

Language and Chronology: Text Dating by Machine Learning. Gregory Toner and Xiwu Han, eds.

Language and Computers: Studies in Digital Linguistics 84. Leiden: Brill, 2019. xii + 183 pp. €88.

The digital humanities is expansive in terms of inquiry, with many areas being actively explored, from statistical text analysis, to data visualizations, to web projects intersecting with the public humanities. Though some machine-learning tools have been used or developed with humanistic scholarship in mind (MALLET, for example, has been used for topic modeling), there is considerable opportunity in the field. Humanists and social scientists have also offered important critiques of machine-learning algorithms, especially around the perpetuation of societal biases in software (for example, *Algorithms of Oppression* by Safiya Noble). In the present volume, Gregory Toner and Xiwu Han employ machine-learning algorithms on historical literary texts in order to date them. They offer a case study on medieval and early modern Irish literature (ca. 700–ca. 1700), with the hope that their efforts may be applied to other literatures and historical contexts.

The study presumes two readers: the humanist and the computer scientist. The introduction frames the historical importance of dating texts, and details challenges. Chapter 1 outlines traditional humanistic approaches to dating, noting the complexities of data, both linguistically in the text and physically in the manuscript. The authors note that dating undated texts is considerably complex “and remains controversial in many disciplines,” arguing that machine learning, like linguistic dating, can be another viable tool for a textual scholar (39). From here, *Language and Chronology* shifts to two chapters of computational discussion, and the authors suggest that humanistic scholars may wish to skip these chapters, though there is no similar indication for computer scientists unfamiliar with the technical details of codicology, despite many assumptions made of domain knowledge.

Chapter 2 mirrors the first chapter, delineating computational approaches to textual dating, considering earlier language modeling research, and highlighting recent approaches falling under regression/ordinal ranking and classification. The advantages and challenges across these approaches are explored, and the authors offer additional solutions arising from their research, indicating that multi-class classification outperforms regression and ranking. Their approaches allow for more flexibility around time segmentation and grouping that account for irregularity and textual thresholds, very relevant in historical texts and still important for contemporary texts. The interdisciplinary development of these solutions demonstrates that humanistic inquiry and historical data are crucial players in computational innovation.

The next two chapters describe their experimentation using the machine-learning software Weka. Chapter 3 takes on English journalistic texts and medieval Irish annals, in one case using a corpus that has been used by other methods to compare efficacy.

Dated English texts offer a control group to the work done on the Irish texts, which are more heterogeneous, inflected, and deal with a larger time period. Worth noting is that the Irish training set is less robust, with fewer total words, which could impact the model. The work on Irish is expanded in chapter 4, which dates long documents from Old Irish to Middle Irish to Early Modern Irish. This chapter discusses building the corpus, pre-processing, the machine-learning results, and how they compare to the scholarly accepted date of composition. Many of the results are promising and correlate well with accepted dates, especially for those texts that are from the period after ca. 1000 CE.

The conclusion underscores that humanists with domain knowledge must interpret machine-learning results, discusses developing a tool that can be used to explore textual temporality, and gestures toward future work, especially within other literatures. *Language and Chronology* assumes an audience who understands that getting to the stage where a historical corpus is ready to be used as training, validation, or test data for a machine-learning model is a significant undertaking; this data challenge alone would make it arduous to replicate in the medieval and early modern periods. Some future considerations for this work include providing an open-source repository with relevant code and cleaned raw text files, and leveraging cloud infrastructure to test across larger corpora. This precise study offers the historical view of dating texts, explores the manuscripts that are considered, and offers some novel and encouraging approaches to dating historical literary texts through machine learning.

Lisa Tagliaferri, *Villa I Tatti, The Harvard University Center for Italian Renaissance Studies*
doi:10.1017/rqx.2021.228

Negotiating Conflict and Controversy in the Early Modern Book World.

Alexander S. Wilkinson and Graeme J. Kemp, eds.

Library of the Written Word 73; The Handpress World 56. Leiden: Brill, 2019. xiii + 287 pp. €126.

This collection of fifteen essays analyzes controversy and its influence on the publishing industry. Organized into three parts, it offers case studies from across Europe, providing a broad geographic understanding of the complexities of the book trade during moments of intense controversy. Part 1 focuses on censorship during the Protestant Reformation. Drew Thomas studies the printing trade in Wittenberg and Leipzig and shows how these cities managed to address both the needs of the market and the restrictions on publishing. Offering insight into the technicalities of censorship, Martine Furno examines the role of the faculty of theology of the University of Sorbonne as the national board of censorship in the case of Desiderius Erasmus and