



A Corpus of Middle English Local Documents - version 2017.1

# Manual

## 1. Preliminaries

### 1.1 The purpose of this manual

MELD 2017.1 is a text corpus that contains transcriptions of 2,017 documentary texts from the period 1399-1525. All transcriptions have been carried out by the team members from digital images or other good-quality reproductions, usually photographs taken at the archive by ourselves. The corpus is produced in four versions, designed for different uses, and is accompanied by the following materials: an Introduction, a Manual, a Catalogue and a spreadsheet containing basic metadata.

This manual describes the basic terminology, transcription conventions and presentation of the corpus. The purpose is to provide users with the practical information needed for making use of the corpus. A more general discussion of the thinking behind the corpus, and of its principles of compilation and classification, is provided in the MELD Introduction.

### 1.2 Basic concepts

The basic informant in the MELD survey is the **scribal text**. The scribal text is defined in LALME (I: 8) as ‘any consecutive written output that is a single text in the literary sense, or a part of such a text, and written by a single scribe’. Typically, a single document constitutes a single scribal text.

However, codices – account books, registers and the like – may contain hundreds of entries, sometimes spanning a long period of time: even if written or copied by a single scribe, these

clearly have to be defined as separate texts. Even a single sheet may contain more than one individual text. Sometimes several drafts or copies of entirely unrelated texts appear on the same sheet, and brief memoranda may be scribbled on any blank space. Texts may also be related: conveyances may be followed by a brief memorandum relating to the transaction, and the person delivering a letter requesting rent payment may have scribbled a receipt for the payment on the dorse of the letter. In these latter cases, even though the texts relate to the same transaction, their function is clearly different, and they are treated as different texts.

It is more difficult to judge cases where additions have been made into a text, so that they form an organic part of it: such additions may be added into margins or blank spaces, or the final lines of a document may have been written in a different hand and/or ink. In the case of minor additions, the added text is normally marked in the Base version with `<add>...</add>` tags, and with a comment if the hand is different from that of the rest of the document. Larger additions, of more than 15 words or so, are included as scribal texts of their own.

Each scribal text included MELD is given a unique code. This consists of a capital D, followed by a four-digit number. If multiple texts share the same archive reference (for example, coming from the same codex or unnumbered bundle), then they also share the same code, suffixed by a hash sign (#) and a number, e.g. D6141#1. Counterparts of indentures also share the same code, irrespective of whether they share the same archive reference, and are suffixed with 'a', 'b', 'c' and so forth, depending on how many counterparts there are, e.g. D6020a and D6020b.

Texts that were mapped in LALME are so far prefixed by a capital L followed by the LALME Linguistic Profile code, made into a four-digit code by adding initial zeros as necessary (e.g. L0007, L0147, L7340, corresponding to the LALME LPs 7, 147 and 7340 respectively). Where a LALME profile has been split, the separated scribal texts are distinguished numbering them (as in L0064#1, L0064#2 and L0064#3, all corresponding to LALME LP 64). However, this system has become highly cumbersome, and will be changed after the initial instalment (which only contains four such texts). This is both because it obscures the connection between texts that are closely related, but which were not all mapped in LALME, and because the LALME practice of defining texts was very different from ours, so that the LALME profile codes sometimes refer to several different texts considered to represent the 'same' dialect.

The texts are generally referred to by a county label and the code, e.g. Beds D4144, Cambs L0282.

### 1.3 The flavours of MELD

The corpus is produced in four versions designed for different uses. The **Base** version consists the original transcriptions, including extensive coding and comments, and gives the fullest information about manuscript reality; it should be used as a reference version together with the others. The **Concordance version** consists of .txt files designed to be analysed using a concordancing programme or other corpus software. This version is tailored to suit a particular concordancing programme, AntConc 3.2.1., but it may be used with other programmes as well. The **Readable** version is produced as .html and .pdf files and is meant

for reading rather than for electronic searching. Finally, the **Diplomatic** version, produced as .html files, provides a reading version that preserves abbreviations and layout features in an iconic form, presenting the material in a format as close to manuscript reality as permitted within our transcription conventions.

## 1.4 Using MELD

All four versions of MELD 2017.1 are provided as .zip archives; in addition, the Readable version is provided as a .pdf file for reading online or printing out.

There is no search function implemented on the web site. The recommendation is that the text files are downloaded and then used with the text processing or corpus software of one's choice. The downloadable text files are UTF-8 encoded and the end-of-line coding follows the UNIX format. However, the files are ASCII compatible: we use only the first 127 characters of the UTF-8 set, which are identical with the first 127 characters in the basic ASCII set. Therefore the text files are suitable for any concordancing program that can digest ASCII, such as AntConc or WordSmith. As the transcription distinguishes between upper and lower case letters for several purposes, it is important that the chosen program support case sensitivity.

## 2. The base version: the actual transcription

### 2.1 Processing the text from manuscript to text file

Most of the samples have been transcribed from a facsimile reproduction: most commonly a digital photograph, but occasionally a photocopy, photostat or microfilm; some texts are transcribed from the manuscript itself. The types of source for each transcription are indicated in the Catalogue. Each transcription has been proofread by at least one other member of the team.

The transcriptions use ASCII characters only, in order to be easily transferable between different platforms. This has been considered a crucial advantage, outweighing all other considerations. It means that the corpus can be easily converted to virtually any other format as desired, and that it can be opened and edited or searched using virtually any software.

The transcription conventions are based upon those used in *A Middle English Grammar Corpus* (MEG-C), which in turn were based on the conventions of the *Linguistic Atlas of Early Middle English*; however, as each project has dealt with very different materials, there are differences in detail.

The following section describes the transcription conventions. The raw transcriptions are provided as the base text files of the Corpus, and provide the most detailed and accurate record of the physical text in the manuscript.

## 2.2 The transcription conventions

### 2.2.1 General

The transcriptions reproduce the text at what might be called a rich diplomatic level. This includes the following features:

- spelling, distinguishing between 31 letters including the sub-graphemic distinctions between <i/j> and <u/v>, but not other variant forms such as different forms of <r>, single and double compartment <a>, and so on.
- capitalization
- abbreviations and some final flourishes/otiose strokes
- accents over i's.
- punctuation, using the full stop, semicolon, colon and slash for the following types of MS punctuation marks: dot, *punctus elevatus*, colon and virgule.
- word division
- line division, initial large capitals and paraphs
- headings/rubrics and otherwise marked text
- folio or page references
- corrections and marginal additions, if plausibly contemporary
- signatures and address lines
- gaps and blank lines

### 2.2.2 Spelling and abbreviations

The transcription is carried out using only symbols belonging to the basic ASCII set. All the letters of the present-day English alphabet are typed in upper case; lower-case letters are used for ME graphs, abbreviations, codes and comments.

The ME graphs <þ, ð, ȝ, æ> are transcribed as the lower case letters or letter combinations <y, d, z, ae> respectively. Of these, only <þ> and <ȝ> are common in the present corpus.

In many texts, <þ> and <y> are not distinguished; in such cases both are transcribed as <y>, irrespective of what the actual letter form looks like. In the great majority of such texts, the letter form looks like y. Where the two letters are distinguished, transcription is strictly according to letter form.

Manuscript capitals are indicated with an asterisk immediately before the letter:

Amen	*AMEN
AMen	*A*MEN

Large or decorated initials that are higher than one line are indicated with two asterisks:

<b>W</b> hen	**WHEN
--------------	--------

<b>W</b> Hen	**W*HEN
--------------	---------

**Abbreviations** are transcribed using lower case letters; however, rather than the ‘expanded’ abbreviations of traditional editions, each abbreviation type is transcribed using a unique code that represents the visual form rather interpreting its ‘meaning’. Purely for reasons of readability, the codes chosen for each abbreviation type are similar to the most common ‘expansions’ of that abbreviation: however, they involve no assumptions about what these marks ‘mean’. Thus, a **macron** is transcribed as a lower case nasal, whether or not this fits our intuitive feel of the linguistic meaning behind.

**Suspension/contraction marks** other than macrons are expanded using a set of lower-case expansions, based on the classification of signs of abbreviation by Hector (1966: 30-35). A complete list of expansions, together with references to Hector’s classification, is given under 2.3 below.

The most complicated question with regard to transcription conventions has been the treatment of the **final flourishes** that Parkes (1979: xxix) characterized as ‘additional strokes which in Latin text would indicate an abbreviation, but which may or may not do so in English’. We take as a starting point Parkes’ (1979: xxx) statement that a transcription can afford neither to ignore final flourishes nor to treat them as abbreviations, but should simply record them as final flourishes.

The problem then arises of the definition of a final flourish: at which point does a long end stroke become a flourish? Are cross bars over h’s or double l’s to be considered ‘flourishes’ even if they occur completely regularly, or are they part of the regular letter shape (*figura*)?

Such questions can often be answered in a fairly satisfactory way for an individual text; however, for the present purpose it is necessary to follow the same guidelines for every text. Recording everything that could possibly be described as a flourish seemed a hopeless task: some scripts tend to involve something flourish-like in virtually every word, and many flourishes, especially of the cross bar type, seem to be best regarded as part of the *figura*. For example, hands that mark final <ll> with a cross bar generally do so completely consistently; this is borne out both by electronic searches of transcribed texts where the cross bar has been marked and by observation of the usage during the transcription and/or proofreading of at least fifty texts. The variation here tends to be between final single <l> and cross-barred <ll>.

In the end, we have decided to record only such types of flourishes that form part of a continuum either with an abbreviation mark or with a final -e (that is, there are borderline cases that could plausibly be defined as either a flourish or as a fully-formed abbreviation

mark or ‘e’). This group includes flourishes on word-final minims as well as on (at least) <r>, <g>, <t> and <k>, as well as up-turned flourishes on <d>. The use of other strokes and cross bars will be noted in the Paleographical notes that is planned to eventually form part of the Catalogue of Sources.

The final flourishes fall into two formal categories. The most common are flourishes/endstrokes made without a pen lift, which may be more or less rounded or looped. Such flourishes are here termed **squiggles**. Squiggles are transcribed with a tilde ~, irrespective of what they ‘represent’. Occasionally, this leads to some rather absurd ‘readings’ such as CUMMYG~ rather than CUMMYnG “coming”

Sometimes, the kind of abbreviations transcribed as <er> or <re> are made without a pen lift and may look identical to squiggles; in such cases the context will usually determine the choice of transcription: OUEr THE MOSSE ‘across the moor’ but THE DAT~ OF THIS INDENTURE ‘the date of this indenture’, rather than OU~ and DATer.

Another type of flourishes are made without a pen lift, but involving a change of direction, so that they end up as a combination of a squiggle and a macron. Their functions also appear to combine those of the squiggle and the macron, in that they seem to commonly represent both a nasal and a potential final *-e*. Such flourishes are here termed **squigrons**, and have been transcribed as @.

The transcription does not aim to preserve graphetic detail: different variant forms for the same letter (e.g. double/single compartment <a>) are not distinguished. As an exception to this general rule, accents over <i> are in the base files indicated with a % sign following the letter. This is because such accents frequently determine the readings of minim clusters; thus, their presence or absence may be of importance in evaluating the justification for our reading. On the other hand, dots over <y> are not included. While they sometimes help to disambiguate <y> and <þ>, they add much less information for the interpretation of the text, and including them has not been deemed economic.

### 2.2.3 Punctuation

Punctuation is indicated using the following signs as appropriate: </ ; : . >. A gap is left between the last letter and the punctuation mark: HE CAME . AND SAW .

### 2.2.4 Word and line division

Manuscript word division is retained. We do not, however, measure the gaps between words: if what we think about as two words are not very obviously written together, then they are deemed to be written apart and transcribed accordingly. However, in clear departures from present-day word-division, the following conventions are used:

Where two words (as defined by the headwords of the *Oxford English Dictionary*) are clearly written together, they are transcribed together, with a + sign indicating that we are dealing

with what might be analysed as two separate words (e.g. A+MAN “a man”). Conversely, when what we (and the *OED*) would consider a single word is divided into two parts, these are combined in the transcription with a hyphen: so WHER-FORE “wherefore”. It should be noted that, while the uses of + and - do preserve the manuscript reality, they also impose an interpretation on the text; using them at all is a purely compilatorial choice, designed to make the next stage of analysis easier.

The text is transcribed line for line, with manuscript line division marked by the Return key. Line numbering can thus be added to the transcriptions when wished. In the base text files, word division at the end of lines is marked by adding # to the end of the first half, e.g.

```
HAP#  
PY
```

If the scribe has marked the division with a hyphen (usually a double diagonal stroke), this is indicated with a = symbol before the hash:

```
HAP=#  
PY
```

The word divisions across lines are retained only in the Base files; in the Readable and Diplomatic versions they have been silently removed, while the Concordance version marks the divisions but brings the words together to make them count as single units.

### 2.2.5 *Foliation*

Foliation (alternatively pagination) is indicated throughout. The beginning of a new folio is indicated within square brackets in the following format: [fol. 8r] (alternatively, [p. 8r] for a paginated text). If the transcription does not begin at the top of a page, line number is indicated as well, in the format [fol. 8r][line 10]. Columns of text are indicated with lower case [a, b...]. Thus, a folio with two columns of text on each page will consist of the following four sections: [8ra], [8rb], [8va], [8vb].

### 2.2.6 *Coding for layout, corrections and commentary*

A set of codes placed within angled brackets are used to indicate specific layout features, corrections and additions, as well as other kinds of commentary:

Rubrics and headings are marked by inserting the following codes before and after the text: <rub>...</rub>. Underlining is marked with the codes <und>...</und>. Expunction or crossing out is marked with the codes <exp>...</exp> and <cro>...</cro> respectively. Partially rubbed-out text is marked with <rbd>...</rbd>.

Added/inserted text may be marked in four ways, depending on where it has been added.

Whether added by the scribe himself or a later corrector, text is most commonly inserted above the line or in the margin; such insertions are marked with `<sup>...</sup>` and `<mrg>...</mrg>` respectively. Occasionally, an addition is made in an existing gap within the text or over a rubbed-out section; in such cases, it is marked with `<add>...</add>`. If the addition is marked with a caret, the code `<ct>` is used. Thus, the sequence `W<ct><sup>H</sup>AT` in the base text indicates a spelling *wat*, with `<h>` added above the line and a caret between `W` and `A`.

The code `<sub>...</sub>` indicates the continuation of a line at the right hand end of the following line. In general, insertions do not appear in this position; if they do, they are marked `<add>...</add>`.

Additions are generally transcribed if they are considered to be at least potentially contemporary. Often it is impossible to tell whether they were carried out by the same scribe or not. Therefore, any text marked with the `<sup>`, `<mrg>` or `<add>` codes should be excluded, or at least viewed with suspicion, when studying the language of a particular scribal text.

Latin words or passages within the text are marked with the codes `<lat></lat>`, and are generally only transcribed when relatively short and easy to decipher; none of the team members claims a specialist competence in medieval Latin. Elements of French are much less frequent in the corpus but appear in some of the earliest texts; they are marked with the codes `<fre></fre>`.

Illegible letters or passages are marked with the code `<ill>...</ill>`. The approximate amount of text missing is indicated within angled brackets: `<ill><c. 2-3 words></ill>`. Sometimes, the last portion of a line may be invisible because it disappears into binding, or it may have disappeared if the pages have been cropped; in such cases, a descriptive comment is placed within angled brackets, e.g. `<binding>`, `<cropped>`.

Finally, any comments may be placed inside the codes `<com>...</com>` and written in ordinary lower case: e.g. `<com><writing slightly smudgy here></com>`, `WUN <com><? four minims></com>`. Such comments appear in the Base text files but are removed from the Reading version unless deemed crucial for the reading of the text; all comments are removed from the Concordance version but an exclamation mark `!` after a word indicates that the Base files should be consulted for more information.

Unexpected written forms, which could be suspected to be transcription errors, are followed by `<sic>` (in the Base version only): e.g. `ALLLE <sic>`

## 2.3 List of symbols used in the transcription

The following list of symbols summarizes the transcription conventions used in the Base files; for a description of their use, see 3.3. above. Abbreviations are defined according to the classification by Hector (1966: 30-35) using his classification numbers. Non-alphabetic symbols are listed first, then letters and finally codes enclosed in angled brackets.

<>	enclose anything that is not to be read as part of the transcription, such as codes and comments
; :	<i>punctus elevatus</i>
.	<i>punctus</i>
/	<i>virgule</i>
&	any symbol used for ‘and’
~	squiggle (= a typically word-final flourish that may either be functionally equivalent to <e> or otiose)
@	squigron (= a squiggle combined with a macron, i.e. a flourish that involves a change of direction)
,	word-final downstroke, historically developed from Hector 9
%	acute accent or ‘dot’ over <i>
\	defines following letter as a superscript one (used only for the systematic use of superscript as in <i>b<sup>t</sup></i> ‘that’; not used for corrections or additions above line)
#	signals word division across the line
=	word division marker in the manuscript (always placed before #)
-	gap between two words that would correspond to a single word in Present-day English usage (e.g. <i>to-geder</i> ‘together’)
+	assumed boundary between two words written together in the manuscript but corresponding to two separate words in Present-day English usage, e.g. <i>a+man</i> ‘a man’
*	defines the following letter as a capital
**	defines the following letter as a large initial capital extending over more than one line
a, ar, ra, ua	abbreviation (derived from superscript <i>a</i> , cf Hector 1966: 34-35)
con, com	abbreviation (Hector 6)
er, re	abbreviation (Hector 3)
es	abbreviation (Hector 9)
ir, ri	expanded abbreviation (derived from superscript <i>i</i> , cf Hector 1966: 34-35)
n, m	macron indicating nasal (Hector 2)
or	abbreviation (Hector 4, the 2-shaped version)
per, par	abbreviation (cf Hector 1966: 34)
pro	abbreviation (cf Hector 1966: 34)
ro	abbreviation (derived from superscript <i>o</i> , cf Hector 1966: 34-35)
sir, ser, sen	abbreviation
ur	abbreviation (Hector 4, the a-shaped version)
us	expanded abbreviation (Hector 5)
y	the letter <b>, ‘thorn’
z	the letter <3>, ‘yogh’ or the similar-shaped <b>variant form</b> of <z>
Z	the zeta-shaped <b>variant form</b> of <z>
<add></add>	enclose text added on the same line, in gap or over erasure, either by the same scribe or by a later corrector (clearly post-medieval corrections are ignored)
<adr></adr>	enclose an address clause, often but not always written on the dorse of a letter
<cor></cor>	enclose text that has been crossed over for deletion (correction of errors)
<cro></cro>	enclose text that has been crossed over for deletion (removing content)
<ct>	caret
<exp></exp>	enclose text that has been expuncted for deletion

- `<ill></ill>` enclose illegible text (approximate amount of text indicated in diagonal brackets between the codes)
- `<lat></lat>` enclose text in Latin
- `<mrg></mrg>` enclose text added in the margin either by the same scribe or by a later corrector (clearly post-medieval corrections are ignored); placed at caret position when marked
- `<pph>` paraph
- `<rbd></rbd>` enclose text that has been rubbed out/erased; if illegible, the approximate amount of text is indicated in diagonal brackets between the codes
- `<rub></rub>` enclose text marked as a heading/rubric or strongly emphasized by means of script type/size or colour
- `<sgn></sgn>` enclose a signature clause, such as ‘by me William Barker’
- `<sub></sub>` enclose text continuing below the line, usually at end of the following line
- `<sup></sup>` enclose text added above the line either by the same scribe or by a later corrector (clearly post-medieval corrections are ignored); placed at caret position when marked. Used only for corrections or additions above the line, not for the systematic use of superscript, as in *b<sup>t</sup>* ‘that’.

## 2.4 Editorial decisions and interpretation in the Base files

On the whole, the transcription aims to record what is visible in the manuscript, rather than giving editorial interpretations. However, any transcription will inescapably involve an element of interpretation. The users of the present Corpus should in particular be aware of the following compilatorial choices in the Base files:

Firstly, the uses of #, - and + entail compilatorial interpretations of word division. A user who does not wish to be influenced by these may download the text and make the following substitutions: zero for # and +, and space bar for - .

The choice between \ and `<sup></sup>` when marking superscript letters is based on the transcriber’s understanding of the distinction between the systematic use of superscript letters as abbreviations (e.g. *w<sup>t</sup>*, *b<sup>t</sup>*, *b<sup>u</sup>*) and the unsystematic insertion of letters above line for the purpose of correction and addition. The latter may be added by a later correctors, and it is often impossible to tell whether this is the case or not, especially from a microfilm reproduction. The compilers have therefore not attempted to distinguish between additions by the same or another scribe in the transcription, with the exception of clearly post-medieval additions, which are ignored. In general, it is therefore advisable to treat with caution all text that appears between the codes `<sup></sup>`, `<add></add>` and `<mrg></mrg>`, and not to take for granted that they represent the same scribal usage as the rest of the text.

The reading of minims often entails interpretation. As accents over `<i>` are recorded in the transcription, the user will be able to determine in which cases they have clarified the reading. Where such accents are absent and the script makes no distinction between `<u>` and `<n>`, a sequence of six minims transcribed as MIN, NIM or NUN is based on the transcriber’s

judgment of what fits the context best. The same applies, in many texts, to the choice between <st> and <sc>.

Finally, in some texts, squiggles (~) and *er/re* abbreviations (Hector 3) may also look identical, and the choice is then a matter of interpretation. Similarly, the choice between superscript <i> and the *ir/ri* abbreviation is often compilatorial: the two are historically identical, and often (if not always) identical in form. As far as this last distinction goes, the user who does not wish to be influenced by our choices may replace lower-case <ir> and <ri> with <\I> in their downloaded copy of the Base Corpus files.

Some of the compilatorial choices have been necessary from the point of view of the use of the data: it is, for example, important to distinguish between such superscript letters that are part of the scribe's spelling system and ones that represent additions that may have been carried out by another scribe. Others would have been possible to avoid, and may be removed by the user, using substitutions such as those suggested above. The possibility of indicating minims by some neutral sign, rather than interpreting them as specific letters, was discussed by the team, but was then abandoned for two main reasons: the shortage of suitable ASCII characters and the distinct loss of readability.

## 3. The Concordance version

### 3.1 Description

The Concordance version is intended specifically for analysis using a concordancer or another similar tool for text analysis. This version lacks the two features of the Base version that make it less suitable for concordancing: large amounts of coding and comments, and word division at line breaks. It is designed with a particular concordancing programme, [AntConc 3.2.1](#), in mind, but should suit most case-sensitive programmes as long as the definitions of word characters (characters that are read by the programme as forming part of words) are adjusted as necessary.

The main principle has been to distinguish strictly between word characters and non-word characters, so that everything that is not meant to be read as a word is either removed or turned into non-word characters. The choice of characters has been based on the requirements of AntConc 3.2.1. The main differences between *MEG-C Base* and *MEG-C Concordance* are as follows:

- All headers, tranche headings and folio/page/line numbers have been removed
- All coding that does not directly affect the readings has been removed, including <und>, <sub> and <brd> are removed, as well as paraps and carets. The '%' sign indicating accents/dots over i's is also removed.
- The <lat>...</lat> sequences indicating Latin text have been replaced with '£'
- Any other string of content that does not represent readable words (illegible or deleted words; symbols) is replaced with the symbol '☐'

- In the case of partially illegible words, the entire word, including the legible letters, is replaced with ‘□’. Letters that are marked for deletion, such as an expuncted I in ABOIUE ‘above’, are removed, but an exclamation mark is added after the word to direct the user to the Base version for information.
- Inserted text, marked with <sup>, <mrg> and <add> codes in the transcription, is indicated by a closing curly bracket after each word: AND} HIS} WYF}. Superscript or added letters within a word are enclosed within brackets: TH{R}OW.
- Text that has been crossed out is treated in different ways dependent on the extent and function of the crossing out. Ordinary corrections of misspellings and the like are removed and replaced with ‘□’. However, more substantial portions of the text, deemed to be removed because of their content rather than their form (as in the case of a record of an outstanding debt, removed when paid) are retained but marked with a following closing square bracket: JOHN] BROUN] OWES] IIIJ] S~]
- Line division is marked with ‘[’ but divided words are written in one, followed by a line break.
- Comments are removed and exclamation marks added to direct the user to the Base version for information.
- Certain characters that are used as wildcard characters by AntConc are replaced by others; for example, both squiggles and squigrons appear as ‘~’, freeing the ‘@’ sign for wildcard use, and capitalisation is indicated by ‘:’ freeing the asterisk for wildcard use. For this purpose, it has been necessary to remove the distinction between ‘:’ and ‘;’ which has been deemed less important (and in any case often difficult to make) than the use of wildcards. Finally, the ‘+’ used for separating words that are written together is replaced by ‘\_’, freeing the ‘+’ sign for wildcard use.

## 3.2 List of symbols used in the transcription

The following conventions are used in the Concordance files (NB that the lower case ‘expansions’ of the abbreviations are exactly the same as in the Base files):

;	<i>punctus elevatus</i> or colon in the manuscript
.	<i>punctus</i>
/	<i>virgule</i>
&	any symbol used for ‘and’
~	squiggle or squigron (see section 3.3.1)
\	defines following letter as a superscript one (used only for the systematic use of superscript as in <i>b<sup>t</sup></i> ‘that’; not used for corrections or additions above line)
[	signals word division across the line; the two parts are written together with line shift following: BI[FORE
=	word division marker in the manuscript: BI=[FORE
-	gap between two words that would correspond to a single word in Present-day English usage (e.g. <i>to-geder</i> ‘together’)
_	assumed boundary between two words written together in the manuscript but corresponding to two separate words in Present-day English usage, e.g. A_MAN ‘a man’

:	defines the following letter as a capital
::	defines the following letter as a large initial capital extending over more than one line
}	enclose text added above line, in the margin or on the same line, in gap or over erasure, either by the same scribe or by a later corrector (clearly post-medieval corrections are ignored). Each word in a sequence of inserted words is bracketed, making it possible to identify them all.
⌘	indicates the presence of a string in the text that is not a readable word; for example a symbol or a wholly or partly illegible word, or a crossed-out word.
£	indicates a string in Latin. Repetition of complete lines of Latin is not marked.
££	indicates a string in French
!	indicates that the user should check the reading in the Base files before using it as a research finding. It often means that a reading is uncertain, but it may also mean that the word contains one or more expuncted letters, that it has been tampered with, or the like.

For best use of the Concordance files, the programme used should recognize the following characters as ‘word characters’ or ‘token classes’, i.e. read them as part of words:

```
abcdefghijklmnopqrstuvwxyzABCDEFGHIJKLMNOPQRSTUVWXYZ1234567890-
&[=~\{\},
```

The following characters should NOT be read as part of words:

```
<>?;./_:\⌘£!
```

## 4. The Readable version

The readable version provides a text that closely resembles an ordinary edition, with coded comments kept to a minimum. It is primarily meant for easy reading, citation, and reference. Any examples are cited from the readable version. The main differences between the transcriptions and the readable version are as follows: the default case is lower case; capital letters are represented in upper case, making the coding for them unnecessary, and the abbreviation codes are rendered in italics and in some cases converted into more traditional expansions. For example, the abbreviation code ‘or’ is converted to *ur* rather than *or*. Paraphs, underlining, superscript letters, deletions, and so forth, are represented iconically. While line division is kept in general, words divided from a line to a new line have been joined silently. Finally, all scribal and compilatorial coding has been deleted, so that paraphs, underlining, superscript, deletion etc. are represented iconically.

The Readable files may be viewed on-line, and they are also available in a .zip archive intended for downloading.

## 5. The Diplomatic version

The diplomatic version contains several symbols and lettershapes not present in the readable flavour and is meant to approximate the manuscript reality as closely as the transcription practises permit. Abbreviations codes are converted into approximations of the medieval marks they represent.

## 6. Updates

The first version of MELD, 2017.1, is published in instalments, consisting of geographically defined subcorpora. Thereafter, each updated version will receive a new version number. Older versions of the corpus will be stored as .zip archives, so that they will still be available after the changes. Information about updates will be posted in the Programme Log on the Middle English Scribal Texts website ([www.uis.mest](http://www.uis.mest)).

## 7. References

Hector, L.C. (1966), *The handwriting of English documents*. 2<sup>nd</sup> edn. London: Edward Arnold.

McIntosh, A., M. L. Samuels & M. Benskin, with M. Laing & K. Williamson (1986). *A Linguistic Atlas of Late Mediaeval English*. 4 vols. Aberdeen: University Press.

Kretzschmar, William A., Jr., and Merja Stenroos. 2012. Evidence from surveys and atlases in the history of the English language. In Nevalainen, Terttu and Elizabeth Closs Traugott (eds.). 2012. *The Oxford Handbook of the History of English*. Oxford: Oxford University Press. 111–122.

Parkes, M (1979), *English cursive book hands, 1250-1500*. London: Scolar Press.

Stenroos, Merja (2007), '[Sampling and annotation in the Middle English Grammar Project](#)' in A. Meurman-Solin and A. Nurmi (eds), *Annotating Variation and Change*. Helsinki: University of Helsinki.