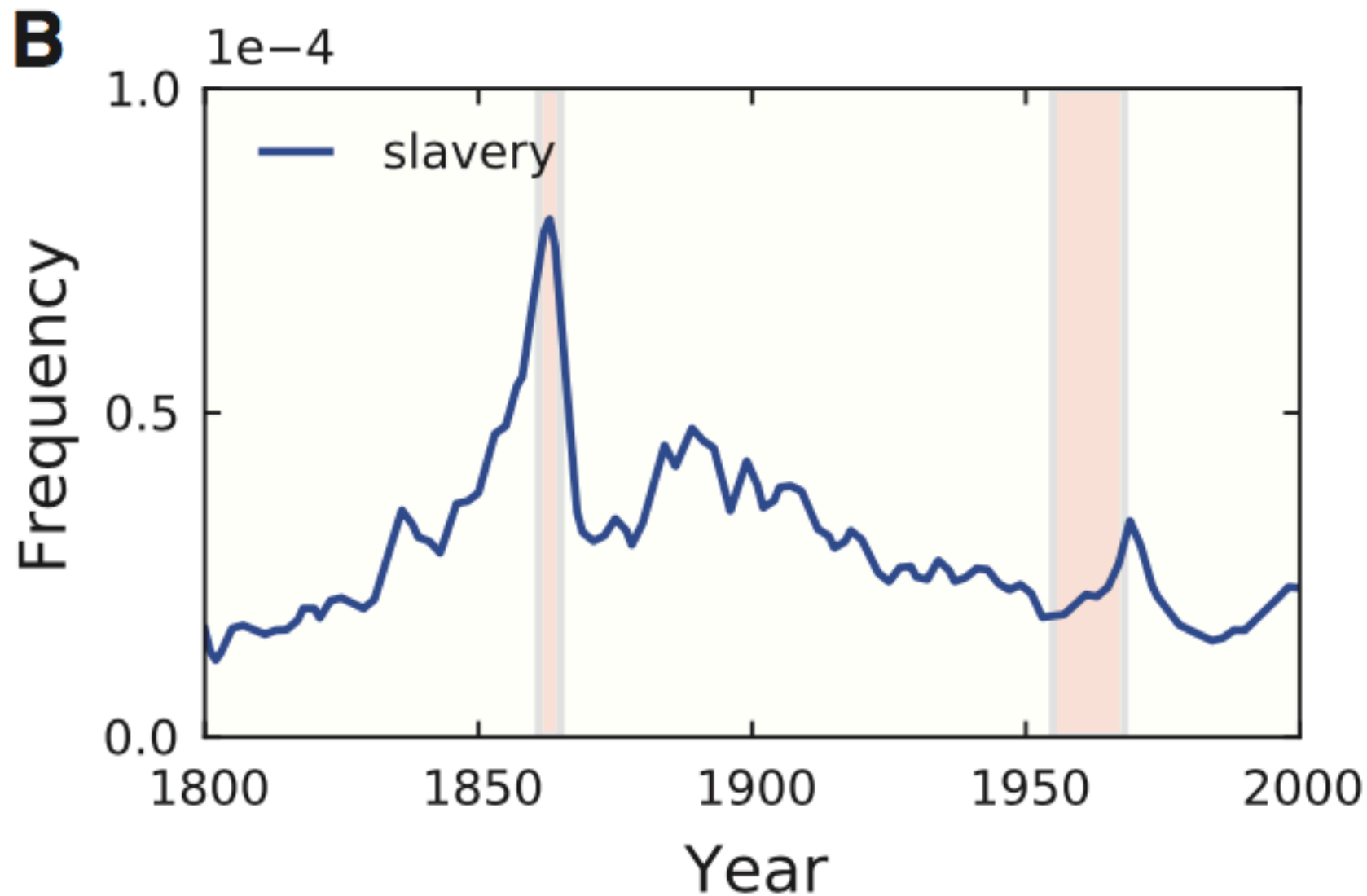


Franco Moretti and Oleg Sobchuk

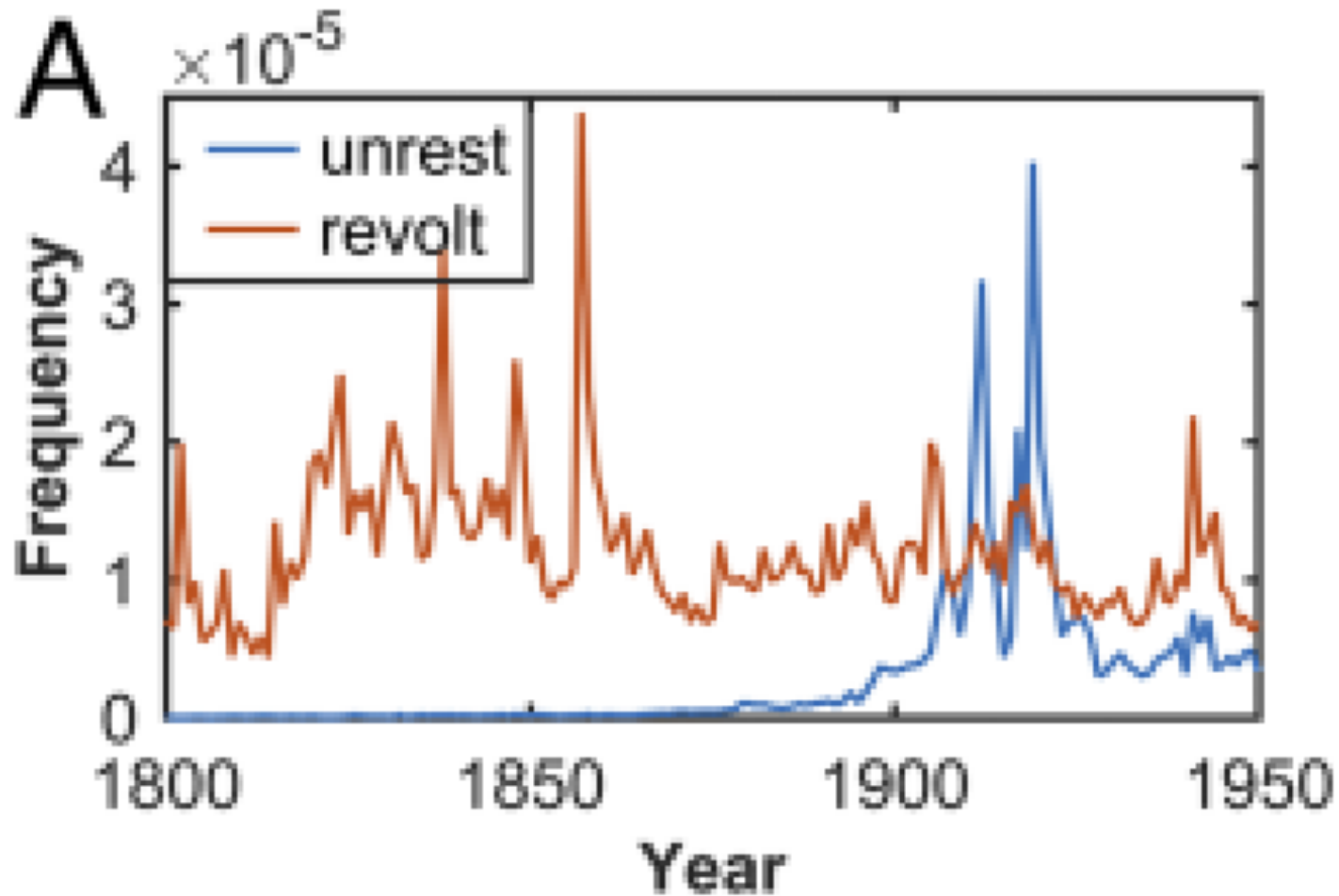
Hidden in Plain Sight.

Thoughts on Data Visualization in the Humanities
(forthcoming *New Left Review*, 2019)

“Quantitative Analysis of Culture Using Millions of Digitized Books” (2011)



“Content Analysis of 150 Years of British Periodicals” (2016)



“Bankspeak” (2014)

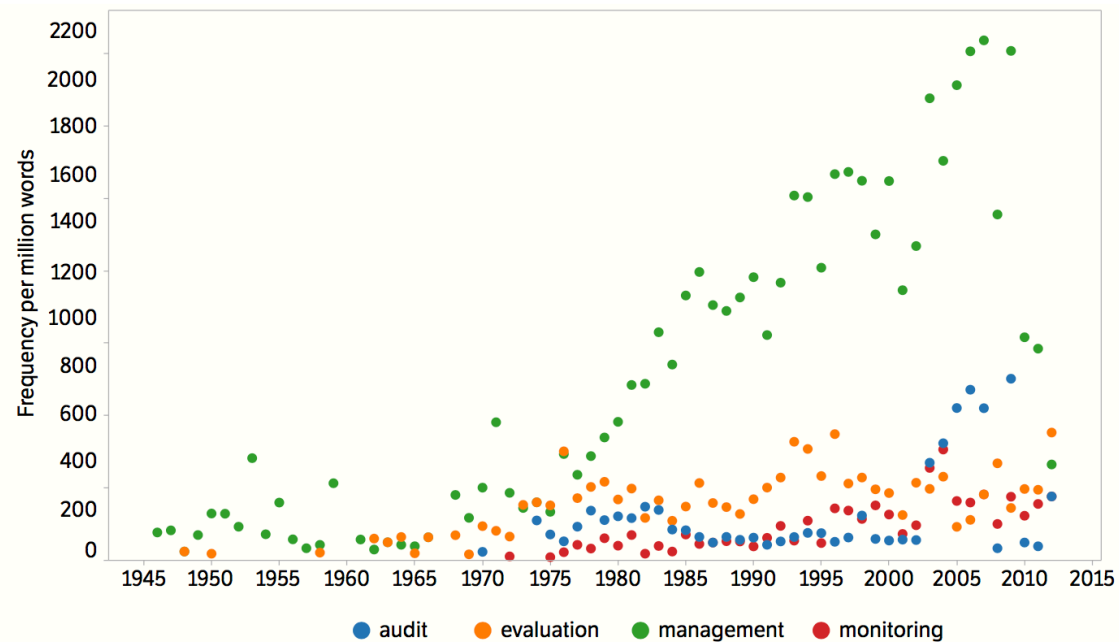


Figure 2 : Management discourse

Though never absent from the Bank’s vocabulary, *management* started its ascent in the late 1970s, when the debt question became central, and was subsequently associated with the drastic policies of “structural adjustment” of the neo-liberal offensive. But it’s only in the 1990s-2000s that management discourse truly flourishes, hinting—at least in subliminal form—that the Bank’s activities are being constantly evaluated and certified by the most advanced tools and the best experts; and that, as a consequence, its investments are the fruit of serious reflection, and their results are as good as they can possibly be.

“A Quantitative Literary History of 2,958 Nineteenth-Century British Novels: The Semantic Cohort Method” (2012)

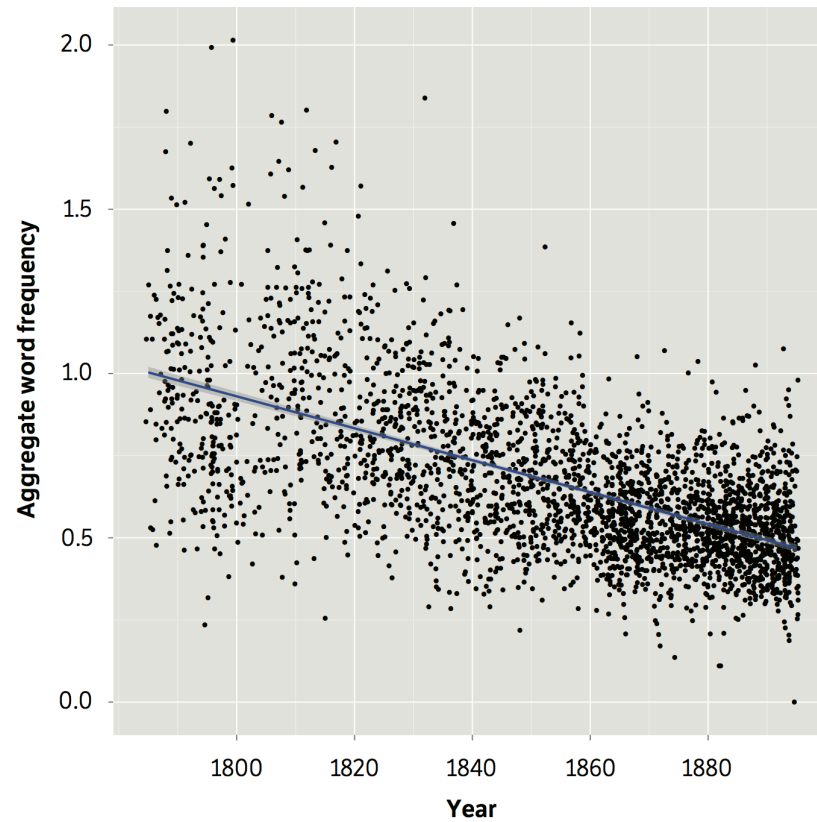


Figure 8: Aggregate term frequencies of the abstract values fields combined in novels, 1785-1900.

“The Quiet Transformation of Literary Studies: What Thirteen Thousand Scholars Could Tell Us” (2014)

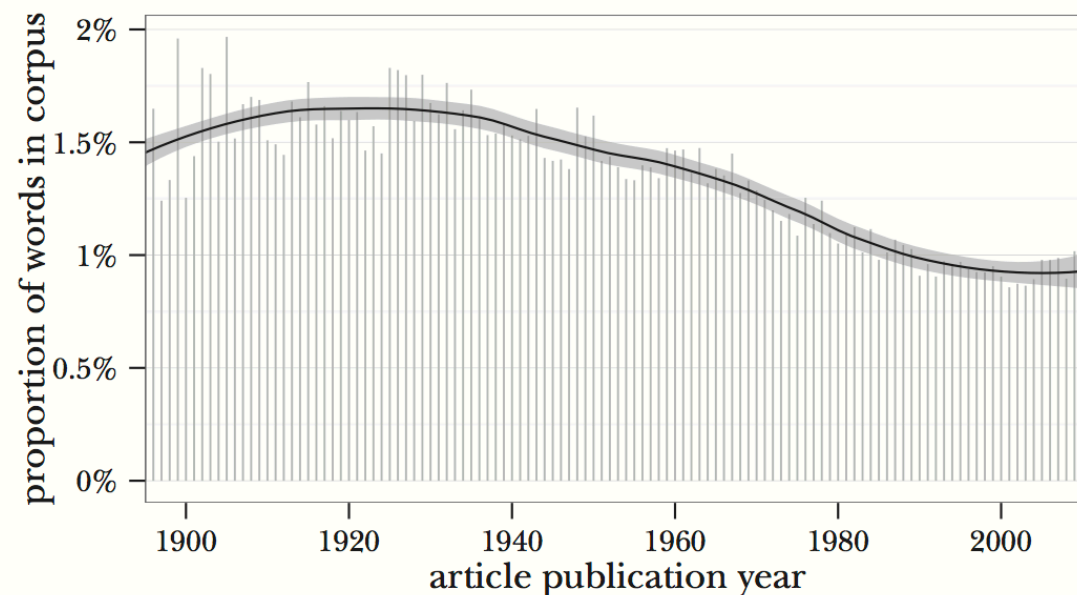
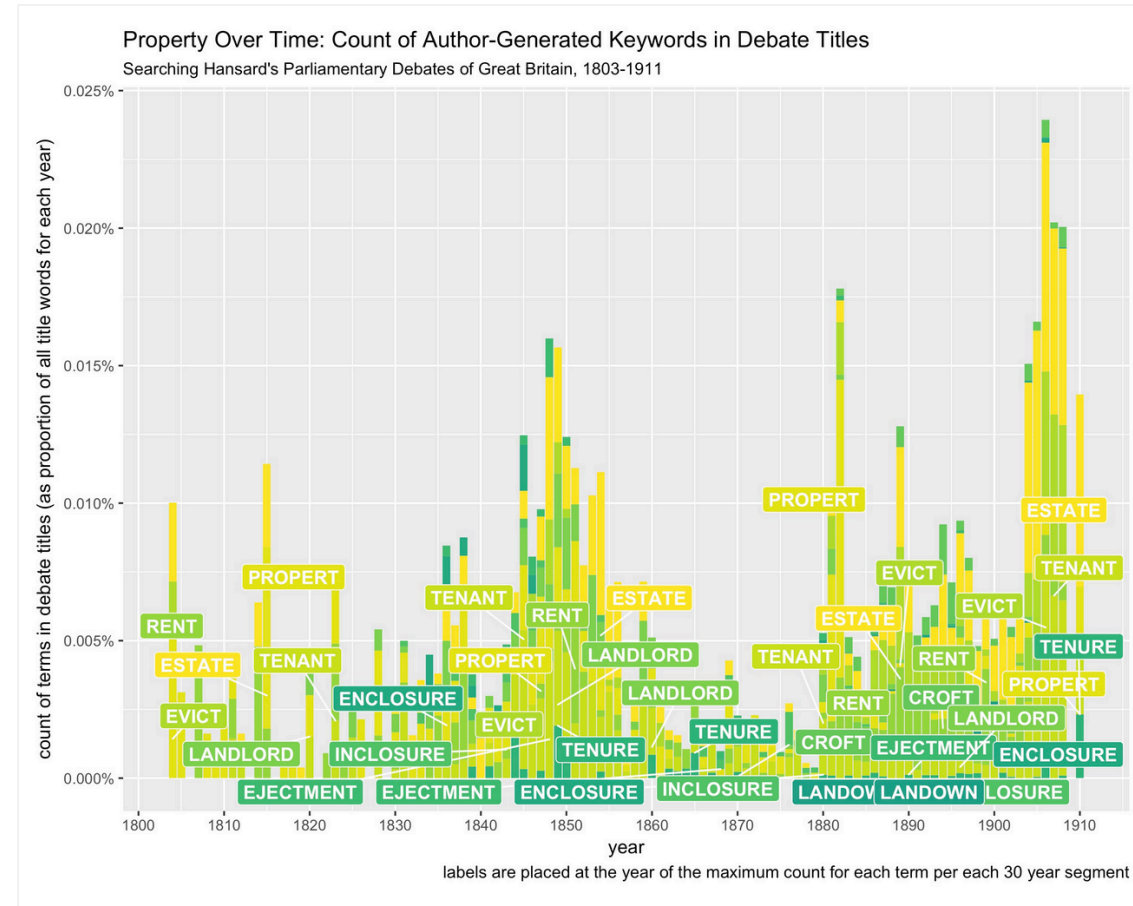


Fig. 1. Yearly frequencies of number words from *two* to *hundred* and *first* to *tenth* in a corpus of articles in seven generalist literary-studies journals. The light gray bars are the data; the dark curve is a smoothed version highlighting the trend.

“Critical Search: A Procedure for Guided Reading in Large-Scale Textual Corpora” (2018)



“The Transformation of Gender in English-Language Fiction” (2018)

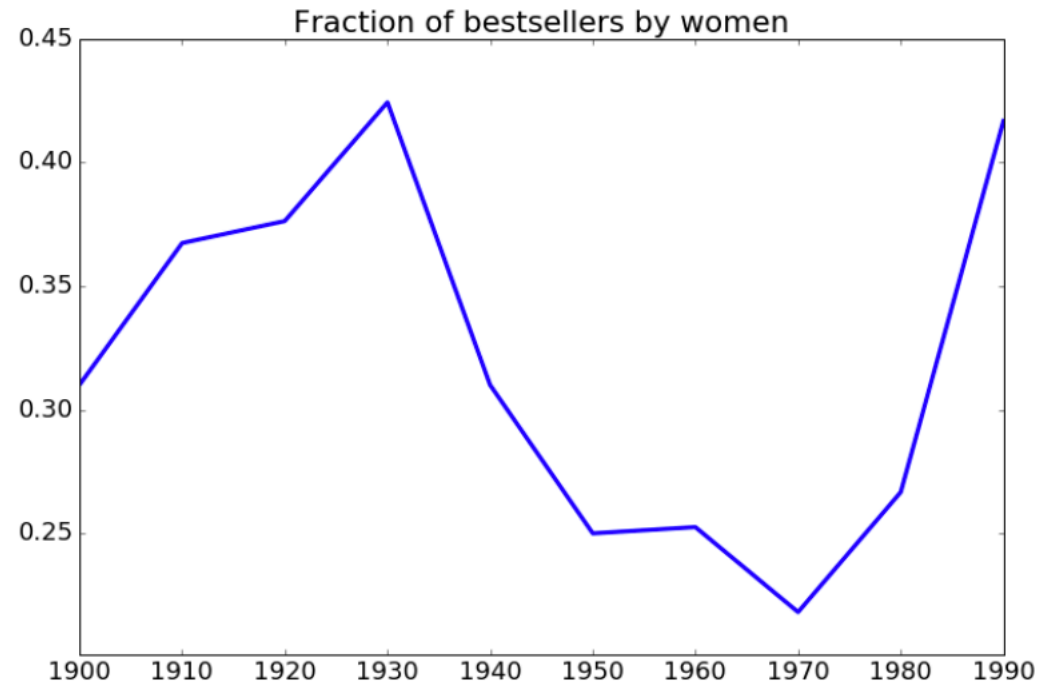
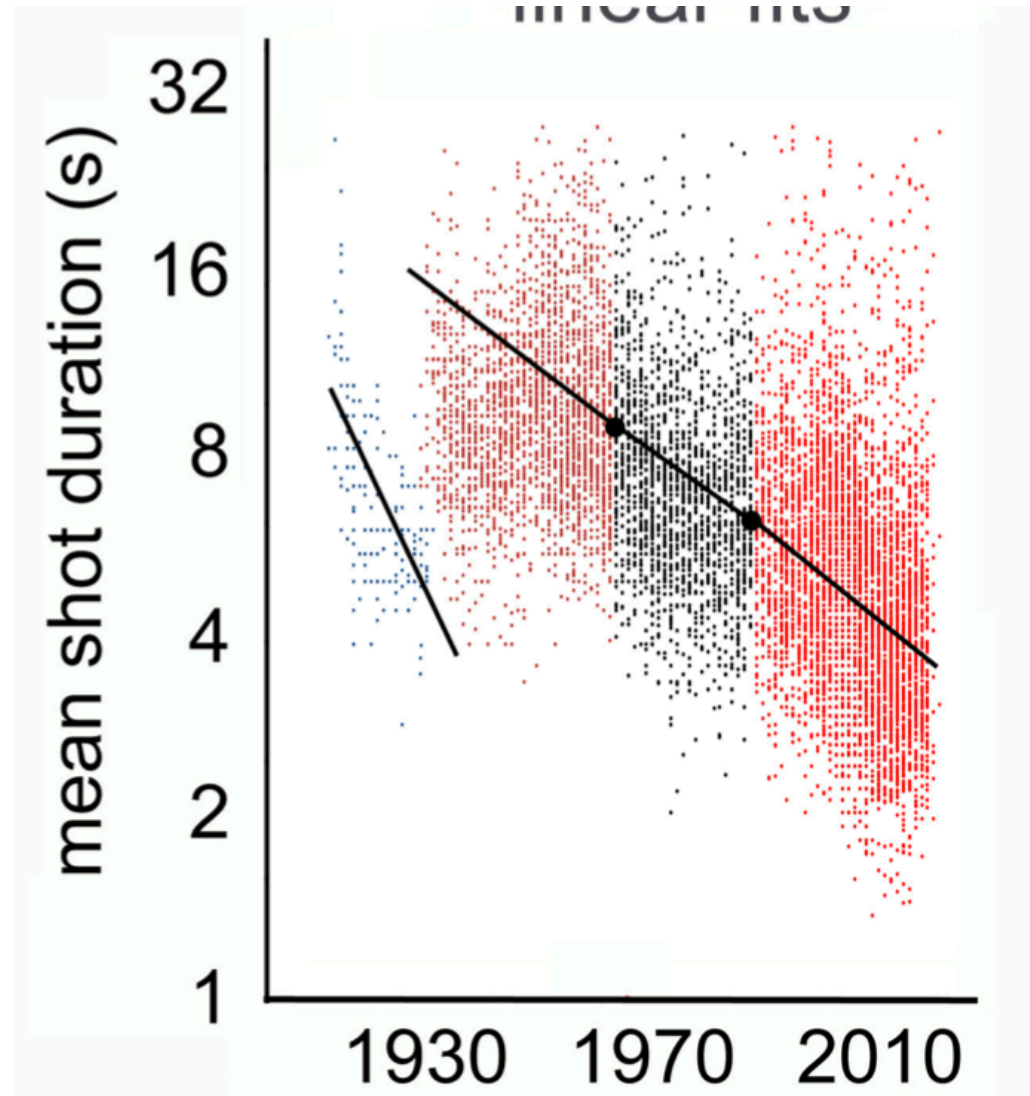


Figure 4. The authors of yearly bestsellers from Publishers Weekly.

“Shot Duration, Shot Classes, and the Increased Pace of Popular Movies” (2015)



“Birth of the cool: a two-centuries decline in emotional expression in Anglophone fiction” (2016)

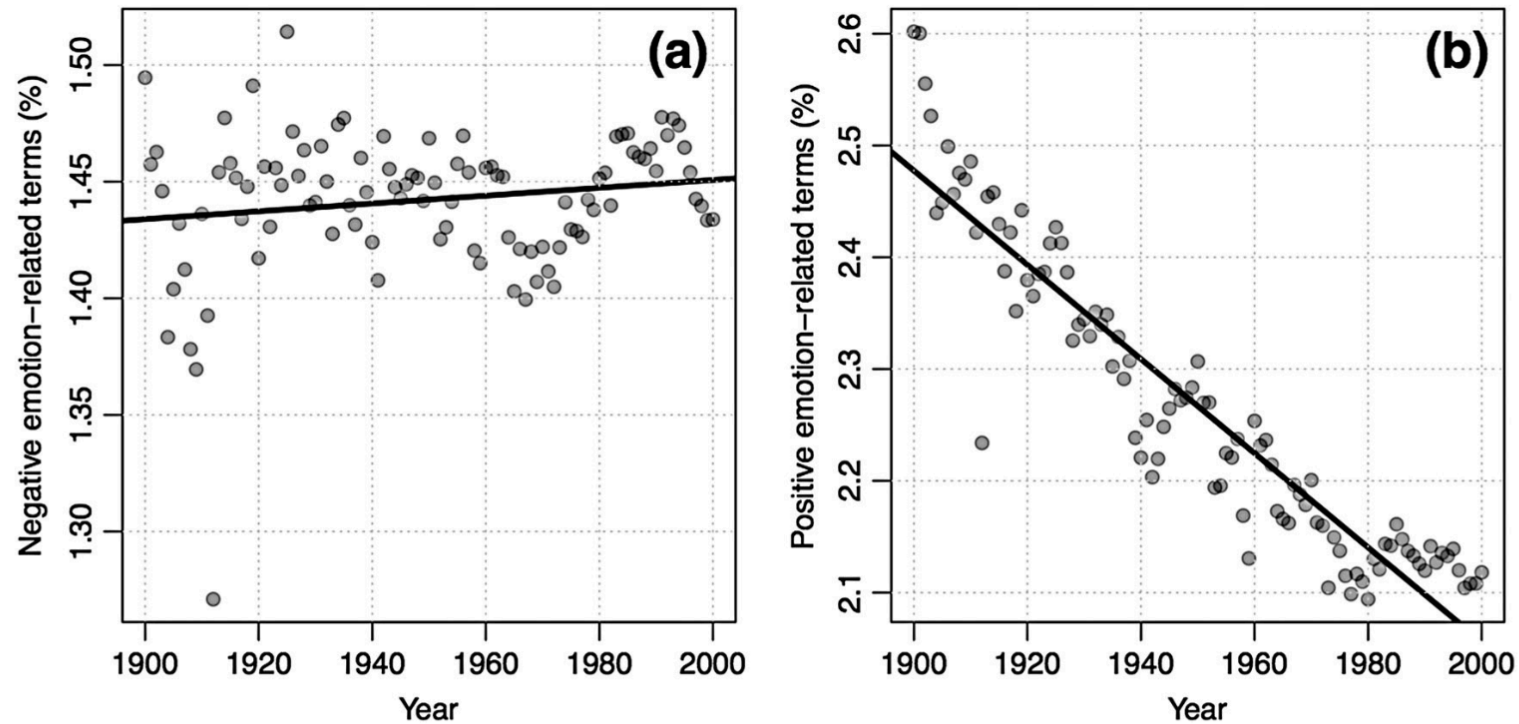
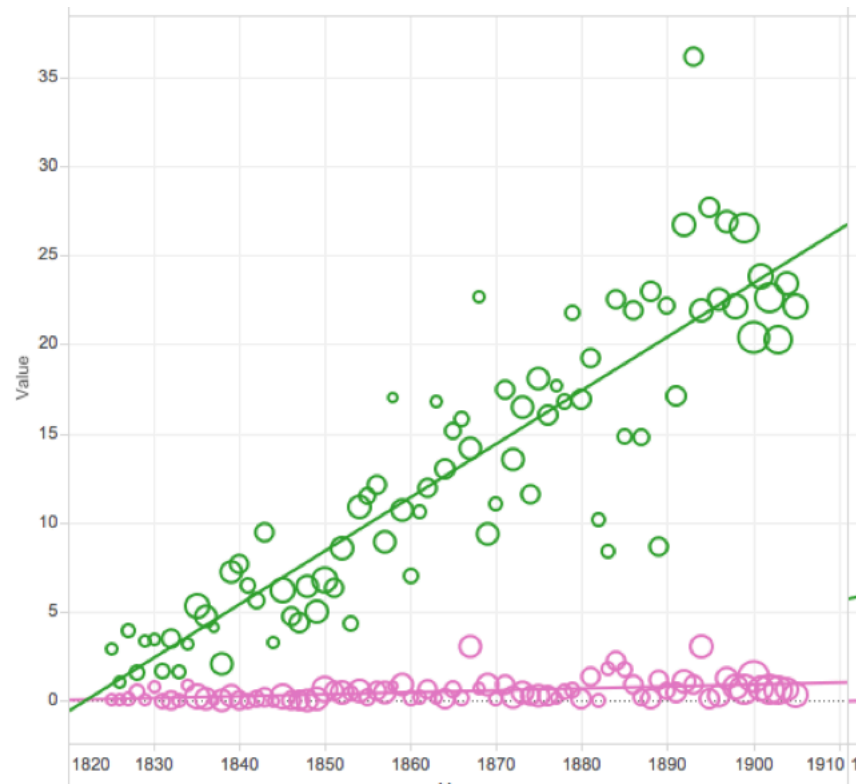


Figure 3. Emotionality changes in Anglophone literature, for the Google Books corpus. (a): negative emotions-related terms. (b): positive emotions-related terms. Solid lines represent linear regressions of the data.

“The Making of Middle American Style” (2016)



“Canon/Archive. Large-scale Dynamics in the Literary Field” (2016)

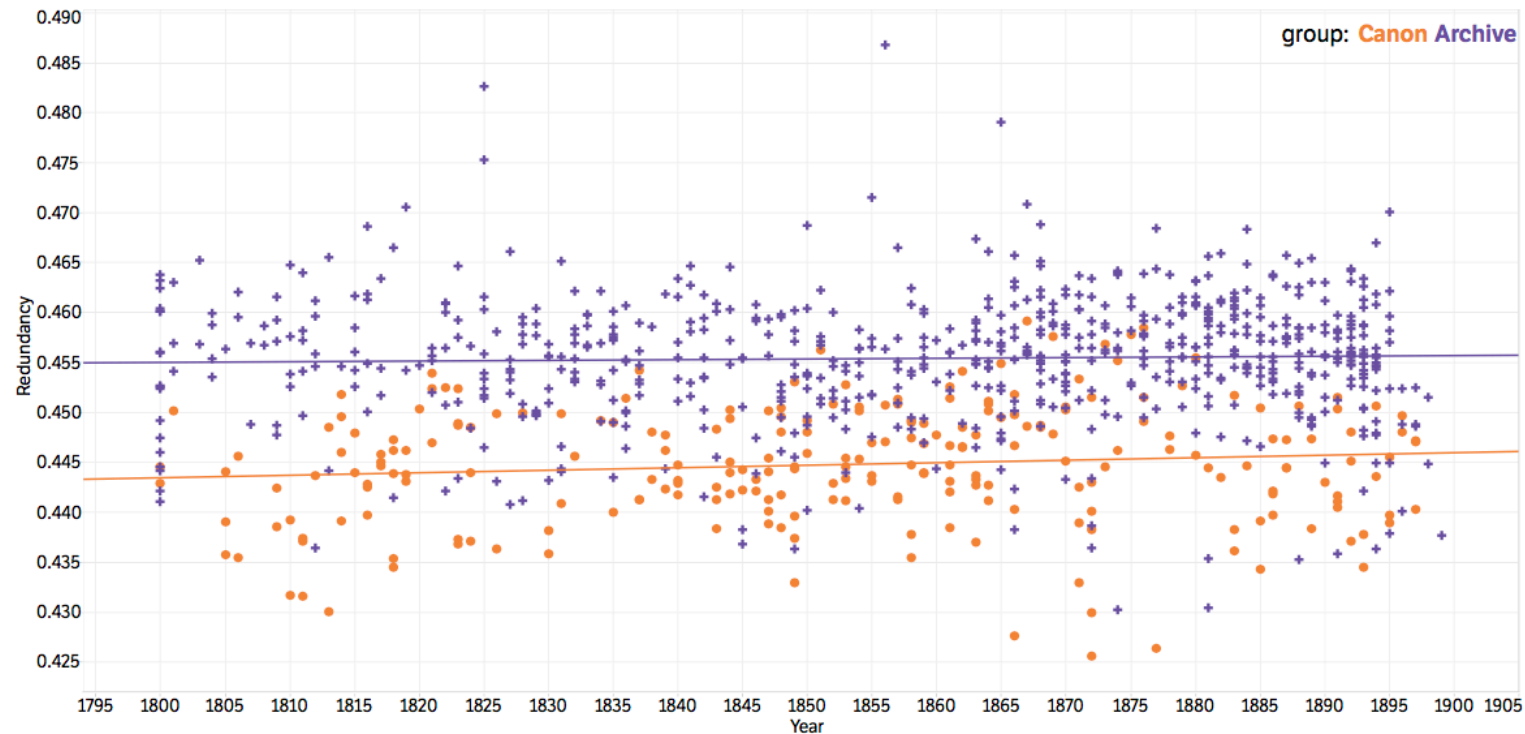


Figure 4.1. Measuring redundancy, 1800-1900

Purple crosses indicate archival novels, orange circles canonical ones

“The Longue Durée of Literary Prestige” (2016)

Underwood and Sellers ■ *Longue Durée* of Literary Prestige

327

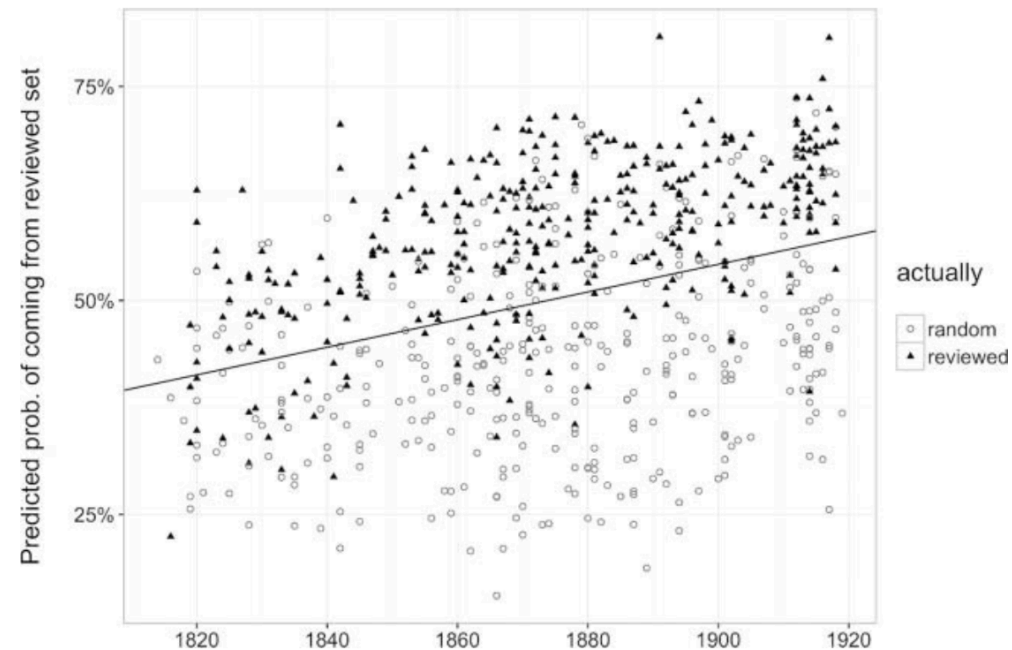


Figure 1. A model of literary prestige from 1820 to 1919

“Dialogism in the novel: A computational model of the dialogic nature of narration and quotations” (2017)

G. Muzny *et al.*

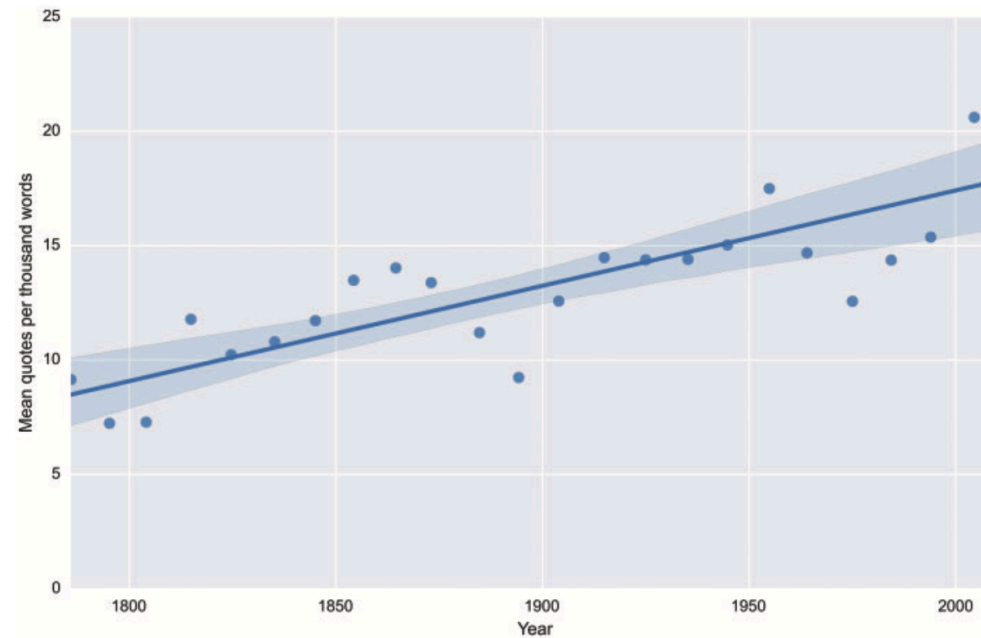
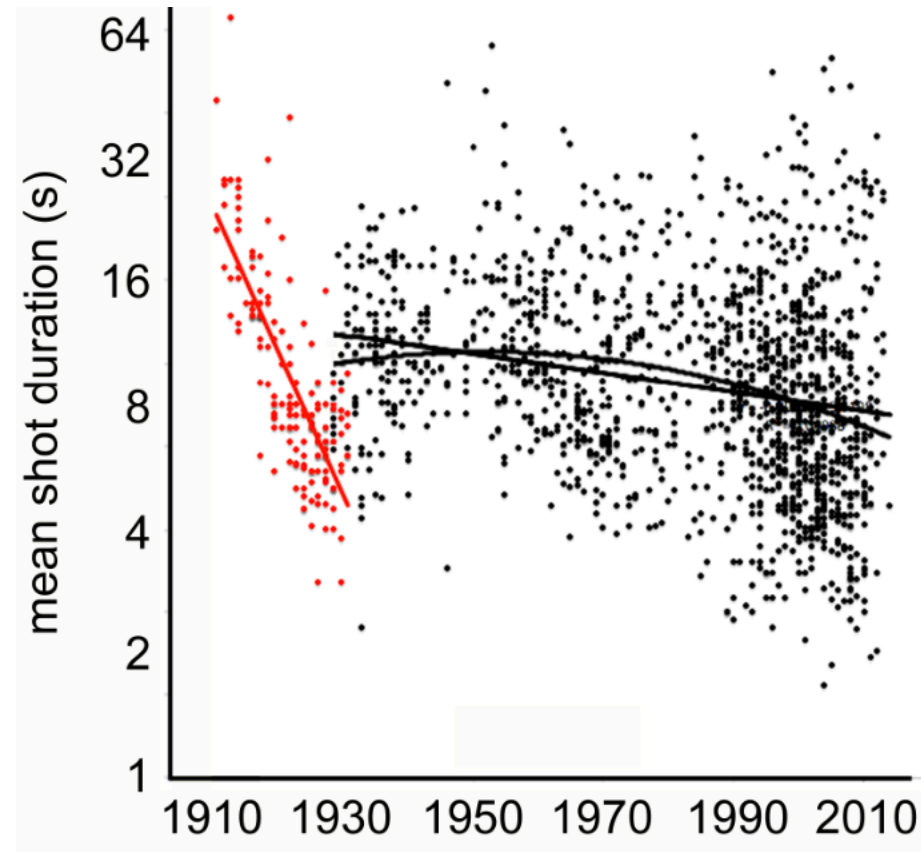


Fig. 2 Number of quotes per thousand words from 1782 to 2011, bucketed per decade, rounded down, so all novels from 1800 to 1809 contribute to the 1800 data point. This approach is used in the all bucketed figures. The line of best fit shows an upward-sloping trend with an r^2 value of 0.65

“Shot Duration, Shot Classes, and the Increased Pace of Popular Movies” (2015)



“The civilizing process in London’s Old Bailey” (2014)

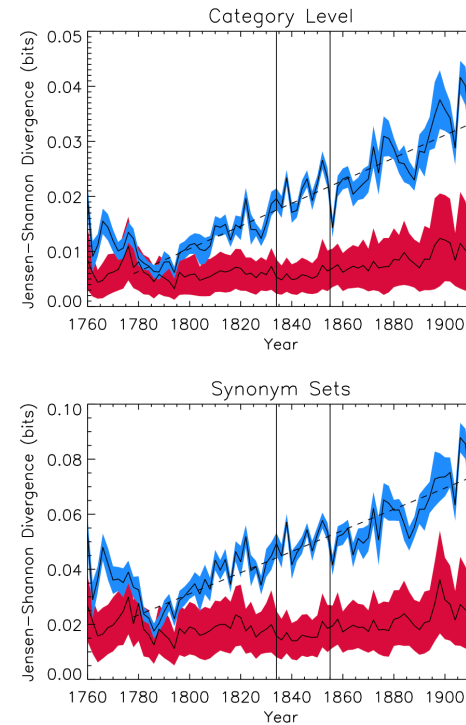


Fig. 1. The emergence of bureaucratic distinctions between violent and nonviolent offenses in the Old Bailey, 1778–1913. Shown is the JSD (Eq. 1) at the category level (*Upper*) and the synonym set level (*Lower*). The blue line shows the JSD (and 1σ ranges) for the violent-nonviolent distinction, whereas the red line shows the ranges for a null hypothesis where the trial genres are defined by arbitrary groupings of indictment classes. Overplot as a dashed line is the maximum likelihood fit for a linear increase in the JSD beginning at 1778. Vertical lines indicate the two major administrative changes in the system in this period.

“The civilizing process in London’s Old Bailey” (2014)

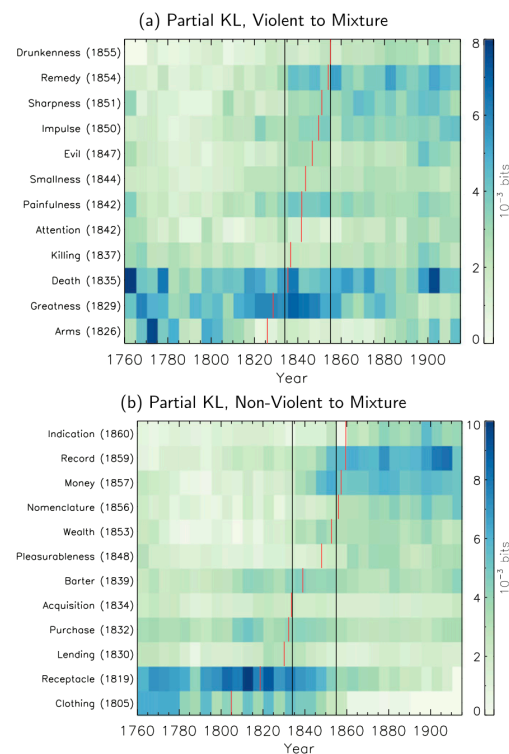


Fig. 2. Distinguishing features of violent and nonviolent trials in the OBC. Shown here is the partial KL, KL_i (Eq. 2), for the 12 most distinctive synonym sets for violent indictments (A) and nonviolent indictments (B). Sets low on the y axis are associated with earlier dates. The centroid for each term is shown in parentheses and marked in red.

Scala della ragione, Verona (c. 1450)



“Loudness in the Novel” (2014)

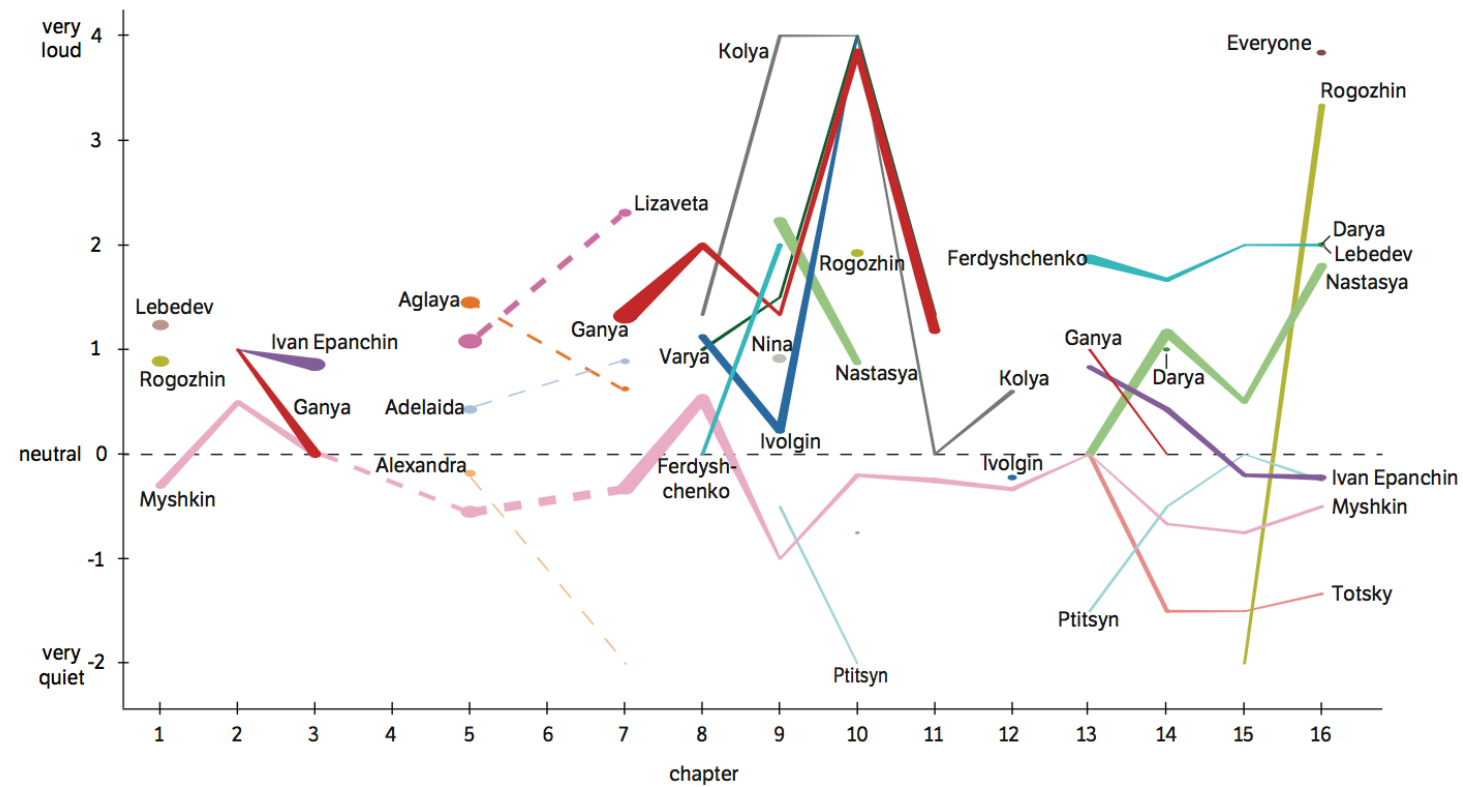


Figure 5: Loudness allocated between characters: *The Idiot*, Book I.

“Computational analysis of the body in European fairy tales” (2013)

S. Weingart and J. Jorgensen

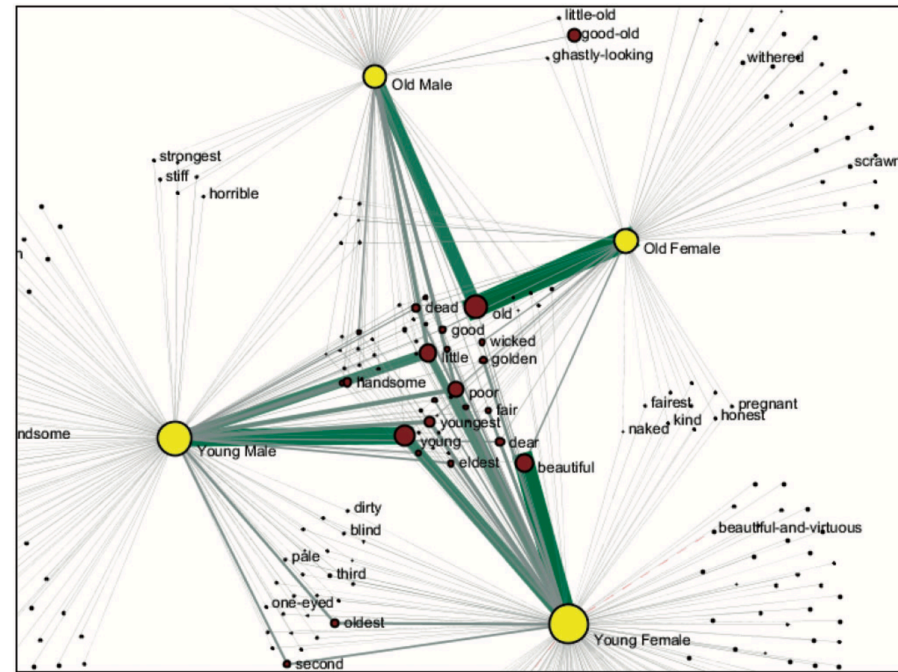


Fig. 4 The four light (yellow) dots represent *old male*, *young male*, *old female*, and *young female*, while each of the dark (red) dots represents an adjective. A line is drawn between an adjective (the dark/red dots) and a body type (the light/yellow dots) if one is used to describe the other. That line gets thicker if it is used more frequently. Thus, a thick line connects *beautiful* to *young female*. Because *beautiful* is in the middle, we can also see it is occasionally used to describe *old females* and *young males*—although never with *old males*.

“Repetition and East Asian Literary Modernity 1900-1930” (2018)

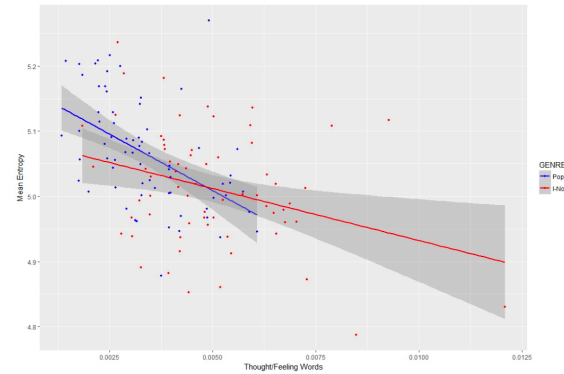
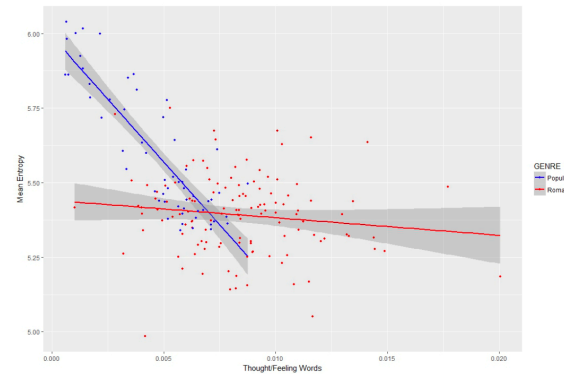
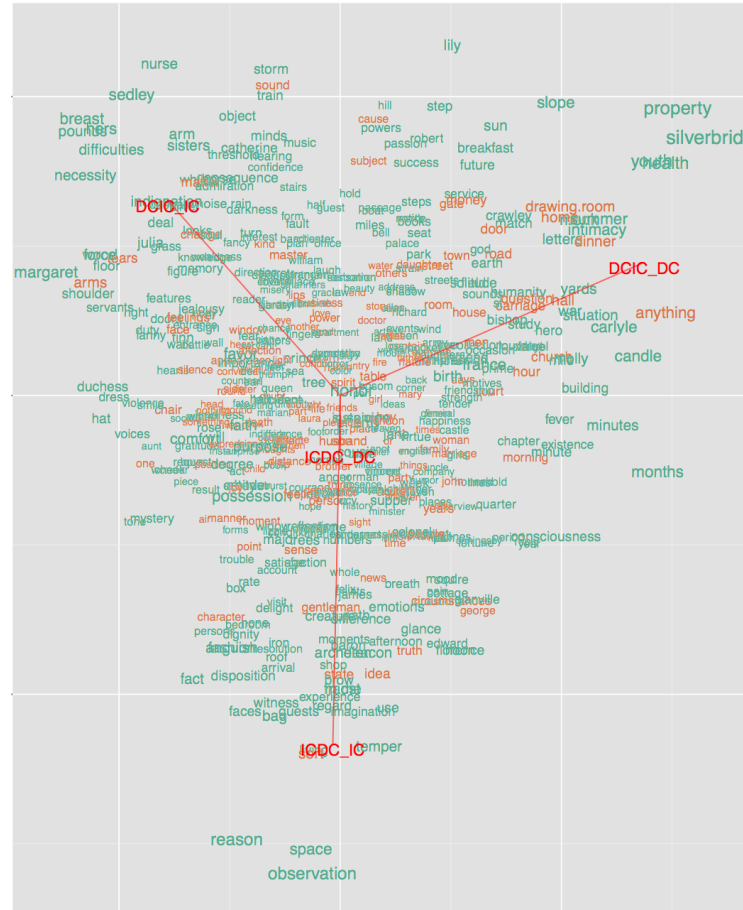


Figure 2. Plots for the ratio of “thought/feeling” words against average entropy for Japan and China, with linear regression lines fitted by genre. In both cases, we can observe that as the ratio of “thought/feeling” words increases (horizontal axis), the mean entropy of the texts decreases (vertical axis), indicating more lexical repetition.



“Style at the Scale of the Sentence” (2013)



“Style at the Scale of the Sentence” (2013)



“On Paragraphs. Scale, Themes, and Narrative Form” (2015)

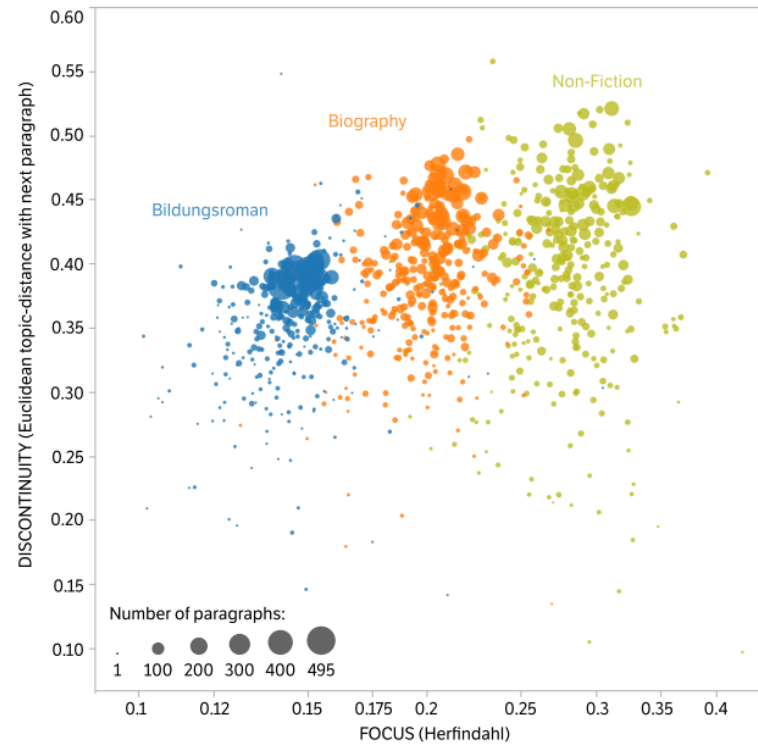
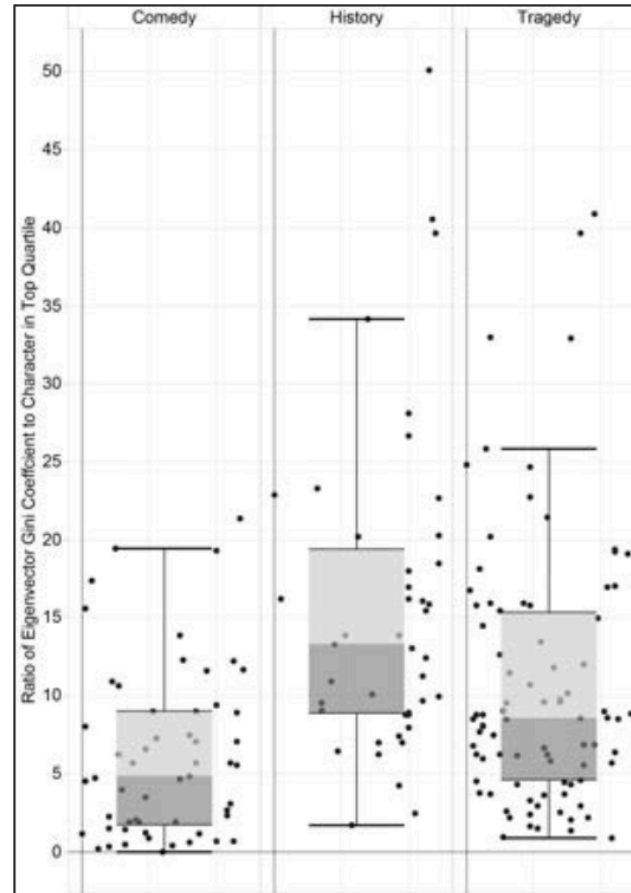


Figure 6.3 Focus and Discontinuity in three Different Registers

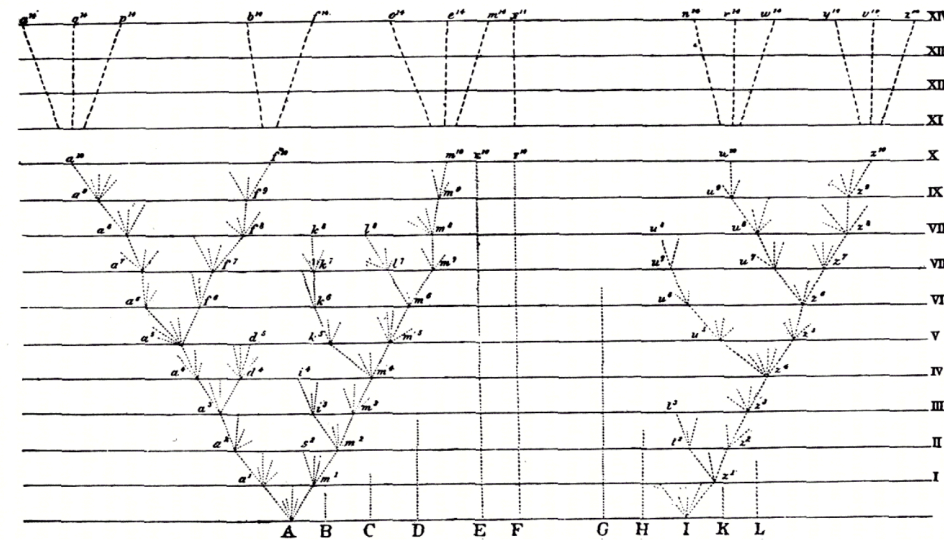
In this chart, the x-axis measures the focus of paragraphs, and the y-axis their discontinuity. The separation between the three discourses – and especially between fiction and non-narrative non-fiction – is unmistakable, and persists even if we topic model each text separately; that is to say, the greater focus of non-fictional texts is not due to the fact that the topics of two books dealing with Kansas and pessimism tend to be more unlike each other than those of two novels (although this is also true); rather, non-fictional paragraphs are both more focused and more discontinuous *even within each individual text*.

“Distributed Character: Quantitative Models of the English Stage, 1550-1900” (2017)



from *The Origin of Species*

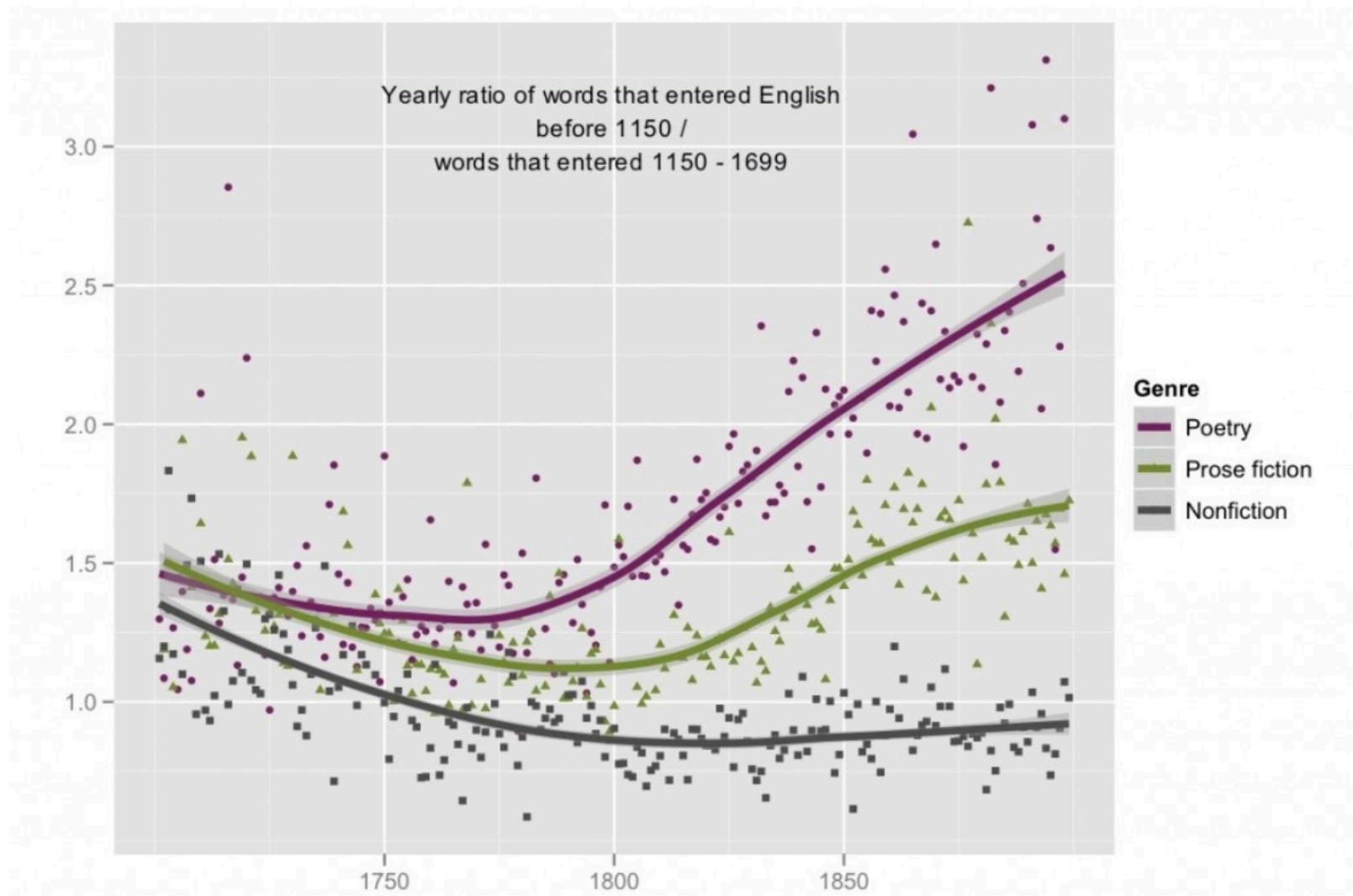
FIGURE 28: Divergence of character



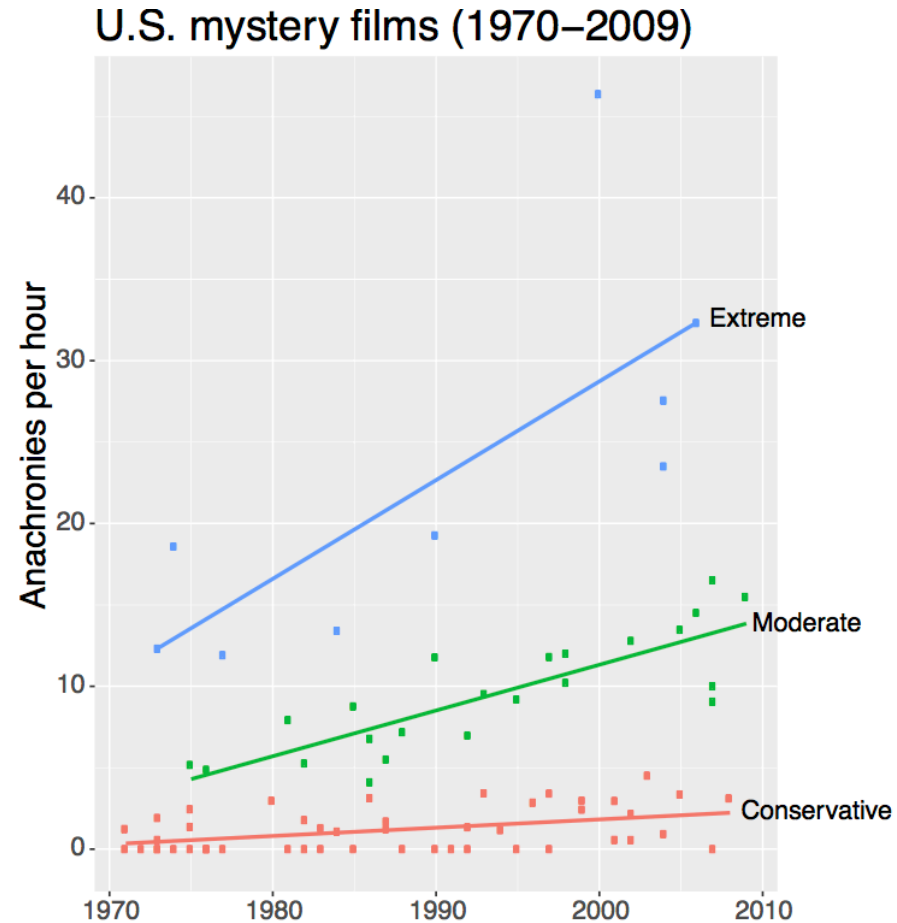
Let A be a common, widely-diffused, and varying species, belonging to a genus large in its own country. The little fan of diverging dotted lines of unequal lengths proceeding from A may represent its varying offspring . . . Only those variations which are in some way profitable will be preserved or naturally selected. And here the importance of the principle of benefit being derived from divergence of character comes in; for this will generally lead to the most different or divergent variations (represented by the outer dotted lines) being preserved and accumulated by natural selection.

Charles Darwin, *On the Origin of Species*

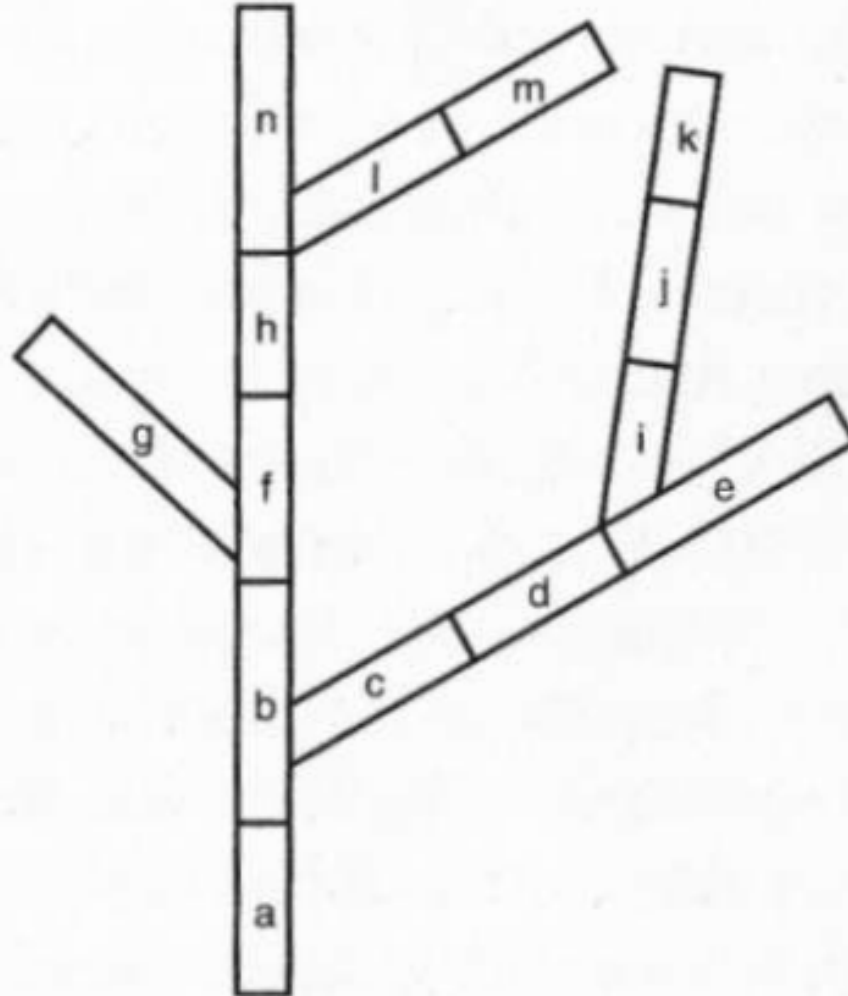
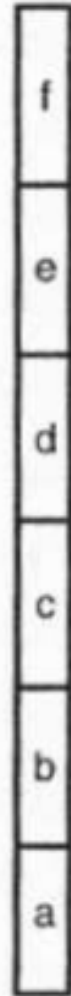
“The Emergence of Literary Diction” (2012)



“Broken Time, Continued Evolution: Anachronies in Contemporary Films” (2017).



Where does this come from?



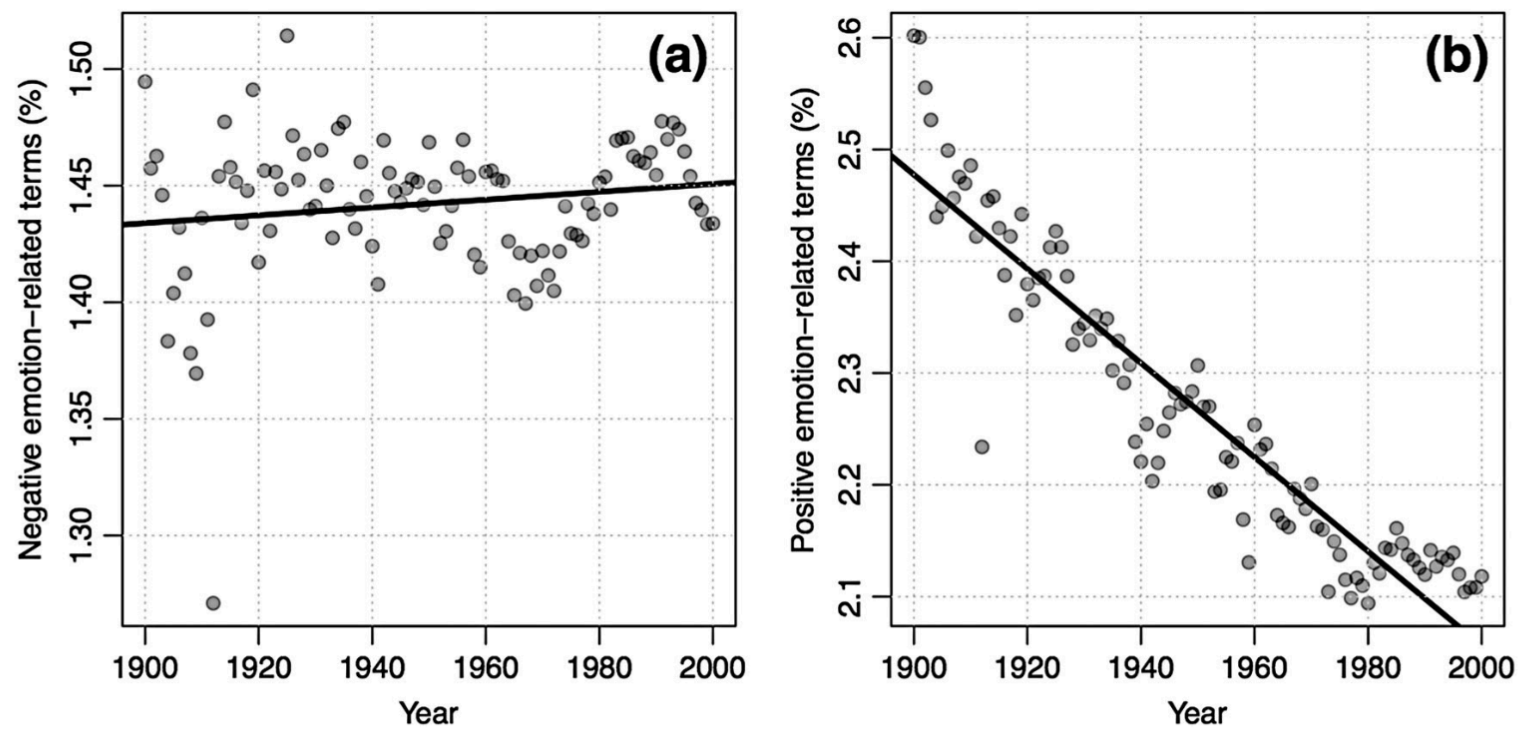


Figure 3. Emotionality changes in Anglophone literature, for the Google Books corpus. (a): negative emotions-related terms. (b): positive emotions-related terms. Solid lines represent linear regressions of the data.