• • • •

Testing a New Spelling Database Created from *A Linguistic Atlas of Early Middle English*

MARIE VAŇKOVÁ

Abstract This paper presents a new tool designed to facilitate linguistic analyses of texts from the early Middle English period, focusing primarily on spelling and its variation. The core of the application is a database which maps correspondences between segments of the numerous spelling variants available in *A Linguistic Atlas of Early Middle English* (LAEME). The interface is suitable for analyses of spelling systems of individual texts and their comparison, as well as analyses with a wider scope such as spelling variants potentially associated with a specific sound change etc. The article is based on a short sample study designed to test various features of the tool. The goal of the study is to describe the distribution of equivalents of 'h', 'ȝ' across multiple copies of three short lyrics. A basic description of the tool is complemented with practical examples of its use taken from the study. The article also relates the results of the study to previous findings in order to check for potential errors in the database, and to assess its advantages or disadvantages in comparison with more traditional research methods.

Keywords Middle English, spelling, *A Linguistic Atlas of Early Middle English*, research tool

1. Introduction

One of the most remarkable things about research into early Middle English (eME) is the change in attitude towards the orthographical and dialectal diversity found in early Middle English texts. Scholars have moved from the (perfectly understandable) dismissal of many early Middle English witnesses as "ill-spelled text" (Dobson, 1979 as cited in Laing & Lass, 2013) to careful examination of the spelling systems. The variation and apparent irregularity of the spelling systems have proved to be a valuable source, which can reveal rather than obscure the developments under way.

The crucial requirement for research into early Middle English texts – to examine each writing system separately – is a very challenging task. Each individual writing system develops in contact with others and copies of texts often contain elements taken from their exemplars. As a result, analyses generate rather long checklists of questions and pieces of evidence to be

examined and other arduous tasks, such as looking for regional patterns of distribution, looking for identical or related spelling variant in other texts, checking a variant against its equivalents in copies of the same text, etc. The questions sometimes lead to a path worth pursuing and sometimes to a dead end, but whatever the case, it is often impossible to explore all the paths within the scope of a single analysis.

New research possibilities have opened with the publication of *A Linguistic Atlas of Early Middle English* (LAEME), which is essentially a corpus of texts from the period 1150–1325. Among other things, LAEME allows the researcher to complement close reading of texts with linguistic and other data obtained from the corpus and to construct maps.

The objective of this study is to present and test a new spelling database created from LAEME data, designed to provide faster access thereto, including data quantification, and easier construction of maps.

The article demonstrates the use of the tool on a series of practical examples, which are accompanied by a commentary on typical problems encountered in research into eME texts. All the examples have been taken from a sample "pilot" study based on multiple versions of selected Middle English lyrics found in several source texts in LAEME. As the development of the tool is at the testing stage, it was considered preferable to choose well researched texts and check the results against previously published findings. The objective of the study was to test what kind of observations about the texts can be made using different functions and features available in the tool.

The first part of the paper briefly outlines the main challenges and problems of research into eME and explains how the tool seeks to respond to them. It also presents the most important findings concerning the examined manuscripts and explains the concept of *litteral substitution* set developed by the authors of LAEME (Laing & Lass, 2013). Litteral substitution provides a useful framework for analysis of eME witnesses. The second part introduces the new tool and describes specific methods of analysis used in the sample study. The final part presents the results and discusses them in the context of previous research. It also comments on the efficiency and limitations of the adopted method and the tool in general.

2. Theoretical background

One of the most prominent characteristics of Middle English, which is highly challenging to researchers, is the seemingly chaotic and variable spelling. The fact that ME spelling appears less regular in comparison with Old English is a consequence of the rupture in the writing tradition, which occurred after the Norman Conquest (Upward & Davidson, 2011). The loss of official support for writing in English resulted in a level of variation in the spelling systems unparalleled by any other period. It is assumed that Middle English is a stage of relatively "close correlation between spoken and written language" (Horobin & Smith 1999: 362). This means that the extant texts allow us to draw inferences regarding phonological developments and dialectal differences which would be obscured by more "standardized" and uniform writing systems. However, drawing such inferences can be very difficult, especially in the case of early Middle English, because surviving material from the period is relatively sparse compared to the later period. As a result, even analyses with a primarily narrow focus, such as the writing system of a single text, a specific sound change or the development of a specific word, need to be carried out in the context of a vast network of possibly relevant pieces of evidence. For instance, the interpretation of the sound value of a single form should ideally take into account our knowledge of sound changes, the spelling system of the given text, spelling systems of related texts (e.g. text copied by the same scribe, if available) or representations of the same word in other texts, etc. Gathering all this evidence may take a very long time.

2.1. *A Linguistic Atlas of Early Middle English*

LAEME is a tool which significantly facilitates access to the possibly relevant pieces of data. For instance, a complete list of all forms of a given word or a list of all texts in which a given form appears can be retrieved in a single search.

The core of LAEME data is a corpus of tagged texts. It covers all the available texts in eME, however, some of the longer texts were not transcribed in full. The basic unit of LAEME is the tag which consists of *lexel* (usually corresponding to lemma), *grammel* (grammatical tag) and *form* (actual word in the MS). There are separate tags for individual words as well as its morphemes. Basic queries like listing all the forms of a specific *lexel* can be combined with

metadata on the dating and localisation of the texts. As a result, it is possible to formulate queries like "list all the forms of FIRE/N and sort them by county". The distribution of different spelling variants can furthermore be plotted on a map. A weak point of LAEME is its online interface, which is not very easy to use and obtaining the desired data sometimes requires more work than it could, as well as a good knowledge of *lexels* and *grammels.*

The purpose of the tool presented here is to take another step towards better accessibility of data, offering relatively simple ways of completing tasks which would require multiple queries in the LAEME search tools.

The main difference between LAEME and the new database on the level of data is that the database maps correspondences between individual segments in a group of spelling variants, which is not available in LAEME. Segments are usually single letters or a digraphs. The online interface designed to access the data in this structure offers a number of clickable links which can lead the user from one piece of data to another without the need to write a separate query. For example, a list of words which contain a given letter is always displayed with links to lists of the actual forms and each form is in turn displayed with a link to Key Word in Context view etc. As a result, the interface invites more of an exploratory approach to data and it can also prompt relevant questions. The functions available in the tool are going to be discussed later on.

The new spelling database owes a good deal not only to the exceptionally rich data but also to the *Introduction* to LAEME, which includes a discussion of some theoretical and methodological aspects of research into Middle English. The most relevant concepts, *littera* and *litteral substitution sets*, are going to be explained in the first part of this section. The second part introduces the texts and manuscripts examined within the sample analysis presented here, and the relations between them.

2.2. The Littera and Litteral Substitution Sets (LSS)[1]

In the *Introduction to LAEME*, Laing & Lass (2013) propose using the model of *litterae* as a framework for dealing with ME spelling systems rather than the more widely known structuralist concepts of *grapheme*, *phoneme* and related terminology. The use of this framework was first advocated by Michael Benskin (1997) and his colleagues responsible for the creation of *A Linguistic Atlas of Late Medieval English* (LALME). The main reason for rejecting the structuralist terminology as expressed by Laing & Lass (2013) is that "such concepts do not always characterise what our scribes appear to be doing", which is why the authors prefer "to use a theoretical framework and notation that cohere more closely with what scribes would have experienced in their education" (Laing & Lass, 2013: 2.3.1). In this framework, *littera* is an abstract object, quite simply a letter, which may be materialized as one of the possible *figurae* (which are a matter of palaeography rather than orthography) and each *littera* may have one *potestas*, literally 'power' or pronunciation, according to the original medieval model, but more *potestates* are allowed in the proposed framework. *Potestas* here essentially refers to the represented sound. For instance, <a> and capital <A> are two different *figurae* of the same *littera* which can have a few possible *potestates*. The term *littera* in the context of the spelling database is applied also to digraphs regarded as relatively fixed (e.g. 'sh', 'sch', 'hw'), etc.

In order to create space for the treatment of variation in eME spelling, Laing and Lass (2013) extended the model, enriching it with two new concepts. *A Litteral Substitution Set* (LSS) is a set of *litterae* which may be used to represent a given potestas. A *Potestatic Substitution Set* is a set of *potestates* which may be assigned to a given *littera.*

Spelling systems may be characterised as *economical* or *prodigal. Economical* systems are relatively close to a biunique representation (one *littera*, one *potestas*), while *prodigal* systems have a number of "unnecessary"

---

1   Litterae as abstract units of a spelling system are written in inverted commas in this article, e.g. 'f', 'v'. Actual word forms or figurae are written in angle brackets or italics, depending on length. Lexels (lexical units) are written in small capital letters, e.g. "final -st in RIGHT/N". These conventions mostly conform to the notation proposed by Benskin (Benskin, 1997, 2001 as cited in Laing & Lass, 2009: 1, note 2).

correspondences (one *littera* for several *potestates* and vice versa (Laing & Lass 2013: 2.3.2). Despite the fact that such systems may appear chaotic due to the multiple non-biunique relations between *litterae* and *potestates*, it is important to bear in mind that the variation is not completely random (Laing & Lass, 2009: 30). "Prodigal" spelling systems in particular can be products of intricate interactions between the scribe's interpretation of the symbols in his exemplar, or other texts he has read, and his approach to copying. Assumed "meanings" of *litterae* can shift in similar ways as the meaning of words, and multiple relations between sound and spelling develop. Such developments were explored by Laing & Lass (2009), who proposed several scenarios from which multiple relations between sound and spelling originate.

The fact that there can be multiple representations of a single sound, and the sound value of a given littera may change from text to text, implies that analyses focusing on sound changes may need to take a range of spellings into account. Furthermore, sound values, as a rule, need to be inferred in the context of the particular spelling system.

2.3. The manuscripts

It has been highlighted that data from the early Middle period is scarce and there is a high level of orthographic variation. One of the things which facilitates our understanding of this material is our knowledge of extralinguistic relations between the texts. The value of ME manuscripts for research increases if there are multiple extant versions of the same text, and also if there are multiple texts in a number of linguistic varieties but written in the same hand (Laing, 1992). The texts to be used in this sample analysis were deliberately chosen to include both of these configurations.

The analysis covers multiple versions of three short ME lyrics – The *Latemest Day* (henceforth LD), *Doomsday* (henceforth DMD) and *Orison to Our Lady* (henceforth OL).

These lyrics belong to a larger group of texts found in four Middle English miscellanies (Laing, 2000: 525), all of which have been tagged for the LAEME corpus. The manuscripts in question are: Oxford, Jesus College 29 (LAEME text #1100, henceforth Jesus), London, British Library, Cotton Caligula A ix

(LAEME #2, #3, #238-#244, henceforth Cotton), Cambridge, Trinity College B.14.39 (LAEME #246-#249, henceforth Trinity) and Oxford, Bodleian Library, Digby 86 (LAEME #2002, #214-#222, henceforth Digby). Jesus, Cotton and Trinity contain all the three lyrics and Digby contains only DMD and LD. OL in Trinity is written in a different hand (hand D) than DMD and LD (hand A). The study also takes into account a version of OL found in London, British Library, Royal 2 F viii (henceforth Royal). The table below presents an overview of the analysed versions, including IDs of LAEME files.

Table 1: Manuscripts overview (numbers refer to IDs of LAEME files).

	Jesus	Cotton	Trinity	Digby	Royal
An Orison to Our Lady	1100	239	249 (hand D)	x	263
Doomsday	1100	241	246 (hand A)	2002	x
The Latemest Day	1100	242	246 (hand A)	2002	x

The individual versions are going to be referred to by manuscript rather than text ID and the distinction between individual texts (e.g. the two hands in Trinity) is going to be explicitly pointed out only where relevant.

According to what is known about textual history, the manuscripts containing the earliest written versions of the texts did not survive (Laing, 2000: 527). There are marked differences between the versions of *The Latemest Day* found in Jesus and Cotton as opposed to Digby and Trinity. Both Jesus and Cotton begin with a short passage which is missing from Digby and Trinity. Digby and Trinity are similar in that *Doomsday* and *The Latemest Day* are copied "without a break". However, Trinity is the version which deviates the most from the other three and Digby is "textually closer" to Jesus and Cotton (Laing, 2000: 528).

The following overview summarizes selected information about the four main manuscripts presented in previous studies, and sometimes also quotes more specific findings to which reference will be made later.

2.3.1. MS Jesus College 29

MS Jesus College 29 contains two longer poems, *The Poema Morale* and *The Owl and the Nightingale* plus a number of shorter pieces. Most of the texts have been tagged for LAEME. The MS is the work of a single scribe in a homogenous language. This is why all the tagged texts are found in a single LAEME 'scribal text' or profile.[2]

A common exemplar (traditionally called X) has been proposed for several texts in Jesus and the corresponding texts in Cotton (see below). Previous research suggests that the scribe of Jesus was a translator, which means that he converted the forms from his exemplar into his own idiolect.

2.3.2. MS Cotton Caligula A.ix

MS Cotton Caligula A ix has two parts, and part II shares a number of texts with Jesus including *The Owl and the Nightingale*. The MS also contains pieces in French. All the texts in part II were copied by the same scribe but their text languages vary from text to text. Accordingly, part II was split into several files in LAEME. *The Owl and the Nightingale* copied in two distinct kinds of language (C1 and C2) is found as scribal texts #2 and #3 and both are placed in Worcesterhire. The remaining texts have a separate file each and all remain unlocalised. The explanation proposed for the varying types of language is that the scribe was a *literatim* copyist (McIntosh as cited in Laing, 2004: 52), i.e. wrote more or less exactly what he found in the exemplar and the differences between the individual texts testify to the fact that the exemplar X (shared with Jesus) was copied by several different scribes, possibly as many as six, according to Cartlidge (1997). His analysis of the text follows the assumption that it should be possible to identify features of the two "languages" C1 and C2 of *The Owl and the Nightingale* in the remaining texts presumably copied from the same source.

The table below presents an overview of the distinctive characteristics of the two types of language adduced by Cartlidge (1997: 254), which are going to be mentioned in connection with results of the present analysis.

---

2   *Poema Morale* was separated from the other texts in the spelling database to enable more precise comparison with the other six extant versions.

Table 2: Features of languages C1 and C2 of *The Owl and the Nightingale* according to Cartlidge (1997: 254).

	C1	C2
the letter 'ð'	absent	in use
historical initial 'f'	'u', 'v', 'f'	only occasional 'u', 'v'
'eo'	'o' over 'eo'	'eo' over 'o'
historical initial 'hp'	'p' over 'hp'	'hp' over 'p'
historical initial 'cp'	'qu' over 'cp', 'cu', 'cw'	'cp' over 'qu'
'h'/'ʒ'	'ʒ' over 'h'	'h' over 'ʒ'

Cartlidge (1997) compared the four texts examined here plus *A Lutel Soth Sermun* and *The Ten Abuses.* He first divided them into two groups based on the proportion of 'w' and 'p' (*wynn*). *The Latemest Day* and *Doomsday* are distinguished from the other texts by a clear preference for 'p', which was the runic symbol for /w/ adopted into Old English. A similar pattern is found with 'cp'/'qu'. This might indicate that the two lyrics were copied into X from a relatively older exemplar (Cartlidge, 1997: 254), which would be in accordance with Laing's (1999) claim that C2 in particular is "traditional and conservative in nature" (Laing, 1999: 253).

Both LD and DMD deviate from C2 very slightly, which led Cartlidge to the conclusion that the versions in X might have been copied by the same scribe (Cartlidge, 1997: 256). As for characteristics distinguishing LD from DMD, the following examples are found in the article: the ratio of 'ð' to 'þ' is about 1 : 1 in LD but 1 : 4 in DMD, 'h' clearly prevails over 'ʒ' in the final position and before 't' in LD, LD has unetymological 'h' (*ibrouhit* (BRING), *forspolehen* (FORWSALLOW), *dihsches* (DISH)). Cartlidge (1997) further claims that all of these features "must have been introduced in X, since they do not appear in MS Trinity 323 or Digby 86" (Cartlidge, 1997: 255-256).

The scribe uses a cline of shapes for <þ> and <p>, which means that they had to be transcribed by context in LAEME (Laing, 2013).

2.3.3. MS Trinity College B.14.39 (323)

The tagged sample of MS Trinity College B.14.39 is in four kinds of language corresponding to four different copyists (A–D). *The Latemest Day* as well as *Doomsday* were copied by hand A and *Orison to Our Lady* was copied by hand D (Laing, 2013).

LAEME description comments on the letter shapes for <y> and <þ>. The two have distinct shapes in the MS but "thorn quite often appears for (consonantal) 'y' = [j], and for vocalic 'y' (especially in diphthongs), but 'y' does not appear for consonantal thorn" (Laing, 2013). In other words, there is a litteral substitution set {*y, þ*} based on the similarity of letter shapes, which enables the use of <þ> in 'y' contexts (Laing & Lass, 2009: 7).

Another important point is that <þ> is identical in shape to <p>, which appears from fol. 29r onwards according to LAEME description. Yogh is very rare in this manuscript. The confusion of letter shapes concerns scribe A as well as D (Laing, 2013).

The works of scribes A and D are examples par excellence of *profligate* writing systems. Laing (2003) mentions two factors which might have contributed to their extreme level of complexity: (a) the systems of the exemplars themselves were complex and (b) the scribes intended to "represent as closely as possible the (perhaps also variable) sound values of the spoken language" (Laing, 2003: 254).

2.3.4 MS Digby 86

As with the other manuscripts, MS Digby 86 is a miscellany and also contains texts in Latin and French. All the English texts were copied by the same scribe but their text languages differ slightly. The tagged sample containing *The Latemest Day* and *Doomsday* was placed in NW Gloucestershire based on extralinguistic evidence. Exceptionally generous information in the marginalia also enabled dating of the MS to "the last quarter of the thirteenth and, perhaps, the earliest years of the fourteenth century" (Laing, 2000: 523).

A detailed analysis of the texts copied by scribe A was carried out by Laing (2000), who concluded that the text language represents

a dialect that is homogeneous and which plausibly belongs to an individual scribe – Scribe A, with the proviso that some of the variants displayed may belong to his passive repertoire of forms rather than to his active repertoire. The degree of internal variation is comparatively slight by Middle English (especially early Middle English) standards (Laing, 2000: 551).

The system of the Digby scribe differs from Jesus and Cotton in being more prodigal and also further removed from "traditional English orthography", which is partly due to the employment of typically French spellings, mainly 'ou' for [u:] and 'o' for [u] (Laing, 2003: 253).

Laing (2000) identified likely examples of forms taken from the scribe's exemplar(s), distinguishing between so-called *constrained selection* and *relict usage*.[3] She examined individual cases of internal variation in the text and determined (a) whether the variation appears across all texts written by scribe A or whether some of the forms are restricted to a specific text and (b) whether similar variation is found in texts localised in the vicinity of Digby 86, i.e. the SW Midlands (Laing, 2000: 532).

2.3.5. Summary

The four manuscripts provide very interesting material because there is a number of possible perspectives and directions which their analyses may take. Multiple levels of copying and the fact that language sometimes varies from text to text in the same MS invite studies focused on stratigraphy.[4] Textual comparison may be combined with analyses of the spelling systems. Furthermore, the scribes employ the full range of previously described copying strategies (Laing, 1999: 252).

The papers referenced in this chapter provide a considerable amount of information, which can be compared with the data obtained from the pilot study. The most extensive and detailed analyses presented here are

---

3   *Constrained selection* refers to the situation when the scribe chooses not to translate a familiar form although it is not in his active repertoire. *Relict usage* refers to untranslated forms alien to the scribe's system (Laing, 2000: 529).
4   Identification of layers of copying in a text.

Laing's (2000) stratification of features in MS Digby 86 and Cartlidge's (1997) orthographical analysis of the lyrics in MS Cotton Caligula A.ix., i.e. studies focusing primarily on the text language of one or two of the manuscripts and discussing multiple litterae and spelling features. Findings concerning comparison of different versions of the same text appear either as general observations or occasional remarks mentioning a single feature, and rarely concern all the four manuscripts.

The perspective adopted in the present study is a comparison focusing on a single group of features – 'h', 'ȝ' and related spellings – in all the extant versions introduced above. Both 'h' and 'ȝ' have been repeatedly discussed in connection with litteral substitution, mainly in Trinity (see Laing & Lass, 2009).

3. Method

The sample analysis was conducted in three stages. First of all, the inventories of litterae from selected texts were compared. The objectives at this stage were to obtain a global picture of the distribution of litterae in the texts, which could be refined at subsequent stages. Comparison of inventories in the tool displays statistical data on frequencies and compares relative frequencies in different texts. It can also list alternatives of a selected littera found in other texts and visualise the data as a network.

The second stage focused only on a comparison of representations of the sounds written as 'h' or 'ȝ' in selected texts from MS Cotton Caligula A.ix with corresponding representations found in other manuscripts available in the LAEME corpus. These specific litterae were chosen because certain words in which they are used are good examples of variation on the level of sound as well as writing. This part of the analysis was structured around "item lists". The concept of an item list is a familiar one in historical dialectology. The usual method is to define a list of units based on a criterion, such as shared historical sound value, e.g. [f] in the initial position, and compare the occurrences of the items in a single text or multiple texts. As yet, the tool does not include data about OE source forms or presumed sound values, which would enable construction of item lists of this sort. It does, however, enable compilation of item lists based on shared littera or a combination of litterae, which can be further filtered by context of the littera, its position or occurrence in

a manuscript or manuscript metadata. For instance, it is possible to define a list of "all items containing initial 'sch' in text #242" or "all items with 'h' followed by 't'". Item lists can be stored and used to search for the items in the manuscripts or to construct maps.

The goal of the second stage of the analysis was to compare instances of items on the item lists found for different versions of the lyrics under examination, and identify patterns of distribution of the variants across texts or possibly related unusual spellings. This was done using a feature of the tool which allows the user to display multiple texts side by side and highlight items from item lists.

The results obtained with item lists served as input for the final stage of the analysis. The objective here was to complement the results with relevant maps and other additional data retrieved from the database.

The next section briefly explains the structure of the spelling database and describes the functionalities which were especially relevant for the completion of the tasks outlined above.

3.1. Spelling database

The purpose of the spelling database discussed here is to contribute an additional layer of data to the LAEME corpus, which would facilitate research into early Middle English texts and dialects. This means that all the data available in the database is originally from LAEME, however, the structure of the database opens up new possibilities for searching and quantifying data. The processing of data from LAEME consisted in "alignment" of the different spelling variants of a word, indicating which segments correspond to one another, as illustrated by the example of selected forms of FIRE/N below:

```
f | ie | r | e
f | uy | r | e
v |  e | r | _
u |  u | r | _
```

The underscores in this notation represent an empty position (slot). The technique is very close to the so-called *grapho-phonological parsing* employed in

other projects, notably *From Inglis to Scots* (FITS)[5], except no explicit sound value is assigned to the individual segments. The alignment is based predominantly on the comparison of the individual forms in the group. Basic queries exploiting this data structure can be formulated in the following manner:

- List all the litterae used interchangeably with 'h' (in text #246).
- List all the slots (and associated items) in which 'g' is used interchangeably with 'ȝ'.
- List all alternatives of 'h' in positions where it alternates with 'ȝ'.

Such queries can instantly return data which would otherwise require reading through a text or running multiple queries in LAEME. Their application in the context of specific tasks will be demonstrated later on.

Efficient querying and use of the segmented data required the construction of an interface tailored to the new data structure. In addition to simple database queries, the interface offers more complex features, namely mapping, network visualisation and so-called *text profiles*, which can be displayed side by side and compared. The presentation of the application in this section is limited to the *text profile* component, because it served as the main tool for the sample analysis and it includes some elements found also elsewhere in the application. A full description of the tool is going to be published along with its electronic version.

3.1.1. Text profile[6]

The purpose of the screen text *profile* is to offer tools for a comprehensive analysis of a text language, but it can also be used as a brief overview of the spelling features of a manuscript. It can serve as a good starting point for analyses, because it can suggest what to focus on and prompt features of the text which require explanation.

---

5   FITS is a project of the Angus McIntosh Centre in Edinburgh, which consisted in the construction of a spelling database from the *Linguistic Atlas of Older Scots* (LAOS), which means that FITS builds on LAOS data in similar ways as the present project builds on LAEME data.
6   The term is derived from linguistic profile developed for LALME (Laing & Lass 2013: 2.5).

The screen combines the inventory of litterae, sets of alternating litterae, and the complete text of the MS. The picture below shows the text profile of the Cotton version of *The Latemest Day* (LAEME #242):
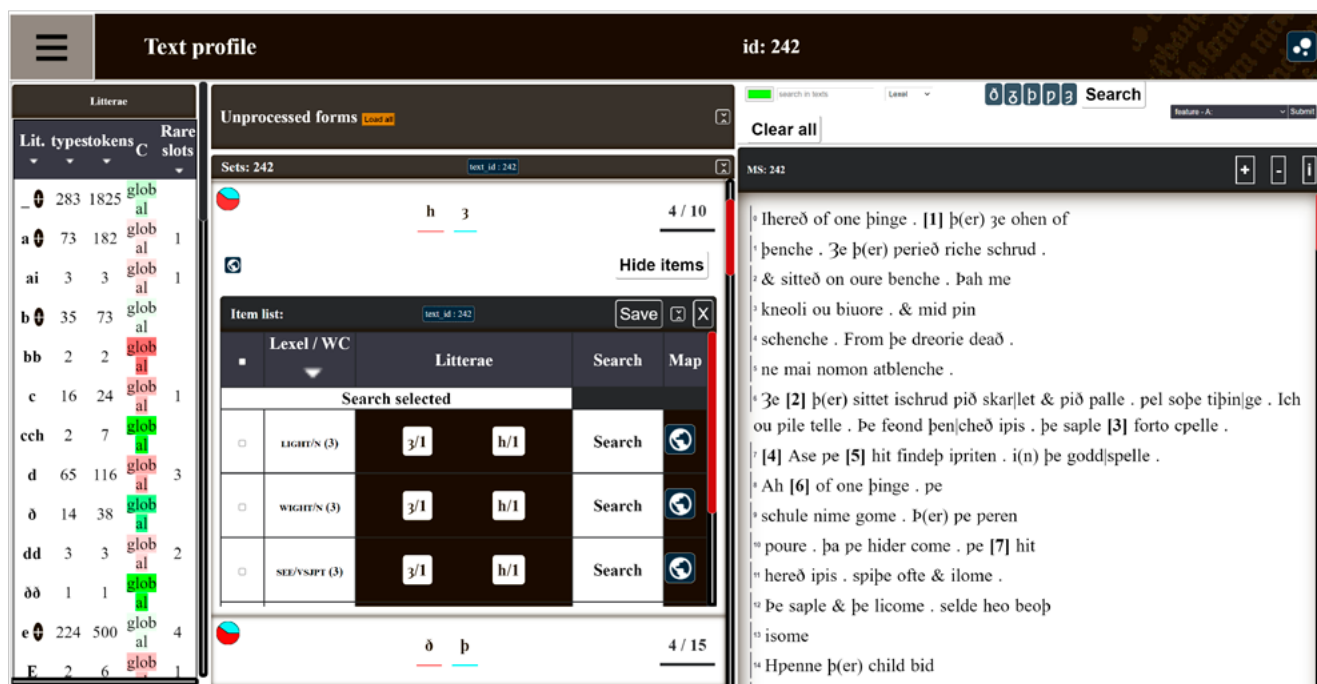

Figure 1: Text profile of the Cotton text of *The Latemest Day*.

The inventory of litterae is found on the left, the middle section "sets" displays sets of litterae used interchangeably (i.e. in the same positions) by the scribe and the right section displays the full text of the MS. The components can interact with each other.

Observations which can be made straight from the inventory and sets of alternating litterae, sometimes correspond to editor's notes on spelling such as "the scribe prefers *wyyn* to 'w', the scribe does not use 'th'". The interactivity of *text profile* makes it possible to relate such observations to specific places in manuscript texts as well as data from other texts in the database.

3.1.1.1. Litterae inventory[7]

The inventory of litterae employed in the given manuscript is displayed in tabular form. The first column gives the littera, and the subsequent two columns indicate its type/token frequencies.[8] The colour of the rectangle in the third column ("C(omparison)") reflects the relative frequency of the littera compared to its average relative frequency in the remaining texts in LAEME.[9] As such, it points to litterae which are either conspicuously rare in the text (marked with red colour) or, contrarily, comparatively more frequent (marked with green colour) and therefore likely to deserve the researcher's attention. For instance, 'p' is relatively frequent in text #242 and therefore it is displayed with a green rectangle, while the relative frequency of 'w' (2 instances only) is clearly below average and is hence displayed with a red rectangle.

Two of the columns in the inventory of litteare are interactive. The first column (littera) can be used to filter the list of sets and show only those containing the littera. For instance, 'sch' in text #242 is found in two sets – {*sch*} and {*sch, sc*}, i.e. it alternates with 'sc'. The column "Rare slots" points straight to a list of items in which the littera rarely appears, but this feature is not relevant here.

3.1.1.2. Sets

Sets show which litterae sometimes alternate with one another in the same slot, i.e. the same position in the same word. One such alternation in text #242, visible in the picture, is the alternation of 'h' and 'ȝ'. The complete list of items relevant for each set can be loaded straight into the *text profile* screen. The picture shows the list for the set {h, ȝ}, which comprises four items – WIGHT/N, LIGHT/VPT, SEE/VSJPT and BRING/VPP. Instances of the individual items can be highlighted in the manuscript text using the link "search" dis-

---

7   The article does not include a full description of this component because not all the functions are relevant for the present analysis.

8   A position (slot) in a specific item (e.g. the initial position in FIRE/N) is counted as one type and a single instance of a littera is counted as one token. For instance, type/token frequency 2/4 for 'f' could correspond to 2 occurrences of 'f' in FIRE/N plus two occurrences in LOVE/N.

9   The word "global" indicates that the frequency used as a reference is calculated from the whole database as opposed to from a specific text or subset of texts.

played with each item. The blue icons serve as links to maps. Any list of items or selected items from the list can be labelled and stored for future use. This feature is especially relevant for the sample study.

3.1.1.3. Manuscript text

The third component of the *text profile* screen is manuscript text displayed along with basic information about the text taken from LAEME (not visible in the picture). Words in the text can be highlighted in different colours either by selecting items or by searching the text by *lexel*, *grammel*, *form* (or a combination thereof). Regular expressions[10] can be used in these searches.

Moreover, it is possible to highlight all items present on a previously stored lists of items (see above). For instance, after storing a list of all items having the alternation of {*h*, *ȝ*} the user can highlight all the items present in any text or group of texts displayed in the application, and examine their realisations. In the case of *The Latemest Day*, items lists taken from the Cotton version of the lyric were used to search the other versions of the text (see below).

3.1.2. Text comparison

Text profiles can be displayed side-by-side. The picture below shows the text profiles of the four manuscripts (Cotton, Trinity, Jesus and Digby):

---

10   Strings or sequences of characters representing search patterns to match text results. The patterns may involve, among other things, the use of "wildcard" characters, e.g. the pattern "[fvu]a" matches "fa", "va" or "ua".
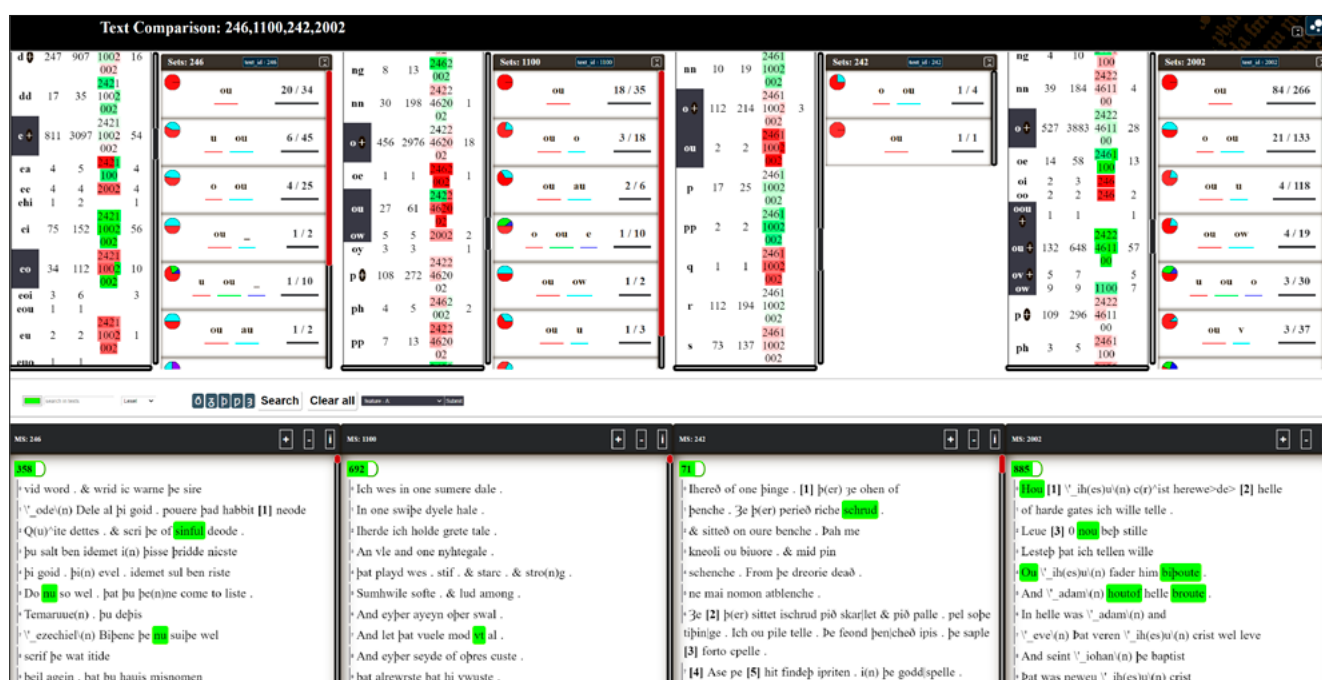
Figure 2: Interface screenshot – text comparison.

The functionalities are very similar to those of the *text profile* of a single text, except any actions (such as queries in the text or filtering of sets) affect all the displayed profiles. Whenever a littera in an inventory is clicked, the corresponding litterae in the other inventories are highlighted. The screenshot above was taken after clicking 'ou' in text Digby (#2002) and it can be seen from the picture (among other things) that Jesus sometimes has 'ow' or 'o' at positions where Digby uses 'ou'.

The visualisation (red-green rectangle) of littera relative frequency is based on relative frequencies in the compared texts instead of the average values for LAEME as a whole, and a separate rectangle is displayed for each of the compared texts. For instance, the data for 'eo' in Trinity (text #246) suggest that 'eo' is relatively less frequent in this text compared to Cotton and Jesus (#242 and #1100) but relatively more frequent compared to Digby (#2002).

3.2. The analysis of {*h, 3*}
This section specifies which pieces of data were analysed at the individual stages of the sample analysis.

3.2.1. Inventories

The inventory of litterae in the Cotton version of *The Latemest Day* was compared with inventories of Trinity (scribe A), Jesus and Digby. This comparison has to be treated with caution because the Cotton version of the *Latemest Day* is the only one which is tagged as a separate file in LAEME. This means that the inventory was in fact compared with inventories compiled for multiple lyrics in a single hand, e.g. all the texts copied by the Jesus scribe. The results were inserted in a table showing the litterae from Cotton, and their alternatives in the other texts.

3.2.2. Item lists

The next step was to retrieve item lists for 'h' and 'ȝ' from each of the texts found in Cotton. These item lists could be further combined with item lists from other manuscripts, but this was not done within the present analysis.

Initially, *The Latemest Day* from Cotton was compared with Trinity, Jesus and Digby and the same was done for *Doomsday*. *Orison to Our Lady* is missing from Digby and it was copied by a different scribe in Trinity, so it was compared with these versions plus the version in Royal.

The previously stored item lists were used to find and highlight all the relevant variants in the compared texts. Next, correspondences between different realisations of a single item were identified, categorized and inserted into a table.

3.2.3. Additional searches

Searches performed at the third stage of the analysis included highlighting in the manuscripts, searches for items and sets of alternating litterae in the database, and mapping. Practical examples presented in this paper also cover the use of item lists. The following tasks will be described:

a) Compilation of an item list from Cotton, plus a search for the items in the compared MSS.

b) Using regular expressions in combination with item lists to examine the distribution of forms with <ii> in Digby.

c) Looking at alternation of litterae in *text profile,* in order to better understand the use of 's' in the spelling system of Trinity scribe A.
d) Searching the database for more items with <st> for historical <ht>.
e) Searching the database for spelling variants (litterae) possibly related to <þt> for historical <ht>.
f) Generating a map from an item list.

4. Results

The presentation of results is partly structured according to the stages of analysis defined in the previous section. The first part discusses the comparison of inventories. The second part merges the results obtained using item lists with additional queries performed at the final stage of the analysis. It first summarizes the patterns of distribution of variants and the most interesting deviations from the patterns. Then it explains how selected tasks were completed using the tool.

4.1. Comparison of inventories

Correspondences between litterae visible in the comparison of inventories are presented in tabular form below. The first column gives the littera in the Cotton version of *The Latemest Day,* which is followed by a bracketed list of litterae which appear at the same positions as the first littera in Cotton. The remaining columns list the litterae from manuscripts Trinity, Jesus and Digby used at the same positions as the littera from Cotton. Although the pilot study focuses on 'h' and 'ȝ', the table covers also other litterae mentioned in previous research in order to better illustrate the possibilities of comparison.

Table 3: Comparison of inventories – version of *The Latemest Day* from Cotton and Trinity, Jesus and Digby; grey background indicates that the littera was marked as relatively more frequent in the tool; black border indicates that it was marked as less frequent.

Cotton	Trinity, scribe A	Jesus	Digby
'ʒ' ('h')	'ʒ', 'þ', 'u', 'th', 's', 'c', 'g'	'y', 'h'	'ʒ', 'u', 'h'
'p' ('þ', 'v', 'u')	'v', 'u', 'w', 'þ', 'p'	'w', 'u', 'v', 'þ', 'p', 'y'	'w', 'v', 'uu', 'þ', 'y', 'i'
'ð' ('þ', 't', 'd')	'þ', 't', 'd', 'z'	'st', 'þ', 't'	'þ', 't', 'd', 'ð'
'w' ('p')	'w'	'w'	'w'
'v' ('u')	'ou', 'v', 'u'	'v', 'u'	'ou', 'oou', 'v'
'sch' ('sc')	's', 'ch'	'sch', 'sc', 'ch'	'sh', 's', 'ch'
'sc' ('sch')	'sc'	'sch'	'sh', 's'
'q'	'q'	'q'	'q'
'ou'	'ou', 'o'	'ou', 'o'	'ou', 'o'
'hp'	'p', 'w', 'v'	'hw', 'w'	'v', 'vu', 'uu', 'u'
'h' ('ʒ')	'c', 'w', 'u', 'þ', 's', 'h', 'sc', 'ch', 'c'	'w', 't', 'h', 'hw', 'c', _	'ʒ', 'þ', 'y', 'w', 'u', 'h', _
'g'	'k', 'h', 'gk', 'g', 'ck', 'cg', 'c'	'g', _	'gg', 'g', _
'ff'	'ff'	'ff'	'ff'
'f'	'v', 'w', 'u', 'f', 'ff'	'v', 'w', 'u', 'f'	'w', 'v', 'f'
'eo'	'u', 'eoi', 'eo', 'ee', 'ei', 'e'	'u', 'o', 'i', 'eo', 'e', 'a'	'u', 'oe', 'o', 'i', 'e', 'ee', 'a'

The table reflects some of the observations quoted in the theoretical part of this article. The most salient finding is probably the clear division between 'p' in Cotton and 'w' in the other texts. The near absence of 'w' and 'q' from Cotton is paralleled by 'v' (3) and 'sc' (1). 'Sc' differs from the other two litterae in that it is an obsolescent variant and its frequency is markedly higher in Trinity than in the other texts. Another rare spelling in Cotton is 'ea' (10 instances in 5 types). 'Ea' in NEVER (*neauer*) has also been quoted among the distinctive features of language C2 (Cartlidge, 1997: 254). Despite the low number of occurrences Cotton has the highest relative frequency of this digraph of all the texts. The instances of 'ea' in Trinity and Jesus are found in different words, which is why 'ea' is not among the alternatives in these two MSS.

On a more general level, the table illustrates the diversity of eME spelling. Trinity appears to be slightly more prodigal than Digby and Jesus, which also agrees with previous findings. The variation may be indicative of sound

changes. For instance, Trinity has 'u', 'w' alongside 'h' as the alternatives of 'h' in Cotton, which could reflect the change of [ɣ] > [w] (Minkova 2014: 83). Observations of this kind can suggest possible directions of further analysis. The table can serve as a reference for the compilation of item lists. The present study covers only 'ȝ' and 'h' for illustration, but the same method could be applied to other litterae within a more comprehensive study.

4.2. Item lists

The results discussed here are limited to items spelled with 'h' and 'ȝ' in Cotton. First of all, the patterns of correspondences between spelling variants are presented. The patterns are primarily based on LD and they were mostly present also in DMD. The commentary on each pattern mentions deviations from the pattern, or interesting forms which required closer examination. The next part gives several practical examples of the application of the tool in searches for answers to some of the questions which emerged during the analysis.

4.2.1. The patterns

Five patterns of correspondences between the texts were identified. The first three patterns concern instances of 'h' ('ȝ') before final '-t(e)' and the other two concern instances of intervocalic 'h' ('ȝ') and initial 'ȝ'. Both 'h' and 'ȝ' are examined together because the patterns of correspondences partially overlap. Items which had 'h' in all of the texts are not discussed.

The first pattern reveals that -*oht* (or -*aht*) (-*oȝt*) in the Cotton version corresponds to -*ouht* in version Jesus and -*out* in versions Digby and Trinity. The table below illustrates this distribution.

Table 4: Distribution of forms, pattern I – selected examples.

I	Cotton	Trinity	Jesus	Digby
THOUGHT/N	þoht	þoutt	þouht	þout
WORK/VPP	iproht	wrout	iwouht	iwrout
BRING/VPP	ibroȝt	ibrout	ibrouht	ibrout
SEEK/VPP	isoȝt	isout	isouht	isohut

A distribution of forms similar to pattern I is found in items ending with
-aht(e) in Cotton:

Table 5: Distribution of forms, pattern II – selected examples.

II	Cotton	Trinity	Jesus	Digby
SEHTAN/VPP-PL	isahte	isaiste	Isauhte	isaute
BETAÉCAN/VPT13	bitahte	bitaiste	bitauhte	bitaute
ǼHT	ahte	haiste	Ayhte	haute
ǼHT	x	x	Ayhte	hayte

The forms in Jesus and Digby are analogical to the first group, except for *hayte*
(ǼHT) in Digby, which has y instead of the usual *u*. This makes *hayte* closer to
the forms in Jesus and Trinity, which differ from the first group. Final *-hte* in
Cotton corresponds to *-ste* in Trinity if preceded by 'a'. The same spelling in
Trinity is found after 'i':

Table 6: Distribution of forms, pattern III – selected examples.

III	Cotton	Trinity	Jesus	Digby
LIGHT/N	liȝte	liste	lyhte	Liȝtte
RIGHT/N	riȝte	riste	ryhte	Riȝtte
SIGHT/N	siht	siste	syhte	siiþe

A possible explanation given by Laing & Lass (2003) points out that "<s> here
does not imply [s], but is apparently an inverse spelling based on the Old
French sound change [st] > [xt~cˌt~ht]" (Laing & Lass, 2003: 263). The only
form deviating from the pattern is *siiþe* from Digby. As for variants in texts
other than LD, Trinity scribe D, who writes *Orison*, sometimes ends the items
of this type in simple <-t>, <-d> rather than <-st¬>. The Royal version of OL has
several forms with <-þt> and others in <-td>.

    Intervocalic 'h' in Cotton mostly corresponds to 'w',' 'u', or 'uw' (Trinity),
'uu' (Digby) in the other texts (pattern IV).

Table 7: Distribution of forms, pattern IV – selected examples.

IV	Cotton	Trinity	Jesus	Digby
ÁGAN	x	x	owen	x
LOW/AV	lohe	louwe	lowe	(stille)
FORSWALLOW/VI	forspolehen	firsuoleuen	forswolewe	forsuoleuen
MAY/VPS21	mohe (þe)	mou (we)	muwe (we)	we mouuen

These spellings are likely reflections of the well-known change of [ɣ] > [w] (Minkova 2014: 83) mentioned above. Although OE source forms of these items had mostly 'g', and the items from group I above had mostly 'h', the two patterns look very similar in eME.

The last group is somewhat less orderly. 'ʒ' in Cotton regularly corresponds to 'y' in Jesus. Digby preserves 'ʒ' initially but it has 'i' in the intervocalic position.[11]

Table 8: Distribution of forms, pattern V – selected examples.

V	Cotton	Trinity	Jesus	Digby
SEE/VSJPT13	iseʒe	loke	iseye	iseie
YAWN/VPS13	ʒeoneþ	gonet	yoneþ	ʒeneþ
GIVE/VI	ʒeuen	þewen	yeuen	ʒeuen
EYE/N	eʒe	eþen (npl)	Eye	eie
SEE/VSJPT13	isehe	seiþe	iseye	iseie

Trinity has initial 'g' in YAWN/VPS13 but initial 'þ' in GIVE/VI. The second occurrence of SEE/VSJPT13 (in the last row) is especially interesting. Firstly, the 'h' / 'ʒ' alternation in Cotton is found intervocalically and not before 't', which would be the usual context in this text language. Secondly, the form in Trinity also has <þ> but this time it is preceded by <i>. If the <þ> stood simply for 'i/y', which the <þ>/<y> confusion postulated for the scribe suggests, the "intended" spelling would have to be *seiie/seiye*, which is not absolutely inconceivable. However, there is only one token of *-eiye-* and two for *-eii-* in

---

11   All the instances of 'ʒ' in Digby are either in the initial position or before 't'.

LAEME.[12] It is at least tempting to read *seiþe* as a hybrid form between *isehe* and *iseie*, where the <i> roughly corresponds to the <i> in *iseie* and <þ> to the <h> in *isehe*. Spellings of the type "*iseihe*" (*ei+h*) are more widely attested.[13] The latter interpretation would imply a different sequence of literal substitutions leading to the use of <þ> in this position.

Orison to Our Lady

As OL is missing from MS Digby 86, the versions in Cotton, Jesus and Trinity were compared with Royal localised in Wiltshire. Also, the version in Trinity was copied by a different scribe (scribe D).

The items in Cotton and Jesus mostly conform to patterns I and III, but corresponding forms in Trinity and especially in Royal are markedly different. All the relevant forms from these two manuscripts are included in the table below.

Table 9: Unusual forms in *Orison to Our Lady* (MSs Trinity and Royal).

	Trinity (scribe D)	Royal
BRING/VPT	brut	brout
GIVE/VSJPS	þef	x
LIGHT/N	licte	lyzt
DIHTAN/VPP	idiit	ydyyt
BRING/VPT12	brutis	brovtest
BRING/VPT	brout	x
NIGHT/N	nitf	nyhyt (rhyming)
BRIGHT/AJ	brit	brytd
MAY/VPS12	mit	mytd
RIGHT/N	rid	ryhyt (rhyming)

---

12   *Greiyed* (GREIÐA) in #1700, *neiiʒ* (NIGH) in #1600 and *leiid* (lay) in #2002. Only the first two examples were retrieved from the spelling DB because it is unable to perform searches crossing the morpheme boundary in *leii+d.*

13   A search in the DB returned e.g. *eiʒe, eihen* (EYE), *dreihen* (DRAW), *sleih* (SLAY), *deih* (DUGAN).

The form *brout* in Trinity (D) is the only one which perfectly agrees with the forms from text copied by scribe A. *þef* (give) is similar in the use of <þ> for expected <ʒ>. *Brit*, *mit*, *rid* and *nitf* do not have A's typical <st> at the end. The most interesting form by far is *idiit* with its direct correspondence to *ydyyt* in R. The fact that both are uncommon and found at the exact same place in the text is strongly suggestive of a shared source. Moreover, the <ii> spelling is reminiscent of *siiþe* (SIGHT/N) found in the Digby version of LD. The same is true of *heyte* (ÆHT), similar to *hayte* in Digby.

The forms *ryhyt* (RIGHT) and *nyhyt* (NIGHT) are probably related to *dyyt* and all of them represent an unusual yet systematic spelling practice. Besides <þ> for historical 'h', the Royal scribe also uses 'h' as the initial littera in give (*hyef*, *hyf*).[14]

4.2.2. Sample tasks

The next part of this section demonstrates the use of the tool on several short tasks associated mainly with the previously mentioned forms in <st> and also the forms with <ii>. The first example demonstrates the generation of an item list and its subsequent use. The following examples focus on searches for additional data in the examined texts within *text profile*, specifically the distribution of forms with <ii> in Digby and alternations involving 's' in Trinity, scribe A. The remaining tasks deal with searches for additional data on the alternatives of <-st> in the database, including the construction of maps.

4.2.2.1. The use of item lists

The patterns of distribution presented above were identified using item lists. This example shows how to generate and use the list of items which are spelled with 'h' in the Cotton version of *The Latemest Day*. First of all, the text was displayed within text profile screen. 'H' in the inventory of litterae was clicked to display sets of litterae alternating with 'h'. The items under the set {*h*} (which also covers all the alternations) were selected and saved under the label "H in #242".

---

14   The confusion of <h> and <þ> is discussed by Laing & Lass (2009).

The four LAEME files containing LD were displayed side by side. The item list was selected from the blue drop-down menu at the top of the screen and submitted. As a result, all the items from the list were highlighted in the texts, as shown in the picture below.
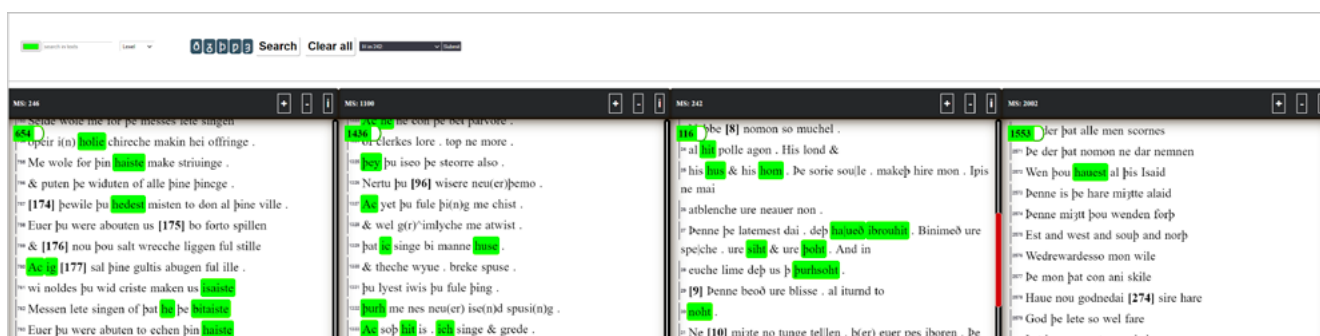

Figure 3: Interface screenshot – applying item list to texts.

This method is very close to simple reading. The main virtue of item lists is that they are easy to compile and manage, which simplifies and speeds up the comparison of multiple texts.

4.2.2.2. Regular expression search in manuscripts

The next task is similar to the first one in that it also involves highlighting in the text. Its objective was to examine forms possibly related to *siiþe* (SIGHT/N) found in Digby, which deviate from pattern III. In order to take a closer look at <ii> in Digby, the relevant words in the text were highlighted and examined. First, <ii> was selected from the inventory of litterae to highlight all the items which sometimes have <ii> in green (light grey). Next, all the *forms* with actual <ii> were highlighted in blue (dark grey) using a regular expression. The picture below shows how the regular expression search was set up, and also some of the highlighted forms in the manuscript.
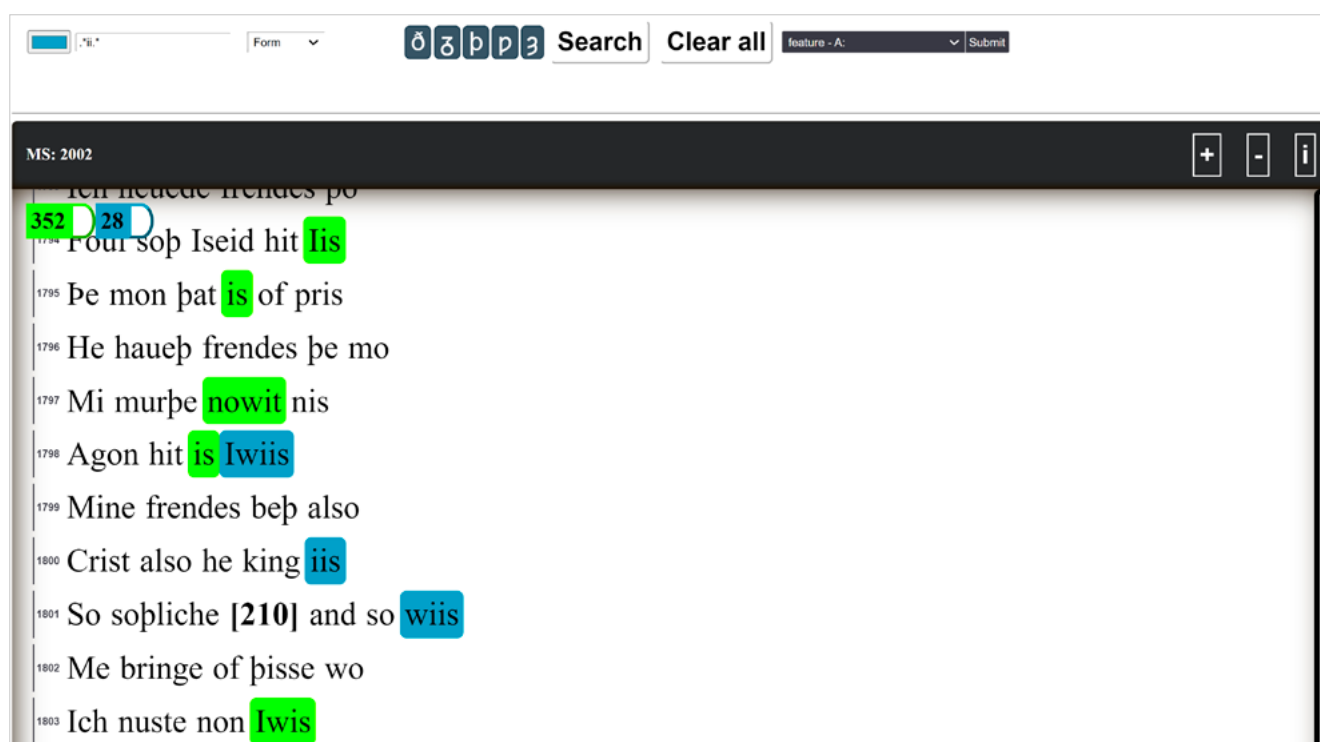
Figure 4: Screenshot – highlighting in text.

This technique allows the researcher to observe the spellings of the examined items as well as the distribution of variants in the manuscript. Highlighting in two colours ensures that instances of the concerned items which have spellings other than <ii> are also noted. If just the forms in <ii> were highlighted it would not be immediately obvious whether their absence from a specific passage in the text is in fact due to the absence of the items in which the spelling appears.

A number of similar forms with double <ii> are found in Digby in the second half of the text. The most conspicuous concentration of such forms appears between lines 2713–2733[15] and the other forms are scattered across the rest of the text. A possible explanation may be that the scribe learned the spelling from his exemplar and either began to use it occasionally or sometimes left the exemplar forms unchanged. This explanation would support Laing's (2000) claim that the scribe of Digby sometimes preserved spellings found in his exemplar.

---

15   Line numbers correspond to the text as displayed within the tool.

4.2.2.3. Alternation within a single text

The next example concerns the analysis of the spelling system. One of the sources which can help us understand the use of a certain littera in a given text is the list of litterae alternating with it. In order to better understand scribe A's use of 's' in <-st>, the list of sets involving 's' was examined. The list can be displayed within *text profile*. The picture below shows the inventory of litterae with alternatives highlighted in grey, a part of the corresponding sets, and also the only item having the alternation {*s, c, sc*}- FLESH/N.

Figure 5: Screenshot – litterae alternations in a system.

The results revealed that, among other litterae, 's' in scribe A's system very occasionally alternates with 'th' (in two different words), 'þ' (1 word) and also 'f' (6 words). The last, however, is probably connected with the similarity of letter shapes for <f> and <s>. Unlike with the previous examples, the tool provides much faster access to data, compared to reading. Unfortunately, there is no straightforward way to check whether the alternations are common or rare in other texts, although the required data is in theory retrievable from the database. Adding such functionality is mainly a matter of extra programming and it is included among possible future upgrades of the tool.

4.2.2.4. *Search for alternations of litterae*

This example demonstrates how to search for additional items with a certain spelling, in this case <-st> spellings for historical <-ht>, originally found in Trinity (scribe A). The advantage of the tool is that it is possible to search for a specific alternation of litterae (in a specific context). In this particular case, the query was formulated as follows "list all the items in which 's' alternates with 'h' before 't'". The screenshot below shows this query in the tool along with its output.

Figure 6: Screenshot – query for the s/h alternation before t.

The input for the query is found in the topmost part of the picture. Only two fields needed to be filled. "*s, h*" was entered into the field "main" and "*t*" was given as the "following littera". The result is a table with all the items in which 's' and 'h' alternate before 't'. The individual variants are listed in the third column. The blue "globe" icons serve as quick links to maps.

A drawback of the tool is that the scope of searches is restricted to single morphemes, i.e. the query above does not return examples of {*s, h*} in morpheme-final position which are followed by the ending *-t(e)*, e.g. *sous+t, soh+t* (SEEK/VPP).[16] It can, however, target morpheme-final position in searches.

---

16   A solution to this issue would be to extract the sequences of litterae at morpheme boundaries and insert them to a special table, however, this procedure is not completely straightforward.

The query for "*s, h*" in this position returned a list of items comprising almost exclusively preterites and participles. Both of the item lists were stored and used to construct maps, which showed that the <-st> variants are found mainly in the Trinity MS, scribes A, B and D.

4.2.2.5. Search for sets

The next example deals with a way of exploring potential connections between spelling variants using queries for *sets*, i.e. lists of litterae appearing at the same position.

The texts in Royal have the forms *bryþt* (BRIGHT), *soþte* (SEEK) and *ariþt* (ARIGHT). It is questionable whether these forms should be interpreted as related to the type *briht*, **bryyt* (analogical to *dyyt*), in which case the second <y> could be a *thorn*, or rather *brytd* (also in Royal). All the forms with <-þt> appear in the second half of the text, while *ryhyt*, *nyhyt* and *dyyt* are found in the first half.

In order to search for similar forms, as well as more variants elsewhere in the database, the following query was used "find all the alternatives in positions where 'þ' alternates with 'h' before 't'".

Figure 7: Screenshot – alternatives of þ/h before t.

The results are presented as a list of all possible sets of litterae found along with 'þ', 'h' before 't'. The numbers on the right give the number of slots (types) at which the exact combination of litterae appears / the number of

tokens. For instance, the second set {*h, _, ȝ, ch, g, þ, ȝh*} occurs at two different positions and the total number of tokens is 72. Only the first three sets are visible in the picture. There were in fact many more combinations, but the only littera which is not found in the sets in the picture was 'ȝ'. The litterae potentially closest to 'þ' are 'th' and 'y'. The next step was to identify the texts in which these variants occur. This task is still relatively time consuming as there is currently no direct link leading from *sets* to *texts*.

The forms in <-þt> were found also in the text of *Laȝamon B* (LAEME #280) localised near Royal in Wiltshire, which very frequently uses <-þt> (alongside <-ht>) for historical <-ht>. There might be a further possible connection with the forms in *-tht* (*lithte, mithte*), as <-th> is the canonical replacement for *thorn*. The forms with 'th' are very rare and found mostly as marginal variants.[17] The forms with *y*, e.g. *fleyt* (FLIGHT/N), *areyt* (ARIGHT/AV), which are also very rare, appear in Oxford, Bodleian Library, Additional E.6, entry 2 (LAEME #161). This manuscript shares the text *The XV signs before Doomsday* with Digby but no instance of a direct correspondence between <y> and <þ> in Digby was found. It needs to be pointed out that *y* is a notoriously problematic *littera* as it has vocalic as well as consonantal uses and it can be very difficult to tell the two apart.

4.2.2.6. Mapping tool

The final example demonstrates the construction of maps. It has been stated that 's' before 't' sometimes alternates with an empty position in forms like *rid* (RIGHT), *mit* (MAY) used by Trinity scribe D. The relevant items were stored as a separate item list, and their variants were plotted on the map. The map generated in this way is displayed below.

---

17   The text with the highest number of such forms is MS Cambridge University Library Ff.II.33 (*Bury documents*, LAEME #1400).
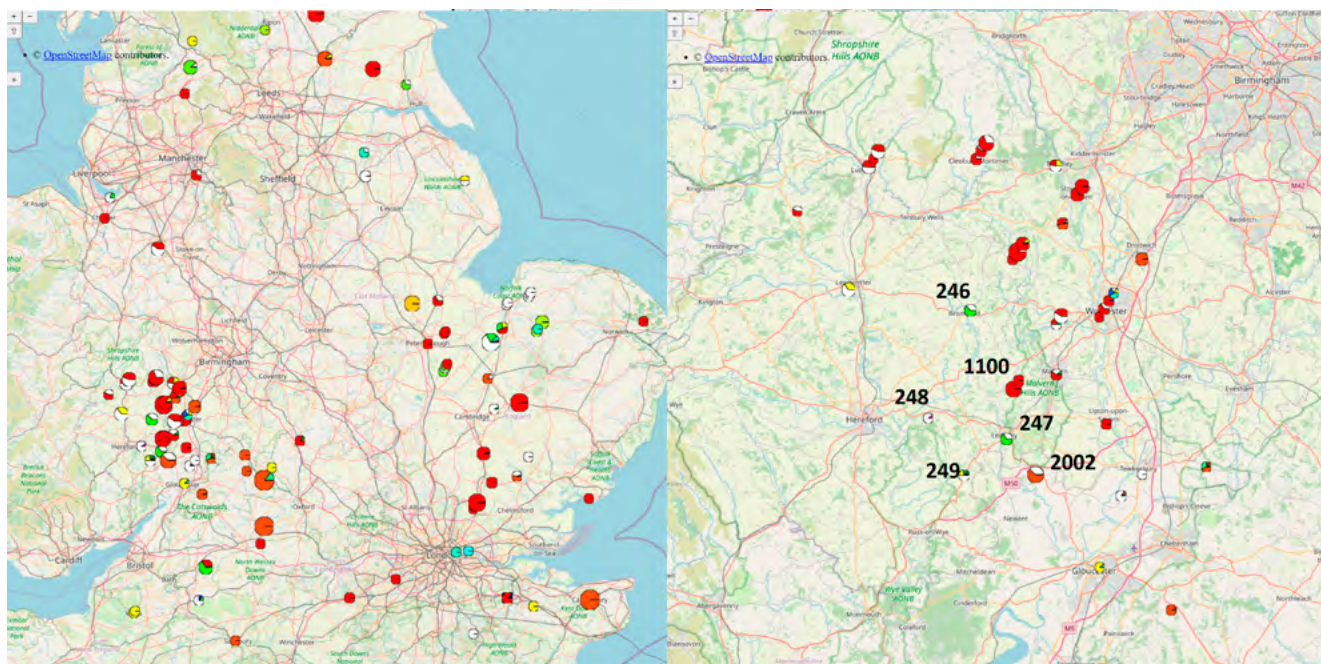
Figure 8: Map – items with \<h\> dropping before \<t\>.

Each pie chart on the map represents one LAEME file, and the colours correspond to different litterae found at the position before the final 't' in the selected items. The forms of the type *rid* have the position empty, which is displayed in white on the map. The map shows that such forms appear in the North-East, the East as well as the West Midlands. The second picture is zoomed in on South West Midlands (SWM), showing the location of the texts from Trinity (#246-#249), Jesus (#1100) and Digby (#2002). The map shows that multiple texts in the area including all the Trinity texts as well as Digby contain instances of an empty slot before 't' but only #248 (Trinity scribe C) uses this variant exclusively. The black colour in #246, #247 and #249 represents 's', blue in Digby represents '3' and red represents 'h'. Jesus (#1100) has consistent 'h', which makes it stand out a little. The likely explanation seems to be that the scribe of Jesus systematically replaced the forms in his exemplar with 'h' whether or not the corresponding sound was pronounced while the scribes of Trinity and Digby probably relied more on their actual pronunciations. The tool can also generate separate maps for the individual items on the list, which may reveal marked differences between them. For instance, the texts in SWM almost universally drop the segment before 't' in NOT but never in LIGHT/VPT.

4.3. Summary

The five patterns of the distribution of variants agree with previous findings concerning copying strategies. The analysis identified several possible connections between similar forms found in multiple manuscripts, related by content or localisation. The most prominent of these groups comprises spellings for historical *-ht(e),* which include <-ʒt>, <-th>, <-þt>, <-td>, <-st>.

The forms with <-þt> are the dominant variants in the text of *Laʒamon B* (where they alternate with <-ht>) and very rare in other texts. The Royal version of OL has the highest frequency of such forms and they appear alongside forms in <-y(h)y->. The <-yy-> spellings are the likely source of <ii> in Trinity and probably also Digby. These connections are very intricate and there are multiple plausible ways in which one of the forms might have been "translated" into another, For example. <-iþt> could become <-yyt> in a system which does not distinguish <y> from <þ>, <-ʒt> might become <-yt> if the scribe read 'ʒ' as [j] etc. Further analysis would be needed to clarify the nature of the connections.

5. Conclusions

The article presents a research tool which seeks to provide new ways of accessing and exploring the data available in LAEME. It is built around an additional layer of data which maps correspondences between segments in a group of spelling variants. Practical application of the tool has been demonstrated on a pilot study dealing with the use of 'ʒ', 'h' in a group of ME texts related by content. The results were discussed in the context of previous research in order to point out the connections between searches in the tool and established methodology and to assess its strong and weak points.

The comparison of inventories undertaken in the pilot study reflected mostly, but not exclusively, spelling features, which have already been pointed out in previous studies. The main advantage of inventories is that they instantly quantify data, which could otherwise take a long time to collect. The comparison of relative frequencies and alternatives would be more precise, if copies of only the same text sharing roughly the same items were compared.

The results obtained with item lists also partly overlap with previous findings in that they mention similar or identical features (the alternation

of 'þ' / 'ʒ', the range of representations of historical *-ht*, possible relations between <y> and <þ> etc.).

The fact that multiple copies of the texts were included in the analysis somewhat broadened its scope. As a result, previous findings could be complemented with specific observations on possible connections between the copies. The main findings concern the distribution of spelling variants with <ii> (<yy>), the <-st(e)> spellings for expected *-ht(e)* plus several isolated forms which seem to stand out, e.g. *hayte* (ÆHT, LD in Digby) and possibly related *þey* (THOUGH, Royal), *bryth* and *bryþt* (BRIGHT/AJ, Royal).

Dynamic item lists are fast and easy to use. As the present analysis focused primarily on comparable items in specific texts, item lists were generated only from the Cotton manuscript. More extensive analyses should ideally combine item lists from different texts or item lists generated from the whole database, e.g. item lists of all items in which 'h' ever alternates with 'ʒ' etc.

As for the main weaknesses of the tool, it is still relatively difficult to access certain useful pieces of data, especially a complete list of texts in which a given variant (e.g. GIVE with initial <þ>) appears. The case of *-th/-tht/-þt/-td* stressed the need to analyse larger segments, like sequences of a few letters, rather than single letters or digraphs, which is yet underdeveloped in the tool, and queries are limited to single morphemes. The analysis also revealed several cases of incorrectly aligned segments in the spelling variants.

The most useful queries, whose output is difficult to match without the tool, are searches for "sets" of similar spelling variants found at a specific position, e.g. 'th', 's', 'þ' before 't' in words like LIGHT or BRIGHT. Instead of searching for the forms of individual items with historical *-ht* it is possible to retrieve all items in which two or more spellings alternate in a single query. Searches for "sets" are compatible with the reasoning behind *litteral substitution sets*. The mapping tool may display the full range of spelling variants in a single image, which makes it easier to think about the likely sound values represented by different spellings in neighbouring witnesses. N

MARIE VAŇKOVÁ

CHARLES UNIVERSITY, PRAGUE

References

BROOK, George Leslie 1972. A Piece of Evidence for the Study of Middle English Spelling. *Neuphilologische Mitteilungen* CXXIII (1): 25–28.

CARTLIDGE, Neil 1997. Orthographical Variation in The Middle English Lyrics of BL Cotton Caligula A.ix. *Neuphilologische Mitteilungen* XCVIII (3): 253–59.

DOSSENA, Marina & Roger Lass (eds.) 2004. *Method and Data in English Historical Dialectology*. Bern – New York: Peter Lang.

HOROBIN, Simon & Jeremy J. Smith 1999. A Database of Middle English Spelling. *Literary and Linguistic Computing* 14 (3): 359–74. https://doi.org/10.1093/llc/14.3.359

KOPACZYK, Joanna, Benjamin Molineaux, Vasilios Karaiskos, Rhona Alcorn, Bettelou Los, and Warren Maguire 2018. Towards a Grapho-Phonologically Parsed Corpus of Medieval Scots: Database Design and Technical Solutions 2018. *Corpora* 13 (2): 255–69. https://doi.org/10.3366/cor.2018.0146

LAING, Margaret 1988. Dialectal Analysis and Linguistically Composite Texts in Middle English. *Speculum* 63 (1): 83–103. https://doi.org/10.2307/2854323

LAING, Margaret 1992. *A Linguistic Atlas of Early Middle English*: The Value of Texts Surviving in More than One Version. *History of Englishes: New Methods and Interpretations in Historical Linguistics: Topics in English Linguistics* 10: 566–81. Berlin – New York: Mouton de Gruyter.

LAING, Margaret 1999. Confusion wrs confounded: litteral substitution sets in early Middle English writing systems. *Neuphilologische Mitteilungen* C (3): 251–70.

LAING, Margaret 2000. The Linguistic Stratification of the Middle English Texts in Oxford, Bodleian Library MS Digby 86. *Neuphilologische Mitteilungen* CI (4): 523–569.

LAING, Margaret 2004. Multidimensionality: Time, Space and Stratigraphy in Historical Dialectology. *Method and Data in English Historical Dialectology: Linguistic Insights 16*, eds Marina Dossena & Roger Lass. 49–96. Bern: Peter Lang Publishing Group.

LAING, Margaret 2013. *A Linguistic Atlas of Early Middle English, 1150–1325*, Version 3.2. Edinburgh: © The University of Edinburgh. http://www.lel.ed.ac.uk/ihd/laeme2/laeme2.html

LAING, Margaret & Roger Lass 2003. Tales of the 1001 Nists: The Phonological Implications of Litteral Substitution Sets in Some Thirteenth-Century South-West Midland Texts. *English Language and Linguistics* 7 (2): 257–78. https://doi.org/10.1017/S1360674303001102

LAING, Margaret & Roger Lass 2009. Shape-Shifting, Sound-Change and the Genesis of Prodigal Writing Systems. *English Language and Linguistics* 13 (1): 1–31. https://doi.org/10.1017/S1360674308002840

LAING, Margaret & Roger Lass 2013. Introduction to LAEME. http://www.lel.ed.ac.uk/ihd/laeme2/laeme_intro_ch1.html

MINKOVA, Donka 2014. *A Historical Phonology of English*. Edinburgh: Edinburgh University Press.

SMITH, Jeremy J. 2007. *Sound Change and the History of English*. Oxford – New York: Oxford University Press.

UPWARD, Christopher & George Davidson
2011. *The history of English spelling.*
Malden, Mass: Wiley-Blackwell.
http://site.ebrary.com/lib/cuni/
Doc?id=10483263