the manuscripts there is a complex hierarchy of letters with different degrees of prominence. Emphatic forms of letters include capital letters of various sizes, differently emboldened letters, and ornamental capitals. In some manuscripts colour was used to give emphasis to particular letters.<sup>14</sup> This system worked together with paragraph marks, layout, and punctuation. In the transcription we use capital letters and also tags to represent bold-face and ornamental letters. This is a simplification of the system found in the manuscripts, as a result of which some information is lost. For example, we do not render different sizes of capital letters. In this discussion, we refer to emphatic and unemphatic forms rather than to upper-and lower-case forms.

Distinguishing between the emphatic and unemphatic forms of some of the letters can be difficult. The difference between them is often slight. Letters for which emphatic and unemphatic forms can be confused differ from manuscript to manuscript, most frequent being h, k, l, w, a, v. In these and other cases many scribes do not have distinct emphatic and unemphatic letter forms. Sometimes what distinguishes a letter at the start of a line from the same letter in the middle is just some degree of emdoldening, or a harder press of the pen. Often, this is not visible in the microfilms from which we must work.

In various instances emphatic and unemphatic forms are distinguished only by height and not by form. This distinction by height (as in a/A, s/S in Hg) is not absolute given the shifting letter heights and wavering baselines of certain scribes, especially that of Hg. Rather, the form must be determined as emphatic or unemphatic by its height relative to the letters immediately about it. For example, in Hg the upper bowl of the characteristic double-story a in the unemphatic form ascends just above the base single-line height of the letters about it, while that of the emphatic form reaches to or almost to the base double-line height of

the letters about it. This may be seen in the initial and of line WBP 18

Abraham WBP 55; in the declamatory Auctoritee of WBP 1; in the sequence Allas, allas in WBP 600. Compare the emphatic and unemphatic forms of s in Hg

WBP 258: 

WBP 258:

(also WBP 83); in the proper name

and in Hg

# sey yoff sooth, by seint Thomas

WBP 644:

clearly made and in cases of ambiguity the transcription is guided by the scribe's usual practice (e.g. and not And within the line El WBP 582).

In some cases this distinction by relative height is not

The distinction by height is further confused in cases where scribes allow more size to an initial where there is more space in the manuscript, as when it is the first line of a page (e.g. Dd). In these cases, we again are guided by the scribe's usual practice. Thus, we transcribe the large h in Dd WBP 157, in the first line of folio 69r, as h, not H, as elsewhere the emphatic and unemphatic forms are not distinguished.

Sometimes the first letter of a word can be different in shape from the same letter inside the word. Both of them can be unemphatic forms. Thus the scribe of La uses two forms of initial m: simple as in mariage WBP 173 and a more ornate form as in mi WBP 175, WBP 169. Both were transcribed as unemphatic. In this manuscript one also finds special initial forms of b (WBP 234, WBP 372) and n (WBP 299) together with ordinary ones.

Lower-case j occurs only occasionally, for example in Roman numerals: iij. Upper-case J is very rare. Usually in words like "Jesus" (Ihus Hg WBP 15,) or "Jacob" (Iacob Hg WBP 56) the same letter form is used as for the first person pronoun. In such manuscripts it was universally transcribed as I. A rare case of distinguishing between I and J is found in Se (thus Jhūs WBP 15, Jobis WBP 436, Jerome WBP 652; cf. Ierusalem WBP 495, Iouynian WBP 653).

Double f was transcribed ff whenever it occurred in the manuscripts.

Some scribes (e.g. Hg) clearly intend to use the emphatic form always at line beginnings, but this intention is obscured by the lack of distinct upper-case forms. In the face of this uncertainty, consistency and accuracy are very difficult to achieve. We discriminate in our transcription between emphatic forms at line beginnings and within the line. Where the scribe's practice shows that he uses separate upper-case forms at the line beginnings for all letters which have such distinct forms, then we elect to transcribe as emphatic all first letters of lines, including those letters for which the scribe has no distinct emphatic form. Note the transcription of the initial Y of Yet as emphatic in the specimen transcript on p. 47. Part of our reasoning was that this would allow transcribers to concentrate on discriminating emphatic forms within the line. In part, this is a practical decision. Scholars investigating the distribution of emphatic and unemphatic forms in the manuscripts should be aware of the arbitrary nature of our choices in particular contexts.

We observe the following concerning capitalization in individual manuscripts:

- Cp: emphatic and unemphatic forms difficult to distinguish include w/W, h/H (but cf. WBP 802-1 in our numbering), l/L (but WBP 624, WBP 726), v/V, y/Y, þ/ƿ, (e.g. WBP 381). A/a are distinguished only by size.
- Ha<sup>4</sup>: forms difficult to distinguish include w/W, h/H (but cf. WBP 802-1 in our numbering), l/L (but L WBP 624, WBP 726; dubious are WBP 724, WBP 726, WBP 731), v/V, y/Y, þ/ƿ (e.g. WBP 381). The closeness of the practice of Cp and Ha<sup>4</sup> supports the argument that the two manuscripts are written by the one scribe (Doyle and Parkes 1978, Robinson 1993: 30-3; cf. Ramsey 1982, 1986).
- El: forms difficult to distinguish include h/H (e.g. he seith WBP 51; he, haue WBP 335), v/V (e.g. vp on, vp WBP 25-6), l/L (e.g. Lameth WBP 54; Lat WBP 143, WBP 476, WBP 501, Lookynge WBP 624 (cf. lake WBP 269), s/S (e.g. statut WBP 198, seint WBP 483, sit WBP 687), d/D (e.g. WBP 320), a/A (e.g. Age WBP 474), k/K (e.g. keep WBP 795), y/Y: distinguished usually by size relative to other letters in the line (e.g. Ythonked WBP 5, Yet WBP 24).
- Hg: forms difficult to distinguish include w/W (e.g. were WBP 2; weddyng, WBP 11), y/Y (e.g. yet WBP 24, dyuyne WBP 26, queynte fantasye WBP 516), v/V (e.g. virginitee WBP 72; venus WBP 594<sup>15</sup> and ff), h/H (e.g. housbondes, had WBP 6; hercules WBP 699), a/A (e.g. and, alle WBP 8, samaritan WBP 22), k/K (e.g. keep WBP 795, kisse, kneled WBP 776, WBP 777), l/L (e.g. lo, Salomon WBP 35, lameth WBP 54, lat WBP 501, Lat WBP 143, Lo WBP 807, Loo WBP 809).
- Fi: the scribe uses two forms of a so it is uncertain in some cases whether to transcribe a double compartment a as emphatic A or to regard it as a different form of unemphatic a: WBP 1, WBP 58, WBP 59, WBP 64, WBP 113-5, bygañ WBP 140. Initial a is usually double compartment in all contexts.
- Ra<sup>3</sup>: The scribe often employs an emphatic form of a within the line, especially for the article (cf. WBP 133, also A wif Allas WBP 166). This form of a is also used inside the line where one may expect a capital letter: Apostill WBP 160.
- 

## 4.9 Punctuation

The transcription will record different marks of punctuation such as punctus, punctus elevatus, virgule, parenthesis, comma. Certain features of layout, such as the use of paragraph marks, ornamental and bold-face capitals, interlinear and other additions, deletions, alterations in the text, blank space left inside the text, underlined words and phrases, will be rendered through the system of tags.

The greatest problem with transcribing punctuation is that very often it cannot be clearly seen on the photographs. The ink and date of the mark can be impossible to assess, and sometimes one cannot be sure that any mark is present. However, we have chosen to transcribe the punctuation that can be seen distinctly in our materials.

We observe the following concerning punctuation in individual manuscripts:

- Ht: virgules are used occasionally. They are not joined to the preceding letters including those with horizontal strokes. In Dd virgules are also common as mid-line punctuation.

- Ch is notable for its use of parentheses e.g. (olde dotard' shrewe) WBP 291; apparently an early instance of their use—the manuscript is dated c. 1450. The scribe also employs punctus, virgules, and commas inside verses and at their ends.
- In En<sup>1</sup> virgules are common as mid-line punctuation. The punctus, punctus elevatus, and virgule also occur at the ends of verses. It is often very difficult to distinguish between tails at the ends of words and virgules (see above). It is easier in the case of d' where the tail is always a curved s-shaped form. On the whole virgules in this manuscript are always thin straight lines and tails are bolder curved lines (cf. tails after lif and wif in WBP 157-8 and a virgule in WBP 159).
- In En<sup>3</sup> virgules occur as mid-line punctuation. The scribe uses a large number of virgules, in many lines after every word. They can be both drawn separately or joined to the preceding letter. They were transcribed up to WBP 134. After this neither virgules nor tails were transcribed.

#### 4.10 Other manuscript features

In the course of our transcription, we seek also to record certain features of manuscript presentation (use of ornamental capitals or other emphasis, etc.), of scribal activity (additions, deletions, underlinings, and the like) and of transcriber annotation (marking of text as illegible, or uncertainly read, etc.) Recurrent phenomena are captured with a tagging system modelled on that used by the established international standard, ISO 8879, of the Standard Generalized Markup Language (SGML). A major international initiative, the Text Encoding Initiative (TEI), is at the time of writing drafting a set of extensions to SGML to facilitate its scholarly use.<sup>16</sup> These will include recommendations for transcription of primary sources and the tagging system adopted for these manuscripts has been designed for compatibility with the TEI draft recommendations, as they existed at November 1993. In time, we expect these transcripts to be translated into SGML/TEI format. The design of the tags here proposed should make this possible without any loss of information. Occasional phenomena in the manuscripts (bracketing of lines, damage, manicules, etc.) are recorded in transcriber annotations, within braces marking them off from the text transcribed.

The tags used in this transcription are:

- [exp]...[/exp]: expansion of an abbreviation (e.g. [exp]ur[/exp]—expansion of abbreviated ur). This is used only for abbreviations which cannot be represented by one of the characters defined in the font.
- [sup]...[/sup]: superscript. Used only for superscripts not in the font.
- [orncp]...[/orncp]: ornamental capital, without further specification of size, style, etc.
- [b]...[/b]: bold-face.
- [emph]...[/emph]: indicates emphasis for letter or word indicated other than by ornamental capital, or emboldening, or underlining. Typically, this is by a two-line (or deeper) capital at the beginning of a line.
- [add]...[/add]: scribal addition, without specification of hand, place of addition, etc.
- [ud]...[/ud]: underdotted by scribe.
- [del]...[/del]: scribal deletion, other than by underdotted, without further specification of hand, manner of deletion, etc.
- [unr]xxx[/unr]: unreadable, for whatever reason (physical damage to manuscript, etc.) The number of xs indicates the number of letters which cannot be read.
- [dub]...[/dub]: the transcriber is not certain of the tagged reading.
- [ul]...[/ul]: indicates underlining in the manuscript.
- [sp]xxx[/sp]: indicates “white space” (for a letter, or word) left in the manuscript. The transcriber encloses within the tag an x for each letter for which there appears to be space; thus “[sp]xxx[/sp]” indicates space left for three letters.

We do not transcribe glosses or other marginalia. However, their presence in the manuscript is noted by the transcriber, thus “{gloss beside line 43}”.

The registration of these manuscript features is capable of infinite elaboration. We particularly welcome advice on refinement and extension of this tagging scheme.

#### 5. Accuracy, consistency, richness

Perfect accuracy is, of course, the aim of every transcription. Over a task of this size, working at this level of detail with the resources we have available, we do not think perfect accuracy in every respect is attainable. Rather, we seek to reduce the effects of error firstly by repeated checking of transcripts, and secondly by defining in these guidelines just what scholars will find most reliable in our transcripts and what they might find less reliable.

Each transcript is checked at least three times, and we expect on the final check to find less than one correction for every four thousand characters. We expect too that on this final check none of the corrections will be “substantive”: that is, they might involve the correct spelling of a word (“her” or “hir”), but would not involve the presence or absence of the word itself: some form of “her” will be present. Scholars who wish to use our transcripts to check whether or not a given reading will appear in a given manuscript should then find these transcripts perfectly reliable. Some of the corrections will be graphemic, as “hir” for “her”; others will be graphic in the sense we have defined it above, as hir for hir. Our experience is that at this final checking stage most corrections are graphic, while the graphemic corrections which have to be made are mostly matters of capitalization. As we point out above, the capitalization of the manuscripts causes particular difficulty in transcription.

In general, the question of accuracy of transcription cannot be separated from its consistency and its richness. Concerning consistency: it is desirable that the same mark in the same context in different manuscripts be rendered both accurately but also consistently, by the same character in each case. With handwritten characters, it may be very difficult to determine that it is indeed the same character, and there may be subtle differences in context. The difficulties with final -

oun/-on discussed above illustrate these problems. In the case of final -oun/-on (as with the analogous problems concerning virgules and tails, and capitalization) we have established rules of transcription that ensure a measure of consistency at one level, with some loss of accurate graphic representation at another. Thus, the final two minims in a -on combination may appear to be clearly written as a u but will be transcribed as an n if there is no macron or flourish above, in obedience to the rule given above. In such circumstances, it is the inconsistency within the manuscripts themselves which causes difficulty. Scholars who wish to apply gross counting techniques to these transcripts should be aware of the interpretative element in our work, and should take account of this in framing their research.

Concerning the accuracy and the richness of these transcripts: the more detailed the transcription, the more possibility there is for error. We have no doubt that we have missed some tails and virgules which are there in the manuscripts, added some which are not in the manuscripts, and misplaced others. We could have eliminated all these categories of error simply by not transcribing tails and virgules. At a stroke, this would permit us to claim a considerably higher accuracy rate. However, we felt that it is better to give a transcription rich in detail, at the cost of some accuracy of rendering that detail, than a transcription which achieves perfection through impoverishment. This bears too on our use of microfilm: these characters do not show up well in microfilm and we are pressing microfilm to the limit in our attempts to discern them. As a partial remedy, we will be checking our transcripts against all the manuscripts readily available to us: at least, those in Oxford, London, and Cambridge. We trust that in time resources will permit checking of this detail against all the manuscripts or (more conveniently) against higher-quality images than are now available to us. In the meantime, the classes of characters for which the possibility of error in our transcription is relatively high will be clear from the preceding discussion.

A few transcripts have been taken through the cycle of a minimum three checks proposed for all transcripts; most transcripts have at the time of writing been checked only once; some manuscripts have not been transcribed at all. Changes in this system of transcription are now relatively easy to implement, but will become progressively more difficult as we proceed. We invite and welcome comments on these guidelines.

## 6. An example of our transcription

This annotated transcript of lines WBP 366-71 of *The Wife of Bath's Prologue*, from Fi, gives a flavour of the decisions we take as we transcribe:

Yet p<sup>̇</sup>chest þ<sup>u</sup> and seist þat a hatefull wyfe  
 Y rekened ys for one of these myschaunx  
 Ben theȝ none othur man<sup>2</sup> resemblauncx  
 That ye may lykken your pables to  
 But yf a sely wyfe be one of tho  
 Thou lykenest woma<sup>n</sup>s love to heff

Line 1:

- Yet: although the scribe does not have distinct emphatic and unemphatic forms of y, distinguishing these (if at all) only by size, we here transcribe it as the emphatic form, and hence the upper-case letter Y, because the scribe's normal practice is clearly to use the emphatic form at line beginnings.
- p<sup>̇</sup>chest: from the context, the mark over the p must be the superscript hook character which stands for re/er, and is transcribed as such.
- þ<sup>u</sup>: from the context, this must be the superscript u and is so transcribed.

Line 2:

- *myschaunsv*: the scribe has written the final two marks in the word (*s*, followed by the characteristic plural abbreviation, as in *resemblaunsv* in the next line) over one another. These are separated in the transcript.

Line 3:

- *Beñ*: the final two minims appear to be a *u*, but from the context must be a *n*: they are so transcribed. The flourish is almost certainly otiose, but is transcribed as elsewhere it is used over *u/n* to signal abbreviation.
- *theṛ*: the mark over the *r* is transcribed as a flourish, as too with *youṛ* in the next line). In one or both these places, this mark might represent abbreviation of final *-e* (as is suggested by the similar graph for the *re-* in *resemblaunsv* in this line) or nothing at all.
- *man<sup>2</sup>*: note the virtual identity of the final two minims and the mark over them with that in *beñ* at the beginning of the line, or of *lykken* in the next line. However, from the context this must be superscript hook standing for *-er*, not the flourish. It is therefore transcribed as the superscript hook.

Line 4:

- *pables*: the final character appears graphically identical with *3*. But from the context it must be simply an unusual graphetic form of *s* and it is therefore transcribed as *s*. This is the only occurrence of this form of *s* found in this manuscript; the transcriber noted its presence in an annotations file.

Line 6:

- *womañs*: the macron extends over all three letters; the rule in our transcription is to transcribe it only over one. We choose to transcribe it over the *n*. The scribe's practice elsewhere is clearly to place marks over *n* in final position (cf. *beñ*, *lykkeñ*; also *womañ* outside this passage) and the intention was probably therefore to mark the *n* of *womans*.

## 7. Notes

1. Versions of this paper were read to the Early Book Society conference in Sheffield in July 1993 and to the Medieval English graduate seminar in Oxford in November 1993, and we have profited from discussion with those present on those occasions. For comments on written drafts of this paper we are grateful to Ian Doyle, Hoyt N. Duggan, Anne Hudson, Malcolm Parkes, Kari Anne Rand Schmidt, Eric Stanley, Ronald Waldron, and Laura Wright. Naturally, the responsibility for the principles here stated remains with the authors.
2. Hereafter, when we refer to the "manuscripts" of *The Wife of Bath's Prologue* and the *Canterbury Tales* we mean both the manuscripts and pre-1500 printed editions. There are fifty-five surviving manuscripts of *The Wife of Bath's Prologue* and four pre-1500 printed editions.
3. Transcribers who worked on the first cycle of transcription included: James McCabe, Hubert Stadler, Cathy Swires, Diana Wyatt, Paul and Maureen Watry, and the authors. Considerable help in devising the system of transcription for this first cycle was given by Laura Wright.
4. The completed transcripts of *The Wife of Bath's Prologue* are published on this CD-ROM.
5. Many scholars, working in widely separated fields, have recently stressed the importance of the physical expression of primary texts. For medieval texts, see the discussion of manuscript *mise-en-page* by Parkes (1976) and the account of the changes in manuscript presentation in the context of the development of research tools in the 12th and 13th centuries by Rouse and Rouse (1991, 191-258); for modern texts see the work of Jerome McGann on "bibliographic codes" (e.g. his 1991 article "What is Critical Editing"), McKenzie's 1986 Panizzi lectures, and the summary discussion by Greetham (1992, 291-4).
6. See Robinson 1993, 30-3.
7. Recent studies which make use of analysis at the graphemic level in Middle English manuscript work include Ronberg 1983 (the two scribes of the *Destruction of Troy* in MS Hunterian V.2.8 are discriminated, and the cursive hand of the manuscript identified with that of Gower's *Confessio Amantis* in MS Chetham 6696, written by Thomas Chetham) and 1985 (two manuscripts of the *Wars of Alexander* are linguistically distinct, despite being so close in their readings as to be copies of a single exemplar, with MS Trinity College Dublin 213 showing evidence of late and Southern provenance); Jack 1991 (the language of the texts in MS Royal 17 A. xxvii represents a single variety of early Middle English, despite there being five different texts and three different scribes). Notable studies outside Middle English include Schipper 1986 (the spellings of Old English prose texts in CUL MS li.1.33 show evidence of the dialect of one of the scribes); McClure 1984 (comparison of the spellings of the Quarto text of Lyndsay's *Ane Satyre of the Thrie Estaatis* with the Bannatyne manuscript suggests a degree of anglicization); Emiliano 1991 (different spelling systems occurring side by side in notarial Latin texts from Spain constitute an "orthographic palimpsest" and reflect the development of a vernacular orthography).
8. The transcription planned for the manuscripts of the *Book of the Duchess* by Murray McGillivray proposes to record certain graphetic distinctions: thus the "three distinct forms of the letter *s*" (1993, 12-13).
9. The parenthesis and comma signs appear only in relatively late manuscripts, for example Ch. The unusual "wedge" punctuation mark, found in the Peterhouse *Equatorie of the Planetis* (Schmidt 1993, 111) and in Hg (Doyle and Parkes 1979, xxviii) does not occur in any of the manuscripts of *The Wife of Bath's Prologue*.
10. Cf. the description of these characters in Benskin 1990 and 1982.
11. Exception will be made for some rare abbreviations which will be expanded and enclosed in tags signalling the presence of a brevisgraph in the manuscript. Certain superscript letters not represented by a separate character in the computer font will be indicated by a superscript tag: thus [sup]d[/sup] or [sup]th[/sup] which sometimes occur with Roman numerals. For lists of abbreviations in common medieval use see Capelli (1990) and Bischoff (1990).
12. For the same conclusion, see the brief discussion of the ligatured *-ll-* in Chaucer manuscripts by Evans (1977).
13. In Ruggiers' transcript of Hg the spelling *Londoū* is used in l. WBP 550.
14. Yellow paint is used to give emphasis to the letters in Cp. In general, because we have to do most of the transcription from microfilm it was not possible to render the use of colour in the transcription because it can not be seen in the microfilms.
15. *V* in *venus* was transcribed as lower-case by Ruggiers.
16. For a description of SGML see Goldfarb (1990, 5-17). The first draft of the TEI guidelines (known as "P1") was published by McQueen and Burnard in 1991; the second draft ("P2") was issued in part-form, with publication of the final draft of this development cycle ("P3") in 1994. One of the authors (PR) served as a member of the "Manuscripts" workgroup of the TEI, dealing specifically with transcription encoding, and as head of the related "Textual Criticism" workgroup. He was primarily responsible for the drafts of these sections for the P2 and P3 publication of the TEI guidelines. Robinson 1994 gives an introduction to the TEI guidelines, as they



apply to transcription of primary sources.

## 8. Bibliography

- Benskin, Michael. "The Letters þ and y in Later Middle English, and some Related Matters." *Journal of the Society of Archivists* 7 (1982): 13-30.
- Benskin, Michael. "The Hands of the Kildare Poems Manuscript." *Irish University Review* 20 (1990): 163-93.
- Bischoff, B. *Latin Palaeography: Antiquity and the Middle Ages*. Tr. D. ÓCróinín and D. Ganz. Cambridge: Cambridge University Press, 1990.
- Bliss, A. J. "Notes on the Auchinleck Manuscript." *Speculum* 26 (1951): 652-8.
- Burnley, D. "Inflection in Chaucer's Adjectives." *Neuphilologische Mitteilungen* 83 (1982): 169-77.
- Capelli, A. *Dizionario di Abbreviature latine ed italiane*. Sixth Edition. Milan: Ulrico Hoepli, 1990.
- Doyle, A. I., and M. B. Parkes. "The Production of Copies of the *Canterbury Tales* and the *Confessio Amantis* in the Early Fifteenth Century." In *Medieval Scribes, Manuscripts, and Libraries: Essays Presented to N. R. Ker*. Ed. M. B. Parkes and A. G. Watson. London: Scolar Press, 1978. 163-210.
- Doyle, A. I. and M. B. Parkes. "Paleographical Introduction." *The Canterbury Tales: A Facsimile and Transcription of the Hengwrt Manuscript, with Variations from the Ellesmere Manuscript*. Ed. Paul Ruggiers. Norman, OK: University of Oklahoma Press, 1979. xix-xlix.
- Emiliano, A. "Latin or Romance? Graphemic Variation and Scripto-Linguistic Change in Medieval Spain." In *Latin and the Romance Languages in the Early Middle Ages*, ed. R. Wright, 233-47. London: Routledge, 1991. 233-47.
- Evans, R. O. 1977. "Whan that Aprill(e)?" *Notes & Queries* cii (1977): 234-7.
- Fix, H. "Production and Usage of a Machine-Readable Manuscript: A Report on the Saarbrücken Version of *Grágás Konungsbók*." In *Computer Applications to Medieval Studies*, ed. A. Gilmour-Bryson. Studies in Medieval Culture XVII (1984). Kalamazoo: Western Michigan University.
- Goldfarb, C. F. *The SGML Handbook*. Oxford: Clarendon Press, 1990.
- Greetham, D. C. 1992. *Textual Scholarship: An Introduction*. New York and London: Garland.
- Jack, G. "The Language of the Early Middle English Texts in MS Royal 17 A. xxvii." *Studia Neophilologica* 63 (1991): 129-42.
- Laing, M. (Ed.) *Middle English Dialectology: Essays on Some Principles and Problems*. Aberdeen: University of Aberdeen Press, 1989.
- McClure, J. D. "A Comparison of the Bannatyne MS and the Quarto Texts of Lyndsay's *Ane Satyre of the Thrie Estaitis*." *Scottish Studies* (Frankfort am Main) 4 (1984): 409-22.
- McGann, J. J. "What is Critical Editing?" *Text* 5 (1991): 15-30.
- McGillivray, M. "Electronic Representation of Chaucer Manuscripts: Possibilities and Limitations." In Ian Lancashire, ed. *Computer-based Chaucer Studies*. CCH Working Papers 3. Toronto: University of Toronto Press, 1993. 1-16.
- McIntosh, A. "Towards an Inventory of Middle English Scribes." *Neuphilologische Mitteilungen* 75 (1974): 602-24. Reprinted in Laing, M. (Ed.) *Middle English Dialectology: Essays on Some Principles and Problems*. Aberdeen: University of Aberdeen Press, 1989. 46-63.
- McIntosh, A. "Scribal Profiles from Middle English Texts." *Neuphilologische Mitteilungen* 76 (1975): 218-35. Reprinted in Laing, M. (Ed.) *Middle English Dialectology: Essays on Some Principles and Problems*. Aberdeen: University of Aberdeen Press, 1989. 32-45.
- McIntosh, Angus, M. L. Samuels, and Michael Benskin., eds. *A Linguistic Atlas of Late Mediaeval English*. 4 vols. Aberdeen: Aberdeen Univ. Press, 1986.
- McKenzie, D. F. *Bibliography and the Sociology of Texts*. The Panizzi Lectures. London: The British Library, 1986.
- Neuman, M. 1991. "The Very Pulse of the Machine: Three Trends Toward Improvement in Electronic Versions of Humanities Texts." *Computers and the Humanities* 25 (1991): 363-75.
- Parkes, M. B. *English Cursive Book Hands*. London: Scolar, 1969.
- Parkes, M. B. "The Influence of the Concepts of Ordinatio and Compilatio on the Development of the Book." In *Medieval Learning and Literature: Essays Presented to R. W. Hunt*, ed. J. J. G. Alexander and M. T. Gibson, 115-40. Oxford: Oxford University Press, 1976. 115-40.
- Ramsey, Roy Vance. "The Hengwrt and Ellesmere Manuscripts of the *Canterbury Tales*: Different Scribes." *Studies in Bibliography* 35 (1982): 133-154.
- Ramsey, Roy Vance. "Paleography and Scribes of Shared Training." *Studies in the Age of Chaucer* 8 (1986): 107-44.
- Robinson, P. M. W. "An Approach to the Manuscripts of The Wife of Bath's Prologue." In Ian Lancashire, ed. *Computer-based Chaucer Studies*. CCH Working Papers 3. Toronto: University of Toronto Press, 1993. 17-48.
- Robinson, P. M. W. *The Transcription of Primary Sources Using SGML*. Oxford: Office for Humanities Communication, 1994.
- Ronberg, G. "Two North-West Midland Manuscripts Revisited." *Neophilologus* 67 (1983): 463-7.
- Ronberg, G. "The Two Manuscripts of The Wars of Alexander: A Linguistic Comparison." *Neophilologus* 69 (1985): 604-10
- Rouse, M. A., and Rouse, R. H. *Authentic Witnesses: Approaches to Medieval Texts and Manuscripts*. Notre Dame: University of Notre Dame, 1991
- Ruggiers, Paul G., ed. *The Canterbury Tales: A Facsimile and Transcription of the Hengwrt Manuscript, with Variants from the Ellesmere Manuscript*. Norman, Okla.: University of Oklahoma Press, 1979.
- Samuels, M. L. "The Scribe of the Hengwrt and Ellesmere MSS." In *The English of Chaucer and his Contemporaries: Essays by M. L. Samuels and J. J. Smith*. Ed. J. J. Smith. Aberdeen: Aberdeen University Press, 1988. 38-50.
- Schipper, W. "Orthography and Dialect in Cambridge University Library MS li.1.33." *Studies in Medieval English Language and Literature* 1 (1986): 53-65.

- Schmidt, K. A. Rand. *The Authorship of the Equatorie of the Planetis*. Chaucer Studies XIX. Cambridge: D. S. Brewer, 1993
- Smith, Jeremy J. "Spelling and Tradition in Fifteenth-Century Copies of Gower's *Confessio Amantis*." In J. J. Smith, ed. *The English of Chaucer and his Contemporaries*. Aberdeen: Aberdeen University Press, 1988. 96-113.
- Sperberg-McQueen, Michael and Lou Burnard. "Guidelines for the Encoding and Interchange of Machine-Readable Texts." Text Encoding Initiative: Chicago and Oxford, 1994.
- Svinhufvud, A. C. *A Late Middle English Treatise on Horses, edited from British Library MS. Sloane 2584*. Stockholm Studies in English xlvii. Stockholm: Almqvist and Wicksell, 1978