

# **A LINGUISTIC ATLAS OF EARLY MIDDLE ENGLISH**

## **INTRODUCTION**

### **PART II: THE CORPUS**

#### **CHAPTER 3**

#### **OVERVIEW**

**Margaret Laing and Roger Lass**

### 3. Overview

#### 3.1. The sample

In chapter 1 we discussed a major problem facing any historical linguistic survey. The data are confined to the written texts that happen to survive. This ‘sample’ is not random but accidental. In the case of early Middle English, the contingent survival of text witnesses is very patchy both spatially and temporally, and in terms of length and of genre. Because of the paucity of data we felt the optimal procedure would be to use all of it, and to treat all examples of early Middle English as potentially equally important components of the corpus. Of course, text dictionaries derived from the scribal outputs representing such a wide range of text types will not provide an even coverage, or present a dialect continuum that entirely speaks for itself. In other words, the text dictionaries and maps cannot always be taken at face value; their assessment depends on an appreciation of the existence and nature of a large number of variables. §1.5.6, on scribal practice, points to the importance of individual textual studies in providing text-specific interpretative commentary.

There is no reason in principle why all the surviving early Middle English materials could not have been included in our corpus: that indeed was the initial intention when we first adopted the corpus-based approach. In the first few years, shorter texts, such as the seven surviving versions of *Poema Morale*, the two versions of *The Owl and the Nightingale*, and *The Bestiary* were transcribed in their entirety; but so were very considerable works, such as *Vices and Virtues* and all the Trinity and the Lambeth Homilies. The corpus methodology, as will become clear later in this chapter, is however extremely labour-intensive and time-consuming. We soon realised that if *LAEME* were ever to be complete enough for publication, some kind of restrictive sampling would have to be employed for the remaining longer texts. Thus, for instance, the five surviving early Middle English versions of *Ancrene Riwle/Wisse* and the two versions of Laðamon’s *Brut* have only been partially transcribed and tagged. With the three textually similar versions of *Ancrene Riwle* (Corpus (A), Cleopatra (C) and Nero (N)), we elected to transcribe and tag corresponding portions to make close comparison possible — a desideratum common to both dialectal and textual studies. In the first instance, parts 1 and 2 of the text for each of these versions have been included in the corpus (ca 15000 words each). The Gonville and Caius (G) text is a much shortened and reordered version, which does not include part 1 and has only bits of part 2. A sample of 8734 tagged words has so far been included to represent it.<sup>1</sup> The Titus version begins imperfectly, the first 13 folios of the manuscript being lost. Moreover, although it is throughout in the same hand (which also contributes versions of *Hali Meidhad*, *Sawles Warde*, *St Katherine* and *Êe Wohunge of ure Lauerd*), the scribe is a literatim copyist the language of whose contributions varies, reflecting different kinds of language in his exemplar(s). Although much of his output is in mixed language, for inclusion in the corpus we were able to isolate a layer of consistent homogeneous usage. It was the actual extent of this type of language in the scribe’s copy that defined our sample in this case (14085 tagged words). In the course of the analytical work necessary to isolate this

---

<sup>1</sup> Note that all word counts here refer to tagged words; this excludes elements such as names, embedded Latin, roman numerals, to which tags are not assigned and which do not feature in the text dictionaries. The actual word counts of these texts and text samples will therefore be somewhat higher than those given.

homogeneous extract (Laing and McIntosh 1995), it was necessary also to transcribe and tag another large sample from *Ancrene Riwe* as well as all the other texts written by the T scribe. Of these, only *Wohunge* proved to be in homogeneous enough language to be included on the maps, but the other texts still form part of the corpus as a whole.<sup>2</sup>

Laðamon A includes the work of two scribes whose contributions are transcribed and tagged as separate text witnesses (see §3.2 (v) below). Scribe A's contribution is short enough to have been used in its entirety (13092 tagged words). Scribe B, however, wrote four times as much: a sample of his contribution of comparable length is used in the corpus (12578 tagged words).

It will be clear that in individual cases the nature of the scribal language can sometimes influence or even dictate the sampling policy. In these cases, references to the relevant explicatory articles may be found in the bibliography attached to that particular text in the Index of Sources. Otherwise, the general principle of transcribing short texts as wholes and of using large samples of longer texts has continued to be followed where possible. Nevertheless, time has not been on our side, and some important texts have still not been transcribed and tagged at the time of writing, e.g. the version of *Cursor Mundi* found in Göttingen University Library. MS Theol. 107r containing two different kinds of language, and three early 14th-century versions of *The South English Legendary* that would be of great interest to compare with the samples of the two versions that have been tagged.<sup>3</sup> Moreover, because of the ever-increasing time pressure, other important texts are represented by smaller samples than originally intended. For instance, the tagged sample for the *Ayenbite of Inwyt*, processed some years ago, is 30562 words, while the recently undertaken tagging of the *Ormulum* has resulted in a sample of only 11342 words. For any text included in the Index of Sources, there will be details of whether or not its language(s) are suitable for dialect mapping, and if so whether they have been tagged for inclusion, and if not whether it is intended that they should be in due course. Fortunately, because of the web format we have chosen, *LAEME* can continue to be a work in progress, and many of these texts may well be added in the future, while samples of included texts may be expanded.

Though we have used the word 'sample', it is important to note that we do not do so in a statistical sense. The universe of *LAEME*, because of historical contingency, is not a statistically well-formed object. What 'randomness' there is in the existing data is due to accidents of survival rather than sampling procedures. In addition, since we do not at this point use all of the surviving data, our corpus is what statisticians would call a 'judgement sample'. That is, a sample in which the purposes of the investigators take precedence over any procedural imperatives.

### 3.2 Text types

It is vital for linguistic study that each scribal contribution to a manuscript is treated separately as an independent witness. If a literatim copyist has copied more than one

---

<sup>2</sup> Texts that turn out to be non-mappable for whatever reason, are still included in the corpus. This is because the corpus functions not only as a quarry for maps but also as a historical linguistic corpus available for many other types of investigation.

<sup>3</sup> The *South English Legendary* texts that have been sampled are those in Oxford, Bodleian Library, Laud Misc 108 and Cambridge, Corpus Christi College 145. Those that have not yet been sampled are British Library, Egerton 2891 and Harley 2277 and Oxford, Bodleian Library, Ashmole 43. These however were all analysed and mapped in *LALME*.

exemplar language, then each of these languages also constitutes an independent witness. In other words it is in principle possible, and in practice happens, that a single scribe can be a witness for more than one survey point: e.g. the two kinds of language to be found in a single hand in the Cotton version of *The Owl and the Nightingale* have been mapped in different places in Worcestershire. In the corpus of tagged texts, each language type is assigned its own index number (the two languages in the Cotton *O&N* are #2 and #3 in the corpus). Where subsequent assessment leads us to the conclusion that contributions from different scribes are in fact linguistically so similar as to make them regionally indistinguishable, then their outputs may be mapped at the same survey point, but they are still kept separate: e.g. the three hands contributing to British Library, Royal 17.A.xxvii mapped in the same place in SE Salop as ##260–262.

Where a single scribe writes a number of different texts, two possible paths have been followed. Where there seems to be no linguistic complexity, the scribal contribution is tagged as a whole and assigned a single text number. Where there is reason to suppose that elements of a scribe's output may have to be treated separately, or where his contribution to any one text is interrupted by the work of another scribe, each text is in the first instance tagged separately and assigned a different number. If on further scrutiny, it turns out that the entire contribution of a single scribe is linguistically homogeneous, then the original single text numbers are still retained, but all the texts are then treated for processing and mapping as a single long text. A superordinate four-figure number is then assigned to it — e.g. the outputs of Scribes A and B of the Trinity Homilies, who alternate with each other in copying the first thirty-three homilies and whose contributions are #1200 and #1300 respectively.<sup>4</sup>

We have already spoken in chapter 1 (§1.5.3) about the few useful examples of documentary anchor texts. The corpus includes copied Old English documentary material of varying lengths from Bury St Edmunds, Suffolk (though its language does not apparently in fact originate from there (#1400)); Benet Holme, Norfolk (#131); Beverley, East Riding of Yorks (#230); Chertsey, Surrey (#184); Coventry, Warwicks (#126); Crediton, Devon (##147–148); Hereford, Herefords (#259); Ramsey, Hunts (##133–135); Sherborne, Dorset (#279); Thorney, Isle of Ely, Cambs (##184–185); Wells, Somerset (##156–157); Winchester, Hants (#143); the Proclamation of Henry III placed in Westminster (two versions #11 and #12); Gospatric's Writ, placed south west of Carlisle in Cumbria (#132). In addition there is the second continuation of the Peterborough Chronicle, which, though not strictly a local document in the usual sense, deals with events specific to the locality and has been placed in Peterborough.

The literary manuscripts comprise a number of different types and vary greatly in length. They fall into the following main categories:

(a) Texts transcribed and tagged in their entirety

(i) single short (i.e. fewer than 500 words) or fragmentary texts (usually lyrics or parts of lyrics) found in manuscripts with local associations but whose other contents are not in English: e.g. a fragment of *Stella Maris* found on fol. 3r of Oxford, Bodleian Library, Rawlinson C 510, associated with Bardney, Lincs (# 130); a version in

---

<sup>4</sup> Note however that Scribe A is a literatim copyist who also copied a version of the *Poema Morale* in the same manuscript (Cambridge, Trinity College B.14.54) but in a somewhat different form of language: it is #5.

English of *Stabat iuxta crucem Christi*, on p. 175 of Oxford, Bodleian Library, Tanner 169\*, associated with St Werburgh's Abbey Chester (# 124).

Such very short texts would not normally be chosen as sources for a linguistic survey done by questionnaire. The vast preponderance of null entries for questionnaire items would render their contribution nearly valueless to the continuum displayed on a series of maps. But for early Middle English such texts often provide the only data for their area of origin. Their contribution to the overall picture of course remains small, but all the available data in English for each particular scribe is recorded in the tagged text and text dictionary, and this small something is better than nothing at all.

(ii) one or more short texts (i.e. fewer than 500 words — usually lyrics) in manuscripts with no local associations. Unless these are found in groups by the same hand, so that their forms can be amalgamated as a single scribal assemblage, these are usually very difficult to localise because there is not enough linguistic material to go on. Many of these texts have nevertheless been included because their contribution is of inherent interest, and their forms may still be compared as the usage of some single witness with those of other scribal witnesses. Some have been included because their texts survive in more than one version: e.g. the nine different texts of the quatrain *Candet Nudatum Pectus* (##13–19, 127, 292). Sometimes a number of different hands each contribute such short texts in a manuscript. Three different hands (one of which writes a version of *Candet Nudatum Pectus* just mentioned) are represented on fol. 9v of Linz, Stiftsbibliothek Sankt Florian XI.57 (## 292–294). Three hands contribute four different varieties of English in Oxford, Digby 2 (## 178–181). In addition to the two scribes providing versions of *Poema Morale* in BL Egerton 613 (## 6 and 7) four further hands contribute short lyrics (## 234–237).

(iii) small scribal contributions to larger texts in a different hand. Sometimes there is clearly a 'main' scribe of a text and a small portion only is copied by a different scribe. One such is Scribe C of the Trinity Homilies who provides only the last homily (# 63) of the thirty-four.

Sometimes a long text will be corrected, expanded or annotated by one or more later scribes. Any contributions that belong to periods later than early Middle English are merely noted in the tagged text of the main scribe. But when extra contributions are in early Middle English they may be tagged separately as scribal witnesses in their own right. Perhaps the most important of these is Scribe B of the copy of *Ancrene Riwe* in BL Cotton Cleopatra C.vi. He makes a number of additions and corrections (# 275) to Scribe A's text (# 273) that for the most part match the readings in the revised version of the text in Cambridge, Corpus Christi College 402 (# 272). Dobson (1972: xciii ff) has convincingly argued that Scribe B of Cotton Cleopatra C.vi is the author of *Ancrene Riwe/Wisse*. Further early Middle English is provided by a somewhat later corrector working on the manuscript in Canonsleigh between 1285 and 1289 (Dobson: 1972: xxv–xxix and cxl ff), who also provided English in Cambridge, Trinity College B.1.45. His scribal contributions have been amalgamated into a single tagged text in the corpus (# 1700). Its language has been localised to North-West Norfolk.

*Vices and Virtues* in BL Stowe 34 was written by two main scribes (## 64 and 65). But a number of contemporary scribes appear to have worked on the text. Of the various correcting hands, one contributes considerably more than the others, and his work has been transcribed and tagged separately (# 302). The section titles were

added after the copying of the main text by another scribe. This scribe was responsible for all but the last two titles, which appear to be in yet a different hand. The main title scribe's work has also been tagged separately (# 303).

(iv) small to medium-sized texts (i.e. more than 500 and fewer than 10000 words) or medium-sized contributions to larger texts. This category includes well known early Middle English texts like the *Interludium de Clerico et Puella* (# 159), the *Proverbs of Alfred* (in Maidstone Museum MS A.13, # 66) the *Bestiary* (# 150) *Jacob and Iosep* (# 158) and the so-called Wooing Group (# 1800), written by scribe B of BL Cotton Nero A. xiv. It also includes the tagged texts formed by the amalgamation of all the different verses written by each of the four main hands of English in Cambridge, Trinity College B.14.39 (## 246–249).

Sometimes medium-sized contributions to a manuscript will be transcribed and tagged in their entirety while longer stretches by more major contributors are only sampled. Thus in BL Royal 17.A.xxvii, containing the Katherine Group, the entire contributions of Scribes B and C (## 261 and 262 comprising 6863 and 5585 tagged words respectively) have been included in the corpus, while for Scribe A (the main scribe) a sample of 13876 tagged words (# 260) includes his section of *Sawles Warde* and his copy of *St Katherine* but omits his copy of *St Margaret*.

(v) Long texts (i.e. more than 10000 words) that have been done completely because of their importance or because of interpretative complexities. These include the Trinity and Lambeth Homilies and *Vices and Virtues* mentioned in §3.1 above. All the Middle English texts in Oxford Bodleian Library, Digby 86 have been transcribed and tagged. All but four were found to be in the same more or less homogeneous language and their tagged texts were amalgamated as # 2002 to give a text sample of over 15000 tagged words. The other four texts were found to be in mixed language, but they remain part of the corpus of tagged texts (## 214, 218, 220, 222). All of the work of the scribe of the Cotton version of *The Owl and the Nightingale* has been tagged. Because he was a literatim copyist, each of his texts has been processed separately (## 238–244) and the two different kinds of language in his text of *O&N* have been mapped in two different locations (## 2 and 3). The whole of *Havelok* (# 285) has been transcribed and tagged (16665 tagged words), both because of its importance as a text in its own right, and because its language belongs in the relatively poorly represented East Midlands.

The portion of *Cursor Mundi* preserved in the Edinburgh, Royal College of Physicians MS is an important witness because of the dearth of northern texts in early Middle English. This manuscript is in three hands. Scribes A and C copy non-continuous and misordered pieces of *Cursor Mundi*. Scribe B copies part of the *Northern Homily Cycle*. The work of all three scribes has been transcribed and tagged in its entirety (## 296–298: 15015, 21811, 13731 tagged words respectively).

(b) Texts not transcribed and tagged in their entirety.

These comprise long texts that do not seem to present linguistic complexity and that we have therefore sampled rather than tagged completely. For instance, Oxford, Jesus College 29 contains most of the same texts as are found in the hand of the Cotton *O&N* scribe, and these have all been transcribed and tagged from the Jesus manuscript too for comparative purposes. But the Jesus scribe is a translator (see Chapter 1, §1.4) and all his texts are in the same homogeneous language, so are treated as a single text

witness (# 1100). He copied a number of other texts, but because his language does not vary very much it was not thought necessary to transcribe them all. *The Poema Morale* was included for comparison with other versions and Thomas de Hales *Love Ron* also forms part of the 18199-word sample.

It can be seen that there is no strict cut-off for sample length even for long texts. We take into account textual content and context when choosing where to begin and end a sample. We also take into account comparability with other versions of the same text. The two versions of the very lengthy *South English Legendary* so far sampled (see §3.1 n. 4 above) in detail contain different material; but some overlap has been assured by including four of the same saints' lives in each sample.<sup>5</sup>

### 3.3. Editorial practice

#### 3.3.1 Use of 'originals' rather than editions<sup>6</sup>

Our primary evidence for medieval language is manuscript texts. In compiling our corpus of early Middle English texts for tagging, we transcribe from originals or (more often) from photographic reproductions, and not from editions. Printed editions can be useful reference tools. They may help in interpretation of manuscript readings, while checking against the texts of editions can help detect possible errors in our own transcriptions. But for any investigation of historical language variation it is crucial for reasons of comparability and authenticity that, where possible, the manuscript be used as the primary source:<sup>7</sup>

... editorial practice varies considerably and for rigorous comparison all corpus texts must be treated consistently. While some editors present a more or less diplomatic version of a text it is often the case that the original is modified in a number of ways, any of which may render it suspect for linguistic study (Laing and Lass 2006: 426).

From the point of view of linguistic study these are the problems that render use of editions deeply problematical (cf. further Laing 2001: 87–91):



(a) Many if not most editors silently expand manuscript abbreviations, taking as the form of the expansion the scribe's 'usual' unabbreviated spelling. If a scribe has more than one 'full' spelling for a word, e.g. *after* and *aftir*, silently expanding examples of *aft<sup>a</sup>* as one or the other may seriously skew numbers as well as suppressing a valid distinct spelling (cf. Lass 2004: 35 n. 14).

(b) Some editors make a virtue of 'normalising' or 'modernising' texts to create easier reading versions for students. Scholars with primary linguistic interests must eschew such bowdlerisations. Very few serious editors nowadays will change <ƿ> and <ð> to <th>, or <ð> to <y> and <gh>. But it is rare to find one who does not substitute <w> for <w>, even though these *litterae* have an entirely different history, and the subtle

<sup>5</sup> The sample for Oxford, Bodleian Library, Laud Misc 108, hand A is 32085 words, that for Cambridge, Corpus Christi College 145 is 29738 words.

<sup>6</sup> For a recent detailed discussion of the problems inherent in treating editions as if they were historical witnesses see Lass (2004).

<sup>7</sup> Printed editions are therefore not used as the source of a transcription unless the original manuscript is lost or inaccessible and there is good reason to believe that the edition is accurate.

and intricate interchange of their use with <u>, <v>, <uu> and <vv> in both consonantal and vocalic functions is a fascinating part of the story of early Middle English (Benskin 1982: 19–20, Laing 1999: 255–260).

(c) Equally suspect for our purposes are conflate editions compiled from numerous different scribal witnesses with the aim of producing some imaginary ‘best text’ that never existed in any time or place.

(d) An editor will often emend a form that he believes to be erroneous to one that he thinks the scribe (or original author) ‘intended’. Of course scribes did make mistakes and some such emendations would probably have been approved by the errant scribe himself. But we cannot *know* this, and we must not suppose that a scribe ‘really meant’ anything that he did not in fact produce. Moreover, some emendations are themselves erroneous and turn out to have removed a form that is a genuine part of the record. For some examples arising out of detailed work on manuscripts for *LAEME*, see Laing (1998, 2001 and forthcoming a).

(e) It might be argued that the editorial conventions practised in (a)–(d) are only harmful if one is interested in the detail of manuscript orthography; historical syntax (and perhaps regionally conditioned syntax) will be unaffected by editorial interference. Most editors of medieval texts, however, add modern punctuation and suppress such manuscript punctuation as exists. Manuscript word division is frequently ‘regularised’ along modern lines. This enables medieval texts to be subjected to the same types of syntactic analysis as modern ones and all too easily allows the assumption that medieval scribes had attitudes towards word, phrase and clause structure similar to our own. The use of diplomatic transcriptions from originals can challenge such assumptions.

(f) As we pointed out above (§3.2), it is vital for linguistic study that each individual scribal contribution be treated separately, and in our corpus each text language is indexed and sorted individually. Conscientious compilers of single text editions will notice any changes of hand in their manuscript. But scholars trawling an edition for linguistic evidence may not always succeed in maintaining the distinction. Even if these broad distinctions are maintained in a printed edition, such care does not always extend to scribal corrections in the text. These may be interlinear, intralinear or marginal insertions. If they are made by the scribe who wrote the main text, whether as he went along or as a separate exercise later, the silent inclusion of the changes may not be too damaging. But it is often difficult, even when dealing with originals (and harder still with black and white photographic reproductions), to be sure whose hand has made a correction. As long as the fact of a correction is noted, the reader is alerted to a possibly extraneous element in the scribal system. There may also be deletions, made by subpunction, erasure, underlining, crossing through or obliteration. The original text may still be wholly or partly legible. The ‘mistake’ may be a truly erroneous and unintended spelling. But sometimes it is the result of misplacing a word, or a longer piece of text, that is otherwise a perfectly good example of the scribe’s usage. Such deletions are usually ignored by editors but we retain them, suitably marked, in the tagged text for possible linguistic analysis (see §3.5.4.1 below).



### 3.3.2 Photographic copies of manuscripts

Most of our ‘originals’ are in fact in the form of photocopies of short texts and microfilms of long ones. Using images rather than the real thing is of course important for manuscript conservation. But it is also much more convenient to have access to the source text at all times. Transcribing and tagging proceed far more quickly when researcher, source, computer and reference books are all in the same room. Instant access to photographs also makes it possible to compare hands from different manuscripts as well as to check and recheck readings. Moreover, once the initial outlay has been made on the reproduction, it is also a much cheaper method of study than travelling to libraries.

It might be assumed that photographic reproductions would be less readable than original manuscripts. This is sometimes true, but by no means always. In fact, a good photograph or microfilm can often be clearer than an original. Parchment is frequently discoloured, stained or blemished and ink is sometimes faded. A reproduction can in some cases create better differentiation between the writing and the background. Using photographs does, however, have a number of disadvantages:

(a) Notwithstanding the observation above, readability is sometimes compromised by not having the original: a picture is taken from one angle only and sometimes text is clearer when lit from a different direction. Moreover, simply being able to make out the words of a text is not the only reason we might want to see the texture of the original materials. Photographs may sometimes clarify what text is actually there, but at the same time they can, for instance, obscure the roughness of an erasure and make it impossible to tell whether or not a particular piece of text is an overwrite.

(b) Most of the reproductions we use are not in colour. The most obvious disadvantage here is that no accurate commentary can be made about use of pigment in the manuscripts of our texts, e.g. use and/or alternation of colour on initial capitals. Coloured ink may turn out the same intensity as ‘normal’ black or brown ink in the text, in which case it is impossible to differentiate it. In some photographs, however, it may appear much fainter in a black and white reproduction than black or brown ink, so that rubricated or other coloured text (often used for embedded Latin) can be difficult to read. Even more serious for our purposes is that lack of colour makes it well nigh impossible to judge whether two black and white images with the same darkness are in fact of the same hue. Change in ink colour can help to signal a change in hand or a different stint by the same hand. It can also draw attention to correction, whether by the same or another scribe.

(c) Even with original manuscripts, some text may be irretrievably lost. Later binders often trimmed the edges of manuscripts and any marginal titles, rubrics, annotations, corrections, or actual text too near the edge could be lost. Bindings are often very tight and this may prevent full opening of the codex, which again causes loss of materials in or near the central gutter. This last problem is likely to be much worse in a photographic reproduction because the photograph misses the gutter altogether or leaves it in shadow.

(d) When manuscripts themselves have suffered damage from damp or rubbing, a photographic reproduction will often exaggerate the effects.

(e) Not all photographs are of top quality. While some may be darker than the original (whether this makes them clearer or the reverse) others may be fainter. This can lead to loss of hairline strokes or small punctuation marks. Some photographs (in certain cases the only ones obtainable) may be out of focus and the text fuzzy.

With these caveats in mind, what do we actually do when transcribing material for the corpus?

### 3.3.3 Diplomatic transcription

Our transcription policy may be described as ‘diplomatic’. In palaeographical terms it should perhaps be referred to as ‘semi-diplomatic’, since abbreviations are in most cases expanded traditionally, though the expansions are always differentiated as such (see further §3.4.5.1 below).

We transcribe at the level of *littera* rather than *figura*. That is, we interpret all varieties of <r>, of whatever shape — Caroline short <r>, 2-shaped <r>, Anglicana long <r> etc. — as ‘r’. Similarly long <s>, short <s>, beaver-tailed ‘Mumpsimus’ <s>, etc. are interpreted as ‘s’. There is one special exception: following the *LALME* praxis, in writing systems that employ the same <y>-shaped *figura* for ‘ȝ’ and ‘y’, we transcribe any <y>-shape as an occurrence of the *littera* ‘y’. This is because there is known regional significance in the employment of this figural merger. In some northern and North Midland varieties <y> for ‘ȝ’ is the norm. Where there is a cline of shapes distinguishable at either end as ‘ȝ’ or ‘y’, but not distinguishable in the middle, we follow *LALME* practice and transcribe the whole range as <y>. In a number of early Middle English writing systems, however, there are distinct *figurae* for ‘ȝ’ and ‘y’, but the functions may still be confused. These are rarer in late Middle English, but do occur, and they fall into Benskin’s (1982: 24) *ȝ + y* category. In these cases (as is the practice in *LALME*) we transcribe at the level of *figura* — <ȝ> for <ȝ> and <y> for <y> whatever the function.

In some early Middle English scripts, there may also be close resemblance in the *figurae* for ‘ȝ’ and ‘w’. In most cases so far examined, there is not complete equation of the two *figurae*, rather there is a cline of shapes formally distinguishable at each end but not in the middle (e.g. London, British Library, Cotton Caligula A ix, *Owl and the Nightingale*). Given the formal distinguishability of the end points, and the fact that there is apparently no regional significance in the distribution of <ȝ>/<w> confusion, we have elected to transcribe all occurrences as one *littera* or the other according to etymology. In this case we are doing no different from interpreting the often identical patterns of two minim strokes as examples of either ‘n’ or ‘u’. In the few cases where the *figurae* for <ȝ>/<w> are always the same shape (e.g. Cambridge, Gonville and Caius 234/120, *Ancrene Riwle*, and Cambridge, Trinity College B.14.39 (323), Hand A), we have not made an exception, but also transcribe according to etymology. This decision is evidently not ideal, but it has been made for the sake of typological consistency and because of the absence of a wider tradition of transcribing confused <ȝ>/<w> usage as either <ȝ> or <w>.

Two other *litterae* whose *figurae* often merge in early Middle English scripts are ‘c’ and ‘t’. The first element of both letters is formed identically in many scripts, and the head stroke may or may not be angled down for <c> or cross the first element for <t>. In some hands, the two *litterae* are formed identically, and (as with ‘ȝ’ and ‘w’) there may be a cline of <t>- and <c>-shapes. In these cases we usually follow

the same practice as for ‘þ’ and ‘w’ (and for ‘n’ and ‘u’) and differentiate the two *litterae* by linguistic function. In some writing systems, however, the functions themselves may be ambiguous, e.g. *-ict* or *-itt* in words with OE *-iht*. In such cases we consider all aspects of the script and spelling system before making a decision.

Medieval writing differentiates the *litterae* ‘v’ and ‘u’, the *figurae* <v> and <u> and the *potestas* [v] and [u]. There was a tradition of using the *figura* <v> word initially and the *figura* <u> word internally whatever the intended *potestas*. However, not all scribes followed this tradition and it was not always adhered to consistently even by those that did adopt it. The mapping of the ‘u’/‘v’ *litterae* onto the <u>/<v> *figurae* may be complex and the story overlaps with that of ‘w’ (Laing forthcoming b):

The origin of ‘w’ in English lies in its use in Anglo-Latin (Benskin 1982: 19–20 and 2001: 211–12). In Old English, the *potestas* [w] was normally realised by wynn — <w>. The Anglo-Latin equivalents of runic wynn were <u> and <uu>, regularly adopted in Latin texts (written in Caroline minuscule script) for the writing of English names containing the element [w]. In the post-Conquest minuscule scripts, the angular version of the *littera*, the *figura* <v> (originating from the square capital script), was adopted as the capital form. It was then increasingly used as the preferred word-initial *figura* (and apparently always for the Roman numeral ‘five’) whether it was intended as a capital or not. <vv> ligatured produces <w>, which in English writing came to be used as the direct equivalent of runic wynn in [w] contexts. But the common litteral origins of <u> and <v>, their doubled forms <uu> and <vv>, and doubled ligatured <v> serving as ‘w’ lead to complex overlapping usages in some early Middle English writing systems, with these *figurae* (and, where used, also <w>) being commutable for the *potestas* [w], [v], [u] and [wu].

With <u> and <v> we therefore take the opposite line from our treatment of figurally identical ‘n’ and ‘u’ and transcribe figurally rather than literally, differentiating strictly by manuscript letter-shape rather than by function.

### 3.4 Internal format

#### 3.4.1 Transcription of manuscript plain text

All transcription and tagging has been done without using anything other than ASCII characters. Over the years, this has made it easier to accommodate updating of mainframe computer systems and transfer of data between co-workers and other scholars. It has also made it simpler to update bespoke programs.

Transcriptions are made using upper case for ‘plain text’ manuscript letters. Thus manuscript *herte* is transcribed HERTE. ‘Capital letters’ in the manuscript are preceded in the transcription by \*. Thus manuscript *Herte* is transcribed \*HERTE. ‘Capital letters’ may be of different kinds. For most *litterae* early Middle English scripts have distinct minuscule and capital (majuscule) *figurae*. This differentiation is thus similar to the difference between upper case and lower case in modern classic printing fonts and most modern handwriting. In medieval scripts, the distinction is often achieved by the addition of a vertical stroke to the basic *figura*, which is also sometimes enlarged. A variation of the single *figura* plus vertical stroke is the ‘doubling’ of ‘f’, and less commonly ‘s’, to indicate a capital.<sup>8</sup> Where a ‘double ‘f’ appears at the beginning of a word it is always transcribed \*F. With ‘s’, it is,

<sup>8</sup> The doubling is, in the opinion of Michael Benskin (pers. comm.) probably a reinterpretation of the single *figura* plus vertical stroke. With <s> the ‘doubling’ is usually (though not always) of the <s>-longa type.

however, possible (in text languages where <ss> stands for [ʃ]) for a double *littera* to appear word initially for phonological rather than graphic reasons. In such cases the double ‘s’ is transcribed SS rather than \*S. Colour is sometimes added to the majuscule *figura* to make a so-called *littera notabilior*. Some *litterae* in early Middle English scripts have only the one basic shape for both minuscule and capital, viz ‘u/v’, ‘w’ (when functioning as a separate *littera*), ‘y’, ‘ð’ and the runic letters ‘þ’ and ‘w’. In these cases, the only way the capital can be distinguished is by size or the use of colour. In the absence of enlargement and when working from black and white reproductions it is often therefore difficult to tell whether a capital is intended for these letters. We have done our best to differentiate but may not always have succeeded.

Capital ‘I’ is a special case. Minuscule ‘i’ in all scripts is made up of the basic minim stroke which lies between the baseline and the so-called headline of the manuscript line whether this be visibly ruled or not. This *figura* is transcribed according to our usual practice as I. Some scribes may distinguish the <i>-*figura* from other minim strokes by means of an oblique hairline stroke or dot (similar to the modern printed dot on ‘i’ and ‘j’). This further specification of the *figura* is especially useful when other *litterae* made up of minim strokes, such as ‘m’, ‘n’ and ‘u’ are immediately next to ‘i’. This practice is rarely systematically employed by any scribe, but when it is we normally treat the stroke or dot as part of the *figura* and do not notice it separately in the transcription.<sup>9</sup> The same goes for dots or strokes on thorn, ‘y’ or wynn. Manuscript capital ‘I’ is bigger than minuscule ‘i’, almost always having an ascender rising above the head-line and sometimes also a descender below the base-line. Its approach stroke is sometimes hooked or looped. It is also sometimes given further differentiation by the use of a punctus (see below §3.4.3) either before and after it or on one side only. The extra differentiation is often, though not always, used to signal the use of the *littera* as the first person pronoun rather than, say, the preposition IN. We transcribe all these variants as \*I, taking any added punctus as part of the *figura* rather than transcribing it as punctuation. The so-called ‘i’-longa, is identical to minuscule ‘i’ apart from having a descender below the baseline, often with a leftward curve or hook. This is the ancestor of modern printed ‘j’. In early Middle English it most often occurs for ‘i’ as the second element of a vowel *littera* cluster, especially when double ‘i’ is used, e.g. HIJ for THEY or LIJF for LIFE. In this its use is like that for the final digit of a Roman numeral, e.g. xvij. It occasionally also appears for simplex [i(:)] and for [dʒ] or [j]. We transcribe it J.

### 3.4.2 Non-Roman letters

We reserve lower case letters in transcriptions for three different functions: the expansion of abbreviations (see below §3.4.5.1), diacritics (see below §3.4.9) and transcription of non-Roman letters:

|                     |                                       |
|---------------------|---------------------------------------|
| y = ‘thorn’ <þ>     | <i>þus</i> is transcribed yUS         |
| d = ‘edh’ <ð>       | <i>seið</i> is transcribed SEId       |
| ae = ‘æsc’ <æ>      | <i>æfter</i> is transcribed aeFTER    |
| z = ‘yogh’ <ð>      | <i>niðt</i> is transcribed NIZT       |
| w = ‘wynn’ <w>      | <i>wiþoute</i> is transcribed wIyOUTE |
| g = insular ‘g’ <g> | <i>geu</i> is transcribed gEU         |

<sup>9</sup> For exceptions to this general rule see §3.4.9 Diacritics, below.

Note that we differentiate yogh and insular ‘g’. The first is a figural development from the second and became perceived as a different *littera* from ‘g’ as a result of a post-Conquest realignment of litteral and potestatic mappings. Because these changes are in progress during early Middle English we have elected to transcribe figurally rather than literally in this case (see further Laing forthcoming b). A special convention has been adopted to deal with Orm’s three different <g>-*figurae*. His insular ‘g’ (<g>) is transcribed, as usual, as g: e.g. *georne* EAGERLY gEORNE, *dagg* DAY DAgg. His peculiar flat-topped ‘g’ that combines insular ‘g’ and Caroline ‘g’ is transcribed as G, because in comparative studies it needs to be seen as equivalent to other writers’ ‘usual’ <g>-shapes: e.g. *ḡoddspell* GOSPEL GODDSPELL, *kinḡess* KINGS KINGESS. The third version of ‘g’ distinguished by Orm, ordinary Caroline ‘g’, is used (usually doubled) for [dʒ]. In the text sample from the *Ormulum* (and nowhere else) it is transcribed as G3; e.g. *seggen*<sup>n</sup> SAY SEG3G3EN^N. (For superscripts see §3.4.7. below.). Orm is not the only scribe who employs distinctions between different <g> *figurae*. The scribe of London, British Library, Arundel 292 who writes *The Bestiary* and before it some religious verses in a similar language, uses two types of <g> with distinct functions. Both have a single lobe and a leftward curving tail as in a normal Caroline ‘g’. One has the usual off-stroke either in final position or linking it with a following *littera*; this ‘hooked’ <g> stands for [g] and the rare occurrences of [dʒ] and, being thus comparable with ‘g’ in other early Middle English writing systems, is transcribed G. The other lacks the off-stroke or ‘hook’ and stands for [j], [ç~x] and [y], i.e. those sounds that in many other early Middle English writing systems are represented by <ō> or <g>. See Gumbert and Vermeer (1971) and cf. Wirtjes (1991: x). Gumbert and Vermeer refer to the hookless <g> as ‘an unusual yogh’ but its shape is nothing like yogh, being identical to this scribe’s normal <g> but simply lacking the offstroke. In the tagged texts from Arundel 292 it is realised as G2, e.g. *ḡurg* THROUGH dURG2, *negge* MAY COME NIGH NEG2G2E.

A special use of lowercase occurs in the transcripts from three scribes: Scribe B of Laðamon A (# 278), the scribe of the Caius version of *Ancrene Riwe* (# 276) and Scribe C of Linz, Stiftsbibliothek Sankt Florian XI.57, fol. 9v (# 294). The script employed by these scribes has ligaturing of <c> and <t>. The form of ligature used is very similar to the commonly used ligature between <s> and <t> — a croquet hoop shaped stroke linking the top of <c> to the stem of <t>. Our normal transcription practice is to ignore ligatures, and such linked *figurae* are simply transcribed as the sequences CT or ST. With the above three scribes, however, it is difficult to separate the sequences <cc>, <ct> and <tt> because the same ligature is used for all three. Rather than normalising according to ‘expected’ spellings we have decided to draw attention to the potential ambiguity (cf. the argument about simplex <c> and <t> above) and transcribe the ligatured sequence as cT wherever it occurs, e.g. LEcTE (expected <tt>) for past 3rd singular indicative of OE *lāetan*, and REcTHE (expected <cc> for 1st person singular present indicative of OE *reccan*). The use of lower case c is so as to distinguish this purely orthographic usage from ‘genuine’ CT = [kt] and from CT = ?[xt] in hands where OE *-ht* words have <ct> as a variant. In these cases we transcribe CT.

### 3.4.3 Treatment of manuscript lacunae and partial *figurae*

Sometimes damage to the manuscript by damp, fire, worms, wear, cropping or other disaster renders a reading impossible or uncertain. Such cases are marked in the

transcription with []. Where whole words are missing this is noted and commented on in the transcription (see further §3.4.6 below). Where a word is partially visible, but so illegible as to make it impossible or injudicious to assign to it a tag, it is flagged in such a way as to cause the tagging program to print what has been possible to transcribe but to skip over it for tagging (see further §3.4.10.2 below). In some cases *figurae* may be only partly decipherable, but sufficiently legible to be able safely to deduce the reading. In these cases the partly visible *figura* is placed within brackets: e.g. RI[w]LE. The brackets are left empty if no part of the affected *figura(e)* is visible or where what is visible could be interpreted more than one way: e.g. S[]G+ES SCOURGES, []LE OWL. If missing words can be deduced or supplied from a reliable edition they are included in a comment. For missing words (whether originally written by the scribe or simply omitted by him), that are conjecturally supplied by the editor or transcriber see §3.5.5.

#### 3.4.4 Special symbols

In medieval Latin texts there are a number of special symbols for commonly used words or expressions e.g. † for *vel* OR, ‡ for *id est* THAT IS, as well as & or & for *et* AND. In early Middle English the only special symbols commonly used are the Tironian sign (&) and more rarely ampersand (&) for AND. In our transcriptions we do not expand these symbols because (a) we would have to make a sometimes arbitrary choice as to whether to expand as e.g. *and* or *ant* or *ond* or *an*; and (b) there is a suitable ASCII symbol that can be employed for the purpose. We use & for the Tironian symbol and &2 for the (much less commonly used) ampersand proper. In some writing systems the Tironian sign is used as a morphograph for the sequence *an(d)*-. We retain & in these instances also: e.g. &LONG for ALONG (< OE *andlang*), &-OyER for ANOTHER.



#### 3.4.5 Signs of abbreviation

##### 3.4.5.1 Abbreviations that are expanded

Abbreviation is far less common in early Middle English writing than it is in Latin. Nevertheless a number of signs of abbreviation were taken over from the practice of Latin writing into the writing of English. Most such abbreviations are conventionally expanded in the tagged texts, but are signalled as abbreviations by being in lower case rather than in upper case.

The bar or titulus over the preceding vowel that indicates a missing ‘m’ or ‘n’ is expanded according to context: e.g. *hī* HIM is expanded HIm, *sūne* SUN as SUnNE.

Bars are also occasionally used over other letters, in Latin loanwords in early Middle English texts, to imply different expansions. In these cases the bar implies the same expansion as it would if used in Latin writing. These abbreviations are for the most part expanded conventionally: e.g. Latin *q̄o* for *que* is sometimes taken over into early Middle English as a segment in a longer word — so *q̄ome* PLEASE is transcribed. QueME. See further §3.4.5.2 below.

The abbreviation sign for <er>/<re>, whether it is shaped as <sup>a</sup> or as ®, is similarly expanded conventionally according to context. So *eft<sup>a</sup>* AFTER is transcribed EFTer, *lau®d* LORD is transcribed LAUerD, *th®e* THREE is expanded THreE. For the expansion of hooks implying other letters, see §3.4.5.2 below.

The abbreviation sign for ‘ur’, whether looped or 2-shaped, is also conventionally expanded, e.g. *at<sup>o</sup>n* ATTIRE is transcribed ATurN, *bett€* BETTER is transcribed BETTur.

In Latin writing <sup>o</sup> can stand for *con-*, *com-* or *cum-* according to context. In early Middle English the use of the abbreviation is uncommon and is limited to Latin and French loans: *°mune* COMMON is expanded as comMUNE, *°fort* COMFORT (from AF *confort*) is expanded as conFORT<sup>10</sup>, *°ceiue* CONCEIVE is expanded as conCEIUE.

When <sup>o</sup> is raised above the baseline, as the abbreviation for ‘us’ (also uncommon in the corpus), it is so expanded: e.g. *v<sup>o</sup>* US is expanded as Vus.

The littera ‘p’ with a line through the descender is expanded conventionally as ‘ar’ or ‘er’ according to context: e.g. *pte* PART is transcribed ParTE, *pril* PERIL is transcribed PerIL

The littera ‘p’ with an extended and recurved lobe is expanded conventionally as <ro>: e.g. *¶cessiune* PROCESSION is transcribed ProCESSIUNE.

The abbreviation for noun plural is not common in early Middle English, but where it occurs it is always expanded ‘es’ not ‘is’ or ‘ys’: e.g. *cnich-ρ* KNIGHTS is transcribed CNICHes. Looped flourishes on final ‘g’ or ‘k’ are comparatively common and these are expanded conventionally as ‘e’ or ‘es’ depending on shape and context: e.g. *bok-ρ* BOOK is transcribed BOKE, *tokenyng-ρ* TOKENING is transcribed TOKENYNGe, *askyng-ρ* ASKINGS is transcribed ASKYNGes. Such expansions may serve wholly or in part as hived off suffixes.

Recurved final ‘r’ for ‘re’ is also not common in the early period, but where it appears it is transcribed Re.

See also §3.4.7 Superscripts and §3.4.8 Nomina sacra below.


### 3.4.5.2 Abbreviations that are left unexpanded

The traditional expansions of the signs of abbreviation come mainly from their use in Latin. For our purposes it is vital that an abbreviated spelling is separable from a form that is fully written out and that is why the expansions are given in lower case. Transcribing *eft<sup>a</sup>* as EFTer need not imply an underlying *efter* form in the mind of the scribe. Many signs of abbreviation had multiple possible expansions in Latin and this is true to a lesser extent for the various medieval vernaculars, as we saw in §3.4.5.1 above in relation to the common sign for missing ‘m’/‘n’ or that for missing ‘er’/‘re’. In spite of these built-in ambiguities, using the conventional expansions for most of

<sup>10</sup> Note that even though Anglo-French has instances of COMFORT with ‘m’, e.g. *comfort*, *cumfort*, in the *LAEME* corpus the initial abbreviation sign is always transcribed con for the sake of consistency.

the signs of abbreviation in the tagged texts seems to us to make for greater clarity and transparency than adopting a system of arbitrary code characters would do.

There are, however, some signs of abbreviation for which it seemed less misleading to signal the fact of the sign of abbreviation without selecting a preferred expansion, because of the great number and variety of different vernacular spellings potentially implied.

For instance, transcribing ~ THAT as yAT in a scribal system for which the commonest fully written variant is *ƿat* might seem reasonable, but more perverse in a system that prefers *ƿet*. Our realisation of ~ must, however, remain consistent across all the tagged texts because we wish to compare its usage beside other variants, whether spelled *ƿat*, *ƿad*, *ƿet*, *ƿæt* or *ƿ<sup>t</sup>*, etc. We treat ~ therefore as a logograph, implying no spell  other than an initial thorn and probably a final dental. We transcribe it y~.<sup>11</sup>

Much less common is a similar abbreviation for AND formed by <a> with a stroke or hook above it: we transcribe this as A~.

In Hands B and C of Edinburgh, Royal College of Physicians MS of *Cursor Mundi* (## 298 and 296 in the corpus) a hook similar to the abbreviation for ‘er’ is also used stem-finally apparently for ‘es’/‘is’: e.g. *ell<sup>a</sup>* ELSE, *wall<sup>a</sup>* WALLS. These are transcribed ELL~ and WALL~ respectively.

It can be seen that these non-literal realisations may also sometimes be found as hived-off endings and are thus representational in the corpus in the same way as they must have been for the scribe who used the non-literal stroke in his text.

We treat similarly the forms %o and q® used for the past tense of OE *cweƿan* SPEAK. These abbreviations are found for (among other things) *quod* in Latin texts and must have been adopted as a formal equivalent in scribal systems using *quod* for QUOTH, SAID, and then transferred as a logograph also into systems that normally spell the word differently. %o and q® QUOTH are transcribed Q~. Sometimes the ‘q’ and abbreviation sign are followed by the rest of the word written in full: e.g. %oat, q®ad are transcribed Q~AT, Q~AD. Sometimes a similar mode of abbreviation is transferred to the earlier native spelling: e.g. cw® QUOTH is transcribed Cw~.

Other Latin loan words may also be realised simply with the initial letter (or first two or three letters) with superimposed bar, a stroke through ascender or descender or an attached hook. These may be preceded and/or followed by a punctus, either as well as the abbreviations sign or instead of it. All such cases are treated as logographs and left unexpanded: e.g. *aḅḅ* ABBOT is transcribed ABB~, *.b®* BISHOP is transcribed .B~, *S.* and *S®*. SAINT are transcribed S. and S~. respectively, etc. Note that *S.* for SAINT survives as a logograph today. The punctus is a commonly used sign of abbreviation in Latin writing. Occasionally it is also adopted to abbreviate native words. In these cases as in the loan words noted above, the punctus is transcribed: e.g. *.M̄*. MAN is

<sup>11</sup> Note, however, the special case of ƿ® in the hand of Part II of British Library, Cotton Caligula A.ix, *The Owl and the Nightingale* etc., which in many of his texts (## 3, 238, 241, 242, 243, 244) is used for both THERE and THAT. Because the abbreviation sign is identical to the scribe’s <er> abbreviation in other words also, and because there are contexts where either THERE or THAT could be read, we have transcribed this as yer in all instances.



transcribed .\*Mn. Here the bar indicating the missing ‘n’ is expanded as normal, but the two punctus are retained to show that the word has been further abbreviated. On the punctus as a punctuation marker see §3.5.1.

Occasionally a scribe will adopt in his English text the Latin syllabic abbreviation for ‘-et’, which is usually yogh-shaped — <ō>. This is transcribed as yogh; that is, with lowercase ‘z’, when it is the same shape as the scribe’s usual yogh: e.g. *hauō* HATH is transcribed HAUz. The figural identity of the two functions within a single scribal system seems interesting enough to preserve. If the abbreviation sign is shaped like a semi-colon it is expanded to ‘et’: e.g. *fall*; FALL *imperative pl.* is transcribed FALLet.

See also §3.4.7 Superscripts below

### 3.4.6 Apparently otiose strokes

~ is also used to represent any apparently otiose stroke when it is separately made (i.e. lifting the pen), whether it is above a letter or through the ascender, or in the case of final ‘r’ or ‘k’, through the limb.

" is used to represent any apparently otiose stroke made without lifting the pen, such as an attached stroke looped back from the second minim of final ‘n’.

### 3.4.7 Superscripts

Scribal superscripts sometimes double as signs of contraction or suspension and sometimes simply imply the value of the superscript *littera*.

‘Full value’ superscripts are transcribed as normal letters but are preceded by ^ to indicate that they are placed above the base line and are usually smaller in size than normal. So †<sup>e</sup> THE and †<sup>u</sup> THOU, etc are transcribed y<sup>E</sup>, y<sup>U</sup> etc.

Some superscripts that signal contractions are not expanded but are treated as logographs: e.g. †<sup>t</sup> THAT and w<sup>t</sup> WITH are transcribed y<sup>T</sup> and W<sup>T</sup>.

Two uses of superscripts in Latin writing were commonly adopted in the writing of the vernacular:

(a) The first usage is in consonant clusters with ‘r’ where the ‘r’ is understood in the following superscript vowel. Thus *g<sup>a</sup>ce* GRACE, *g<sup>e</sup>de* SHOUT, *c<sup>i</sup>st* CHRIST, *f<sup>o</sup>* FROM, *p<sup>u</sup>de* PRIDE. In these cases the understood <r> is transcribed in lowercase and the superscript is as usual preceded by ^. So the forms cited above are transcribed Gr<sup>A</sup>CE, Gr<sup>E</sup>DE, Cr<sup>I</sup>ST, Fr<sup>O</sup> and Pr<sup>U</sup>DE respectively.<sup>12</sup> With superscript <i>,

<sup>12</sup> Note that with superscript <a> the superscript letter has often become highly stylised in shape and may not always match any of the <a> figurae employed by the scribe for his normal littera ‘a’. In some cases superscript <a> may not be recognisably ‘a’-shaped at all, sometimes being ‘u’-shaped or ‘cc’-shaped and sometimes being finished with horizontal stroke at the top. In late Middle English scripts the <a> superscript may be reduced to a horizontal squiggle. (For the history of these shapes see Johnson and Jenkinson 1915: 3–4.) All the types of superscript <a> evidenced in the corpus are subsumed in transcription under ^A.

sometimes the ‘r’ to be understood follows the vowel: e.g. *u<sup>t</sup>tue* VIRTUE, *f<sup>i</sup>ste* FIRST. These are transcribed U<sup>^</sup>IrTUE and F<sup>^</sup>IrSTE. Sometimes a scribe writes out the required <r> and also makes the vowel superscript: e.g. *fr<sup>o</sup>ggen* FROGS, *gr<sup>e</sup>t* GREAT, etc. For the sake of direct comparison with other ‘implied ‘r’ superscripts, these are transcribed as FRr<sup>^</sup>OGGEN, GRr<sup>^</sup>ET, etc.

(b) The second commonly found superscript usage is after ‘q’ where the following superscript vowel understands preceding ‘u’: e.g. *q<sup>e</sup>ne* QUEEN, *q<sup>a</sup>rtene* PRISON, *q<sup>i</sup>lk* WHICH, etc. These are transcribed as Qu<sup>^</sup>ENE, Qu<sup>^</sup>ARTENE, Qu<sup>^</sup>ILK, etc.

In the writing of Latin, there are numerous other cases where a superscript letter is used to stand for a longer string, e.g. *t<sup>i</sup>* for *tibi*, *m<sup>i</sup>* for *mihi*, *n<sup>c</sup>* for *nec*, *m<sup>o</sup>* for *modo* and many others. This practice is occasionally taken over for the writing of Latin loans in Middle English texts. In these cases transcription policy depends on individual word shapes. For instance, the word APOSTLES may be abbreviated in a number of different ways: *ap<sup>o</sup>les*, *ap<sup>l</sup>an*, *ap<sup>o</sup>*. These would be transcribed AP<sup>^</sup>OstLES, APostLAN and AP<sup>^</sup>O respectively. In the first two cases the superscript ‘o’ and the bar through ‘l’ are signs of contraction and the plural ending is transparent. The missing letters are clearly ‘st’ and ‘ost’ in each case and the expansion reflects this. In the form *ap<sup>o</sup>*, however, the superscript ‘o’ is a mark of suspension. The form may be used for the word APOSTLE in any number or case and the ‘correct’ expansion of the ending is opaque; the form is logographic and is therefore left unexpanded. For transcription and expansion policy of text in Latin (not tagged) see §3.5.1 below.

In the tagged texts there are some special uses of the ^ flag:

(a) in some scripts an <e>-*figura* may be attached to the *figura* of a preceding ‘d’ near the top of its back. It is most often a small 2-shaped *figura* formed by omitting the first element of the ‘e’ and joining the lobe and an extended horizontal hasta to the ‘d’. This has the effect of making the two-part *figura* look somewhat like an elongated figure 8. This practice may well have originated from the use of the ligature in Latin scripts as a form of the word *de* FROM, OF. Unlike in Latin and French, in early Middle English the segment *de* does not itself normally form a complete word. However, in the work of the hands that adopt it for writing English, the two-part *figura* may be used for the *de* segment within a word. In some writing systems it can occur in any position in the word; but its use is more often than not word final.

This two-part *figura* has been referred to, with reference to Continental scripts, as ‘the *de* nexus [nesso]’ (Ciarelli 1998), and as ‘the *de* monogram’ (Short 2005: I/16).<sup>13</sup> In the *LAEME* corpus it appears in only nine hands. There is one example in each of BL Additional 27909 (# 232, ca. 1300, proto-gothic non-cursive script), Oxford, Bodleian Library, Add. E. 6, hand B (# 161, last quarter of the 13<sup>th</sup> century, cursive early Anglicana script) and language 1 of the Lambeth Homilies (# 2000, written in a protogothic book hand of ca. 1200). There are two examples in the sample transcribed from BL Cotton Cleopatra C. vi, *Ancrene Riwle*, hand A (# 273, a protogothic book hand of the second quarter of the 13<sup>th</sup> century, with some elements of contemporary documentary script), and two also in the sample from *The Ormulum* (# 301). Orm was probably writing in the last quarter of the twelfth century. His script is idiosyncratic, heavy and compressed, with strong resemblances to Anglo-Saxon minuscule. There are a dozen examples in the work of Scribe A of the Trinity

<sup>13</sup> I owe these references to Philip Bennett. For a discussion of the *de* monogram in relation to certain Anglo-French texts see Bennett (forthc.),

Homilies (# 1200) and three times as many in the work of Scribe B (# 1300). Both hands are dated to late 12<sup>th</sup> century; B is a protogothic book hand while A is mixed, showing elements of protogothic and English Caroline minuscule. Hand A of BL Stowe 34, *Vices and Virtues* (# 64), has 100 examples of the *de* nexus. It also appears commonly in the work of the scribe of Oxford, Jesus 29 (# 1100).<sup>14</sup> It is transcribed as <sup>^E</sup>: e.g. *aq<sup>o</sup>ld<sup>e</sup>* KILLED is transcribed as AQu<sup>^OLD</sup><sup>^E</sup>, *bi-hynd<sup>e</sup>* BEHIND is transcribed as BI-HYND<sup>^E</sup>, *d<sup>e</sup>me* DEEM is transcribed as D<sup>^EME</sup>, *ƿrow<sup>e</sup>* PERIOD, TIME is transcribed as yROW<sup>^E</sup>. The examples given here are all from Jesus 29. It can be seen that the ligature is not confined to word final instances, nor is it here always used with preceding ‘d’, though the Jesus scribe is the only one that I have recorded to use this form of superscript with *litterae* other than ‘d’.<sup>15</sup>

(b) in some scripts a flourished final ‘s’, in the shape of a reversed question mark, is drawn out from (or drawn back into) the top of the preceding letter. This attached, raised version of ‘s’ is transcribed <sup>^S</sup>: e.g. *ri<sup>s</sup>* BRANCH RI<sup>^S</sup>

(c) in Orm’s special writing system, the famous double letters are sometimes made with the two *figurae* side by side. But sometimes the *figurae* are stacked one on top of the other. Doubled letters are transcribed as they appear in the manuscript, either as e.g. SS, when Orm writes the *figurae* side by side, or as e.g. R<sup>^R</sup> when he stacks them on top of each other. Some stacked *figurae* are merged into a single symbol. In the transcription, y<sup>^y</sup> and w<sup>^w</sup> represent the doubled vertically stacked thorn and wynn that appear on a single ascender.<sup>16</sup>

### 3.4.8 Nomina sacra

In Latin writing, the sacred names *dominus*, *deus*, *iesus*, *christus* and *spiritus sanctus* were not normally written out in full but were heavily abbreviated. In early Middle English this tradition was not followed with the names for God or the Holy Spirit. The native words GOD, LORD, FATHER, ALMIGHTY, HOLY and GHOST were written transparently. But the title Christ and the name Jesus are, in some hands, exceptions.

It was traditional with the word CHRIST to employ the Greek abbreviation  $\chi\rho\varsigma$  (chi, rho, sigma, for  $\chi\rho\iota\sigma\tau\omicron\varsigma$ ), which might come out in medieval scripts, using the Latin alphabet, as either *XPC* or *XPS* with or without a bar above. Sometimes the sigma is dispensed with (*XP*) sometimes both rho and sigma are missing (*X*). When these forms are adopted within text in early Middle English they are all transcribed as christ (i.e. all letters in lower case). Depending on the grammatical case of the word in context, it could in Latin appear as *Xe* (in the vocative), *Xm* (in the accusative), *Xi* (in

<sup>14</sup> I have recorded altogether 40 examples in the sample transcribed from Jesus 29.

<sup>15</sup> Though only in five instances, all in the *Owl and the Nightingale*. The Jesus scribe attaches the superscript <e> to the ascender of ‘b’ in *stubb<sup>e</sup>* fol. 159va, and to the ascender of ‘h’ in *clench<sup>e</sup>*, fol. 164vb, and *wrech<sup>e</sup>*, fol. 167va. He attaches the superscript <e> to the second long stroke of ‘w’ in *iknow<sup>e</sup>* and *ƿrow<sup>e</sup>*, fol. 159va. Ker (1963: xvi) refers to the Jesus scribe’s hand (which he dates to the second half of the 13<sup>th</sup> century) as “amateur”, admirably plain and simple, and, when possible spacious; not essentially different from a twelfth-century hand’. He goes on (Ker 1963: xvii): ‘To save space *e* is often attached to the top of the back of *d*’, and to refer to this practice as ‘a well-known device’, though he does not give any other references to it.

<sup>16</sup> Note that missing double ‘m’ or double ‘n’ are represented by Orm with stacked abbreviation bars not with the bars placed side by side. The expansion in these cases is simply doubled, e.g. HImm, not HIm<sup>^m</sup>.

the genitive) or Xo (dative or ablative). In these cases the final letter is very often superscript. The Latin genitive abbreviated form X<sup>i</sup> appears in the early Middle English of hand B of the Trinity Homilies for the anglicised word CHRIST in all grammatical contexts. It is there transcribed chr<sup>i</sup>st. When an abbreviated form is alphabetically mixed, e.g. *Xist*, the transcription will reflect this: chrIST.

The Greek abbreviation for JESUS was ιης (iota, eta, sigma, for ιησους), which tended to come out in medieval scripts, using the Latin alphabet, as *ihc* or *ih̄s*, usually with a bar above it, running through the ascender of the <h>-shaped *figura*. The abbreviation for the commonly used vocative, JESU (also used in English when addressing Jesus), was *ihu*, and for the Latin accusative *ih̄m*, both with a bar above. While chi and rho were not clearly transparent letters in the Roman alphabet, ‘i’, ‘h’ (despite its ultimate origin as eta) ‘s’ and ‘u’ were transparent. The <h> could apparently be reanalysed as ‘h’, in both Latin and English writing, as the word is sometimes written out in full including it: *ihesu(s)*. In the transcriptions I therefore expand as follows: barred *ih̄s* is transcribed IHesuS, barred *ihc* as IHesus, barred *ihu* as IHesU. As a personal name, JESU(S) is not assigned a tag and is not processed with the other linguistic data. For the spelling, marking and retrieval of place and personal names see §3.4.12 below.

### 3.4.9 Diacritics

In some early Middle English hands an oblique stroke may be added to a vowel. As intimated above (§3.4.1), such oblique strokes on ‘i’ and ‘y’ (as well as thorn and wynn) seem to have no special significance other than perhaps further to distinguish the *figura* from that of similarly shaped *litterae*. On these *litterae* the stroke or dot is not therefore normally separately noticed, other than in the two exceptional text languages mentioned below. Oblique strokes over other vowels are taken to be not integral to the *figurae*, and these accents, which are for the most part explicable as length markers, are indicated by a lower-case x following the vowel: e.g. *téne* TEN is transcribed as TExNE, *aróas* AROSE is transcribed as AROxAS. Sometimes the accent may have the extra function of differentiating a content word from a grammatical, unstressed word: e.g. *á* EVER, ALWAYS (transcribed Ax) as opposed to *a* A, indefinite article (transcribed A), *ṛé* THEE (transcribed yEx) as opposed to *ṛe* THE, definite article (transcribed yE).

Orm (# 301) has his own system of accents to indicate vowel length, involving single, double, and occasionally even triple oblique strokes over some long vowels. In the transcription these are realised by x, xx, and xxx respectively, always placed after the vowel; although in the manuscript the accents are often placed over the following consonant rather than over the vowel itself.<sup>17</sup> So *All-áne* ALONE is transcribed \*ALL-AxNE, *write* WRITE is transcribed wRIxTE, *ṛwertú't* COMPLETELY is transcribed ywERRTUxxT, *he't* HE IT is transcribed HExxT. A breve over a vowel, indicating shortness, in contexts where the vowel length would otherwise be ambiguous, is transcribed as a lower case ‘v’ following the vowel: e.g. *written* WRITTEN is transcribed wRIvTEN^N. It will be seen from the above examples that the normal policy of ignoring oblique strokes on ‘i’ is breached in the special case of Orm’s

<sup>17</sup> Note that the black-and-white microfilm from which the transcription was made does not show up different coloured inks. No attempt has been made therefore in the transcript to differentiate accents possibly added by others than Orm.

usage. There is one other scribal witness in the *LAEME* corpus that seems to have a detailed accent system on vowels: the scribe of fols. 64r–70v of BL Egerton 613, *Poema Morale* (# 6). His system is not as transparent as Orm's, but in the circumstances of his detailed use of oblique strokes on all vowels, I have elected in his case (the only other one apart from Orm) to transcribe all oblique strokes as x, even those on 'i'.

### 3.4.10 Flags

#### 3.4.10.1 Flags which control aspects of tagging

Within the transcriptions, a set of non-alphabetic characters has special significance for the operation of the tagging program (Williamson 1992/3, Laing 1994). For the most part, these flags are stripped out by the program in the process of tagging and do not appear in the resulting tagged texts. But two flags remain visible because they have morphological significance or because they demarcate elements of a compound word. These are + and -. + is used when there is no space in the manuscript between the elements that come before and after it. - indicates a space in the manuscript between the elements on either side.

+ is used in four ways:

(a) to flag an inflection (plural, genitive, verb ending, etc.: e.g. BOK+ES, SCHO+N, \*ABBOT+es (noun plural), SUSTR+ES, NADDR+E (noun genitive), GOD+E, MEONUR+^S (adjective plural), FIND+ES, HA+y (verb third singular present indicative), FALS+INDE, VLEOT+InGE (verb present participle). When the inflection is separately listed in a text dictionary, the + introduces the suffix: +ES, +N, +es, etc.

(b) to mark off a derivational affix in relation to the text-word stem or to another affix: e.g. BI+yURFE, BEARD+LEAS, wILL+FUL+NESSE. When the affix is separately listed in a text dictionary, the + follows a prefix (or first element of a compound suffix) and introduces the suffix: BI+, +LEAS, +FUL+, +NESSE.

(c) to divide elements of a compound: e.g. \*TWELF+MONyE, CHIRECHE+DURE, yER+TOgEINES. When the second element of the compound is separately listed in a text dictionary, the + introduces it: +MONyE, +DURE, +TOgEINES.

(d) to signal when two words that are normally separate in modern English have been run together as one in the manuscript. For the purposes of tagging we normally separate such cases into their constituent parts: e.g. manuscript *ðungemen* would appear as zUNG+E+ and +MEN, the trailing + in these instances indicating the joining of the two elements in the manuscript.

- is used to mark a manuscript space between two elements of a text-word where one or more of the elements is to be treated separately for tagging in addition to the whole text word. It is therefore used in same contexts as + is used in (a) – (c) above, except that - indicates that there is a space between the two linked elements in the manuscript. Although inflectional suffixes are rarely separated from their stems, it does occasionally happen: e.g. WIT-STAND-AND. Derivational affixes are

frequently separated visually, as are compounds: \*BI-CLUTE, CLEAN-NESSE, MON-SLACHT.

- is also used to mark a space between two elements of a text word where the combined elements are treated as inseparable: FOR-dI-dAT, IN-TO.

#### 3.4.10.2 Flags for specific elements

Not all elements of a text are to receive a lexico-grammatical tag. However non-tagable elements are still taken over into the tagged texts and some are marked for retrieval with a tag substitute.

' is used to mark personal names: e.g. '\*IHesU, 'ADAM, '\*DAUI. These are skipped by the tagging program but the forms are printed out in the tagged text preceded by ' and they come out in the tagged text as: '\_\*IHesU, '\_ ADAM, '\_\*DAUI, '\_AYLMer. When there is more than one separate element to a personal name the two elements are linked with a hyphen, e.g. '\_\*ROGer-\*BIGOD.

; is used to mark place names: e.g. ;BROMLEGE, ;\*NORTHFOLC, ;EDEN. These are skipped by the tagging program but the forms are printed out in the tagged text preceded by ; and they come out in the tagged text as: ;\_BROMLEGE, ;\_\*NORTHFOLC, ;\_EDEN.

It may be desirable to identify personal and place names by using modern equivalent names as tags. However, it seems best to treat name tagging as a separate task.

! is used for miscellaneous other elements that are not to receive a tag:

(a) Roman numerals: e.g. !.XIX., which comes out as !\_XIX. in the tagged text. These can be retrieved, if desired, for comparison with the native number names that are written out and do receive tags.

(b) other non-verbal indexing or formatting labels used by the scribe of the text: e.g. !\_1.A., !\_1.B., !\_2.A., !\_2.B. etc. used by Dan Michel in the *Ayenbite of Inwyt* (#291).

(c) illegible, semi-legible or partial readings that cannot be assigned a tag (see §3.4.3 above): e.g. (from *Ayenbite*) !\_UO[+L[] after which appears the textual comment {=Letters obscured by stain or blot. Morris (1866) supplies UO[RLET]=}. For the treatment of textual comments see §3.6 below.

#### 3.5 Further elements that are not tagged

A number of other elements are not subject to the tagging process. The transcriptions may contain comments or contextual information, e.g. folio references, indications of line ends, notice of insertions or deletions. Extra information of this kind is placed within braces in the transcription. Any material within {} is ignored by the tagging program but is preserved embedded in the resulting tagged text. Sometimes the notices of line ends, or insertions occur within a form that is to be assigned a tag. In these cases the indicatory flags do have to be included in the tagged element within the tagged text, but they are stripped out in subsequent sorting and analysis (see further §3.5.3 and §3.5.4.2 below).

### 3.5.1. Punctuation<sup>18</sup>

In the *LAEME* transcriptions no editorial punctuation is added. In early Middle English, and in verse texts especially, punctuation can be minimal, but where present it is normally preserved in the transcription (but see §3.7 below). Punctuation is, however, not subject to the tagging process, so in the transcriptions it is put within braces. Manuscript punctuation is recorded as follows:

- . or · = punctus, whether it appears on the baseline or is raised, is transcribed as {.}
- / = virgula is transcribed as {,}
- ≥ = punctus elevatus is transcribed as {.'}
- :
- ≤ = punctus interrogativus is transcribed as {?}
- ¶ or ♂ = any form of paraph, paragraphus or capitulum is transcribed as {para}
- † = any form of obelus, is transcribed as {obelus}

Orm (# 301) has some extra marks of punctuation not found in any of the other *LAEME* text witnesses:

- positura is used between sections and is transcribed as {;}
- a dash, used in the same way as an em or en dash in modern English to indicate a pause or parenthesis, is transcribed as {-}.

Note that manuscript hyphens, sometimes employed (whether single or double) at line ends to indicate that a word has been broken in the middle, are *not* transcribed (see further §3.5.2 below). This decision was made to avoid confusion with the hyphen used as a special transcription flag.

### 3.5.2 Line ends

\ is used to indicate the end of a line in the manuscript text. \\ is used to indicate the end of a text (e.g. a poem or a homily) when the corpus sample continues with more text(s) written in the same hand and language. When a word is broken between lines (and whether or not a hyphen is used by the scribe to indicate this) the \ is simply embedded in the word in the transcribed text: e.g. GIF+EN^N. This remains so in the tagged form of the text, but is stripped out in subsequent processing such as text dictionaries or in text placed on maps. Otherwise, \ and \\ are treated as comments and placed within braces: {\}, {\\\}.

### 3.5.3 Folio references

Manuscript column references and folio or page references are normally given exactly where they occur in the manuscript text and are placed between {~}: e.g. {~p89~} (where p = page), {~f13va~} (where f = folio, v = verso and a = first column), {~f53rb~} (where r = recto and b = second column). When a word is broken between pages, folios or columns, the reference is placed immediately after the broken word and the exact position of the column or folio break is observable from the \ within the previous word:

- e.g. Cr^IST ALL\MAH^HTIg {~f10vb~}
- (not Cr^IST ALL\{~f10vb~}MAH^HTIg).

---

<sup>18</sup> On medieval manuscript punctuation see Parkes (1992).

### 3.5.4 Deletions and insertions

Sometimes a scribe deletes unwanted text or inserts additional text (see also §3.3.1 (f) above). Deletions may be of single *figurae* (or even of parts of *figurae*), or of whole words or of longer stretches of text. They may be made by erasure (scraping the ink off the parchment), crossing through, underlining, subpuncting (underdotting, cf. ‘expunge’), obliteration (covering the whole with ink — a method favoured by Orm). Insertions may be interlinear, intralinear or marginal.

#### 3.5.4.1 Treatment of deletions

If a deletion is completely illegible its presence is simply noted in the transcription (labelled {=del=}, with or without any further comment. Such a note is treated like any other textual note (see further §3.6 below). If a single *figura* or only part of a word has been deleted and replaced, by the same scribe, with a different *figura* or segment (see §3.5.4.2 Treatment of insertions, below) the deletion is again noted and described in a textual comment. If the deletion and insertion are thought to have been made by a scribe different from the text witness himself, again the fact and the insertion will be noted, but the original text will, if legible, be preserved for tagging.

When a simple scribal deletion is completely legible, it is transcribed and is placed between <<. If (as is usually the case) the deleted text is in the same hand as the surrounding text, a decision is made as to whether or not to include it for tagging. If part of a word has been deleted, only in very unusual cases is the word tagged with the deletion still in place. Normally the deleted *figura* or *figurae* are omitted from the transcription of the word, as being unwanted by the scribal witness, and the deletion is described in an accompanying note, e.g.:

FLIz+T {=S erased before z and partially overwritten with it=}

In the case of Orm (# 301), however, it is known that he began by writing certain Old English *eo*-words using the traditional *eo*-spelling and then revised them by erasing the <o> in each case. Both spellings are Orm’s own but belong to different phases of his spelling system. In this case, Orm having been responsible both for writing the <o> and for erasing it again, the deletion is included in the transcription for tagging: e.g. E<O<RyE, TRE<O<+S.<sup>19</sup>

If a deleted form is deemed to be truly erroneous (i.e. not a sensible form in context in the scribe’s language), it is placed between {<<} and will be skipped by the tagging program. Similarly, if the deletion is of an incomplete word, perhaps because the scribe has misspelled and immediately realised the error before completing the word, the letters written are still transcribed between {<<}. e.g.:

AND BET+ERE {<ME<} {=del, subpuncted and crossed through=} {} MAY

In the above case, *me* could have been written erroneously for the first two letters of the word MAY, which is then, after the deletion, spelled ‘correctly’ according to the scribe’s own system (in this case Dan Michel in the *Ayenbite*, # 291); or it could have

<sup>19</sup> Note that in Cambridge, Corpus Christi College 145, *South English Legendary* (# 286), <o> has frequently been erased in word final <eo> combinations. It is assumed in this case that the erasures were made by a subsequent ‘corrector’ and in that text therefore the transcription is, e.g. TRE[O], BE[O].



been a complete word, written for ME or for MAN or MEN. In any event it is not here possible to assign it a tag.

If the deleted text has simply been copied in the wrong place or is an example of dittography, it may be possible to analyse it as running text and assign to it a plausible lexico-grammatical tag.<sup>20</sup> In such cases the deletion is placed between {<} and {<}. The textual note about the deletion (prefaced by del, for ‘deletion’) immediately follows the first {<}. The tagging program skips the {<} and the textual note, but reads the form(s) in between, e.g.:

INE ALL+E {<} {=del, crossed through, dittography=} yE {\} GUOD+ES {<} yISE  
GUOD+ES OF KENDE

The case above might have been just a simple repetition, subsequently corrected. But given the minor change in wording, it seems most likely that Dan Michel wrote *ŕe guodes* before realising straight away that the text should read *ŕise guodes*. Rather than emending *ŕe* to *ŕise* he chose to delete the first attempt and continue with the second. Both versions are well-formed text and both may therefore be tagged. The same is true in cases of exact dittography. The decision has been made to tag all such cases where they are legible. Where spellings of repeated words or phrases differ in repetitions, both versions can be taken (at least in the first instance) as belonging to the repertoire of the scribal witness. Where the spellings are identical, tagging of repetitions (whether deleted or not) will lead to extra tokens for the relevant items being counted in the sample. Recourse to the tagged text itself, and removal of deleted words, will make it possible for such repetitions to be excluded from statistical counts if desired.

#### 3.5.4.2 Treatment of insertions

Insertions are placed between >>. They may occur within a word to be tagged, e.g. (both examples from hand B’s contribution to the Trinity Homilies (# 1300):

HE>RE> {=RE interlined above by Scribe B himself=}

wRA>d>dE {=First edh interlined above by Scribe B himself=}

Sometimes the placing of such intra-word insertions is indicated by the scribe with an insertion siglum or with a line or caret. Very often, however, the inserted *figura(e)* are simply interlined by the scribe. In such cases it is occasionally difficult to determine whether an interlined *figura* is a *post hoc* insertion (to be transcribed between >>) or a planned superscript (to be preceded by ^ in transcription). Judgements are made in individual cases, bearing in mind the scribe’s usual practice and also superscript traditions.

If an insertion is of a whole word or of more than one word, and is to be tagged, it is placed between {>} and {>} with any note or comment (prefaced by ins, for ‘insertion’) being made immediately after the first {>}. If the insertion replaces a deletion that will also be noted, e.g. (from *Ayenbite* # 291):

<sup>20</sup> This may even, in simple cases, be possible where the syntax after the deletion turns out to be different from what has to be assumed to have been intended before the deletion.

AND yE {>} {=ins, in right hand margin in different ink=} HER+YINGE {.} {>} {\}  
 {<} {=del, crossed through to be replaced by HERYINGE at end of line above=}  
 BLISSE {<}

If a piece of text has been inserted within an already inserted piece of text, this is placed between {>>} and {>>} so, e.g.:

{>} {=ins, heading, underlined to right of main text=} {} {para} \*yE EzTENDE BOz  
 {>>} OF {>>} AUARICE {.} {} {>}

If a word or sequence of words has been inserted and is not to be tagged (usually because it is in a different hand from that of the scribal witness — see further §3.5.8 below) then it is placed between {>>} and is skipped by the tagging program, e.g. (from Hand A's contribution to *Vices and Virtues*, # 64):

{<} {=del, by subpunction probably by another hand=} dER-OF {<} {>} LEAN)>}  
 {=ins, interlined above deletion in the main correcting hand=}

In the above case, *der-of* is in the hand of the relevant scribal witness and although deleted (probably not by Scribe A himself) is to be retained as part of the tagged text. *lean* has been substituted by a correcting hand, and it is not to be included in the tagged text for Hand A.<sup>21</sup>

### 3.5.5 Missing words

Sometimes a text will seem from the sense to have a word or words missing, whether this be from damage to the manuscript or from scribal omission. In order to help with interpretation of a text, missing words may sometimes be conjectured and supplied by the transcriber, or from a previous edition. Such conjectural words cannot, of course form part of the tagged text; they are placed within {[ ]}.

### 3.5.6 Identification of headings

{'} {'} or {' '} are placed round headings or titles, depending on whether the title text is to be tagged or not.

### 3.5.7 Glosses to text words

{" " } contain glosses to text words in cases where the form of the tag may not reveal, or may mislead, as to the precise meaning of the text-word, e.g.: CHEKER {"chess board"}, where in the tagged text the form will carry the tag \$checker/n (for tagging conventions, see Chapter 4).

---

<sup>21</sup> The main correcting hand's contributions have in fact been tagged, but they have been transcribed separately and they form a separate tagged text, # 303, in the *LAEME* corpus.

### 3.5.8 Treatment of text not to be tagged as part of a *LAEME* corpus sample

#### 3.5.8.1 Text in English but not in the hand of the scribal text witness

Text in English in a different hand from the scribe of the tagged text, whether in the form of commentary, glosses, corrections or additions, is excluded from the tagged text. Such text is placed within `{ } }` and normally carries a separate textual comment (for which see further §3.6 below). Here are two examples (the second as an insertion), from the transcription of Hand A's contribution to *Vices and Virtues* (# 64), excluded by the bracketing from text to be tagged:

```
{ }para *OF wISDOM)} {=Written by the title scribe to the right side of the line,
separated from the text of the next section by the paraph=} {>)*OF *WISDOM)>}
{=ins, in right margin in a modern hand=}
```

#### 3.5.8.2 Text in languages other than English

Text in Latin or in French embedded in the early Middle English text being transcribed is normally also transcribed, but is bracketed so as to be skipped by the tagging program. Non-English text, is marked by being enclosed between `{( ( }` if it is in the same hand as the text witness and by `{)( ( }` if it is in a different hand. Here are two examples from the transcription of Hand B's contribution to the Trinity Homilies (# 1300), the second as an insertion in a different hand:

```
{(*A*DUERSARIus UESTer DIABOLUS TAmQu^Am LEO RUGIENS CIRCUIT
QueRENS \ QUEM DEUORET .{ }
```

```
URN+EN {>)(PRECIPITAUERUnT)>} {=ins, interlined in the glossing hand above
URNEN underlined=}
```

Note that transcription policy for text in Latin is much the same as that for the early Middle English text. However there are some differences, because Latin text tends to be much more heavily abbreviated than Middle English text. The Latin text is supplied not for the purposes of linguistic analysis but for information and for reasons of contextual clarity. Therefore logographic abbreviations are all expanded traditionally for ease of comprehension, even if the manuscript 'word' is simply an initial letter and a punctus. Where a punctus is used as a sign of abbreviation it immediately follows the expansion, e.g. `{(Scilicet. GAUDIUm PLENUM .{ }`, where *scilicet* appears in the manuscript as *s*. Where a punctus (or other punctuator) is used as a punctuation sign within the already bracketed Latin text, it is not additionally 'bracketed out'. But unlike in modern punctuating practice, a space is left between it and the preceding word to indicate that it is not here being used as an abbreviation sign. For illustration see the punctus after `DEUORAT` and `PLENUM` in the examples above.

### 3.6 Textual notes

Textual notes are of two kinds: linguistic and miscellaneous.

`{* * }` are placed round short simple comments that relate specifically to a linguistic form or structure and do not include non-linguistic information. The most commonly

occurring of these is `{*sj context*}` following a form that is not formally distinguished from the indicative (present or past) but which one might expect to have been in the subjunctive, whether because it follows a particular conjunction or for other contextual reasons. Longer comments that may include linguistic information alongside other commentary are usually treated as miscellaneous notes, see below.

`{= =}` are used for miscellaneous, general notes and comments. The miscellaneous category is large and varied. It includes all the various comments on deletions and insertions and on different hands exemplified in the citations above. It also includes any textual notes on readings or palaeographical commentary. The bracketing conventions allow for embedding of different types of commentary, so glosses within " " may appear within a general comment inside `{= =}`, e.g.:

LOR+yEw+ES `{>)LORDES)>}` `{=ins, interlined in the glossing hand above LORyEwES underlined. This is a mistaken gloss - it should be "teachers"=}`

### 3.7 Summary and apologia

Our aim has been to make the *LAEME* corpus consistent in the way that the transcriptions have been made and in the use of the bracketing conventions described above. However, since starting the work in the late 1980s, our transcription policies have evolved:

when I began transcribing the early Middle English texts for tagging, I did not include textual ‘details’ such as punctuation, accompanying Latin tags and quotations, notes of corrections or additions by other hands, or even — at the beginning — manuscript line ends. Gradually, in the course of building up the corpus, I began to rectify these omissions, but as a result of the early failure, I am still, at the time of writing, in the process of going back to the microfilms and adding manuscript punctuation, embedded Latin text and marginal notes to a corpus of nearly 650,000 tagged words (*LAEME* Preface: 6).

For each tagged text in the corpus there is a note as to the status of the text in relation to the addition of punctuation, embedded Latin, and fuller textual notes with designated bracketing. At the time of writing, only 30 texts remain to be brought up to standard in this way. It is hoped that in the course of the ensuing months all the tagged texts will be standardised for these categories. In the meantime all tagged texts are nevertheless usable for almost all kinds of linguistic study.

This chapter has described how the corpus is transcribed. The next stage after this is tagging which is treated in Chapter 4.

## References

- Beadle, R. and Piper, A.J. eds. 1995. *New science out of old books: studies in manuscripts and early printed books in honour of A.I. Doyle*. Aldershot: Scolar Press.
- Bennett, P. forthc. A new look at the *Gormont et Isembart* fragment: Brussels, Bibliothèque Royale Albert 1er, MS II, 181. In: Proceedings of the 17<sup>th</sup> International Conference of the Société Rencesval, University of Connecticut, July 2006.
- Benskin, M. 1982. The letters <†> and <y> in later Middle English, and some related matters. *Journal of the Society of Archivists* 7.13–20.
- Benskin, M. 2001. The language of the English texts. Chapter 4 of T. Hunt (Ed). *Three receptaria from medieval England* (pp 193–230). *Medium Ævum Monographs*, NS 21. Oxford: The Society for the Study of Medieval Languages and Literature.
- Ciarelli, A. 1998. Per lo studio del nesso *de* e per datazione di O (Oxford Bodl. Libr. MS Digby 23): note paleografiche. *Scrittura e Civiltà* 22: 31-118.
- Dossena, M and Lass, R. eds. *Methods and data in English historical dialectology*. Bern: Peter Lang.
- Gumbert, J.P. and Vermeer, P.M. 1971. An unusual *yogh* in the *Bestiary* manuscript — a palaeographical note. *Medium Ævum* 40: 56–57.
- Johnson, C. and Jenkinson. H. 1915. *English Court Hand A.D. 1066 to 1500. Illustrated chiefly from the Public Records*. New York: Frederick Ungar Publishing Co.
- Ker, N.R. (ed.) 1963. *The Owl and the Nightingale. Facsimile of the Jesus and Cotton Manuscripts*. EETS OS 251. London: Oxford University Press.
- Laing, M. 1994. The linguistic analysis of medieval vernacular texts: two projects at Edinburgh. In: M. Kytö, et al. (eds.), 121–41.
- Laing, M. 1998. Raising a stink in *The Owl and the Nightingale*: a new reading at line 115. *Notes and Queries* 243.276–284.
- Laing, M. 1999. Confusion *wrs* confounded: litteral substitution sets in early Middle English writing systems. *Neuphilologische Mitteilungen* 100.251–270.
- Laing, M. 2001. Words reread. Middle English writing systems and the dictionary. *Linguistica e Filologia* 13.87–129.
- Laing, M. forthcoming a. ‘*The Owl and the Nightingale*: five new readings and further notes’ *Neuphilologische Mitteilungen* 108/3, 2007.
- Laing, M. forthcoming b. The Middle English scribe: *sprach er wie er schrieb?* To be offered to the Conference Proceedings of the 14th International Conference on English Historical Linguistics, Bergamo Italy, August 2006.

- Laing, M. and Lass, R. 2006. Early Middle English dialectology: problems and prospects. In: A van Kemenade and B.L. Loos 2006: 417–51.
- Laing, M. and McIntosh, A. 1995a. The Language of the *Ancrene Riwe*, the Katherine Group texts and *Êe Wohunge of ure Lauerd* in BL Cotton Titus D.xviii. *Neuphilologische Mitteilungen* 96.235–263.
- Laing, M. and McIntosh, A. 1995b. Cambridge, Trinity College, MS 335: its texts and their transmission. In: Beadle and Piper 1995: 14–52.
- Lass, R. 2004. Ut custodiant litteras: editions, corpora and witnesshood. In: Dossena and Lass 2004: 21–48.
- Parkes, M.B. 1992. *Pause and Effect: an Introduction to the History of Punctuation in the West*. Aldershot: Scolar Press.
- Short, I. (ed.) 2005. *La Chanson de Roland. The Song of Roland. The French Corpus* (general editor J.J. Duggan), part I, *The Oxford Version*. Turnhout: Brepols.
- van Kemenade, A. and Loos B.L. (eds.) 2006. *Handbook of the History of English*. Oxford: Blackwell.
- Williamson, K. 1992/3. A computer-aided method for making a linguistic atlas of Older Scots. *Scottish Language* 11/12: 138–57.
- Wirtjes, H. (ed.) 1991. *The Middle English Physiologus*. EETS OS 299. Oxford: Oxford University Press.