Visualizing English Print from c. 1470 to 1800

Dr Jonathan Hope, Strathclyde University, Glasgow, UK

Note: This project is a collaborative one with the following co-investigators:
Prof Michael Witmore, Director, Folger Shakespeare Library, Washington, USA
Prof Michael Gleicher, Computer Science, University of Wisconsin-Madison, USA
Prof Robin Valenza, English Department, University of Wisconsin-Madison, USA

The hypothesis that drives this project is the claim that computer analysis of linguistic variation in massive collections of texts will allow us to track the development and emergence of cultural forms in the period under study. We seek not only to answer important historical and cultural questions (When does a recognizable American literature emerge? When does scientific discourse appear and from what sources? Do human-identified genres like the novel have consistent linguistic fingerprints over time?) but also to establish a methodology for digital humanities enquiry which we believe will structure humanities research for the next generation. In addition to our own questions, we also seek to map out a series of problems and questions for researchers in other domains.

As the full range of Anglophone printed material becomes digitized, our subject changes. We can ask different questions – and we can answer the old ones with new methods. Someone planning a book on the history of English genres in 1960 could legitimately have laid out a research plan that would have been equally legitimate in 1860, or 1760: a wide range of reading, certainly, but selective reading, and even more selective writing when it came to evidence. The narrative would arise out of the reading, but that reading could be only partial, so the claims made would have to be rhetorical: rhetorically framed, and rhetorically judged. Evaluation would of necessity be based on the questions, "Are they persuasive? Seductive?"

The narrative of the new humanities lies in the data: claims stand or fall on the size of the sample, the statistical significance of the results, and the care with which the procedures have been applied. Franco Moretti has written such a book for the genres of the novel. We are planning a study of the genres of English in the period 1450-1800. In the past, such studies would have been magisterial in the sense of broad surveys which constructed a grand narrative, judged on the quality of the studies' prose and organization of materials. Now we are faced with something more like an epidemiology of literary populations, light sluicing across a map where the virus has passed. Only now, with the vast quantity of data available in varying quality, are digital humanities researchers developing analysis software and methodologies capable of making sense of the data. Our work so far has demonstrated the ability of digital analysis to complement and extend traditional humanities techniques. It has also shown that as data sets increase, methodologies need to be adapted: domain expertise is crucial in interpreting the results of digital analysis. Humanities 2.0 does not replace Humanities 1.0 but incorporates it.