The *Reference Corpus of Late Middle English Scientific Prose*

Javier Calle-Martín
Universidad de Málaga, Dpto. Filología
Inglesa, Francesa y Alemana, Campus de
Teatinos s/n, 29071 Málaga, Spain
jcalle@uma.es

Laura Esteban-Segura
Universidad de Murcia, Dpto. Filología
Inglesa, Campus de la Merced,
30071 Murcia, Spain
lesteban@um.es

Teresa Marqués-Aguado
Universidad de Murcia, Dpto. Filología
Inglesa, Campus de la Merced,
30071 Murcia, Spain
tmarques@um.es

Antonio Miranda-García
Universidad de Málaga, Dpto. Filología
Inglesa, Francesa y Alemana, Campus de
Teatinos s/n, 29071 Málaga, Spain
amiranda@uma.es

Abstract

This paper presents the current status of the project *Reference Corpus of Late Middle English Scientific Prose*, which pursues the digital editing of hitherto unedited scientific, particularly medical, manuscripts in late Middle English, as well as the compilation of an annotated corpus. The principles followed for the digital editions and the compilation of the corpus will be explained; the development and application of several specific tools to retrieve linguistic information within the framework of the project will also be discussed. Our work joins in with worldwide initiatives from other research teams devoted to the study of medical and scientific writings in the history of English (see Taavitsainen, 2009).

1 Introduction

Digital editing has been much debated for more than a decade since the advent of the first projects in English like *The Canterbury Tales* (started in 1993) and *The Electronic Beowulf* (in 1994). The active scholarly thinking is corroborated not only by the publication of a plethora of *ad hoc* monographs discussing the nature of digital editions from different perspectives (Sutherland, 1997; Burnard, O'Brien O'Keefe and Unsworth, 2006; Deegan and Sutherland, 2009), but also by the special-themed issue published by *Literary and Linguistic Computing* (2009), approaching the topic from theoretical and empirical domains.

It is now a common practice among many manuscript holders to digitise and to publicise previously unpublished texts and/or manuscripts, offering not only an edition in itself but also the foundations for further work. The benefits of this activity are manifold to such extent that "digital editions of manuscripts […] have opened new possibilities to scholarship as they normally include fully searchable and browsable transcriptions and, in many cases, some kind of digital facsimile of the original source documents, variously connected to the edited text" (Pierazzo, 2009: 169; also Ore, 2009: 114).

In the light of this trend, the present paper describes the model of electronic editions followed in the *Reference Corpus of Late Middle English Scientific Prose*, an on-going research project developed at the universities of Málaga and Murcia with the following objectives: (a) the implementation of on-line electronic editions of hitherto unedited *Fachprosa* written in the vernacular in which the manuscript high-resolution images are accompanied by their diplomatic transcriptions; and (b) the compilation of an annotated corpus from this material facilitating Boolean and non-Boolean searches, both word- and lemma-based.

The justification for our work lies in the need for faithful transcriptions for research purposes, avoiding the use of published editions. Lavagnino's definition of electronic/digital

424

edition has been adopted. This is described as an archive offering "diplomatic transcriptions of documents, and facsimiles of those documents. And it should avoid […] the creation of critically edited texts by means of editorial emendation […] What readers need is access to original sources" (1998: 149). The rationale behind the project is then to offer faithful reproductions of the originals which can be eventually used as primary sources for research in Linguistics and other side areas like Palaeography, Codicology, Ecdotics or History of Science. This material also serves as the input for the compilation of a corpus of late Middle English scientific prose, thus allowing for the synchronic study of the language from different perspectives, such as phono-orthographic and morpho-syntactic. The internal coherence of the project derives from the contribution of two variables, one qualitative and the other quantitative. On the one hand, the project exclusively focuses on 14th- and 15th-century scientific treatises and, on the other, it displays complete texts, not samples.

The present paper addresses the concept of electronic editing as applied to the *Reference Corpus of Late Middle English Scientific Prose* in order to (a) describe the editorial principles and the theoretical implications adopted; (b) present the digital layout along with the tools implemented for information retrieval; and (c) evaluate our proposal for linguistic research.

2 The collection

The manuscripts have been primarily taken from two holders, the Hunterian Collection at Glasgow University Library, and the Wellcome Library. These two repositories have been chosen on account of (a) the number of scientific treatises from the 14th and the 15th centuries available; (b) their unedited status; and (c) their availability because they have provided us with digitised images of the manuscripts as well as with permission for on-line publication.[1] In its present form, the following items have been transcribed amounting to 471,143 running

words, which have also been annotated. The treatises are listed below:[2]

- Glasgow, University Library, MS Hunter 307, *System of Physic* (ff. 1r-166v), including an anonymous Middle English treatise on humours, elements, uroscopy, complexions, etc. (ff. 1r-13r); the *Middle English Gilbertus Anglicus* (ff. 13r-145v); an anonymous Middle English treatise on buboes (ff. 145v-146v); a gynaecological and obstetrical text (ff. 149v-165v); a Middle English version of Guy de Chauliac's *On bloodletting* (ff. 165v-166v); and an *Alphabetical List of Drugs with their Properties* (ff. 167r-172v).
- Glasgow, University Library, MS Hunter 328 including Gilles of Corbeil's *Treatise on Urines* (ff. 1r-44v); an *Alphabetical List of Remedies* (ff. 45r-62r); and *An Alphabetical List of Medicines* (ff. 62v-68v).
- Glasgow, University Library, MS Hunter 497, translation of Macer's *Herbary* (ff. 1r-92r).
- Glasgow, University Library, MS Hunter 509, *System of Physic* (ff. 1r-167v), including an anonymous Middle English treatise on humours, elements, uroscopy, complexions, etc. (ff. 1r-14r); and the *Middle English Gilbertus Anglicus* (ff. 14r-167v).
- London, British Library, MS Sloane 404, *Medicine: A General Pharmacopoeia* (ff. 3v-319v).
- London, British Library, MS Sloane 2463, *Antidotary* (ff. 154r-193v).
- London, Wellcome Library, MS Wellcome 397, *A Treatise of Powders, Pills, Electuaries and Plasters* (ff. 52v-68v).
- London, Wellcome Library, MS Wellcome 404, *Leechbook* (ff. 1r-44v).
- London, Wellcome Library, MS Wellcome 542, *Leechbook* (ff. 1r-20v).
- London, Wellcome Library, MS Wellcome 799, William de Congenis's *Chirurgia* (ff. 1r-23v).

---

[1] The British Library and the Bodleian Library witnesses are offered without the digitised images as a result of their expensive copyright prices. The Ryland manuscript, in turn, presents the digitised images freely offered by the holder (http://www.library.manchester.ac.uk/).

---

[2] The list relies on the data and collation provided by Young and Aitken (1908), Moorat (1962), and Cross (2004), among others.

- London, Wellcome Library, MS Wellcome 5262, *Medical Recipe Collection* (ff. 3v-61v).
- London, Wellcome Library, MS Wellcome 8004, *Physician's Handbook* (ff. 113r-133v).
- Manchester University Library, MS Rylands 1310, Gilles of Corbeil's *Treatise on Urines* (ff. 1r-21r).
- Oxford, Bodleian Library, MS Rawlinson C. 81, *The Dome of Urines* (ff. 6r-12v).

The catalogue is being currently enlarged with the addition of the following treatises:

- London, British Library, MS Sloane 340, Henry Daniel's *Liber uricrisiarum* (ff. 39r-62v).
- London, Wellcome Library, MS Wellcome 226, Henry Daniel's *Liber uricrisiarum* (ff. 1r-70v).
- London, Wellcome Library, MS Wellcome 537, *The Practice on the Sight of Urines* (ff. 15r-46v).
- London, Wellcome Library, MS Wellcome 8004, *Bloodletting* (ff. 18r-32v); *Celestial Distances* (ff. 49r-96v).
- Oxford, Bodleian Library, MS Add. A.106 (ff. 244r-258v).

3 The platform

The *Reference Corpus of Late Middle English Scientific Prose* is hosted at http://referencecorpus.uma.es.[3] The homepage contains three main items: a table (on the left), a door with the sign "Reading room" on it (on the right) and flags (on the right, above the door). The flags provide links to the websites of the institutions taking part in the project, each one represented by its emblem in the following order: University of Málaga, Junta de Andalucía (Autonomous Government of Andalusia) and University of Murcia.

On the table there are several objects: a picture, which supplies information about the members of the research team; documents, which include the transcription policy; several letters, which give access to the tool "Words & Phrases" (see 4.1); an envelope with a link to contact details; a diary with information about the project ("Description", "Copyright", "Acknowledgements", and "Sitemap"); and a globe, through which the "Guided tour"[4] can be reached.

By opening the door the user enters the "Reading room", which holds the manuscripts and provides access to their digitised images, transcriptions, and brief physical descriptions. Five volumes, with the words "Hunter", "Rawlinson", "Rylands", "Sloane" and "Wellcome" (from left to right), are showcased on a table. By clicking on the spine of the desired collection,[5] all the manuscripts belonging to that particular collection appear.[6]

Special characters for symbols such as runes and punctuation marks are employed, and so users need to have a Unicode compliant font in their computers (a link from which it can be freely downloaded is supplied).

As far as linguistic analysis is concerned, three different tools for the retrieval of linguistic information are available, namely *Word Search Tool*, *Text Search Engine* (*TexSEn*) and *Concordance Manager*. The first of them (see 4.1) allows creating word and lemma lists. *TexSEn* has been especially developed for the extraction of morpho-syntactic and statistical information from annotated corpora (see 4.2). The *Concordance Manager* (see 4.3), in turn, serves as an aid to view the concordances generated by *TexSEn*.

3.1 Editorial principles

More often than not, modern editions are guided by the editorial principles of publishers, which may deny the reader an immediate access to the source text as it was copied by the mediaeval writer. Aspects such as abbreviations,

---

[4]This supplies information about how to interact best with the icons, animations and visual elements displayed in the site, including the following points: (i) "Interaction", (ii) "Pop-up messages", (iii) "The cursor" and (iv) "Help messages".

[5]It must be borne in mind that the images of the spines arrayed do not correspond to the spines of any real manuscript.

[6]Only MS Hunter 328 (ff. 1r-68v), MS Wellcome 404 (ff. 1r-44v) and MS Wellcome 542 (ff. 1r-21r) are open for public consultation. The other manuscripts are freely available after registration, a required process to control the use and integrity of the resources offered.

---

[3] The site is best viewed using Firefox (or a compatible web browser) and JavaScript. It requires Flash Player 10, with a resolution of 1280x1024 for best visualisation.

marginalia, emendations or punctuation are usually a *desideratum* in modern editions, leaving aside physical details such as lemmata, decoration, illuminated capitals, etc. As the main goal here is the compilation of an annotated corpus of late Middle English scientific material in the vernacular, the principles of a semi-diplomatic edition have been adopted so as to provide an accurate reproduction of the source text. Our transcription follows the guidelines presented next, partially adapted from Petti (1977: 24-35):

a) The spelling, capitalisation and word division of the original have been retained, including the use of <u> for <v> and <y> for <i>. The different spellings of a same consonant, however, have been regularised, particularly in the case of the letter <s>, which may be represented by <β> and <ſ>, among others, since the occurrence or choice of separate graphs depends on the position of that letter within the word.[7]

b) The punctuation and paragraphing of the original have been retained. Marks of punctuation have been represented by the symbols they stand for, e.g. the paragraph mark (¶), the virgule (/) and the caret (^).

c) Lineation has been preserved and, for reference purposes, the lines have been numbered accordingly (every five lines).

d) Abbreviations have been expanded with the supplied letter(s) italicised.

e) Deletions are retained preserving the scribal practice.

f) Insertions have been included in the body of the text, enclosed in square brackets. The caret mark, if used, has been placed immediately in front of the first bracket.

g) Catchwords are given at the bottom of the page.

3.2 The electronic edition

The electronic editions of the manuscripts can be freely consulted once the user has registered. Each edition consists of the digitised images of

_____

[7]A graphetic transcription has been discarded on account of the research interests which lie behind the edition itself, as we pursue the compilation of an annotated corpus. A graphemic transcription has been adopted for linguistic reasons, as already noted by Robinson and Solopova (1993: 24–25), and Robinson (2009: 45).

folios of a given manuscript and additional features, including its transcription. This fully adheres to the principles of the semi-diplomatic editorial method, in which intervention is kept to a minimum (see 3.1). The only exception is the inclusion of the number of folio and lines, meant as a help to locate information.

The design recreates an environment that brings visitors as close as possible to the experience of consulting the original manuscripts without the need to move from their computers and with other added assets. In order to consult a specific manuscript, the volume that represents the collection housing it should be selected first. When this is done, a pop-up window, in which all the manuscripts available for that collection, together with their different treatises, are listed in alphabetical order, appears. Two icons appear next to the name of the treatise and the range of folios/pages in which it is held: one which shows "image & text" and one which displays information about the treatise selected. Concerning the latter, another pop-up window either with data extracted from the catalogues in which the manuscripts have been described (see footnote 2) or providing a link to the online description of the library catalogue, emerges. For consulting the desired treatise, two possibilities are given: either the images on their own or the images and their transcriptions. After loading, the original cover of the manuscript comes into view. The pages or folios can be browsed either manually or automatically and zoomed for a more exhaustive view by means of a magnifying glass (the icon is on the top-left side of the screen). A particular page or folio can be searched for, a helpful and time-saving option with long treatises. Those visitors unfamiliar with mediaeval scripts can read the transcribed text, which is displayed in a pop-up window next to the image –on the right-hand side in the case of versos and on the left-hand side in the case of rectos (see figure 1). In order to do so, the user only needs to click on the feather icon on the top-right side of the screen.

It is possible to go back to the "Reading room" at any time, just by clicking on the icon "Return to Library", on the bottom-right side of the screen.
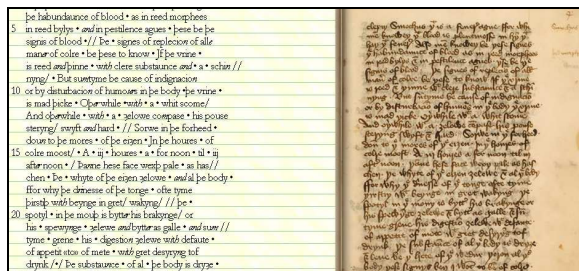
Figure 1. Digitised image of folio recto with transcription (f. 4r, MS Hunter 509)

3.3 Corpus annotation

Lemmatisation and morphological tagging are the two main stages followed in the linguistic annotation of our corpus, which currently comprises around 1,200,000 tokens. Once the manuscripts have been transcribed, the texts are pasted into spreadsheets and every individual item is assigned a row. Each word form occurring in the corpus is then paired with a corresponding base form or lemma. Thus, inflected forms of a word are identified as instances of the same lemma. Lemmatisation also helps to overcome the difficulties posed by orthographic variation, a salient characteristic of Middle English. For the selection of the lemmas, the main headword found in the online version of the *Middle English Dictionary* (*MED*) has been taken as the reference because it provides a standard form that can be used for all the texts. The task of lemmatisation is not exempt from complications, which include, for example, choosing an adequate lemma when a particular lexical item is not gathered in the *MED* (mainly Latin terminology) or deciding whether combinations of words should form part of the same or different lemmas.

Morphological tagging consists of tags including information about words, such as part of speech (noun, adjective, verb, etc.), tense, number, case, gender, etc., as shown in Figure 2. The texts have also been labelled so that each item includes reference to the folio and range of five lines in which it occurs.[8]

One of the most important advantages of such annotated corpora is that they allow extracting linguistic information easily. They can also be used as input to be exploited by machines or computer-driven tools, which can provide the researcher with detailed analyses. Our corpus consists of full texts which have been transcribed following the original manuscripts closely and faithfully, and therefore constitutes a reliable resource for the study of Middle English (for instance, of aspects dealing with morphology, dialectology and punctuation, etc.).

The main disadvantage is that annotating is a time-consuming and taxing process, since the information in the tags for each particular item has to be manually recorded. This leads to another weakness: errors, which may compromise the accuracy of the results, can be made. For this reason, the annotated corpus has undergone several phases of revision.

4 Information retrieval

Various types of information may be retrieved from a corpus complying with the features explained in section 3. For the purpose, the tools offered in the platform may be used, namely *Word Search Tool*, *TexSEn* and *Concordance Manager*.

4.1 *Word Search Tool*

By clicking on the "Words and Phrases" icon, users access the *Word Search Tool*, which allows them to obtain a list of the variant spelling forms of the selected text, along with the number of occurrences and the reference where each of them can be found. It is even possible to perform a word- or a lemma-based search.

---

[8]Further specifications concerning the method of annotation are discussed in Moreno-Olalla and Miranda-García (2009), and Esteban-Segura and Marqués-Aguado (forthcoming).

	Word	Lemma	Class	Subclass	Type	Tense	Number	Person	Case	Gender	Folio	Line
1	Als[o]	alsō, b	Adve	Affirm							149	1
2	.	Pmark	Pmark								149	1
3	we	wē, r	Pron	Pers			Plur	1st	Nom		149	1
4	schulen	shulen, v (1)	Verb		Pret-P	PrsInd	Plur	1st			149	1
5	vndirstonde	understōnden, v	Verb			Infin					149	1
6	þat	that, c	Conj	Compl							149	1
7	wy*mme*n	wŏmman, n	Noun				Plur				149	1
8	han	hāven, v	Verb			PrsInd	Plur	3rd			149	1
9	lesse	lēs(se, a	Adje				Compa				149	1
10	hete	hēte, n (1)	Noun				Sing				149	1
11	in	in, p	Prep						RegDat		149	1
12	her	hēr(e, d	Dete	Poss			Plur	3rd		Fem	149	1
13	body	bōdī, n	Noun				Sing				149	1
14	þan	than, c	Conj	Compa							149	1
15	men	man, n	Noun				Plur				149	1

Figure 2. Sample screenshot of morphological tagging (f. 149v, MS Hunter 307)

The first tab requests the selection of the text/treatise to be analysed. If the user is logged in, all the texts are displayed for analysis; otherwise, four texts are shown as samples. After choosing the text in the first tab, two options are possible, i.e. a list of words or a list of lemmas may be provided. For the purpose, the user should click on the preferred option under 'Show word / lemma', the default setting being 'Words'. The second tab then presents all the units of analysis of the chosen text in alphabetical order, i.e. either all its words (taking into account spelling variation) or else the lemmas, which are in turn retrieved from the database including all the entries for all the texts. Since the word-class is shown alongside the lemma, the variant forms linked to each lemma can be distinguished according to their word-class (e.g. *either* is tagged as a pronoun and as a conjunction in the corpus).

The results are automatically shown as a KWIC (Key Word In Context) index with 6 lexical units preceding and following the unit under scrutiny, plus the reference. If there is more than one occurrence, the full list (arranged in order of appearance) is shown and, by clicking on the reference, the user may view the manuscript folio/page where that occurrence is found.

4.2. *Text Search Engine*

TexSEn, which is available at http://texsen. uma.es and linked to the research project presented in this paper, may perform both simple and complex (i.e. Boolean) searches using annotated corpora complying with its

requisites, together with several statistical calculations (Miranda-García and Garrido-Garrido, 2012).

Simple or non-Boolean searches comprise lists and indexes which, if the user is not logged in, will refer either to MS Hunter 503 or else to the ophthalmologic treatise or the antidotary housed in MS Hunter 513, all of which are used as samples.[9] Once the text under analysis has been picked out from the list under 'Bookshelf', the requirements of the search have to be set by clicking on the corresponding tabs: 'Selection' refers to the page or folio range; 'Item' presents the units of analysis available (words or lemmas); 'Output' offers three possible types of lists (either a complete list with all the items; or else a reduced list with only the different items, allowing also for the addition of the number of hits); and, finally, 'Save as/open', which enables the user to select the filename for the list produced. Within a few seconds, the user is presented with the results in order of appearance according to the search criteria. The default configuration shows 50 results per screen, although the tab 'Output' allows changing this figure, along with the field used to sort the results (sequential order or alphabetical order of the unit of analysis, in ascending or descending order).

Boolean searches include the options to generate complex searches, lemma-sorted KWIC concordances and glossaries.[10] Complex searches work following the same procedure as simple searches, since the text has to be first selected from the list under 'Bookshelf' and then the choices regarding 'Selection', 'Item', 'Output' and 'Save as/open' must be specified. Additionally, the 'Search parameters' have to be set with the aim of determining the type of unit upon which the search is going to be performed. First, the type of unit under analysis or accidence (word, lemma, word-class, number…) must be picked from the 'Column' section. Then, the Boolean operator must be added in the 'Condition' tab, followed by the values for that unit/accidence under 'Value' (e.g. singular, plural, both and dual are the options provided). Another possibility is using several values together, hence increasing the potentialities of the complex search. The parameters have to be validated by clicking on the box icon, which changes from red to green when the former are suitably arranged, and then needs to be dragged to the suitable column in the KWIC arrangement, i.e. from six words preceding to six words following the keyword, the latter of which can also be specified, as shown in Figure 3. If the user is interested in typing in a particular word or lemma to carry out the search, a window with Unicode symbols is shown on the right, from which letterforms can be added by a simple click.

The tool lemma-sorted KWIC concordance builder replicates the same structure ('Bookshelf' to choose the text under scrutiny, and tabs to set the requirements concerning 'Selection', 'Output' and 'Save as/open'). In the 'Output' tab users may choose the span of words preceding and following the keyword, which can range from 1 to 20 each, as well as the word types for analysis (nouns, verbs, adjectives, adverbs, all function words, all content words or all items). The results are shown on the screen using the default configuration, that is, showing the concordances arranged by lemmas, whose spelling forms (and therefore, the corresponding lines of concordance) are presented in order of appearance. Yet, this presentation may be modified by choosing a different parameter to arrange all the results, or by altering the ascending/descending icons in some/all of the columns of the results page.

4.3. *Concordance Manager*

The *Concordance Manager* is an online application that allows viewing the concordances generated by *TexSEn*. As with the previous tool, if the user is registered, the concordances to a wider range of texts are offered (those in the Hunter manuscripts only); otherwise, only MS Hunter 503 is available for consultation as a sample.

In order to access the concordances of a particular lemma, the manuscript on which the search is to be performed first needs to be picked from the list provided, along with the lemma, selected in turn from a predictive list including all the lemmas for the text chosen.

The results screen displays the list of lines of concordances (i.e. rows) in tabular format, in which five words typically precede and other five follow the keyword, which is highlighted in red. Hence, the information rendered is structured into: a) lemma; b) preceding context; c) keyword; d) following context; and e) reference. The order in which the concordances are presented follows that of occurrence in the text (page or folio number, side and line span), as signalled in the column for the reference, rather than alphabetically. The default number of lines of concordance shown per page is 200, although this figure may be modified by the user. Likewise, the results can be re-arranged according to the reference (from end to beginning), to the word preceding or following the keyword, and to the keyword itself (thus grouping all the examples of particular spelling variants together, for instance).

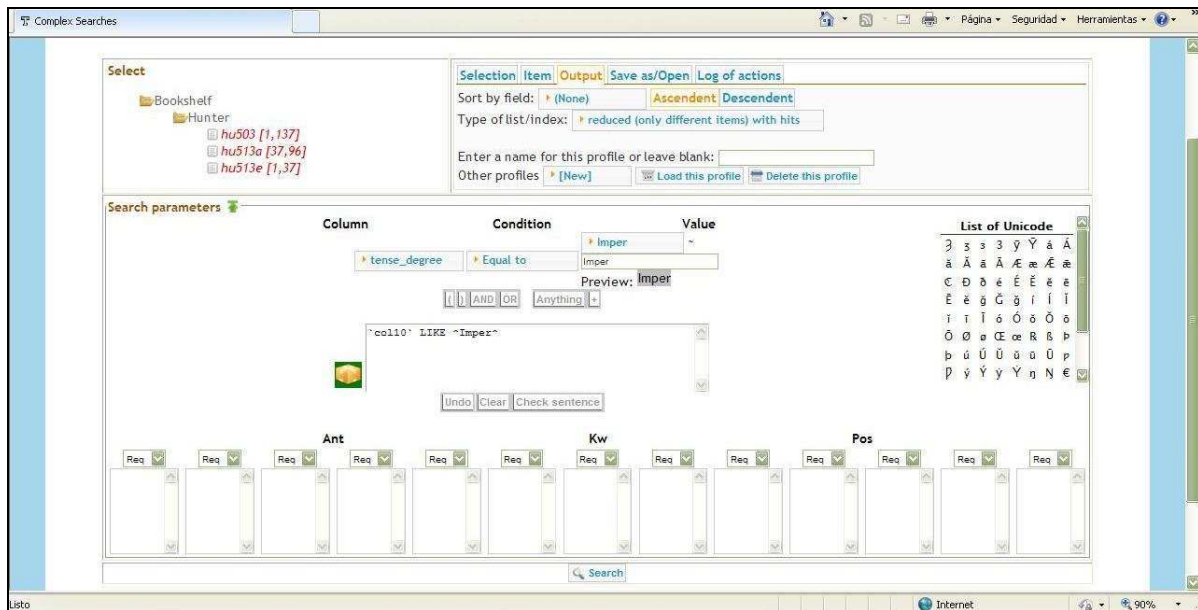[10]The tool Glossary builder is not available at the moment.

Figure 3. Sample screenshot of complex searches with *TexSEn*

Beyond purely statistical data, concordances lend themselves well to other types of studies. For instance, they are very helpful for the analysis of allomorphs of a particular inflection or morpheme, especially if the latter depend on the ensuing context. Likewise, comparative studies are possible if the same search is carried out on various texts with a view to analysing whether certain lemmas are common to all the texts under scrutiny, of if they are only used in some texts and alternatives are sought for other texts. These results may even suggest some kind of authorial fingerprint.

Conclusions

This paper has presented the highlights of the project *Reference Corpus of Late Middle English Scientific Prose*, discussing the motivations for electronic editing and the principles followed. It has also shown the potential of lemmatised and annotated corpora, especially when they are used in conjunction with specific-purpose tools. The ones presented in this paper are particularly aimed at extracting relevant data from annotated corpora in Middle English, hence allowing for a wide variety of quantitative studies on the language of the period. Furthermore, by being available online, most results can be made public easily. Future research will focus on the expansion of the corpus so as to include manuscripts from the Early Modern period, which will eventually allow for more comprehensive diachronic studies.

Acknowledgments

References

Lou Burnard, Katherine O'Brien O'Keefe, and John Unsworth (eds.). 2006. *Electronic Textual Editing*. New York, Modern Language Association of America.

Rowin Cross. 2004. *A Handlist of Manuscripts Containing English in the Hunterian Collection, Glasgow University Library*. Glasgow, Glasgow University Library.

Marilyn Deegan and Kathryn Sutherland (eds.). 2009. *Text Editing, Print and the Digital World*. Surrey, Ashgate.

Laura Esteban-Segura and Teresa Marqués-Aguado. Forthcoming. New Software Tools for the Analysis of Computerized Historical Corpora: GUL MSS Hunter 509 and 513 in the Light of *TexSEn*. In Vincent Gillespie and Anne Hudson

(eds.). *Editing Medieval Texts from Britain in the Twenty-First Century*. Turnhout, Brepols.

John Lavagnino. 1998. Electronic Editions and the Needs of Readers. In W. Speed Hill (ed.). *New Ways of Looking at Old Texts II*. Tempe, AZ, Renaissance English Text Society Publications:149–156.

Robert E. Lewis *et al.* (eds.). 1952–2001. *Middle English Dictionary*. Ann Arbor: University of Michigan Press. Online version in Frances McSparran *et al.* (eds.). 2001–. *Middle English Compendium*. University of Michigan Digital Library Production Service. Available at http://quod.lib.umich.edu/m/med.

Antonio Miranda-García and Joaquín Garrido-Garrido. 2012. *Text Search Engine* (*TexSEn*). Málaga, Servicio de Publicaciones de la Universidad de Málaga. Available at http://texsen.uma.es.

Samuel Arthur Joseph Moorat. 1962. *Catalogue of Western Manuscripts on Medicine and Science in the Wellcome Historical Medical Library. Vol. 1. Catalogues Written before 1650 AD*. London, Publications of the Wellcome Historical Medical Library.

David Moreno-Olalla and Antonio Miranda-García. 2009. An Annotated Corpus of Middle English Scientific Prose: Aims and Features. In Javier E. Díaz Vera and Rosario Caballero (eds.). *Textual Healing: Studies in Medieval English Medical, Scientific and Technical Texts*. Bern, Berlin, etc., Peter Lang:123–140.

Espen S. Ore. 2009. ... They Hid Their Books Underground. In Marilyn Deegan and Kathryn Sutherland (eds.). *Text Editing, Printing and the Digital World*. Surrey, Ashgate:113–125.

Anthony G. Petti. 1977. *English Literary Hands from Chaucer to Dryden*. Cambridge, Mass., Harvard University Press.

Elena Pierazzo. 2009. Digital Genetic Editions: the Encoding of Time in Manuscript Transcription. In Marilyn Deegan and Kathryn Sutherland (eds.). *Text Editing, Printing and the Digital World*. Surrey, Ashgate:169–186.

Peter Robinson. 2009. What Text Really is Not, and Why Editors have to Learn to Swim. *Literary and Linguistic Computing*, 24(1):41–52.

Peter Robinson and Elizabeth Solopova. 1993. Guidelines for Transcription of the Manuscripts of the Wife of Bath's Prologue. In Norman Blake and Peter Robinson (eds.). *The Canterbury Tales Project. Occasional Papers Volume I*:19–52. Available at http://www.canterburytalesproject.org/pubs/op1-transguide.pdf.

Kathryn Sutherland. 1997. *Electronic Text: Investigations in Method and Theory*. Oxford, Clarendon Press.

Irma Taavitsainen. 2009. Early English Scientific Writing: New Corpora, New Approaches. In Javier E. Díaz Vera and Rosario Caballero (eds.). *Textual Healing: Studies in Medieval English Medical, Scientific and Technical Texts*. Bern, Berlin, etc., Peter Lang: 177–206.

John Young and P. Henderson Aitken. 1908. *A Catalogue of the Manuscripts in the Library of the Hunterian Museum in the University of Glasgow*. Glasgow, Maclehose.