

PROBLEM

Statistical methodologies for identifying the place of issue for historical documents.

MOTIVATION

The Documents of Early England Data Set (DEEDS) contains records of property holding and transmission largely from the late 11th to the early 14 centuries. Determining both the accurate dating and location of the place of issue of these documents are important for studying the social and economic dynamics of a particular time period.

CHALLENGE

To find statistical methodologies for estimating the place of issue of the documents.

PRESENTATION OUTLINE

- Description of the data
- Discussion of a statistical method for dating documents, and the extension of this method for estimating their place of issue.
- The results, which are determined by a test set of 700 documents.

CONCLUSIONS

Dr. Tilahun has described the methodology, developed by her & Professor Feuerverger in the Dept. of Statistics, in collaboration with the DEEDS Project at the University of Toronto, for measuring chronological “distances” between medieval charters of the twelfth and thirteenth

centuries in order to date them, and for assessing their topographic origins. The dating algorithm has enabled us to render dates with an average absolute error of 12.5 years and a median absolute error of 6 years, both of which are far more useful to the historian than assigning dates to undated documents to, for example, “the last half of the twelfth century” or “the first half of the thirteenth century,” as has been the general practice when the content of a charter contains no further chronological evidence.

The documents selected for the related study of a document’s topographic origin were chosen because all contained the expression “datum apud”, that is, “issued or dated at.” For England, a territory measuring approximately 1,000 by 480 kms, a median error of 75 kms can be considered reasonably accurate, but only ‘reasonably’. What, we wonder, is preventing an accuracy similar to that achieved in dating the documents? Like the dating algorithm, the locational algorithm is based on the recurrence of common word patterns. We have found, however, that the date of the document has virtually nothing to do with determining its location. It is the word patterns alone, then, which can serve as topographical locators, but, since the Latin used in the medieval charter has a fairly standardized usage, not only in England but across Roman Catholic Europe as a whole, the differences in vocabulary and word order are nowhere nearly as great as they might have been were regional vernaculars used in charter composition.

We began our study imagining that most monasteries would have issued their own charters written by resident scribes who would also have composed conveyances for local lay society. We also anticipated that we might identify differences between house styles, and possibly even between religious orders. Such differences do not appear commonly to have been the case, however. The evidence points rather to a society in which itinerant scribes carried the custom of charter composition from the place they had learned it to the places which required their services.

As our research progresses, we now expect to be able to identify regional centres of learning and the scribal diaspora which issued from them.

In closing, let me introduce you to the new DEEDS website and the 'DATER' which will enable you to assign a fairly secure circa date to an undated document. It uses a different but related methodology, which depends on the statistical variation of word strings across time. As an example, I chose from the on-line charter editions prepared under the enlightened leadership of Olivier Guyotjeannin, a charter from the Abbaye S. Pierre in Montmartre dated 1179. I dropped it into the DEEDS DATER, which applied a degree of standardization to content and spelling (the words in red on the right do not occur in our database of word patterns). The result, with no further editing, provided a date of 1178. Despite differences in charter diplomatic, we were not surprised that the database composed of patterns from English charters worked equally well on French counterparts.

The website currently contains some 25,000 charters from England, France and Germany, and we welcome contributions of others, preferably in html.