A Bilingual Digital Edition of Trinity College Cambridge MS O.1.77.

Honkapohja, Alpo
alpo.honkapohja@helsinki.fi
University of Helsinki

The poster will present my work-in-progress PhD project of a 15th-century bilingual medical manuscript, containing Latin and Middle English. The edition is designed with the needs of historical linguistics in mind, and will have some corpus functionalities. My long term aim is to use it as a pilot study of sorts in contrastive investigation of Latin and Middle English medical writing.

1. Background

Medieval medical writing for a long period of time received fairly little attention. For instance, Robbins described it, in 1970, as a "Yukon territory crying out for exploration". In the 1990s and 2000s, the situation has changed, and the field is becoming filled with tiny flags stating the claims of various research projects and individual scholars. There are now large electronic corpora such as the *Middle English Medical Texts* (MEMT), published 2005, and *A Corpus of Middle English Scientific Prose*, currently being compiled in collaboration between the University of Malaga and Hunter Library in Glasgow.

These resources do, however, have one inherent bias. They focus on Middle English material, which gives a distorted view of the linguistic situation in England in the late Middle Ages. England, after the Norman conquest, was a trilingual society in which educated members of the society were likely to have at least some degree of literacy in Latin, Anglo-Norman French as well as English. This shows, for instance, in the fact that manuscripts containing texts in more than one language outnumber monolingual ones. (cf. Voigts 1989). Moreover, marginal comments also suggest they had a readership competent in more than one language.

My PhD project is intended as the first genuinely bilingual online resource of medical manuscripts in late Medieval England, and will hopefully pave the way for similar resources in the future. It is designed for both historical linguists and historians, but paying special attention to the needs of linguistics.

2. Trinity College Cambridge, MS O.1.77.

Trinity MS O.1.77. is a pocket-sized (75 x 100 mm) medical handbook, located in Trinity College Cambridge. It contains 10 to 18 texts on medicine, astrology and alchemy. It is usually treated as a sibling MS of the so-called Sloane-group of Middle English manuscripts, which is a group of late Latin, English and French MSS originating from London or Westminster in the late Middle English period (cf. e.g. Voigts 1990). James assigns MS Trinity O.1.77 an exact date 1460, based on astrological markings in the final flyleaf (1902), although it may not be entirely accurate. (see Honkapohja 2010, forthcoming)

Roughly 4/5 of the manuscript is in Latin and 1/5 in English, that is, out of slightly less than 30,000 words, c. 24,000 words are Latin and 5,500 in English. There does not appear to be a clear-cut division between prestigious Latin texts and more popular English ones. Latin, however, is used almost exclusively for metatextual functions such as incipits and explicits. Nearly all marginal comments in the manuscript are in Latin.

3. The digital edition

The digital edition which I am preparing will be designed in such a way that it will function as reliable data for historical linguistics. This involves encoding a sufficient amount of detail on linguistic variants without normalising, modernising, or emending the data, and keeping all editorial interference transparent (see e.g. Kytö, M., Grund P. and Walker T. 2007 or Lass 2004)

On the technical side, I am using TEI P5 – conformant XML tagging built on stand-off architecture. Things included in the base-level annotation are a graphemic transcription of the text (cf. e.g. Fenton & Duggan 2006), select manuscript features such as layout, and

information about the manuscript and hand. Each word will also be tagged with a normalised form, useful for linguistic research, and an ID which allows the addition of additional tagging by means of stand-off annotation – including, for instance, POS tagging, semantic annotation or lemmatisation.

The edition will have an online user interface, which will allow the user to select the level of detail he or she wishes. It will be possible to use it with either normalised text or diplomatic transcription. It will be released under a Creative Commons license. The user will have full access to the XML-code, including all levels of annotation, and will be allowed to download and modify it for non-commercial purposes.

4. DECL

The development of the edition will take place in collaboration between the Digital Editions for Corpus Linguistics (DECL) project based at the University of Helsinki.

The DECL project was started by three post-graduate students in 2007. It aims to create a framework for producing online editions of historical manuscripts suitable for both corpus linguistic and historical research. DECL editions use a more strictly defined subset of the TEI-guidelines and are designed especially to meet the needs of corpus linguistics. The framework consists of encoding guidelines compliant with TEI XML P5. The aims of the project are presented in more detail in our article (Honkapohja, Kaislaniemi & Marttila 2009).

5. Digital Edition of O.1.77 as a resource for the study of bilingualism

My PhD project has both short and long term goals related to the study of multilingualism. The short term aim is to design the edition in a way that is of maximum use for scholars working with medical texts and especially multilingualism. I am especially putting a lot of effort into interoperability and making the encoding as flexible as possible.

Hypothetical research questions for the edition will include, for instance:

- *Spelling variation.* Using the edition will enable getting information on spelling variation in English and Latin, in order to see whether the accepted general view that Latin was more regular is supported by quantitative data.

- *The use of brevigraphs and contracted forms.* Manuscript abbreviations are an extremely common feature in the Latin texts of the manuscript. They are also applied in the Middle English sections, but with less frequency. The edition will make it possible to obtain exact statistical information on which manuscript abbreviations carry into the vernacular, and with how much variation and frequency.

- *Syntactic complexity:* Do sentences in Latin contain a greater number of sub clauses and other signs of syntactic complexity than Middle English ones?

- *Textual Functions:* The use of English and Latin in various text types, recipes, metatextual passages (in which Latin very much dominates). The type of structural and background information which is being annotated in the edition will enable the user to perform the searches on different level textual passages, including marginal comments and metatextual passages.

After the completion of the PhD project, the edition will be expanded with other related multilingual medical and alchemical manuscripts in the Sloane group, which will increase the usefulness of the database, by allowing, for instance, comparative study of the same text in different manuscripts. I am also planning to make use of the available corpora on Middle English medical writing for comparisons to Middle English.

References

A Corpus of Middle English Scientific Prose. `http://hunter.filosofia.uma.es/manus cripts` (accessed 13 March 2010).

Fenton, E. G., Duggan, H. N. (2006). 'Effective Methods of Producing Machine-readable Text from Manuscript and Print Sources'. *Electronic Textual Editing.* Burnard, L., O'Brien O'Keeffe, K., Unsworth, J. (eds.). New York: MLA.

Honkapohja, A., Kaislaniemi, S., Marttila, V. (2009). 'Digital Editions for Corpus Linguistics: Representing manuscript reality in electronic corpora'. *Corpora: Pragmatic and Discourse. Papers from the 29th Internationl Conference on English Language Research on Computerized Corpora (ICAME 29). Ascona, Switzerland, 14-18 May 2008.* Jucker, A. H., Schreier, D., Hundt, M. (eds.). Amsterdam/New York: Rodopi.

Honkapohja, A. (2010: forthcoming). 'Multilingualism in Trinity College Cambridge Manuscript O.1.77'. *Studia Anglica Posnaniensia.*

James, M. R. (1902). *The Western Manuscripts in the Library of Trinity College, Cambridge. A Descriptive Catalogue.* Cambridge: CUP V. III, Containing an Account of the Manuscripts Standing in Class O. .

Lass, R. (2004). 'Ut custodiant litteras: Editions, Corpora and Witnesshood'. *Methods and Data in English Historical Dialectology.* Dossena, M., Lass, R. (eds.). Linguistic Insights. Bern: Peter Lang.

Robbins, R.H. (1970). 'Medical Manuscripts in Middle English'. *Speculum.* **No.3 Jul 1970**: 393-415. `http://www.jstor.org/stable/285350` `0` (accessed 13 March 2010).

Taavitsainen, I., Pahta, P., Mäkinen, M. (eds.) (2005). *Middle English Medical Texts.* Amsterdam: John Benjamins, CD-ROM.

Text Encoding Initiative (TEI). `http:/www.tei-c.org` (accessed 13 March 2010).

Voigts, L. E. (1989). 'Scientific and Medical Books'. *Book Production and Publishing in Britain 1375-1475.* Griffiths, J., Pearsall, D. (eds.). Cambridge: Cambridge University Press.

Voigts, L. E. (1990). 'The "Sloane Group": Related scientific and medical manuscripts from the fifteenth century in the Sloane Collection'. *The British Library Journal.* **16**: 26-57.

Kytö, M., Grund P., Walker T. (2007). 'Regional variation and the language of English witness depositions 1560-1760: constructing a 'linguistic' edition in electronic form'. Pahta, P., Taavitsainen, I., Nevalainen, T., Tyrkkö, J. (eds.). Towards Multimedia in Corpus Studies. Studies in Variation Contacts and Change in English. Helsinki: Research Unit for Variation, Contacts and Change in English (VARIENG), University of Helsinki . `http://www.helsinki.fi/varieng` `/journal/volumes/02/kyto_et_al` (accessed 13 March 2010).