

Introduction

In his often-quoted 1959 Rede Lecture (Snow, 2012), English novelist and physicist Charles P. Snow lamented the growing incompatibility between scientific and humanistic modes of thinking and the negative influence of this phenomenon on societal development. In recent decades, however, this “split between ‘literacy’ and ‘numeracy’, between quantifying-empirical and narrative-interpreting cultures of knowledge” seems to have been increasingly mended. Partially thanks to the diffusion of ‘mixed methods’ research frameworks, it is now accepted both that “the sciences do not only collect their data but also have data to interpret, be it by theory or narration”, and that “numbers and counting also play an important role in the Humanities” (Krämer, 2023b, 2).

A prominent role in deconstructing any oppositional thinking has also been played by the Digital Humanities (DH) research community. Within the subfield of Computational Literary Studies (CLS), for example, qualitative and quantitative methods are increasingly interwoven in approaches such as Martin Mueller’s ‘scalable reading’ (Mueller, 2020; see also Krautter, 2024) – a more balanced alternative to the much-discussed ‘distant reading’ and ‘macroanalysis’ paradigms brought forth respectively by Franco Moretti and Matthew Jockers (Moretti, 2013a; Jockers, 2013).

At the same time, the perception of the research objects themselves has changed as well. According to the traditional distinction between ‘idiographic’ and ‘nomothetic’ approaches to knowledge, popularised in the late nineteenth century by Neo-Kantian philosopher Wilhelm Windelband (see Windelband, 1998), humanities research is generally devoted

to studying individual and unique phenomena, while sciences try to derive general rules and patterns from objective fact (Sobchuk, 2018, 26).

Such a perspective has now become quite outdated, as it has become clear how “pattern-seeking practices have always been part of the humanities” (Bod, 2018, 24; see also Bod, 2013). Furthermore, *explicitly* nomothetic approaches to culture have become more common: scholars of literary studies, for example, have been increasingly attempting to find “law[s] of literary evolution” (Moretti, 2013a, 50), often using theoretical models from biology and other natural sciences (see e.g. Carroll, 1995, 2018).

This book is devoted to one such problem of literary evolution, namely the development of European dramatic literature in the early modern age. Drama seems indeed an especially productive test case for pattern-seeking approaches, insofar as it combines highly conventionalised, easily recognisable formal features – such as act-scene divisions, character systems, speech distributions – with a strong regional variation, which appears especially pronounced in the early modern period.

A seminal essay by Moretti, ‘Modern European Literature: A Geographical Sketch’ (Moretti, 1994), already devoted significant attention to this matter, producing a convincing account of how extremely different dramatic traditions have grown out of earlier common models. Nevertheless, while remaining widely cited, this work and its conclusions have never been tested empirically, while the author himself, despite having later shifted towards more quantitative methods, never revisited its argument using computational techniques.

However, advances in corpora availability and the development of new techniques within the ever-growing field of computational drama analysis¹ now make it possible to attempt such an empirical assessment, and to explore other holistic, data-driven explanations for the peculiar evolution of early modern European drama. This book suggests possible avenues for approaching the issue with computational tools while prototyping replicable workflows to that end. While it cannot fill this gap

1 Defined as “the field of research that attempts to model, analyze, and interpret dramatic texts using computational methods” (Andresen and Reiter, 2024b, 1).

completely – a daunting task, requiring collaborative, large-scale effort – it provides a foundation for such future work.

Chapter One establishes the theoretical backdrop for the work, focusing on the evolution of European theatre during the early modern period. It begins by examining Franco Moretti's evolutionist take on the development of drama in that age. Complementing his view, I then briefly assess some competing theories that emphasise the cross-border similarities of different dramatic traditions, such as Joachim Küpper's cultural net theory (Küpper, 2018) and Louise Clubb's theatergrams (Clubb, 1990). Before transitioning to the experimental section of the work, the chapter concludes with a concise summary of the history of dramatic literature in the five regions under study (English-, French-, German-, Italian-, and Spanish-speaking areas), which is meant to aid the reader in navigating early modern poetics.

Chapter Two introduces the Early Modern Drama Corpus (EmDraCor), a multilingual collection of texts specifically assembled for investigating my research question. EmDraCor was developed as an informal spin-off of the larger *Drama Corpora* (DraCor) project (Fischer et al., 2019b), an open platform and a growing network of resources for hosting, accessing, and analysing theatre plays,² and thus follows its structure, best practices, and technical requirements.

DraCor can be described as an ever-growing digital archive for TEI-XML-encoded dramatic texts, which are organised into several 'programmable corpora' (mostly language-, but also author-based). The project also features a suite of applications, packages, and tools for computational literary studies which can be employed on each corpus – mostly revolving around the use of the API to compute various types of network and textual metrics. The collections themselves are mostly born from conversion and adaptation of existing digital sources, following the pipelines outlined in Börner et al. (2023a), and are contributed both by the DraCor team and by external scholars.

Before discussing the actual composition of EmDraCor, Chapter Two outlines the selection criteria for the corpus in terms of timeframe

2 See <https://dracor.org/doc/what-is-dracor>.

covered, geographical scope, and size. I then describe the process of retrieving suitable texts for my goal and the actual corpus building workflow, elaborating on the different pipelines for sources with high-level, medium-, and low-level markup. Editorial practices adopted, especially in terms of textual (post)correction and segmentation, are openly discussed as well. Some final remarks address how the corpus was deployed within the DraCor infrastructure through the Docker technology. Additional details on the specific challenges in building each sub-collection are provided in Appendix A, while a full list of the plays included in the corpus is presented in Appendix B.

Chapter Three deals with the methods employed to investigate EmDraCor. As a preliminary step, it discusses some options for operationalising drama, starting from pioneering approaches from the formalist and structuralist tradition, such as those by Boris Yarkho and Solomon Marcus, and considering early computational attempts such as those by Hartmut Ilseman. An excursus is then devoted to the analysis of dramatic networks, which has emerged in the last twenty years as one of the major formal approaches to drama.

Afterwards, the actual strategy for modelling EmDraCor plays is outlined. Here, I propose *vectorising* each play, i.e. transforming it into a mathematical object (vector or embedding) containing various metrics related to textual features (characters, interactions, overall plot, etc.). By performing mathematical operations on the vectorised texts, systematic comparison of dramatic structures across the corpus becomes possible. A typology of the features employed, according to the different aspects of drama they embody, is also sketched, with a detailed overview being later provided in Appendix C.

Chapter Four illustrates the implementation and results of several experiments conducted on the corpus, grouped according to the 'epistemic objects' they investigate (distances, clusters, patterns). First, the existence of a general trend towards formal diversification within the corpus is probed through repeated measurements of vector distances through two complementary options (pairwise and centroids-based computation). To gain further insight into the phenomenon, I then

move on to plotting the vectors on a Cartesian plane and examining their clustering in successive temporal snapshots.

Since these two methods do not allow for a clear picture of possible evolutive patterns, I later attempt to compute shifts in absolute values for each metric (i.e. each vector component) across the time span, trying to see if the development of each individual feature helps in distinguishing one dramatic tradition from another. In doing so, I am able to construct quantitative profiles for each of the five theatrical cultures considered. As a coda, I present a reproduction experiment, conducted on the larger FreDraCor and EngDraCor corpora. It aims at providing a benchmark for previous findings, but also at demonstrating how the methods developed in the work are generalisable beyond the specific EmDraCor use case.

The concluding chapter recapitulates the main findings of the work, discusses its epistemic limitations, and outlines some directions for future research. Across these pages, I argue that early modern dramatic traditions exhibit measurable formal divergence, but that this divergence proceeds unevenly across structural features, suggesting that 'national' dramatic identities are better understood as clusters of tendencies than unified stylistic programs. Finally, the interested scholar is invited to build upon these preliminary findings, as all materials related to this work – including EmDraCor texts, scripts, and metadata – are available in a public *GitHub* repository³ and archived on *Zenodo*⁴ for long-term preservation.

3 <https://github.com/lucagiovannini7/emdracor>.

4 <https://doi.org/10.5281/zenodo.13827542>.