

Tracing and Classifying Social Frustrations Over Time: A Computational Text Analysis of Social Problems in Early Modern England (1475-1700)

Alejandro Abisambra – Northwestern University

This paper leverages computational text analysis methods using a corpus of English-language texts published in England between 1475 and 1700 to identify the distribution of social problems or frustrations identified by the authors in the corpus and how they change over time. The sweeping structural social changes in Western Europe since the industrial revolution and the consolidation of modernity (starting circa. 1770-1840) are well documented (Collins 1980; Scott 1998; Weber 1927; Wrigley 2016). Moreover, studies have sought to explain the occurrence of these transformations in the West as opposed to other regions (Mokyr 2017; Pomeranz 2009), and the literature has documented the rise of scientific and technical discourse over this period (Almelhem et al. 2026; Erikson 2021). What is less well documented is the extent to which pre-industrial, pre-modern societies in these geographies were concerned with the transformations that were to come in the following centuries. Did modernity and industrialization address the major social frustrations and social problems of pre-industrial times or were these social transformations a process of simultaneously creating solutions to newly manufactured social problems that were not necessarily considered priorities by their ancestors? If divergent, within which time window does the stream of social problems voiced in the corpus begin to approach the array of macro-social changes that occurred in the late eighteenth and early nineteenth centuries in the west?

To address these questions, the project relies on a vast and representative corpus of approximately 60,000 texts published in England that covers the period between 1475 and 1700 (Early English Books Online-TCP). I chose this collection for its broad coverage and because it offers an entry point into English society in the early modern, pre-industrial period; the social order that immediately preceded the society that then led the industrialization and macro-social transformations of the modern western world. Using computational text analysis techniques, I perform a series of analyses to (i) establish a classification of the corpus based on the main topics being treated (i.e. topic models) and how they change over time; and (ii) train the models to identify and retrieve the array of social problems discussed by the authors within each of these topics.

Theoretical Framework

For the purposes of this project, I define a social problem or social frustration (which I will use interchangeably) as instances where an author presents an idea that meets the following 3 characteristics. First, it describes a phenomenon that goes beyond the individual or immediate close-kin context. For example, 'my cow died from a disease' is a statement at the individual level, while 'many cows in the village are dying from disease' speaks at the collective or social level. Second, the described phenomenon is characterized as having a negative valence, something that is an undesirable state of affairs. Third, the phenomenon is characterized as something addressable: that is, something towards which human efforts should be directed to change or rectify.

After identifying the universe of social problems discussed in the corpus, I will classify them way into a typology that has two dimensions. The first dimension identifies whether the problem is framed as a redistributive or a ‘growth’ problem. The second dimension identifies whether the problem is framed as one that is solvable within the existing means/technologies of the time (present-oriented) or one that requires the creation of new means and technologies (future-oriented), much like the industrial changes generated later on.

[Table 1]. Brief examples of hypothetical problems in each quadrant from the year 1600 perspective.

	Redistributive	Growth
Present-Oriented	The riches of the king and the court are too large in comparison to the meager lives of peasants. Land rents need to decrease.	There are not enough vessels in the navy’s fleet. Construction of new vessels is needed to remain competitive against foreign powers.
Future-Oriented	Institutional change: The king and courts system is a system that is too closed, non-representative. We need to move to systems of broader representation in political power and decisions.	The navy’s fleet is too slow and inefficient. Wind-powered vessels are not good enough; new breakthroughs are needed.

The typology is grounded in the broader framework of my dissertation, asking to what extent two defining features of modernity appear in pre-modern social discourse. The first feature concerns the unprecedented economic and demographic growth associated with technological and institutional change, as well as the cultural centrality of growth and profit as social logics (Erikson 2021; Collins 1980; Weber 1927). The second feature involves the modern turn toward future-making and large-scale social engineering, characterized by efforts to master nature and reshape the environment to human aims (Giddens 1990; Scott 1998). Because these dimensions are widely seen as distinguishing the modern industrial era from its predecessors, the typology is designed to assess whether pre-industrial societies articulate concerns in ways that echo –or not– these distinct modern features.

Data (Early English Books Online EEBO-TCP)

The paper uses the data from the Early English Books Online – Text Creation Partnership (EEBO-TCP) collection, which offers strong alignment with the research aims of the project. EEBO-TCP provides rich coverage of printed English-language works from 1475 to 1700. It captures more than two centuries of social, political, religious, and economic discourse, through ~60,000 transcribed works spanning a vast array of authors, genres, and domains. It prioritizes first editions and typically excludes later reprints or new editions. It also excludes periodicals, which is a constraint for my project but does not undermine the dataset’s value for broad analysis

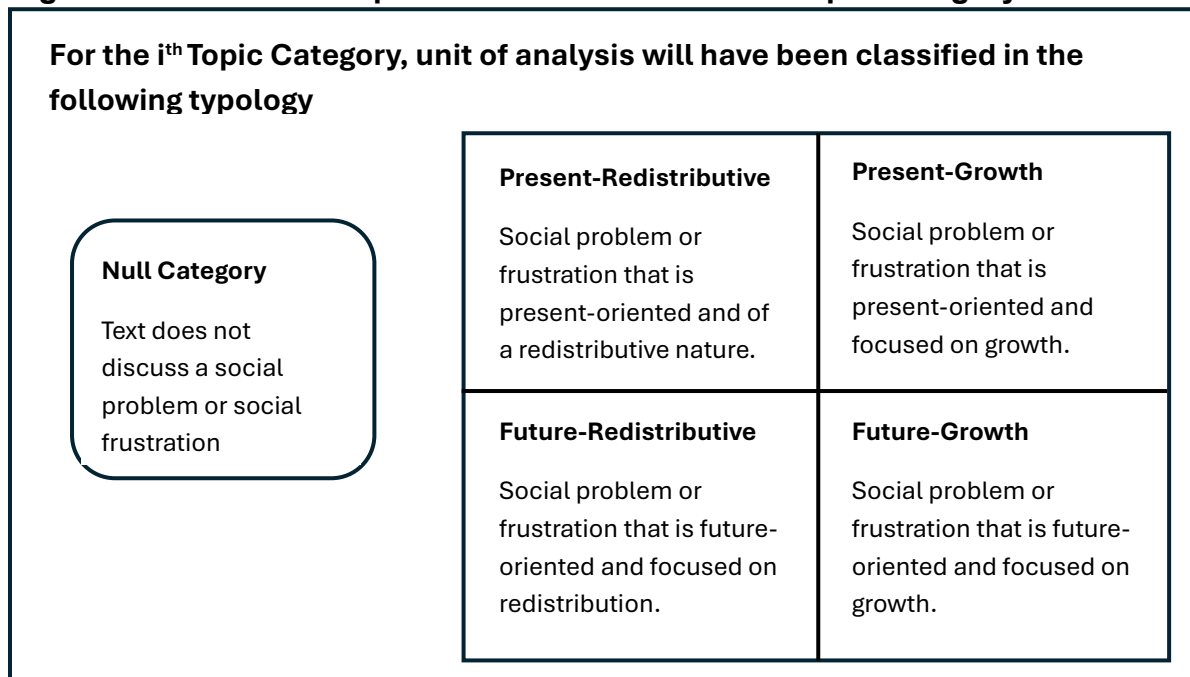
of printed discourse. Part of its contents have already been used by sociologists for historical studies (Erikson 2021).

Methodological Approach

The methodology can be divided in 3 stages, following the structure suggested by Grimmer et al (2022). First stage consists of identifying the main topical groupings in which the corpus can be divided. In this stage of discovery, I plan to use clustering, topic models, and potentially PCA. The goal at this stage is to discover the main topics that are being discussed in the corpus. I expect there will be substantial presence of health and medical topics, but also presence of governance, political economy, and aesthetics and morality.

The second stage consists of annotating the corpus. Within each of the topical areas identified in the first stage, I will identify the segments that discuss social problems or frustrations (following my definition above), and then classify those that do into one of the four quadrants of my typology (future vs. present-oriented, and redistributive vs. growth oriented). To do this, I plan to use supervised learning models in a cycle of human labeling, model training, and model validation.

Figure 1. Annotated Corpus Classification for each Topic Category



The final stage of the methodology focuses on interpreting the results in relation to broader historical dynamics. I will provide a description of the nature of the social frustrations that appear within each quadrant of the typology and trace how these forms of frustration evolve across time. Particular attention will be devoted to assessing whether, and at what point, the content of these frustrations begins to converge with the new technologies, practices, and forms of social

organization associated with the industrial revolution and the consolidation of modern state institutions in Europe.

The analysis will also compare patterns across the four quadrants of the typology to establish if and how the relative prevalence of different types of social frustration –present- vs future-oriented, as well growth vs redistribution-oriented– varies across the period of analysis. By situating these shifts within the typological framework, the interpretation will highlight both the continuities and discontinuities in how social grievances were articulated across pre-modern and early modern Europe.

References:

- Almelhem, Ali, Murat Iyigun, Austin Kennedy, and Jared Rubin. 2026. "Enlightenment Ideals and Belief in Progress in the Run-up to the Industrial Revolution: A Textual Analysis." *The Quarterly Journal of Economics* 141(1):263–314. doi:10.1093/qje/qjaf054.
- Collins, Randall. 1980. "Weber's Last Theory of Capitalism: A Systematization." *American Sociological Review* 45(6):925–42. doi:10.2307/2094910.
- Erikson, Emily. 2021. *Trade and Nation: How Companies and Politics Reshaped Economic Thought*. 1st ed. The Middle Range Series. New York: Columbia University Press.
- Giddens, Anthony. 1990. *The Consequences of Modernity*. Cambridge: Polity in association with Blackwell.
- Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. 2022. *Text as Data: A New Framework for Machine Learning and the Social Sciences*. Princeton University Press.
- Mokyr, Joel. 2017. *A Culture of Growth: The Origins of the Modern Economy*. The Graz Schumpeter Lectures. Princeton: University Press.
- Pomeranz, Kenneth. 2009. *The Great Divergence: China, Europe, and the Making of the Modern World Economy*. Vol. 6. The Princeton Economic History of the Western World. Princeton, NJ: Princeton University Press.
- Scott, James C. 1998. *Seeing like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale Agrarian Studies. New Haven [Conn: Yale University Press.
- Weber, Max. 1927. *General Economic History*. Adelphi Economic Series. London: Allen & Unwin.
- Wrigley, Edward Anthony. 2016. *The Path to Sustained Growth: England's Transition from an Organic Economy to an Industrial Revolution*. Cambridge University Press.