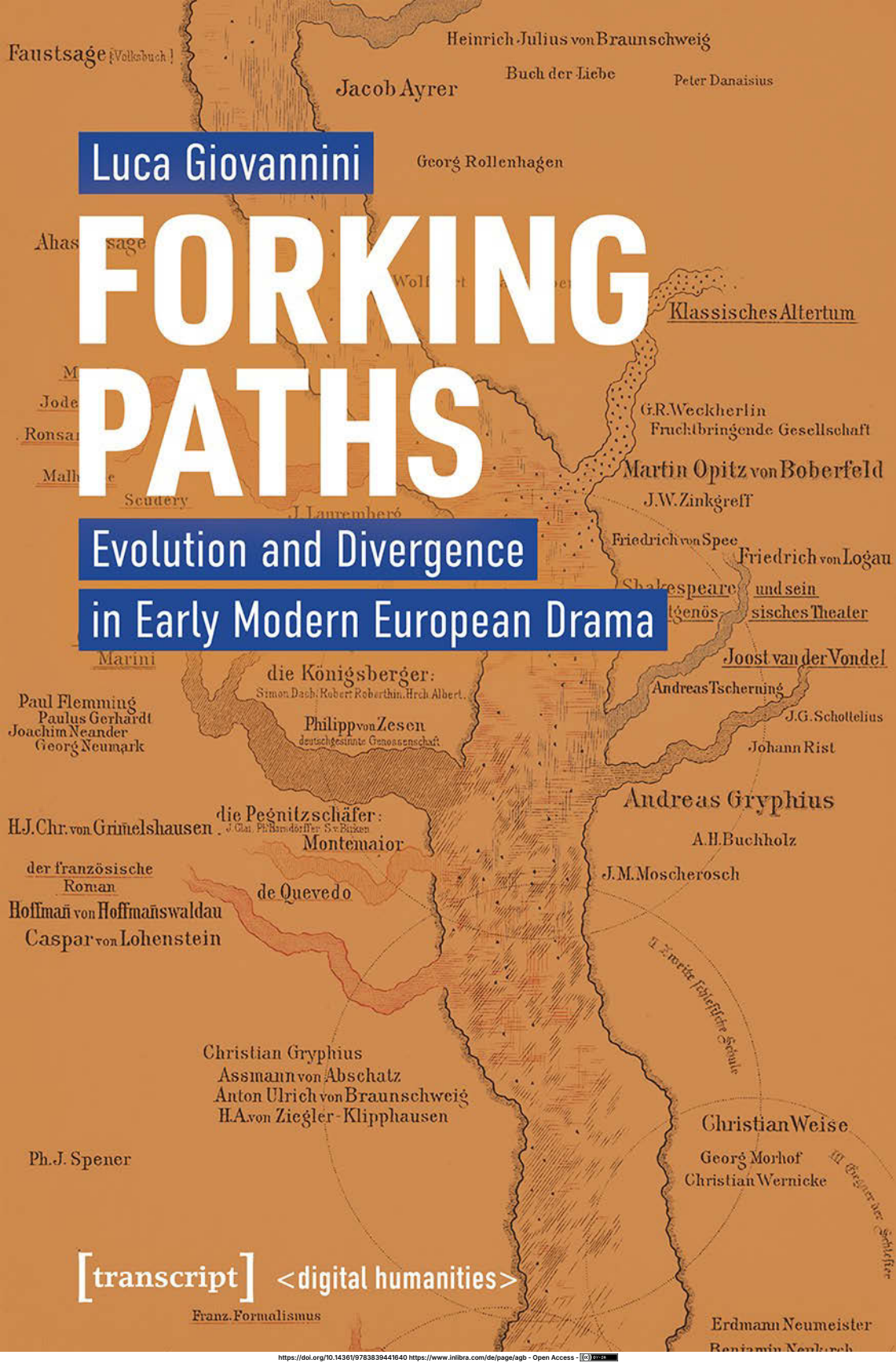




Luca Giovannini

FORKING PATHS

Evolution and Divergence

in Early Modern European Drama

[transcript] <digital humanities>

Franz. Formalismus

Luca Giovannini
Forking Paths

Luca Giovannini, born in 1995, works as research coordinator for the Network for Digital Humanities at the University of Potsdam. After studying comparative literature in Turin, Warwick, and Cologne, he got a joint PhD in the same field from the University of Potsdam and the University of Padova. His research mostly deals with the computational investigation of literary and cultural artefacts.

Luca Giovannini

Forking Paths

Evolution and Divergence in Early Modern European Drama

[transcript]

This publication was supported by funds from the Publication Fund for Open Access Monographs of the Federal State of Brandenburg, Germany, the Potsdam Graduate School, and the Network for Digital Humanities at the University of Potsdam.

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available online at <https://dnb.dnb.de>



This work is licensed under the Creative Commons License BY-SA 4.0. For the full license terms, please visit the URL <https://creativecommons.org/licenses/by-sa/4.0/>.

Creative Commons license terms for re-use do not apply to any content (such as graphs, figures, photos, excerpts, etc.) not original to the Open Access publication and further permission may be required from the rights holder. The obligation to research and clear permission lies solely with the party re-using the material.

2026 © Luca Giovannini

transcript Verlag | Hermannstraße 26 | D-33602 Bielefeld | live@transcript-verlag.de

Cover design: Maria Arndt

Cover illustration: Cäsar Fleischlen, Graphische Litteratur-Tafel (1890), detail.

Niedersächsische Staats- und Universitätsbibliothek Göttingen (8 HLU II, 4532). Public Domain.

Printing: Elanders Waiblingen GmbH, Waiblingen

<https://doi.org/10.14361/9783839441640>

Print-ISBN: 978-3-8376-7962-5 | PDF-ISBN: 978-3-8394-4164-0

ISSN of series: 2628-4995 | eISSN of series: 2703-0415

Printed on permanent acid-free text paper.

*Alla mia famiglia;
e, in particolare,
a mio fratello Davide (弟)*

Contents

Acknowledgements	9
Introduction	11
Theory	17
Diversity out of unity: Franco Moretti's evolutionist account	18
Similarity beyond diversity: cultural nets and theatergrams	20
Early modern poetics: A panorama	25
Italian drama	27
French drama	29
Spanish drama	32
German drama	33
English drama	37
Corpus	39
Corpus selection	41
Time frame	41
Scope	43
Extension	44
Corpus retrieval	46
Corpus building	49
Encoding pipelines	49
Editorial practice	52

Deployment	57
Methods	59
Operationalising drama	60
Stages in the development of computational drama analysis	63
Excursus: dramatic network analysis	66
Modelling EmDraCor plays	68
A typology of features	76
Experiments	83
Distances	84
Clusters	90
Patterns	93
English drama	97
German drama	101
Spanish drama	103
Italian drama	106
French drama	107
Coda: A reproduction experiment	109
Conclusion	117
Appendix A	125
Appendix B	131
Appendix C	141
Bibliography	151

Acknowledgements

This work is a substantially revised version of a doctoral dissertation titled “Evolutive Dynamics in Early Modern European Theatre: A Computational Approach”, submitted to the Faculty of Arts of the University of Potsdam and the Department of Linguistic and Literary Studies of the University of Padova in October 2024, and defended in March 2025. The original dissertation is accessible on the University of Potsdam’s publication server.¹

In this book, I have tried to address flaws and limitations of the original text while streamlining my argument and making it more accessible to a broader audience of computational and non-computational literary scholars. I will not reproduce the dissertation’s acknowledgements here, but I would like to thank again Peer Trilcke and Rocco Coronato not only for supervising my doctoral research, but also for providing valuable comments and suggestions for the preparation of this manuscript. Much of the computational work presented here would not have been possible without the help of members of the DraCor group, notably Daniil Skorinkin, Ingo Börner, and Carsten Milling. As always, I remain deeply grateful to my family and friends for their support throughout these years.

Both my doctoral studies and this publication have been supported by the Federal State of Brandenburg, Germany. Thanks are also due to Luisa Bott and the team at *transcript* for agreeing to publish this book and following the publication process throughout.

1 <https://doi.org/10.25932/publishup-69516>.

All errors and omissions remain of course my responsibility. All quotations from primary and secondary sources in languages other than English, unless otherwise specified, were translated by me.

*

Many things have changed since the day of my thesis defence, both personally and professionally. To some extent, that day represented a false start, or a 'false movement', to borrow the expression a disillusioned critic used to describe the latest turn in digital humanities.² And yet, in both cases, I would argue that something has eventually moved forward. Theories may indeed be grey, as an early modern devil once remarked, but the golden tree of life keeps growing.

2 See Moretti (2022).

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>.

Theory

Formulating overarching theories about any *longue durée* phenomena is a thankless, often unfruitful task and, given the sheer scope of the European dramatic tradition during the early modern era, any attempt in this sense might fall short of actually offering a coherent narrative. A notable exception, nonetheless, was represented by ‘Modern European Literature: A Geographical Sketch’ – a short, but significant essay which argued for a progressive differentiation process within early modern drama out of an earlier formal unity, and analysed its implications.

After recounting Moretti’s narrative, this chapter brings up a different perspective by introducing some competing theories, informed by concepts such as the ‘cultural net’ (Küpper, 2018) or the ‘theatergrams’ (Clubb, 1990), which stress instead the constant flux of interchanges between European dramatic spaces. Ultimately, a succinct summary of the evolution of European early modern drama within the five different linguistic areas under investigation is proposed, with particular attention to issues of poetics and playwriting which contributed to shaping the texts’ formal features.

Diversity out of unity: Franco Moretti's evolutionist account

As it often happened with many of Moretti's early essays, 'Modern European Literature'¹ "[took] a branch of empirical social or natural science as a freestanding model for a new literary history" (Eyers, 2017, 38). More specifically, he began there to blend three approaches which would remain central in his later career, i.e. evolutionary theory, geography, and formalism (cf. Moretti, 2013a, 2). This can be clearly seen in works such as *Atlas of the European novel, 1800–1900* (1998) or *Graphs, Maps, Trees: Abstract Models for a Literary History* (2005).

The conceptual framework for this early essay was provided by Ernst Mayr's *Systematics and the Origin of Species* (1942), a book which introduced the idea of allopatric speciation – the genesis of new biological species through their movement into new spaces. As Moretti himself later recounted, the argumentative device he developed out of that idea was relatively straightforward: he "took forms as the literary analogue of species, and charted [several] morphological transformations triggered by European geography" such as "the differentiation of tragedy in the seventeenth century, the novel's take-off in the eighteenth, the centralization and then fragmentation of the literary field in the nineteenth and twentieth" (Moretti, 2013a, 1).

Moretti's account of European literary evolution begins indeed with the genesis of the baroque tragedy, which he describes as a movement away from a previously unified formal model and towards a multitude of competing dramatic traditions. In the sixteenth century, he argues, figures like Scottish humanist George Buchanan – wandering across the Continent, writing acclaimed Latin tragedies² – embodied "the lasting unity-across time and across space of European drama". From London

1 Originally written in Italian for the first volume of Einaudi's *Storia d'Europa* (1993), the essay appeared in English in the *New Left Review* (208, 1994) and was eventually reprinted in *Distant Reading* (2013), with a new preface by the author which sheds some light on its inception. Unless otherwise specified, I hereafter quote from this latter version (= Moretti, 2013a).

2 See Cardinali (2009).

to Naples, “cultivated tragedy” was “almost always [modelled on] Seneca, while medieval traditions, rooted in popular religion, tend[ed] to be very similar everywhere” (9). Even on a thematic level, elements like the tragic hero (i.e. the absolute sovereign at the centre of his court) and his downfall were repeated in endless variations.

Then, during the early modern era, something happens:

[...] within a couple of generations, the stable, common features of European drama are replaced by a rapid succession of major formal mutations. By the mid seventeenth century, the tragedy of western Europe has branched out in three or four separate versions, where everything has changed: the relationship between word and action, the number of characters, stylistic register, temporal span, plot conventions, spatial movements, verse forms. In fact, not even the name of ‘tragedy’ is shared any longer. (11)

This passage effectively sums up most of Moretti’s argument and is worth a closer look. On one side, the focus is on a specific genre (tragedy) and its alleged local mutations – even though the essay often switches to the more general term of ‘drama’, suggesting that the model it proposes could be extended to other theatrical genres as well. On the other side, the elements where this change is visible touch upon several aspects of drama, including performance (“spatial movements”), style (“stylistic register”, “verse forms”), and content (“plot conventions”). His perspective, however, is decidedly oriented towards form, as the continuation of his argument (12) shows.

Furthermore, while the topic is not explicitly addressed by Moretti, the ‘branching’³ (or speciation, in evolutionary terms) of European

-
- 3 “We should borrow a different metaphor from evolutionary theory, and think of literary development as a large bush: branches that coexist and bifurcate, that overlap and at times obstruct each other – but that, whenever one of them withers away, are ready to replace it with an ever thicker and stronger organism” (18). More than thirty years later, Moretti will still refer to “the age of Shakespeare and Racine” as “the great bifurcation in the history of tragic form” (Moretti, 2024, 8).

drama is linked, as one could expect, to the political fragmentation within the Continent: the progressive consolidation of (some) state entities during the early modern age creates ‘protected’ spaces where experiments in form, deviating from previous models, are allowed to blossom. Within the large European space, each local ‘mutation’ is able to occupy its own milieu of influence – “cultural ‘niches’ for mutations to take root, and later contribute [back] to literary evolution” in other regions (11).

According to this narrative, formal variety is not the product of history, but rather of geography: “form needs time in order to reproduce itself; but in order to arise it is space that it needs most” (13). The divided European space, “where the exploration of formal possibilities may be allowed, and in fact encouraged as a sort of patriotic duty” (13), offers such a playground for dramatic imagination, and can thus be described as a “discontinuous, fractured, [...] archipelago of (national) sub-spaces, each of them specializing in one formal variation” (12) – the optimal breeding ground for the diverse dramatic forms of early modernity.

Similarity beyond diversity: cultural nets and theatergrams

This account of European drama history seems, however, to downplay a significant aspect, i.e. the possible transnational exchanges between literatures. According to Moretti’s perspective, the process of ‘branching’ seems to imply a progressive compartmentalisation of the various dramatic traditions: even if contaminations with other dramatic spaces are still possible, they tend to build stable models where the same formal features are repeated over and over until they are exhausted (13). In recalling the work of one of the first systematic historians of literature, Henry Hallam, and his *Introduction to the Literature of Europe in the Fifteenth, Sixteenth and Seventeenth Centuries* (1837), Moretti underscores how, within the European space, “spatial proximity never turns into functional interaction” (14) between dramatic cultures. Only later will French-led neoclassicism present itself as a meaningful supranational phenomenon, but in a homogenising fashion (15–16).

While the idea that the early modern age “marks the birth of original national aesthetics” in theatre (Lochert, 2009, 28) is fairly uncontroversial among scholars, emphasising the divergence between dramatic spaces is not, however, the only possible approach. As Levi (1988) points out, the problem of the plurality/unity of early modern culture can indeed be tackled from multiple angles, from social to symbolic, and actually goes beyond literature itself. Moretti’s model, however powerful, does indeed raise the question of how one could account for the possible cross-pollination between dramatic traditions during this same period.

In this vein, calls have been made to at least problematise the category of national dramatic literatures and adopt instead a more holistic perspective. A notable example of such a stance is Joachim Küpper’s proposal to understand early modern drama as a ‘cultural net’: according to this theory, the “[p]arameters informing a given artifact (a “work”), such as plot, genre, ideological frames (explicit or implied), message, style, etc., are [...] items that are conceptually separate from, and in most cases indeed independent of specific vernaculars” (Küpper, 2018, 105), and circulate freely within a common, continent-wide network structure. From time to time, such elements are extracted from the ‘information flow’, refashioned, and employed in individual works, thus accounting for transnational similarities.⁴

Although Küpper does not explicitly engage with Moretti, his approach seems to implicitly challenge the evolutionary model proposed by the Italian scholar. As an example, his constant warning about “the nonproblematized (over)interpretation of the bonds that literary texts of any kind by necessity maintain to the extra-literary ‘reality’ from

4 Evident common traits could be explained as originating from the participation in one common network structure [...]. Transculturality would thus become the standard case, “nationality” of cultural artefacts would be considered the particular case to be explained according to the control logics outlined above (mainly: power- or belief-driven attempts to restrict the floating material available at a given place at a given moment)” (Küpper, 2018, 21).

which they originate" (252) could be read as a jab at Moretti's belief that geography represents a key component of drama history.⁵

Furthermore, the very concept of a consistent cultural evolution is treated with some scepticism, even in its non-teleological variants. In the cultural net, Küpper argues, "most artistic development is caused by the random, or at least nonsystematizable combination of pre-existent material", due to "external needs and demands, or conscious processes of active transculturation" (299), and thus narratives of linear evolution are not sustainable.⁶

From the outset, Küpper's theory of the cultural net appears indebted to the relational framework developed by Pascale Casanova in her *World Republic of Letters* (Casanova, 1999). Although not always in agreement with her approach (see e.g. 46 ff.), he nevertheless shares her core premise that "specific features of [...] cultural elements [...] can be explained on the basis of their position" (Erlin et al., 2021, 83) within a complex socio-cultural environment, be it the 'world literary space' or the 'cultural net'.

Furthermore, his attempt to map early modern drama according to an intercultural perspective is not unprecedented, and rather builds on earlier conceptualisations of transnational dramatic exchange. Most notably, the 'parameters' Küpper envisions as floating within the cultural net could be seen as more abstract versions of the modular dramatic units, or 'theatergrams', first singled out by Louise George Clubb.⁷ This

5 While Küpper indeed accepts that language and topography *can* act as 'bonds' in that sense, he also proceeds to deconstruct them through a short contrastive analysis of Flaubert's *Madame Bovary* and Fontane's *Effi Briest* (see 252–54).

6 This view, however, would be consistent with one of Moretti's most famous essays, 'The Slaughterhouse of Literature' (*Modern Language Quarterly* 61, 1, 2000, 207–227), where the (in)voluntary exploitation of a formal device (clues) is considered the key to the success of Conan Doyle's detective stories against contemporary competitors.

7 "Constant as a principle from the time of Ariosto on was construction by contamination, the meditated and usually explicit combination of pre-texts. But in addition to the mere fusion of borrowed plots, this demanded the interchange and transformation of units, figures, relationships, actions, topoi, and framing

concept, indicating “elements of dramatic composition such as character, situation, genre, and scene out of which narrative arcs and climaxes, catastrophes and dénouements of plays are assembled” (West, 2010, x), remains influential in contemporary theatre research.

Examples of its lasting productiveness can be found in the work of the *Theater without Borders* research collective (see e.g. Henke and Nicholson, 2008, 2014b, Katritzky and Drábek, 2019). Devoted to investigating the “transnational and comparative impact of Early Modern Drama”⁸, this group has often been more interested in performance-related issues, but they still retain an understanding of cultural dynamics akin to Küpper’s: as Henke and Nicholson (2014a, 6) postulate, “[t]he transnational dissemination of European theatre resembled an amorphous net or web much more than a line”.⁹

Globally, their work is useful insofar as it maps the main avenues of influence within the European dramatic space, going beyond individual translations, reprises, or imitations. Such contributions help indeed to understand the fundamental asymmetry of intercultural relations, which might be characterised by direct influence (e.g. Italian on English theatre, or English on German) or non-influence (English on early modern Italian, Spanish, and French theatre) (Henke and Nicholson, 2014a, 5).

Such an account of flow and circulation, nonetheless, requires a structuring framework. Patterns of exchange between dramatic traditions have been traditionally understood by scholars within the context of a pan-European ‘rift’, which Clubb describes as follows:

Most polarizations divide European theater into regular, ‘Aristotelian’, theoretical Italian and French on one hand and irregular, pragmatic

patterns, gradually building a combinatory of theatergrams that were at once streamlined structures for svelte play making and elements of high specific density, weighty with significance from previous incarnations” (Clubb, 1990, 6).

8 <https://theaterwithoutborders.org>.

9 It should be noted that, while referencing Küpper on this point, they also raise the possibility of combining theatergrams and distant reading by name-checking Moretti (1–2).

Spanish and English on the other, whether the antipodal concepts are labeled classical and romantic, or have been further dehistoricized as learned and spontaneous, or artificial and natural, or evolved politically into conservative and progressive or elite and popular. (Clubb, 1990, 3)

While Clubb herself views with suspicion such dichotomies, especially when cast as the stale opposition between a ‘classical’ and a ‘romantic’ mode, this interpretative scheme still holds some value in sketching a general picture of the field. Tracing its roots, Véronique Lochert shows how early modern theatrical aesthetics across Europe can actually be defined in their relationship with the two main preceding models, i.e. the performance-centred medieval theatre and the reading-centred humanist theatre:

In France and Italy, the medieval tradition was rejected by theorists and modern dramaturgy was developed as an extension of humanist theatre through imitation of ancient theatre. In Spain and England, on the other hand, the theoretical reflections of a few readers of Aristotle and the erudite imitations of humanist playwrights remained marginal in the face of the development of new theatrical practices, which were part of the continuity of medieval theatre. (Lochert, 2009, 93)

Again, this account should not imply an overly rigid partition, nor rule out any instance of internal strife within individual literary cultures,¹⁰ but remains useful in summarising a series of oppositions¹¹ between two

10 Some examples, offered by Clubb and Lochert (*passim*), are the French debate between *réguliers* and *irréguliers* (starting with the *querelle* about Corneille's *Cid*), the contrast in England between Elizabethan (Shakespeare) and Jacobean dramatists (Jonson), the ‘War of the Theatres’ between Jonson and Dekker/Marston, the Italian contraposition between *commedia erudita* and *commedia dell'arte*, the more general gap between baroque and classicist theatre, etc.

11 Lochert again: “The tension between these two poles covers a number of other divides: between text and performance, within the dramatic work itself; between reading and performance, scholarly public and popular public, at the

main modes of drama, centred either on the performance or on the text, which were actually attested in literary history.

To sum up, the two frameworks introduced – which one might term, with some approximation, *differentiation-based* and *circulation-based* – offer contrasting readings of the history of early modern European drama. The former, exemplified by Moretti's evolutionist account, emphasises the progressive branching of dramatic forms into distinct national traditions, shaped by geography and political fragmentation. The latter, informed by concepts such as Küpper's cultural net and Clubb's theatergrams, foregrounds instead the continuous movement of dramatic elements across linguistic boundaries, and the resulting structural similarities between otherwise distinct traditions. While the first captures an evident increase in formal divergence visible by the end of the seventeenth century, the second accounts for the persistence of links that complicate any narrative of simple compartmentalisation. The tension between these perspectives must then always be taken into account for a global understanding of the landscape of early modern theatre.

Early modern poetics: A panorama

This last section presents a concise summary of the evolution of dramatic literature within five key European linguistic areas (the Italian-, French-, Spanish-, German-, and English-speaking territories); the reasons for their choice will be outlined at the beginning of the next chapter. This purposely minimalistic account of drama history is clearly not meant to offer a comprehensive overview of authors and literary movements,

level of its reception; between absolute rules and historical relativity, in the context of the quarrel between the Ancients and the Moderns; between theory and practice, with France and Italy producing far more treatises than Spain and England, where theoretical reflection is generally integrated into the works themselves" (93). The notable absence of German-language drama from this narration, while justified by the focus of Lochert's book, is also telling of the difficulty in placing such dramatic tradition inside this binary scheme.

but rather to highlight key aspects in the development of dramatic traditions which may help contextualise Moretti's depiction of European drama as an archipelago of formally distinct literary experiences. As discussed extensively in later chapters, my approach here is decidedly text-centric, and thus I do not explore here the dimensions of staging and acting, which would be otherwise crucial for a global understanding of early modern theatrical texts. Here, I prefer to offer a necessarily limited overview of questions of poetics, citing relevant theoretical texts and critics.

This focus seems warranted by the fact that, during the period under investigation, theatrical practices were increasingly the object of a theoretical reflection which often led to the establishment of normative aesthetics. To employ Jurij Lotman's notable categories, one could indeed read the evolution of early modern drama as a transition from a 'work- or 'example-based' landscape (i.e. largely devoid of theorisation, such as the medieval one) to a 'rules-based' or 'grammaticalised' one (Allegri, 2000, 1183–84). Within this framework, it is therefore imperative to take into account the poetological discourse which shaped the playwrights' activity and eventually peaked in seventeenth-century Classicism – an exemplary instance of 'rules-based' culture according to Lotman and Uspensky (1978, 218–19).

While Lotman's ideas are mostly compatible with Moretti's general theory on dramatic evolution, there is an important distinction to be noted. While Moretti sketches a monodirectional evolution (*ex uno plures*), Lotman envisions rather a cyclic process, where the two cultures "can be considered stages of a single evolution that takes place as a pendulum alternation of different constitutive principles", with none of them establishing itself once and for all (Lotman, 1985, 81).¹²

-
- 12 Lotman describes such a cycle in these terms: "When an original civilization constituted as a system of customs ossifies itself to the point that its redundancy assumes catastrophic dimensions, the need arises for self-codification, which takes the form of the adoption of a grammar of culture. At this stage, grammaticality intervenes as a revolutionary principle and leads to the abrupt complication of the internal structure of the code of culture. Even rules, however, tend to ossify: their redundancy increases, and the ability to effectively

At the same time, however, both authors agree that geographical space is a decisive factor to be taken into account: while Moretti sees it as the actual engine of literary evolution, Lotman underlines that the two types of culture he describes could also be considered not only stages of a process, but also “stable national or area features”. Thus, cultural material moving from a ‘work-based’ to a ‘rules-based’ environment could from time to time be subjected to a “functional reinterpretation” to adapt to the new context (Lotman, 1985, 80–81).

Actually, this interpretative framework fits neatly with the traditional understanding of European theatrical cultures with regard to their poetics: to reprise Clubb's terminology, ‘regular’ Italian and French theatre could be seen as an expression of a ‘grammaticalised’ culture, while ‘irregular’ Spanish and English ones seem to still belong to an ‘example-based’ one – with German theatre in a somewhat intermediate position.

Such a systematisation, however, should not be assumed as conclusive, since in all areas the dominant aesthetics – be they ‘prescriptive’ or ‘unmediated’ – were not completely uncontested. The following summaries trace these dynamics within each linguistic area, attending particularly to how normative poetics emerged, were contested, or remained marginal – and to the formal consequences of these different trajectories.

Italian drama

According to Küpper (2018, 29), the first instance of a true ‘post-medieval drama’ emerged in the Italian territories. From there, its modules were gradually exported beyond the Alps to the other major dramatic cultures.

own and retain information begins to expire. Although the problem can be solved within this or that cultural cycle by changing grammars, this nevertheless discredits the very principle of grammaticality. At this stage, the irruption of the textual principle into the construction of culture decisively increases its capacity for informational content” (Lotman, 1985, 81). – Since no English translation of the original Russian essay (Lotman, 1971) is extant, I translate here from an Italian edition.

This could be said specifically in terms of literary norms, since the Italians were the first protagonists in the rediscovery of Aristotle's *Poetics* and the dissemination of its principles – *vraisemblance*, morality, universality, genre differentiation, the three units of space, time, and action, etc. – through several commentaries,¹³ which also incorporated prescriptions from Horace's *Art of Poetry* and later treatises.

Despite this regulatory framework, pastoral drama – a 'third', conventions-breaking genre – soon became the dominant genre of the late sixteenth century theatre, marking the passage from Renaissance to Baroque (Angelini, 2000, 255); notable examples were Torquato Tasso's *Aminta* (1573) and Giovan Battista Guarini's *Pastor Fido* (1580). Criticism from purists, who saw it as a perverse cross-breed of the established Aristotelian genres,¹⁴ did not prevent the pastoral – and, more broadly, tragicomedy – from assuming a central role by the end of the century, which was recognised and defended in contemporary theorization by Guarini himself (*Compendio della poesia tragicomica*, 1601) and Angelo Ingegneri (*Della poesia rappresentativa e del modo di rappresentare le favole sceniche*, 1598).

The fortune of the pastoral continued in the following century, even though it soon found a competitor in another 'mixed' genre, the melodrama, whose first example has been traditionally recognised either in

13 Influential commentaries included those by Francesco Robortello (1548), Vincenzo Maggi (1550), Giulio Cesare Scaligero (1561), Antonio Minturno (1563), and Lodovico Castelvetro (1570). On this topic see, among others, Conte (2003) and Weinberg (1961).

14 The most famous expression of this conflict is often acknowledged in the querelle between University of Padua lecturer Jason Denores and Guarini himself. A devout Aristotelian, Denores criticised tragicomedy and pastoral as 'monstrous' and 'disproportionate' hybrids, trying to stitch together comic and tragic without success (*Discorso intorno a que' principii...*, 1586). Guarini perceived it as an attack on his *Pastor Fido* and sharply replied with his *Verato* (1588). Further exchanges in 1590 and 1593 were eventually cut short by Denores' death. For more details, see Artico (2020) and Perella (1973).

Jacopo Peri's and Ottavio Rinuccini's *Euridice* (1601) or in Emilio de' Cavalieri's and Agostino Manni's *Rappresentazione di anima et di corpo* (1600).¹⁵

Even writers working within the neo-Aristotelian framework did not shy away, to some extent, from blending different genres, as evident in the production of two leading writers of the time, Giambattista Andreini and Giacinto Andrea Cicognini (Angelini, 2000, 245–246). More generally, comedy did develop on a double track, reviving classical Platine and Terentian models (as in Guarini's *Idropica*, 1584) but also embedding innovations from the *comici dell'arte*, whose themes and topics, as outlined in their *canovacci* (content outlines), were summarised in Flaminio Scala's *Il Teatro delle favole rappresentative* (1611). As usual in Italian early modern theatre, distinct regional cultures contributed to the development of different local variations of comedy (e.g. the *commedia ridicolosa* in Rome).

High tragedy, for its part, still enjoyed some success as “the preferred form of expression” (Angelini, 2000, 228) of the Baroque zeitgeist, with authors like Federico della Valle, Carlo de' Dottori, Giovanni Delfino, in addition to those writing classical tragedies (Carletti, Alberi, Bonarelli) or operating in the milieu of religious colleges (Santi, Scammacca, Tuccio, Pallavicino, Stefonio).¹⁶ In their works, the general adherence to the prescriptions of Aristotelianism was often bent to ethical and ideological needs, i.e. to the necessity of offering moral teaching in the context of the Counter-Reformation (Angelini, 2000, 242).

French drama

Like the Italian one, French theatre in the second half of the seventeenth century was heavily influenced by classicist models, first upheld by the Pléiade movement.¹⁷ Writers belonging to this group, such as Étienne Jodelle and Jean-Antoine de Baïf, and later like-minded authors (Jacques Grévin, Jean de la Taille, Robert Garnier, Pierre Larrivey, etc.) set out to

15 On the birth of melodrama see among others the considerations by Leopold (2003).

16 Cf. Russo (1959), Trivero (2015).

17 See Faisant (1998).

develop a classicist theatre, often taking the Italian *commedia erudita* as a model (Brockett and Hildy, 2013, 179). De la Taille himself authored an essay (*De l'art de la tragédie*, 1572) in which, despite not referring directly back to Aristotle, he popularised some of his prescriptions within French theatre, through the usual mediation of Horace and the Renaissance theorists (Refini, 2009).

Political upheavals marked the development of the French dramatic tradition, with the production slowing down at the time of traumatic events such as the religious wars (1572–1598). At the turn of the century, tragedy started again to develop under Antoine de Montchrestien and Alexandre Hardy, but the dominant genre remained the farce influenced by the *commedia dell'arte*. Only with the renewed political stability of the 1620s and 1630s did further developments occur, including the rise of notable figures like Jean de Mairet, Pierre du Ryer, Jean de Rotrou, and Pierre Corneille. These authors were instrumental in consolidating classical aesthetics in drama, “link[ing] back to the humanist theatrical tradition (Jodelle, Garnier), mov[ing] away from Seneca’s tragedies to get closer to Sophoclean and Euripidean models, and begin[n]g to read, translate and discuss Aristotle’s *Poetics*” (Biet, 2016, 303).

This process concerned first and foremost the tragedy. As argued by Georges Forestier, it took authors just a few years to move from shunning the whole genre as too bound to classical models to keenly accept the need for rules to steer dramatic creativity:

Classical tragedy had been rejected by moderns because it symbolised slavish obedience to the rules of the ancients; the modern introduction of ancient rules presented as necessary in the name of reason and the effect produced on the spectator led to the reintroduction of tragedy. [...] All reasoning was overturned: while classical rules had been contested because the rigour of the *dispositio* seemed to restrict the richness of the *inventio*, now, conversely, the modern rules were said to only hinder the ‘sterile genies’ incapable of finding ‘matter that was rich enough’ to produce ‘beautiful effects’ within the framework of a regular *dispositio*; while classical tragedy had died because it was an unreasoned imitation of the tragedy of the ancients, now modern

tragedy appeared, beyond any principle of imitation, to be the most suitable genre for expressing the new dramaturgy resulting from the reasoned acceptance of the rules. (Forestier, 1993, 127–128).

Another important stage in this process was represented by the so-called *querelle du Cid*, which raged around Corneille's eponymous piece and its transgression of some Aristotelian norms. The final deliberation of the Académie Française¹⁸ was instrumental in defining rules to be applied across the stages, but did not end the debate, which was carried on by Corneille himself in the prefaces to his collected works, and by his opponent, the Abbé d'Aubignac. The latter's *Pratique du théâtre* (1657), written under the auspices of Cardinal Richelieu, also established itself as a key reference for playwrights, summing up a series of recommendations for writing and staging dramatic pieces,¹⁹ a similar role, though limited to textual aspects, was later played by Nicolas Boileau's didactic poem *L'Art poétique* (1674).

In the heyday of French classical theatre, the different genres found also specific production niches, with tragedy shifting back to the “declamation of a simplified poetry” and assuming the form of a text-centric “theatrical poem” – as evident in the works of Scarron first and then, after another period of relative stagnation corresponding to the Fronde crises, in those of Thomas Corneille and Jean Racine. On the opposite side of the dramatic spectrum, mixed forms came to play the “hyper-spectacular role” previously reserved to tragedy (Biet, 2016, 316).

Opera eventually spread from Italy into France, where early hybrid forms of theatre, music, and dance, such as the *comédie-ballet*, were joined by truly ‘operatic’ (i.e. completely sung) compositions such as the *tragédie lyrique*. Popularised among others by Jean-Baptiste Lully and François Quinault, such compositions soon became the standard form

18 Documents on the *querelle*, including the Académie's official response penned by Jean Chapelain (*Des sentiments de l'Académie française sur la tragi-comédie du Cid*, 1638), are collected with a rich introduction and commentary in Civardi (2004).

19 See the still useful account by Arnaud (1970).

of French opera, capable of rivalling spoken theatre (Norman, 2001, 7). At the same time, comedy achieved the same dignity of tragedy thanks to the efforts by Molière,²⁰ and flourished within the context of the now professionalised theatre companies, such as those of the Palais Royal, the Hôtel de Bourgogne, and the Marais – later fused in the world's first national theatre, the Comédie-Française.

Spanish drama

While Italian and French drama represent an instance of a transition toward 'rules-based' culture, with theoretical codification preceding and shaping dramatic production, Spanish theatre, by contrast, exemplifies the persistence of 'work-based' principles, where practice resisted (to some extent) theoretical prescription.

The early political situation in the Iberian Peninsula, shaped by the Reconquista, contributed to the prominence of religious drama, which served the State's agenda of reasserting Catholic cultural hegemony. While before 1500 such representations had shared features with other postmedieval European analogues, they then developed some idiosyncratic features which granted them a large fortune until 1765, when they were suppressed. The most successful genre, practised by Pedro Calderón de La Barca and others, was the auto sacramental, recovering themes and modes from the morality and cycle plays but also abiding by contemporary theatrical practice.

Cultural exchanges with Italy gradually led to importing humanist models, but Aristotelian norms never gained the same traction as in Italy and France, due to a limited reception of the *Poetics*, which was translated only a couple of times, and the presence of other composition models (Sánchez Laílla, 2000, 9). This could be seen in the tragedies of the time, such as those by Cristóbal de Virués, Juan de la Cueva, or Francisco Agustín Tárrega (Capra, 2000, 691 ff.), in the comedies and *entremeses*

20 On Molière's relationship with the Aristotelian models see in particular Calder (1993, 2–5).

(short comic interludes) by Miguel de Cervantes, but also in theoretical writings.²¹

Perhaps the most famous rejection of neo-Aristotelian tenets can be read in *El Arte nuevo de hacer comedias en este tiempo* (1609), the artistic manifesto of the age's main playwright, Félix Lope de Vega Carpio. There, drawing some conclusions from his long and successful stage career, the author defended his unorthodox playwriting style by openly rejecting principles such as the separation between comic and tragic and the unities of time and space, and proposing instead a style tailored to the tastes of the public, without pedantism.

In the end, Lope set the example for the dominant form of Spanish drama, the *comedia nueva* – where *comedia*, out of habit and diffusion, had come to indicate a 'work to be staged' in a neutral sense, beyond its actual 'comic' content (Sullivan, 2018, 33; Newels, 1959, 17). Genre distinctions were not however completely blurred; rather, one witnessed a multiplication of specific labels for its different types (such as *comedia heroica*, *palaciega*, *de figurón*, etc.).²² Another distinction had to do with the author's operating milieu: while some wrote for public theatres or *corrales* (Juan de la Cueva and Miguel de Cervantes early on, and later Lope, Guillén de Castro, Tirso de Molina, Juan Ruiz de Alarcón), others were more active in court theatres (such as Francisco de Rojas Zorrilla, Agustín Moreto, and Calderón himself, who wrote autos and light dramas called *zarzuelas*).

German drama

On Lotman's spectrum, German theatre would occupy an intermediate position – open to theoretical frameworks imported from abroad, yet

21 See, e.g., the works by Alonso López Pinciano (*Philosophia antigua poética*, 1596), Luis Alfonso de Carvallo (*Cisne de Apolo*, 1602), Francisco Cascales (*Tablas poéticas*, 1617) and Jusepe Antonio González de Sala (*Nueva idea de la tragedia antigua*, 1632).

22 For this, see the useful scheme by García Berrio and Huerta Calvo (1992, 210), based on Vitse (1988).

sometimes resistant to strict codification. This ambivalence could, of course, be explained in social terms. As noted by Niefanger (2012, 231), the political and cultural milieu of the German-speaking lands, still split in a myriad of mostly autonomous small states within the nominal Empire, prevented the birth of a uniform theatre as happened in the early nation-states (cf. also Levi, 1988, 231–32). The lack of a national theatre – which should not be assumed to imply, however, a lack of idiosyncratic dramatic forms – translated thus into an openness to innovations from neighbouring countries, with a large influx of translations and adaptations dominating German scenes.

In terms of poetics, however, the influence of Aristotle was as decisive as in other countries, representing an inescapable theoretical reference to be either followed or rejected.²³ His influence was mediated through Renaissance poetics, such as Scaliger's fundamental *Poetices libri septem* (1561), and later contributions by Daniel Heinsius, Gerard Vossius, and Nicolas Boileau. While German theorisations of theatre were quite uncommon,²⁴ some dramatic principles were enunciated directly in the plays' prefaces or in broader treatises on poetics, like the seminal *Buch von der deutschen Poeterey* (1624) by Martin Opitz. Those remarks, however, were mostly reserved to tragedy, while comedy and the mixed genres, lacking a solid theoretical tradition, did not get much attention (Alexander, 1984, 49–50).²⁵

Across the wide spectrum of dramatic possibilities found in German early modern theatre, it is possible to identify some major strands: the religious drama (both Protestant and Catholic), the *Wanderbühne* movement, and the Silesian school of drama. In addition, the role of music theatre cannot be discounted completely: Italian opera was imported

23 Schings (1980, 21) considers “the deviation from Aristotle” as the “guiding heuristic principle” of German Baroque drama, where new ideas and concepts were introduced under the disguise of interpretation of established theories.

24 An example is Balthasar Crusius' short Neolatin treatise *De Dramatibus* (1609); see Alexander (1980).

25 As Bonfatti (2000, 838) notes, however, an attempt to introduce mixed genres into the Aristotelian binomy of tragedy and comedy was done by Jesuit Jacob Masen in his *Palaestra Eloquentiae Ligatae* (1654–1657).

at the beginning of the century into the courts of Bavaria, Saxony, and Austria, and its models and texts remained long dominant. By contrast, smaller courts in the Protestant North developed relatively soon a German-language opera, which found its main expression in the *galante Stil* centred in Hamburg.²⁶

Religious drama with didascalical purposes was produced on both sides of the Reformation, finding its epicentres first in Protestant colleges (such as the gymnasium of Strasbourg) and then in schools run by Catholic orders like the Jesuits, and could be actually considered “the backbone of German theatre in the XVI and XVII century” (Bonfatti, 2000, 806). Plays written in these contexts were usually in Latin, but German often surfaced in intermezzi and paratextual materials (such as the *periochae*, i.e. booklets containing information on the play’s staging and contents).

This dramatic tradition lacked prescriptive poetics, but developed manuals on staging practices, starting from the early comments by Jakob Willich and later culminating with Franz Lang’s *Dissertatio de actione scenica* (1727) (Bonfatti, 2000, 811; Niefanger, 2012, 238–39). Towards the end of the century, Protestant school drama gained again some prominence with authors like Sebastian Mitternacht, Johannes Riemer, and especially Christian Weise, whose eclectic work however failed, according to Bonfatti (2000, 864), to create a German-language comedy with original features; the same could be said, outside the school environment, of Christian Reuter.

An alternative to the school drama was represented by the *Wanderbühne* movement, made up of theatrical companies travelling across German lands and staging their plays in various venues (courts, markets, cities). Initially dominated by foreign actors, such as the English ones fleeing the Puritan shutdown of playhouses, these groups offered the public a theatrical experience where the non-verbal elements and body language prevailed over the actual text. Their international repertoire creatively remixed major coeval authors, from Calderón to Molière and

26 See the edition of relevant *galante* texts by Wiedemann (1969).

Shakespeare, and was later decisively shaped by *commedia dell'arte* modules imported through France, while touching upon all genres and often featuring stereotypical buffoon figures like Hanswurst or Pickelhäring.

An important role was also played, in the second half of the century, by the Silesian school of drama, whose representatives, such as Andreas Gryphius and Daniel Casper von Lohenstein, wrote both tragic (*Trauerspiel*) and comic works (*Lustspiel*). Especially the first type, a composite blend of classicist (Seneca) and allegorical influences informed by neo-Stoic ethics, was considered an *unicum* within the European horizon by Walter Benjamin, who also underscored its differences from the classical tragedy.²⁷

More recent scholarship, however, has somewhat tempered this claim, alleging that “the genre emerged into something unique [...] within a matrix that contained both the appropriation of alien forms and autochthonous developments”, i.e. “working with templates lifted from the broader European context” (Lande, 2016, 320 and 342) – with the relation between Gryphius and Dutch dramatist Joost van den Vondel being a case in point. In any case, both *Trauer-* and *Lustspiel* still deferred to the example set by Opitz in his double role as theorist (with his already cited *Buch von der deutschen Poeterey*) and translator of Seneca (*Troades*, 1625).

27 See Benjamin (1963, 45 ff.). The key differences Benjamin sees between *Trauerspiel* and *Tragödie* are thus summarised by Bloemendal and Smith (2016, 29): “[Benjamin] distinguishes *Trauerspiel* with its origins in history, its earth-bound stress on this world from the more transcendental *Tragödie*, which originates in myth, acts out a rite of sacrifice and centers on hamartia, a human mistake. Whereas *Tragödie* features a tragic hero and is characterized by reticence, the protagonists of *Trauerspiel* are tyrants and martyrs, often embodied by one and the same character, and its signature is ostentation. This *Trauerspiel* was rooted in Lutheranism and the Counter-Reformation. Behind all ‘mourning-play’ protagonists is Christ, the ultimate king and martyr”.

English drama

Like Spanish drama, English theatre largely resisted theoretical codification, though neoclassical ideals found occasional advocates. Examining the situation of English theatre around the mid sixteenth century, it appears that the main Aristotelian genres were still sufficiently distinct both in terms of locations and contents. On one side, comedies were mostly acted in the houses of nobles or in schools, and the need for entertainment was reflected in the persistence of medieval and farcical models, albeit tempered by a composition technique inspired by Latin 'regular' comedy (as exemplified, early on, by the production of Nicholas Udall).

On the other hand, tragedies were often conceived in the milieu of professional colleges, such as the Inns of Court in London, and took up a more experimental character, creatively reassembling a variety of source materials while emulating Seneca's tragic style. As the starting point of this latter wave, critics have traditionally cited Norton and Sackville's *Gorboduc* (1561), which also introduced iambic pentameter as the standard tragic verse.

The fusion of elements from classical and medieval theatre with diverse themes and topics from history, mythology, and folklore was furthered by the next generation of authors, the so-called University wits, and by William Shakespeare and his contemporaries. Nonetheless, such hybridism was not uncontested, with Philip Sidney criticising it in his *Defence of Poesy* (1583) and expounding instead neoclassicist ideals which were then taken up, to some extent, by Ben Jonson (Marenco, 2000, 314–15). Sidney's text was also embedded within the large controversy on the liceity of theatre, attacked by Puritans as immoral;²⁸ incidentally, it ought to be noted that the playwrights themselves rested their hopes for social advancement more in poems and translations

28 See, e.g., the pamphlets by John Northbrooke (*A Treatise against Dicing, Dancing, Plays and Interludes*, 1577) and Stephen Gosson (*The School of Abuse*, 1579), countered by Sidney's apology and by Lodge's *A Defence of Poetry, Music and Stage Plays* (1579). On the issue see Marcus (2017).

than in their drama, still suffering “the stigma of [being] a minor genre” (Marenco, 2000, 415–16).

The turn of the century – and the dynastic succession from Elizabeth I to James I – brought again some diversification of drama according to its staging occasion: while public theatres, already consecrated to a central role in English society, hosted the bourgeois drama of Thomas Dekker and Thomas Heywood, private ones gave authors like George Chapman and John Marston the occasion to critically reflect on society with their satirical plays (Marenco, 2000, 428). Further playwrights, such as John Fletcher, Francis Beaumont, Philip Massinger, John Webster, and John Ford, united polished dramatic skill with an already Baroque gusto for the pathetic and the sensational, especially visible in their tragicomedies (Brockett and Hildy, 2013, 115).

The English civil war (1642) and the Puritan revolution led to an 18-year-long closure of theatres, which did not, however, stop dramatic activity in full – musical ‘entertainments’ were still permitted and under this guise William Davenant was able to present his *Siege of Rhodes* (1656). With the Stuart Restoration, new acting companies (led by Davenant and by Thomas Killigrew) rose to fame while renovating their repertory to meet the mutated tastes.

Heroic tragedy modelled after continental authors (such as Corneille) was soon followed by ‘neoclassical’ plays which combined some superficial respect of Aristotelian norms with echoes of the Shakespearian heritage. Together with lyrical opera, comedy was also highly successful and was developed across a range of subgenres (comedy of humours, of intrigues, of manners, London and Spanish comedy, etc.), even though the end of the century with the Glorious Revolution saw a renewed Puritan-based attack on theatre.²⁹

29 I refer here to Jeremy Collier’s *Short View of the Immorality and Profaneness of the English Stage* (1698), which elicited responses from dramatists such as John Vanbrugh and William Congreve. On this see Cordner (2000).

Corpus

The issue of corpus is crucial for any investigation of problems of literary history which aims at moving beyond the canon-centred paradigm and deploying ‘distant’ (Moretti, 2013a) or ‘scalable’ (Mueller, 2020) readings. Accordingly, scholars in computational literary studies have devoted considerable energy to the identification and description of ‘good practices’ one should follow to assemble corpora for digital investigation (cf. Desagulier, 2017; Mrugalski et al., 2022; Schöch, 2017a). Key criteria to be fulfilled in literary corpus building, as outlined by Herrmann and Lauer (2018, 136 ff.), include representativeness, balance, and fidelity, but one should also reflect upon the theoretical framework of the research, the possible inclusion of translations, and the need to adhere to FAIR principles (Wilkinson et al., 2016).

Furthermore, researchers need to consider another factor, which plays an even more decisive role when working on early modern materials: text availability. Mechanisms of cultural transmission are notably complex, and it is often difficult to reliably estimate how many works – in my case, how many plays from the early modern age – have survived the ‘slaughterhouse of literature’ (Moretti, 2000), and how many have been subsequently digitised, encoded, and annotated to be fully machine-actionable.

An example might help illustrate this point. In 1755, a group of Venetian intellectuals, coordinated by librettist Apostolo Zeno, completed the revision of the *Drammaturgia*, an authoritative catalogue of Italian drama first compiled two centuries earlier by a Roman erudite, Leone Allacci. Thanks to the recent conversion of the *Drammaturgia* into a

database (Giovannini and Gallucci, 2024; see also Gallucci and Giovannini, 2025),¹ it is now possible to use this resource to roughly gauge the extension of early modern Italian theatre.

The catalogue lists about 6,000 plays from 1449 to 1751, but out of its more than 2,000 authors, fewer than one third have a corresponding item in *Wikidata*, suggesting that most writers and titles are likely ignored by modern scholarship. While some texts might be at least digitised in large digital archives such as *Google Books* or *Internet Archive*, only a tiny fraction of them is encoded in TEI-XML and available in the standard digital library for Italian literature, *Biblioteca Italiana*,² which was used as the source for the Italian Drama Corpus (ItaDraCor).

Given this often-unsatisfying data landscape – partially shared by all early modern European literatures, though in highly varying proportions – assembling a corpus suitable for exploring the evolution of early modern drama presented multiple challenges. On a general level, the corpus building strategy was tailored to Herrmann and Lauer's (2018) advice (based on Mueller, 2020), according to whom the corpus building process needs to be oriented first and foremost by the research question one wants to address: “the size of the corpus on which a study is based and the degree of abstraction, the research design and the research methodology can only be determined on the basis of the respective research question” (138).

Therefore, the following pages report on all steps leading to the creation of the so-called *Early Modern Drama Corpus* (EmDraCor), including the procedures for text retrieval and the pipelines for corpus building and deployment. I will start, however, by stating some general principles behind the selection of the actual texts, pertaining to their chronological and linguistic collocation and to their coverage.

1 The *Allacci Digitale* database can be accessed at <https://allacci-digitale.github.io/index-en.html>.

2 The search engine (<http://www.bibliotecaitaliana.it/catalogo>) returns as few as 162 plays between the fifteenth and eighteenth century.

Corpus selection

Time frame

As famously pointed out by Lotman, “[i]t is virtually impossible to divide periods according to dates”, since “human culture is a dynamic system”, and “[a]ttempts to locate stages of cultural development within strict temporal boundaries contradict that dynamism” (Bassnett, 2002, 48–49). As a consequence, literary periodisation can be indeed seen as “the product of a process of generalization and synthesis, within a historical and social horizon, of the small-scale critical and interpretive practices that characterize the study of literary text” (Ciotti, 2022), whose validity is however confined to the hermeneutic framework one assumes.³

In his essay, Moretti spoke explicitly of ‘Baroque’ tragedy, employing a somewhat vague label which originated in art history before finding its way into music studies and eventually literary criticism.⁴ As it often happens with categories borrowed from other disciplines, the notion of Baroque is far from uncontested, experiencing relevant semantic shifts in different countries, and appears unsuitable as a catch-all term for literary movements as different as the French *classicisme*, the Spanish *Siglo de Oro*, or the English theatre from its Elizabethan inception to its later offshoots.

Therefore, I prefer in this work to employ the broader notion of ‘early modern drama’. Within the usual time span it is often associated with (roughly from mid sixteenth century to the end of the seventeenth), the choice of the year 1561 as the corpus’ start date was admittedly informed

3 On this topic see also Wellek and Warren (1949, ch. 19) and Meneghelli (2013, 3).

4 The history of the concept of ‘Baroque’ – first appearing in the writings of Jakob Burckhardt and later popularised by Heinrich Wölfflin in his *Renaissance und Barock* (1888) – cannot be fully outlined here. For its specific significance in literary criticism, see at least Hatzfeld (1955), Wellek (1946).

by symbolic reasons. Indeed, during the Christmas and New Year festivities of 1561–62, the Inner Temple in London hosted the inaugural staging of *The Tragedie of Gorboduc*, a collaboration by Thomas Norton and Thomas Sackville, which is often considered both the first printed English tragedy and the first tragedy in blank verse (Winston, 2019, 185).

Inaugurating the ‘golden age’ of the Elizabethan theatre, this play could be assumed as exemplary for the broader, Europe-wide movement towards autonomous dramatic forms with some degree of independence from classical models. In this sense, *Gorboduc* assumes a paradigmatic role, explicitly acknowledged in Moretti’s first version of the essay (Moretti, 1993, §3), and its staging year seems therefore fit as the lower chronological boundary for this investigation.

The choice of the cut-off date for the corpus proved challenging as well. On one side, Moretti’s essay does not stop at early modern drama, and after examining the ‘polycentrism’ of Baroque tragedy it moves forward to investigate the birth of the ‘Republic of Letters’ concept and the spread of French literary hegemony and neoclassicism. On the other side, the ‘branching’ process is described as quite swift, but no precise time frame for it is established:

The most significant transformations do not occur because a form has a lot of time at its disposal: but because at the right moment – which is as a rule very short – it has a lot of space. [...] [Form] achieve[s] rather quickly a stable structure, which remains unchanged for decades, until it becomes sterile and disappears”. (Moretti, 2013a, 13)

To shed clarity on this, a *longue durée* analysis of Moretti’s claims would require including texts covering not only the decades where the branching happened, but also the following ones, where national form consolidated. Accordingly, a 150-year arc (i.e. from 1561 until 1710), was considered long enough to capture effective variations in dramatic form while remaining within the boundaries of the common definition of ‘early modern’ age. Furthermore, the endpoint signals, if not the end of the differentiation process, at least a first relevant caesura, marked

by such a major shift as the triumph of French neoclassicism all across Europe.

Scope

While Moretti's narrative of the evolution of European early modern drama moves seamlessly between authors and works from different cultures, supporting his argument with examples from Calderón, Racine, and Shakespeare, his focus always remains on the "three great western nation states" (i.e. Spain, France, and England) and "the German and Italian territories" (11).

Now, reducing the whole of European literature to some main 'centres of prestige' is not uncommon in comparative studies, but nevertheless problematic. At least in the case of early modern written culture, such simplification might find some justification in raw statistics such as the size of book markets. According to data provided by Buringh and Van Zanden (2009, 419), for example, between 85% and 93% of the books printed between 1550 and 1700 stemmed from one of the five linguistic areas explicitly mentioned by Moretti.⁵ While partial and not specifically drama-oriented, such data offer a rough proxy for cultural infrastructure and readership, and thus support focusing on such a selection of linguistic areas.

From another perspective, however, such choice might still be questionable if one aims at constructing a comprehensive narrative of European literary history. For example, it would be fair to object to the exclusion of the Netherlands, both from a quantitative – book production from 1600 onwards there begins to dwarf the Spanish one,⁶ – and qualitative perspective (ignoring the likes of Vondel, Hooft, Vos, or the *Nil Volentibus Arduum* society).⁷ The contributions to early modern drama made

5 Buringh and Van Zanden's figures are derived from the *Global Historical Bibliometrics* database (<https://www.iisg.nl/bibliometrics>).

6 On the momentous growth of the Dutch printing industry in the seventeenth century see Hoftijzer (2001).

7 One could also argue that excluding Dutch drama from a comparative analysis of European drama may bias results toward Moretti's differentiation thesis, as

by other ‘peripheral’ language spaces, such as the Portuguese or the Polish one, seem less prominent, but could have also helped in getting a clearer picture of the whole.

Nonetheless, pragmatic considerations and stricter adherence to Moretti’s framework, again following Herrmann’s and Lauer’s advice, suggested limiting this investigation to texts from the Spanish, German, English, Italian, and French traditions. At the same time, extending the investigation to further literatures remains a desideratum and could provide more layers of depth to any general argument about early modern drama.

Extension

Having defined the time frame and the linguistic scope of the corpus, the last step required determining its size. Such an operation translates into a balancing act between ‘representativeness’ and ‘practicality’ (Reppen, 2022, 32), i.e. between the desire to comprehensively cover a phenomenon and time- and resource-related constraints. These last issues are particularly relevant outside of the context of larger projects, such as EMOTHE – *Classics of Early Modern European Theatre* (Oleza and Tronch, 2020) or DraCor itself, which rely on distributed teams which collaboratively create and manage their corpora.⁸ Furthermore, comparability of collections (Schöch et al., 2021, 15) has to be taken into account, since the five subcorpora I aimed to build needed to be homogeneous enough to allow productive cross-readings.

As anticipated, the question of availability of early modern works in digital formats plays a decisive role in defining the corpus’ size. Despite massive advances in digitisation, many minor early modern texts are still

Dutch theatre is notable for its heavy international influences (e.g. from the Spanish *comedia nueva*, see Vergeer, 2026). On the role of the Netherlands as an exchange hub for dramatic models see also Bloemendal and Smith (2016, 17–21).

8 On the concept of ‘distributed corpus building’ in DraCor see Giovannini et al. (2023).

not available either in a standard scholarly format for critical editions, such as TEI-XML, or in simpler formats such as HTML or TXT. Furthermore, their distribution displays a noticeable variety across literatures, following obvious patterns of cultural prestige: for instance, properly encoded seventeenth century plays in Italian, beyond a handful of canonical authors, are rare, while there is an overabundance of digital resources on Shakespeare (Estill, 2019).

Given all these elements, multiple strategies for corpus building were taken into consideration; following the advice by Biber (1993, 243), I first focused on “a thorough definition of the target population and of the method of sampling” itself, before deciding the sample size. A first approach, for example, could have involved collecting wide lists of early modern plays from old and new bibliographic resources⁹ and then performing chronologically stratified sampling on the titles to obtain the desired number of titles. This option, while having the advantage of countering canonical bias, was excessively time-consuming: drawing a truly random sample would have returned texts whose digital availability, even as scanned items without text layers, was far from assured.

Another option would have been relying on literary histories to identify works which, being already recognised as relevant and thus partially ‘canonised’, were likely to be already available in digitised formats.¹⁰ While easier, such an approach would have had a negative impact on representativeness; if one follows a truly quantitative line of thought, any narrative of the evolution of European drama cannot be built only on a sequence of exemplary works, but rather requires engaging *at least to some extent* with what Cohen (2002, 23) famously termed ‘the great unread’.

9 E.g., the *Cambridge Bibliography of English Literature*, vols. 1–2, for English.

10 See, e.g., *Canon 60*, a collection of Siglo de Oro Spanish plays edited by Joan Oleza and the EMOTHE team for the *Biblioteca Virtual Miguel de Cervantes* (<https://www.cervantesvirtual.com/obra/canon-60-la-coleccion-esencial-del-tc12-teatro-clasico-espanol>).

A third method, perhaps the most pragmatic one, consisted in exploiting existing digitised resources to their full extent and reducing new encoding tasks as much as possible – in other words, assembling what Mrugalski et al. (2022, 21) call “an opportunistic corpus”. This translated into retrieving as many digitised plays as possible and then filling the gaps by manually encoding further texts from scratch. The negative implications of this strategy are clear, especially as it necessarily reintroduces some canonical bias. At the same time, as Annelie Ädel acknowledges, “it is [...] very difficult not to include some element of opportunism in corpus design, as we do not have boundless resources”, and therefore such an approach can be accepted where time- and resources-related constraints are present, and provided that “criteria for selecting material for the corpus be clear, consistent and transparent” (2020, 6).

Ultimately, the corpus size was set at 150 plays – a manageable size that ensures temporal coverage without exceeding practical constraints. While such an extension situates the corpus in an intermediate position – too big for an accurate close reading, too small for data-driven distant reading or cultural analytics (Manovich, 2020) – it nonetheless meets the requirements for a pilot study on evolutive trends within early modern drama.

Corpus retrieval

Once the preliminary corpus parameters were set, I moved to identifying and retrieving the actual texts. To do so, I first checked the relevant DraCor collections together with a number of databases offering open access to TEI-XML-encoded early modern plays, including the *Deutsches*

Textarchiv,¹¹ *TextGrid*,¹² *EMOTHE*,¹³ *EarlyPrint*,¹⁴ the *Oxford Text Archive*,¹⁵ and *Théâtre Classique*.¹⁶

Figure 1: The number of plays written between 1561 and 1710 and available as TEI-XMLs in selected datasets.¹⁷

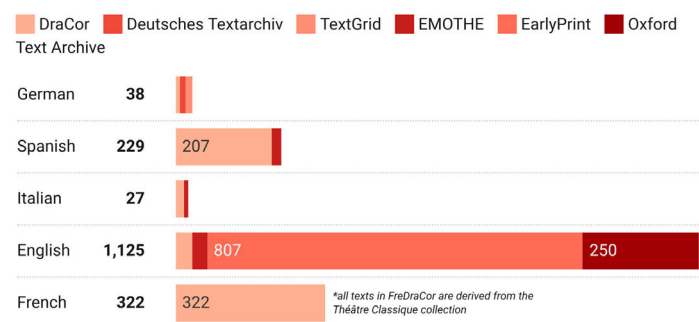


Figure 1 conveys immediately the imbalances in text availability across different collections: while there is plenty of TEI-XML plays in French and English, cutting the need for other sources almost to zero, the other language corpora are less resourced. Therefore, the next step required retrieving texts which were already available online but in formats other than TEI-XML (e.g. HTML, TXT, DOC).

11 <https://deutschestextarchiv.de>.
12 <https://TextGridrep.de>.
13 <https://emothe.uv.es/biblioteca>.
14 <https://texts.earlyprint.org/works>.
15 <https://ota.bodleian.ox.ac.uk>.
16 <http://www.theatre-classique.fr>. This was cross-checked with the *Dramacode* index (<http://dramacode.github.io>), compiled by Frédéric Glorieux.
17 These statistics refer to the time at which this research project started, i.e. April 2022. Since most of these corpora are 'living', the number of early modern texts has meanwhile increased: as of January 2026, for example, GerDraCor hosts 17 plays written between 1561 and 1710.

Most of these plays came from scholarly projects (such as the *Biblioteca Virtual Miguel de Cervantes*¹⁸ or individual philological editions), public-oriented archives (like *Wikisource*¹⁹), or commercial databases (such as the ProQuest/Chadwyck-Healey's *Teatro Español del Siglo de Oro* collection²⁰). Only as a last resort did I turn to performing OCR, correcting, and encoding scanned texts provided by platforms such as *Google Books*²¹ or the *Internet Archive*.²²

As far as the actual choice of titles was concerned, the main concern was ensuring that the whole 150-year-long time frame under investigation was evenly populated, with each period adequately represented. In practical terms, this translated into employing random stratified sampling to select ten titles for each decade (two for each language) from the available pool of texts. In the decades where no texts were immediately available, I used bibliographic sources (listed in Appendix A) to identify suitable titles to retrieve and encode from online archives; in this phase, some degree of hand-picking while filling these gaps was unavoidable.

Some additional criteria were enforced during the selection process. Where possible, I tried to avoid including anonymous works, and I checked the authors' distribution so that there were no more than three works for an individual writer. I also excluded operas and other dramatic texts which feature extensive singing or music (e.g. the *Singspiele* of the late German Baroque, the Venetian *drammi musicali*, etc.), preferring to focus on more established dramatic genres.

Finally, some issues common to all corpora might warrant some words. On one side, I was often faced with a lack of complete and/or reliable bibliographical information on some less-known works, as it often happens in the field of early modern studies. For example, there

18 <https://www.cervantesvirtual.com/>. The *Teatro clásico español* (TCE) collection from the BVMC contains 4388 items between primary texts and critical literature, available sometimes as HTML and sometimes as PDF (scanned images).

19 <https://wikisource.org>.

20 This database, accessible at <https://www.proquest.com/teso>, contains more than 800 plays by 16 different dramatists.

21 <https://books.google.com>.

22 <https://archive.org>.

might be difficulties in dating with certainty some works, such as plays published in collections only years after their composition or representation, and this would negatively impact any attempt at reconstructing an accurate genealogy of theatrical evolution. To address this issue, I consulted critical literature to collect the years of composition, first representation and publication for each play, which are used to determine the ‘normalised’ or reference year DraCor employs.²³

On the other side, the issue of copyright needed to be taken into account. While most plays are in the public domain, some are derived from scholarly editions and thus still protected by copyright (Margoni and Perry, 2011). Since EmDraCor texts were from the start intended to end up in the main DraCor corpora – released under the *Creative Commons* o (CCo) license – I re-used scholarly editions only when their copyright had expired, or employed them only as reference while independently conducting transcriptions from public domain sources.

Corpus building

Encoding pipelines

Figure 2 schematically summarises the various procedures I developed for converting texts with different layers of markup (from well-formatted TEI-XMLs to plain texts from OCRs) to the standard DraCor format as described in its ODD encoding scheme.²⁴

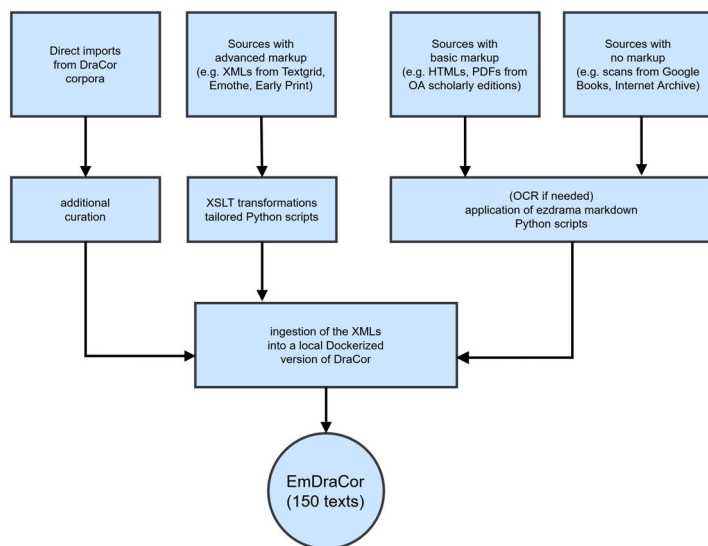
To begin with, extensive bibliographical research made it possible to locate more than one hundred texts already in TEI-XML format. Among them, about 40 were already featured in some DraCor corpus, and thus were only downloaded and subjected to some additional curation to ensure homogeneity. Other files, originating from various scientific projects on early modern drama, needed instead a complete refactoring of their markup to make them DraCor-ready.

23 See <https://dracor.org/doc/faq>.

24 See <https://dracor.org/doc/odd>.

Generally speaking, the transformation procedure consisted in two main tasks: the removal of excessive layers of semantic and/or graphical annotation (e.g. lemmatisation, part-of-speech tagging, facsimile layouts etc.), which are superfluous to my purposes, and the attribution – if not already assigned – of a speaker tag to each speech instance, which serves as a prerequisite both for literary network analysis and speech distribution statistics.

Figure 2: *Corpus onboarding pipeline.*



Tailored ‘DraCorisation’ pipelines were developed for converting the TEIs obtained from archives which provided a sizeable number of texts to the final corpus, such as *TextGrid*, *EMOTHE*, and *EarlyPrint*. The *EarlyPrint* corpus, in particular, presented a larger opportunity, as I collaborated with other DraCor team members not only to transform the thirty

plays needed for this project, but also onboarded the whole collection to create the *English Drama Corpus* (EngDraCor).²⁵

The main challenge there lay in speaker disambiguation, i.e. identifying which character speaks each line. This problem proved difficult to take on, as early modern printing conventions were inconsistent, with speakers identified variously by name, role, abbreviation, or description, sometimes shifting within a single text. Resolving these ambiguities was essential for any network analysis or speech distribution metric, since the DraCor algorithms are based on the speaker tags each character is marked with. Once speaker attribution was stabilised through a dedicated tool,²⁶ the texts underwent XSLT transformation to produce DraCor-conformant TEI files. The EmDraCor texts were subsequently forked from the main project for independent curation, though they remain compatible with the larger EngDraCor architecture.

About one-third of EmDraCor plays, however, were derived from files in formats other than TEI-XML. Some of them still contained some semantic markup, such as the HTML files from *Wikisource* or the PDFs from various digital scholarly editions. Other texts, instead, were preserved only as scans in the digitised collections of *Google Books* and the *Internet Archive*, which allow downloading PDFs and plain text outputs of optical character recognition (OCR). While these transcripts were still messy, and thus required extensive post-correction, tests with other OCR tools such as *OCR4all* (Reul et al., 2019) did not show dramatic improvements in text quality; accordingly, I decided to work directly with the raw outputs from the aforementioned platforms.

For these lower-quality sources, I adopted a modified version of the conversion utility *EasyDrama* (Skorinkin, 2024). The principle behind its functioning is simple: rather than wrestling with inconsistent markup, one strips the text to plain TXT and annotates structural elements (such as act and scene divisions, speaker attributions, and stage directions) using a lightweight, Markdown-like syntax. The script then transforms this annotated plaintext into valid TEI-XML. This approach trades granular

25 <https://dracor.org/eng/>; on its inception see Börner et al. (2023a).

26 <https://dracor-org.github.io/epdracor-whois/>.

markup for speed: while some basic markup information from the original HTML or PDF is lost, the conversion process becomes fast enough to handle dozens of texts. Given these advantages, I decided to process all non-XML sources through this pipeline, regardless of their original format.²⁷

Editorial practice

The general goal of the DraCor project is not to produce flawless philological or critical editions, but rather to make dramatic works both readily available to readers and suitable for computational analysis, i.e. machine-actionable. Accordingly, editorial interventions on texts are most often cut to a minimum and confined to markup adjustments. In the case of EmDraCor, however, its comparatively small size made possible some minor fixes and improvements. Such editorial interventions chiefly regarded low-quality plain texts derived from OCRs, but some of them, like the improvement of segmentation, were applied to all texts.

On one side, even though the methods later applied to the corpus are language-independent, improving text quality still remained necessary to ensure the correct computation of certain metrics. Accordingly, I applied an extensive clean-up pipeline to all OCRed text. This started with the selection of the most recent copyright-free edition (often from the 19th century) as a source, in order to reduce OCR noise as much as possible. Since I did not have philological interests and I wanted to avoid remaining embroiled in the intricacies of digital scholarly editing,²⁸ I also settled for a normalised/modernised transcription as unobtrusive as possible, using (when available) the *Wikisource* style conventions as guidelines.

Among a few other interventions, I restored correct punctuation when lost during the OCR, normalised speaker names to ease their

27 Some novel approaches, not yet developed at the time of the corpus preparation, include semi-automatic, LLM-powered TEI encoding (Giovannini and Skorkin, 2024b) and direct OCR-to-TEI pipelines (Dennerlein et al., 2025).

28 For an overview see Pierazzo (2020).

subsequent disambiguation,²⁹ and corrected some notable errors in markup (e.g. songs clearly sung by a character not being attributed to him/her, but rather encoded in a separate element, etc.).

On the other side, since most calculations performed by the DraCor algorithms are based on textual segmentation, the operation of dividing a text into units appeared crucial as well. Together with speaker disambiguation, it has indeed been often singled out as a major issue when trying to operationalise dramatic structures.³⁰ More generally, scene segmentation represents a long-standing issue both in natural language processing and computational literary studies, with the concept of 'scene' blurring into adjacent ones such as 'plot', 'topic', and 'event' (see Zehe et al., 2021, 3167–69).

Drama knows essentially two types of segmentation: the high-level 'act' division and the low-level 'scene' division. The usage of the former type, in particular, has fluctuated according to the poetics of each age, with the criteria for marking act breaks shifting over time: the *Oxford Encyclopedia of Theatre and Performance* speaks in this respect of "some combination of (a) an interruption of chronological continuity, (b) a change of setting, (c) a significant change in thought, mood or plot development" (Vince, 2003).

As far as literary theory is concerned, the genealogy of the concept of act starts with Aristotle's comments about the five stage units of Greek drama, upon which Horace later grounded his peremptory recommendation: "neve minor neu sit quinto productior actu fabula"³¹ (Ruggerio, 1973, 173–74). Horatian influence, mediated through later grammarian Aelius Donatus, eventually came to dominate classicist-oriented theatre in the early modern age, but was not readily accepted everywhere.

29 Early modern texts often feature several different spellings and abbreviations for the same character. Thus, *Richard* might appear as *Rich.*, *Ri.*, *R.*, etc.

30 See e.g. the comments by Brainerd and Neufeldt (1974, 35 ff.) on the work by Salomon Marcus discussed in the next chapter.

31 "the play should not be either longer or shorter than five acts" (Hor. *Ars Poetica* 189–190, trans. T. S. Dorsch/P. Murray).

Here again, it is possible to see a split along the fault line previously introduced: French and Italian theatre, following the classical models mediated by the humanists, made use of acts, while other dramatic traditions followed the rather relaxed segmentation from medieval theatre, where pauses within the action were mostly linked to stage reasons (Lochert, 2018a, 191).

It is not surprising, then, that competing four- and three-act structures, based on different readings or willing departures from previous theorists, became available especially (but not only) in the English and Spanish milieus. Three-act structure, in particular, became the norm in Spanish drama following the precepts set by Lope in his *Arte nuevo de hacer comedias* (1609).³² Even no segmentation at all remained an option: the majority of English printed plays between 1565 and 1594 had apparently no act or scene division (Howard-Hill, 1990, 142), and the segmentations later introduced in Shakespeare's plays were little more than artificial "editorial encrustation[s]" (Hirsh, 1981, 12). German theatre, for its part, was in the first phase quite irregular as well, with extra acts being freely added after the fifth when the text's length required it (Jordan, 1939, 399).

Given this heterogeneity of practices, it is fortunate that the standard segmentation span used by DraCor algorithms³³ is not the act, but rather the scene, a perhaps more 'natural' partition born out of the necessity to deal with changes in the characters' configuration on stage or to move around theatrical props. The number of characters which needed to be changed in order to identify a scene break, however, was again dependent on local dramatic conventions, as Lochert (2018b, 201–4) shows.

French neoclassical theatre, for example, often saw a character remaining on stage and being the *trait d'union* between different configurations or scenes, in a practice called *liaison des scènes*,³⁴ while the sub-

32 See vv. 211–212: "El sujeto elegido [...] en tres actos de tiempo le reparta" ("Distribute the chosen subject in three acts of time") (Vega, 1914).

33 Technically speaking, they use the lowest-order (i.e. most nested) <div> element, which most often corresponds to a <div> with <type="scene">.

34 On this see Lochert (2018b, 204 ff.) and especially the full-length study by Douguet (2023).

stitution of a full set of characters with another marked an act change. Shakespeare, on the other side, tied scene sequence not to characters, but to changes in location (Ranke, 2010, 710, qtd. in Krautter, 2023, 273). These two concurring options, effectively tying scene division to character groups or to places, were variously received in other dramatic traditions.³⁵

Most texts in EmDraCor already featured first- and/or second-level segmentation, usually from their original source. Works digitised and transcribed from historical editions, however, did not always have this markup in their printed editions, often for the historical reasons outlined above. Maintaining part of the corpus without a proper segmentation, however, would have hindered any meaningful comparison through computational tools, and thus required some fix.

Previous work in scene segmentation offered some inspiration. For example, Gius et al. (2019), using categories from Genette (1972), proposed this broad narratological definition:

A scene is a segment of the *discours* (presentation) of a narrative which presents a part of the *histoire* (connected events in the narrated world) such that (1) time is equal in *discours* and *histoire*, (2) place stays the same, (3) it centers around a particular action, and (4) the character constellation is equal. (Zehe et al., 2021, 3169)

Drama scholars introduced other segmentation modes specifically tailored to theatre practice. For instance, Marcus (1970) introduced the concept of dramatic ‘configuration’, defining it as “the time frame during which no entry or departure from the stage takes place” (291). Jansen (1968, 77) proposed instead a new unit, the ‘situation’, whose boundaries

35 This crucial distinction was already noted by critic Heinrich Wilhelm von Gerstenberg in his *Briefen über Merkwürdigkeiten der Literatur* (1767): “It should be noted in advance that the English divide the scenes in a different way than the French. The former indicate a change in the stage, but the latter only indicate a change in the characters, for which the English have no special made-up word” (qtd. in Jordan, 1939, 403).

were defined by the change of characters and/or scenic elements; in doing so, he was again reprising, to some extent, the notion of scene as in French classical theatre.

Drawing on this, I opted for a simpler, pragmatic solution. Close reading of the unsegmented plays suggested that single entrances or exits rarely signal major dramatic shifts, while the movement of two or more characters typically marks a genuine change in stage configuration. Accordingly, a full scene change was added in the markup when both the following criteria were satisfied:

- At least two speakers left or entered the stage
- There is a stage direction unequivocally signalling such action (often shown by verbs of motion such as *salen*, *vanse*, *exeunt*, *escono*, etc.)

Such general principles, which try to establish a common baseline for texts from very different dramatic cultures, remain of course highly subjective. As Bartsch et al. (2023, §56) underline, “segmentation is by no means a purely formal matter but must always be thought of as a hermeneutical practice” which has a noticeable influence on text modelling.³⁶ Whichever solution is chosen, however, describing and motivating the criteria employed for segmentation has positive epistemological effects, favouring correct interpretation of the results and later re-evaluation of the methodology.³⁷

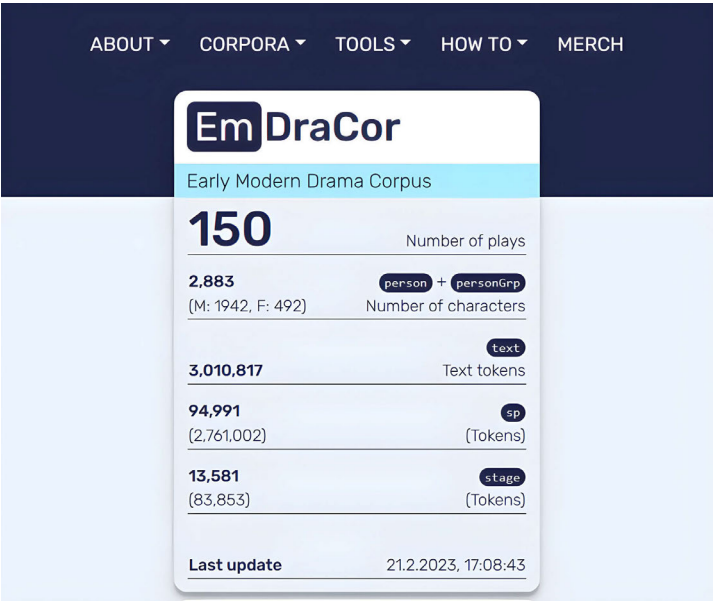
36 Bartsch and colleagues divide the segmentation process in three phases: “the cutting of a segment from the text continuum (zoning), the assignment of the segment to an analytical category (subsumption) and the determination of the depth of mental access on the basis of more or less contextual information (interpretation)” (§32).

37 “The digital access forces the disclosure of segmentation decisions, so to speak, and strongly promotes the definition and explication of criteria for these decisions. The algorithmic formulation as well as its implementation as a clear and discrete segmentation decision thus sharpens my concepts” (Bartsch et al., 2023, §74).

Deployment

Once all TEI-XML files were clean and properly encoded, I loaded them into a dockerised version of DraCor, following the procedure described by Börner et al. (2023b). This workflow uses the Docker technology to recreate a fully functional, locally hosted DraCor environment, complete with its API, frontend, and metrics service. This container can then be populated with TEI-XML plays from any sources, exported as an image, and later reactivated as needed.

Figure 3: Model card for the Early Modern Drama Corpus.



This procedure contributes significantly to addressing the ‘reproducibility problem’ often raised within the digital humanities (O’Sullivan, 2019; see also Chapter Four, ‘Coda’), insofar as it freezes the corpus in a stable version on which one could later re-run the same original

experiments run here. Accordingly, the final product of this pipeline, EmDraCor (Figure 3), can be analysed using the full range of DraCor tools.

Methods

The pipelines described in the previous chapter produced a highly curated, machine-actionable corpus of early modern plays, whose standardised features now enable meaningful comparison across traditions. Before proceeding with the analysis, however, it is imperative to clarify which methodology is to be followed. Since, as Gius and Jacke (2022, 16) rightly underline, “computational approaches [should be] aligned with the theoretical background guiding the [original] analysis”, I wanted to use an approach consistent with the framework originally employed in ‘Modern European Literature’.

As evidenced in Chapter One, Moretti’s focus there lay very clearly on the morphological variety of early modern drama – the study of “the component parts [...], their relationship to each other and to the whole”, as Vladimir Propp (1968, xxv) would put it in the opening pages of his *Morphology of the Folktale*.¹ While still lacking a computational dimension, his approach foregrounded, to some extent, the paradigm of ‘quantitative formalism’ (Allison et al., 2011) in which he would spend the remainder of his career, and which he thus later summarised:

Quantitative formalism: taking an aesthetic form and dismantling it down to its basic elements; transforming – or rather, let’s say, reducing – a novel to its paragraphs, or a drama to a sequence of linguistic exchanges. Having done this, one is confronted with ‘a second-level form’, as Alex Woloch once wrote to me; a squared, abstract form, a

1 On the relation between morphology and literature see also Gilodi (2021) and Marx (2013).

form of forms [...]. This is the eminently analytical vocation of quantitative formalism: to make everything rest on explicitly “local” interactions such as [...] the back-and-forth between one character and another. (Moretti, 2021, 385)

This book situates itself in the same tradition by treating plays as configurations of speech turns and scenic units rather than as linguistic content, i.e. by following a form-centred approach. Furthermore, such a strategy also makes it possible to circumvent two major features of the corpus, namely its multilingualism and historical character. EmDraCor contains texts in five different languages, spanning more than one century, with considerable linguistic variation within each tradition. Such inter- and intralingual differences would have represented a hurdle for directly applying natural language processing methods (see Piotrowski, 2012). To achieve reliable results in most tasks, it would often have required training multiple language-specific models from scratch rather than simply fine-tuning existing ones (cf. Manjavacas and Fonteyn, 2022), considering also that domain adaptation for literary corpora typically demands more than fine-tuning (Bamman, 2024).

In the following pages, I first briefly outline some possibilities for operationalising the concept of drama, recalling some older and newer approaches, and then present my own methodology for modelling EmDraCor plays.

Operationalising drama

From Moretti (2013b)’s seminal formulation – influenced, among others, by Percy Bridgman’s *The Logic of Modern Physics* (1929) and by Thomas Kuhn’s essay ‘The Function of Measurement in Modern Physical Science’ (1961) – to the most recent, DH-specific debate (see especially Krautter, 2022; Pichler and Reiter, 2022), operationalisation represents a core task in any quantitative approach to literature, and its role within every research pipeline cannot be overstated. “[B]uilding a bridge [...] from the concepts of literary theory, through some form of quantification, to lit-

erary texts” (Moretti, 2013b, 1), operationalisation is a delicate procedure which requires a balancing act between the specificity of the literary text and the algorithmic requirements of the machine.

In the context of computational drama analysis, perhaps the most important operationalisation practices concern the identification and quantification of some crucial dramatic features and the modelling of their relations. Defining what the key components of a play are, however, represents a complex theoretical problem in its own right, with its first discussion tracing back at least to Aristotle’s comments on the six parts tragedy is made of.² One could say, with some approximation, that most scholarship has been working with what Kretz (2012, 105; cf. Kafitz, 1989, 32–40) called the three ‘elementary components’ of drama:

- *Characters*: the entities to be represented on stage, to be understood as deliberate literary constructs (cf. Pfister, 1988, 160–161, which for this reason prefers to use the word ‘figure’);
- *Plot*: ‘the arrangement of the incidents’ (*synthesis* or *systasis ton pragmaton*) happening during the play, which for Aristotle represents the most important component of drama;³
- *Dialogue*: the alternation of speech acts, which is often considered the first ‘external’ feature of drama one immediately recognises (Schlegel, 2018, 19) and thus a “primary distinctive feature” of the whole genre (Veltruský, 2017, 248).

The same tripartition can also be found in the Russian Formalist tradition, with Boris Yarkho – a major theoretical referent for this book, as the reader will shortly see – indicating ‘dialogue’ (*dialog*), ‘characters’ (*personazhi*), and ‘action’ (*dejstvie*) as the three distinctive elements of drama

-
- 2 “Necessarily then every tragedy has six constituent parts, and on these its quality depends. These are plot [*mythos*], character [*ethe*], diction [*lexis*], thought [*dianoia*], spectacle [*opsis*], and song [*melos*]” (*Poetics* 1450a, 5–10, trans. W.H. Fyfe).
 - 3 “[...] most important of all is the structure of the incidents. For Tragedy is an imitation [*mimesis*], not of men, but of an action and of life” (*Poetics* 1450a, 15, trans. S.H. Butcher).

(Yarkho, 2006, 410). Of course, such narratology-oriented perspectives sideline once again the performative dimension of drama, which other researchers tried to recover,⁴ but which also remains outside the scope of this work.

In any case, it can be argued that these three high-level components find a concrete expression in different textual aspects (content, style, structure...). Such aspects can in turn be investigated through a range of computational methods: as Willand et al. (2017) acknowledge, the most popular options in this sense have been the topic modelling of dramatic genres (e.g. Schöch, 2017b), the exploration of network structures (see *infra*), and the analysis of character speeches (e.g. through stylometry, see Šeĭa et al., 2024; Savoy, 2023).

As mentioned earlier, my approach to operationalising the EmDraCor corpus is both content- and language-agnostic, and my methodology will be therefore grounded in form-oriented approaches. Before presenting my own strategy, I will briefly review previous similar approaches to drama analysis. Following and integrating the reconstruction by Fischer (2024), I will begin by outlining the exemplary contribution of three significant twentieth-century scholars – Boris Yarkho, Solomon Marcus, and Hartmut Ilseemann – whose pioneering work laid the ground for modern computational study of drama.

After briefly examining the current scholarly landscape, I will then devote a methodological excursus to literary network analysis, which has been central in past and present DraCor-based scholarship and will feature prominently in the experiments presented in Chapter Four. For this reason, it seemed helpful to briefly summarise its origins and discuss its strengths and weaknesses in modelling dramatic texts.

4 Bernhard Asmuth (1980, 4), for example, singles out the Aristotelian *mythos*, *lexis*, and *opsis* as the three most important formal elements of drama, while considering the characters (*ethe*) part of the plot itself. Accordingly, its own tripartite systematisation is based on the concepts of plot (*Handlung*), dialogue or character speech (*Figurenrede*), and 'sensory' performance (*sinnliche Darbietung*).

Stages in the development of computational drama analysis

Modern scholars agree in considering the formalist school, and more specifically its Muscovite incarnation,⁵ a primary source of inspiration for quantitative formalist endeavours; the links between the two movements have been increasingly scrutinised in recent years (see e.g. Pilshchikov, 2020; Lvoff, 2021; Pilshchikov, 2022). This renewed interest within the digital humanities,⁶ as Fischer et al. (2019a, 1) put it, is motivated by pragmatic reasons, since “we are now able to operationalise and automatise formalist research ideas, to reproduce them, to scale them up and to further develop methods along those lines”. In other terms, modern methods seem now in a position of nurturing the seeds of ‘computational thinking’ which were already planted as early as the 1920s.

Moscow scholar Boris I. Yarkho (1889–1942), in particular, has been rediscovered as “a very early forerunner of contemporary research practice in the Digital Humanities” (Fischer et al., 2019a, 7). A multifaceted scholar well-versed in various languages and philologies, Yarkho made

-
- 5 As Pilshchikov (2020, *passim*) points out, the formalist movement has often been equated tout court with the St. Petersburg-based OPOJAZ (Society for the Study of Poetic Language) helmed by Shklovsky, Tynianov, and Eichenbaum, while its Moscow counterpart, i.e. the Moscow Linguistic Circle founded by Jakobson and frequented by Tomashevsky and Yarkho, has often been neglected. The difference between the two flavours of Formalism – which remained nonetheless intertwined thanks to the constant exchange of people and ideas – can be roughly summarised as follows: the OPOJAZ theorist advanced a more ‘morphological’ or ‘functionalist’ formalism, focused on texts (and thus to some extent ‘qualitative’), while the Muscovites were rather involved with methodological issues, adopting a ‘quantitative’ perspective which is somewhat nearer to that adopted, in modern times, by the Stanford LitLab.
 - 6 Two milestones in this sense can be recognised: the 2015 Stanford conference on ‘Russian Formalism and the Digital Humanities’ (program: <https://digitalhumanities.stanford.edu/russian-formalism-digital-humanities>) and the 2019 special issue of the *Journal of Literary Theory* on ‘Moscow Formalism and Literary History’, featuring new translations of works by Boris I. Yarkho, Mikhail L. Gasparov, and Maksim I. Shapir.

a significant contribution to drama analysis in two studies written between the 1920s and 1930s (“The Comedies and Tragedies of Corneille. A Study in the Theory of Genre” and “Speech Distribution in Five-Act Tragedies. A question of Neoclassicism and Romanticism”⁷), which showcased an original, quantitative-driven approach to the poetics of genre and of literary trends.

Another pioneer in the quantitative treatment of drama, from a different perspective, was Romanian mathematician Solomon Marcus (1925–2016). His approach, he claimed, went beyond the “simple statistics” done by previous authors, including the Russian Formalists, but aimed rather at constructing a full ‘mathematical-linguistic’ model of drama to which “various mathematical methods coming from game theory, graph theory, theory of sets, algebra of binary relations, probability and information theory, [and] formal grammars” could be applied (Marcus, 1999, 294–95).

Among his contributions, he notably developed the idea of a Boolean matrix to keep track of the changes in character configurations, based upon the virtual model of an ‘unusual spectator’ “who is only capable to observe the entrances and exits of the actors and monitor their co-occurrences on stage without listening to what they say” (Fischer and Skorinkin, 2021, 520). Marcus’ whole approach to drama analysis was also notable in its strong functionalism: “the whole mechanism of a dramatic work is seen as a machine, with whose help I can understand better the functionality of a dramatic work” (Marcus, 1971, 150).

Marcus’ configuration matrix was later commented upon in the work of leading German structuralist Manfred Pfister (Pfister, 1988, 171 ff.),

7 Both studies are appended to the modern Russian edition of Yarkho’s magnum opus, the *Methodology for the Exact Study of Literature* (Yarkho, 2006). “Speech Distribution in Five-Act Tragedies” has also been translated into English (Yarkho, 2019) and its workflow has been computationally reproduced and tested by Wendell (2021). An English translation of “Corneille’s Comedies and Tragedies” is still missing but its workflow has already been outlined and/or digitally reproduced by Sonkin (2022), Wendell (2021), and Seĵa (2020).

and later played a central role in the work of one of the first actual practitioners of digital drama studies, Hartmut Ilseman, as exemplified in his paper ‘Computerized Drama Analysis’ (Ilseman, 1995). There, he introduced an experimental but already quite complete set of programs for the quantitative study of plays, which translated ideas by Marcus, Pfister, and Wilhelm Fucks⁸ into (computational) practice. In particular, software developed by Ilseman and his collaborators⁹ allowed for computing several statistics related to stage configurations and speech acts, and visualising them in several formats.

While more ‘technical’, Ilseman’s 1995 paper touched also upon some key points which remain debated in computational drama research to this day: the criticism against statistical methods, which seem to refute the uniqueness and irreproducibility of the work of art (12), the necessity of “a large data background” (13) to establish “a reliable typology of dramatic forms” (21), the belief that “quantitative statistical relations are capable of expressing qualitative aspects of dramatic plays” (14).

Inspired by structuralist and formalist thinkers such as Yarkho, Marcus, and others, and heralded by the path-breaking work of Ilseman and colleagues, computational drama studies have become throughout the years an important component within the larger ecosystem of digital literary studies, both in quantitative¹⁰ and in qualitative terms.

This vitality has also manifested in the establishment of several research projects dedicated to creating and annotating corpora and developing reusable programming tools. In the German milieu, for exam-

8 German physicist who worked on the application of empirical-quantitative methods to the humanities (see e.g. Fucks, 1968, 1971).

9 For a brief technical account of the inception of Ilseman’s original program see <http://www.shak-stat.engsem.uni-hannover.de/progdoku.html>. For its successor, the *Internet Drama Analysis Program* (IDAP), whose main purpose was providing statistical data on English Renaissance plays, see Ilseman (2006).

10 See e.g. the increase in the share of drama-related papers in the *Index of Digital Humanities Conferences*, (Weingart et al., 2024).

ple, the long-standing DraCor initiative¹¹ has been over time joined by projects such as *Quantitative Drama Analytics* (QuaDrama/Q:TRACK, see Pagel et al., 2023¹²) and *Emotions in Drama* (EmoDrama, see e.g. Dennerlein et al., 2023¹³). Furthermore, specific fora have been established to allow scholarly exchange within the subfield, such as the ongoing series of *Computational Drama Analysis* workshops (Cologne 2022, Berlin 2025).¹⁴

Excursus: dramatic network analysis

Assuming the *Computational Drama Analysis* workshops as barometers for recent trends in computational drama research, one immediately notices how the methodology of (literary) network analysis is still playing a central role in current investigations. First popularised by Moretti himself in ‘Network Theory, Plot Analysis’ (now in Moretti 2013a, 211–40), and then more thoroughly developed among others by Trilcke (2013), the modelling of dramatic structures as networks remains indeed one of the most popular practices in computational literary studies, with a rich scientific production exploiting its potential for tasks such as investigating characters and plots, assessing literary theories, and more (see Labatut and Bost, 2020, 26–27; Trilcke et al., 2024, 2).

The remarkable success of this methodology in the subfield of drama studies – part of a wider ‘network turn’ (Ahnert et al., 2021) in the humanities – can be explained by pragmatic and conceptual factors. On the one

11 Emerging from the grassroots DLINA project (<https://dlina.github.io>), DraCor has later been supported by the pan-European *Infrastructure for Computational Literary Studies* (<https://clsinfra.io>).

12 <https://quadrama.github.io/>. See also the software (*DramaAnalysis*, *DramaNLP*) they developed for drama analysis: <https://quadrama.github.io/tools.en>.

13 The same research group also developed *Katharsis*, a “tool for drametrics” (cf. Romanska, 2014) which is based on Marcus’ and Ilseman’s work and available at https://lauchblatt.github.io/Katharsis/sa_selection.html; see Schmidt et al., 2019.

14 The proceedings of the first workshop are collected in Andresen and Reiter (2024a); selected papers from the second one are forthcoming in the *Journal for Computational Literary Studies*.

hand, theatrical texts are well suited for computational network analysis, since they almost invariably present a rigid structure which can be easily transformed into a network where characters are represented as nodes and their relations as edges. Such relations could be modelled with different layers of precision, ranging from the creation of basic co-presence networks¹⁵ to networks based on interactions, family relationships, or other complex connections (e.g. information flow, see e.g. Andresen et al., 2022).¹⁶

On the other, network representation offers a “particular kind of abstraction”, which is “uniquely capable of working across scales” (thus supporting both micro- and macro-analyses) and at the same time “[i]ntuitive to grasp and yet mathematically complex” (Algee-Hewitt, 2017, 752). In this perspective, network properties are assumed to mirror some crucial features of the works – density may correlate with plot complexity, or centrality distribution may reflect character hierarchies. Accordingly, a comparison between different networks can yield insight into literary history and cultural development.

Furthermore, this methodology proves especially effective when deployed on multilingual corpora, because its “abstract modelling and the possibility of a (more or less) ‘purely’ topological comparison makes the comparative analysis highly independent of any domain- or language-specific origins of the networks” (Trilcke et al., 2024, 2).

Applications of network analysis to dramatic texts, however, raise methodological challenges as well. In an early phase, it was not uncommon to find papers which considered drama only as a test ground for verifying mathematical or statistical principles, and thus disregarded the works’ literary specificity. These investigations, which often seemed to be little more than “a challenge for computer scientists” (*ibid.*, 7), often

15 This type of network – the easiest to extract automatically, since it only requires character identification and scene segmentation within the text – is the one natively featured in DraCor.

16 Such networks require further preprocessing of the source text (e.g., involving NLP pipelines to recognise speakers and addressees) or task-oriented annotation.

failed to offer significant contributions to computational literary criticism.

Similarly, one needs to problematise the role of social networks analysis (SNA), which initially acted as a bridge from (pure mathematical) network analysis to literary network analysis (LiNA). While providing an important inspiration, concepts and theories from SNA should not be applied slavishly to fictional texts: even though some properties of real-life networks can be found also in literary ones (Labatut and Bost, 2020, 47–48; see also Waumans et al., 2015), “[t]here is no reason to suppose that the writer tries to mimic actual social relationships when producing the work of fiction”, and “numerous character networks extracted from fictions do not exhibit realistic topological properties” (Labatut and Bost, 2020, 5).

Accordingly, one should always assess the significance of network metrics within the context of the literary work one investigates, and not at some abstract level. Otherwise, as Peer Trilcke has warned,

a literary network analysis which, when reflecting on its foundations, disregards the objection that it may miss the specific ontology and historicity of its object [will] hardly be able to deliver reliable textual empirical results from the viewpoint of literary theory. (Trilcke, 2013, 207)

In concrete terms, this also means that any attempt at literary network analysis should start from a reflected, thorough operationalisation of dramatic structures. It should, in other words, define clearly some key formal features of a play, namely, what constitutes an exit from or entrance to the stage, what constitutes a character, and what does it mean for two characters to have an interaction (Jokisch and Rojas Castro, 2025).

Modelling EmDraCor plays

In accordance with these methodological reflections, the first step in modelling dramatic texts would consist in identifying the features most

suitable to describe them. Past scholarship has often focused on developing new measures for capturing what they consider key aspects of textual structures, but it is important to remember that creating such measurements should not be a goal in itself. Introducing novel metrics should rather provide an objective hermeneutical advantage over existing ones or conventional non-computational methods. Furthermore, it ought to be kept in mind that

[m]easurements can be taken of any quantifiable aspect of a text, but figuring out the significance of that metric to an understanding of the text, or better, mapping that metric onto a preexisting critical concept (such as style, plot, or theme), is crucial to making sense of what is being measured. (Algee-Hewitt, 2017, 759)

Some examples from previous work might help frame the question. Yarkho's study of Corneille's plays, for example, was based upon 30 features describing various structural aspects, which he later cut to 15 measures he considered the most significant for his research goal, i.e. effectively differentiate between comedies and tragedies.¹⁷ His subsequent paper on Classicism and Romanticism relied instead on only four features, all related to characters. Interestingly, Yarkho's quantitative-oriented, but still 'analogic' (i.e. pre-digital) approach led him to employ in his analysis eminently qualitative features as well. One of the parameters he employed in his essay on Corneille, for example, was 'saturation' – a blend of metrics indicating, among other things, "the percent of personages that act in their own interest" and "the ratio of the number of actions on stage to the number of actions that occur behind-the-scenes" (Gasparov, 2017, 141–42).

Other Russian (neo)formalists, reviewed in Wendell (2021, 31 ff.), followed Yarkho's example and developed their own features for the study of specific works. Revzina and Revzin (1971), for example, introduced metrics such as 'character connectedness', 'dramatic engagement rank', and 'probabilistic link', which anticipated several later DH endeavours;

17 The full list of features can be read in Wendell (2021, 39–40).

their approach was later updated by Sapogov (1974), whose proposal for calculating speech distribution is also implemented in the DraCor frontend (cf. Börner and Trilcke, 2023, 64 ff.). Sperantov (1998), for his part, focused on stage directions in Russian verse tragedy, which he explored through measures like frequency, average length, presence within verses, lexical diversity, etc.

Modern computational approaches have kept developing their own features as well. Work done by the DraCor group, for instance, has sought to tackle the progressive structuration of plot, following earlier attempts by Xanthos et al. (2016). In particular, Trilcke et al. (2017) devised several ‘event-based’ measures, which “are used to identify or characterise a particular point in time within a drama”, and ‘progression-based’ ones, which “allow a general characterisation of the transformation of a dramatic network” (Fischer et al., 2017, 2–3). Thanks to these metrics, they were ultimately able to identify low- and high-dynamic plays within DraCor.

In his quantitative study of English early modern drama, Algee-Hewitt (2017) tried instead to build upon standard network measures of centrality, like eigenvector and betweenness, and developed two advanced metrics for the analysis of drama: ‘protagonism’ and ‘mediatedness’. While the first one is meant to “measur[e] the tendency of the play to concentrate the function of the protagonist in a single character [...] or to distribute it among multiple characters”, and to show the existence of an alleged “core/periphery divide” among the plays’ characters (763), the second one embodies the “need for a character to bridge the factions and to mediate between the subnetworks” (767). Gauging the chronological development of these two metrics allows the author to draw some conclusions about literary evolution and genre history as well – e.g. tragedies and historical dramas becoming more morphologically similar to comedies (771).

While these last two examples are representative of the tendency to centre metric development on network-related aspects of drama,¹⁸

18 The reasons for this are thus spelled by Algee-Hewitt: “If the network, as a visualization, compresses the dynamic action of the play into a schematic morphol-

it could be argued that only a broader pool of features, able to cover as many aspects of a play as possible, might be truly useful for a deeper comprehension of literary history and cultural evolution dynamics. While an all-encompassing operationalisation of drama is virtually impossible, it is possible and advisable to employ multiple methods to capture different structural aspects of drama – not only network analysis, but also scores capturing size, speech distribution, plot development, etc.

In modelling the plays from EmDraCor, I thus propose to go beyond the comparison of a restricted number of measures and rather collect as many of them as possible in feature vectors, which I will then use as hermeneutical tools for examining intertextual relationships and literary history at large. Vectors, as is well-known, are mathematical entities that represent a quantity characterised by both magnitude and direction. In computer science, and by extension within the digital humanities, vectors are more generally defined as a set of numbers arranged in a specific order or orientation (e.g.: $v = \{1, 2, 3, 42\}$). The most commonly used vectors within computational literary studies are word embeddings, which aim to capture the meaning of words or text based on their distributional properties.¹⁹

In the context of computational drama analysis, one can follow a similar logic to build *play embeddings*, i.e. vectors containing a variety of measures embodying different textual aspects of the dramatic text. While not completely novel, this approach has in the past concerned only individual components of drama, such as characters. Solomon Marcus, for example, stored the appearances of characters as vectors

ogy in which all of the interactions (relationships, conflicts, romances, murders) are part of a static representation, then the metrics from the distributed character network reduce these multidimensional visualizations into statistics that represent, in minute detail, a single aspect of the morphology of the drama" (Algee-Hewitt, 2017, 756–57).

- 19 Following the famous precept by Firth: "You shall know a word by the company it keeps" (Firth, 1957, 179). For an introduction on applications of word embedding models in digital literary studies see Schöch (2022); some theoretical considerations in Dobson (2021)

and computed their ‘difference’ via Hamming distance (Marcus, 1970, 293; 300). More recently, there have been at least two notable attempts to computationally model characters in German drama:²⁰ Fischer et al. (2018) used count- and network-based character measures to describe quantitative dominance relations, while Krautter et al. (2023) employed a composite array of features²¹ to investigate character types such as ‘the virtuous daughter’ or ‘the tender father’.

To my knowledge, the only recent works which attempted formal vectorisation of *whole plays* are Wendell (2021), which however deployed it only where necessary to digitally replicate Yarkho’s study on Corneille, and Szemes and Vida (2024), which applied it to solve a genre classification task. Their pioneering investigation, aimed at distinguishing comedies from tragedies within GerDraCor and ShakeDraCor, employed both standard network analysis metrics and some new features developed for the specific task.²²

While the supervised classification procedure the authors performed did not find a “striking [structural] difference” (177) between the two genres, they could nonetheless single out some properties which were highly predictive of generic alignment – as speculated already by Ilseman (1995, 16). Their results appeared thus to “confirm the existence of a ‘genre fingerprint’ that shapes the dramatic structure of tragedies and comedies” (Szemes and Vida, 2024, 175), and that playwrights may (un)consciously choose to adhere to or depart from. The algorithm was

20 For a similar, exploratory attempt on all DraCor corpora see Giovannini and Skorinkin (2025).

21 Their features cover nine different character-related aspects: text statistics, network measures, stage presence, semantic fields, topics, action verbs, sentiment scores, character information and drama information.

22 The metrics Szemes and Vida (2024, 171–175) employed were average clustering coefficient, density, average path length, diameter, maximum betweenness, average degree – maximum degree ratio, maximum degree – number of characters ratio, speech distribution (high/medium/low), weighted degree distribution (high/medium/low), average character speech, average character per scene, number of connected components. Some of these measures will be reused and explained here as well.

also rerun removing the plays' last act, which is generally treated as a strong genre predictor,²³ but the variation in results was limited.

In Giovannini and Skorinkin (2024a), we reprised Szemes and Vida's method and fine-tuned it in order to explore the features of opera libretti. In this study, we were not interested in a pure classification task, i.e. sorting out libretti from non-libretti, but we aimed at assessing whether libretti possessed some peculiar formal properties and how they evolved across the centuries. While results of this early experiment were not conclusive, confirming known challenges of modelling dramatic texts through structural features *only*, they also revealed some formal patterns which were typical of libretti, especially those with a non-comic background. Even more importantly, that paper represented a test ground for workflows and methodologies later employed here, especially in terms of feature selection protocols and visualisation workflows.

Before moving to describe the actual composition of my vectors, however, a brief reflection on the epistemic validity of this method is warranted. Compressing texts into a series of numerical values purported to represent some of their salient aspects, and then comparing these 'bags-of-numbers' to make assertions about literary history, could be easily called out as another example of the – at the same time – hubristic and reductionist mindset often attributed to digital humanities.

23 Comedies often end with a wedding where all the characters are present, thus making the network denser, while tragedies often end with the death of many characters, originating a less dense network. This observation is often traced back to Bernard Shaw, who did however complain about this simplistic view: "Now neither tragedies nor comedies can be produced according to a prescription which gives only the last moments of the last act. Shakespear [sic] did not make *Hamlet* out of its final butchery, nor *Twelfth Night* out of its final matrimony" (Shaw, 1921, xciv). The first evidence of this *topos* I could find is in Byron's *Don Juan*: "All tragedies are finish'd by death, / All comedies are ended by a marriage" (III, 9).

I would argue, however, that drama vectorisation²⁴ is just another expression of the ‘cultural technique of flattening’ German philosopher Sybille Krämer assumes as foundational for DH endeavours (Krämer, 2023a; see also Krämer, 2023b), and which is actually shared both by ‘standard’ humanistic inquiry and modern computational approaches. According to her, the opposition between traditional modes of scholarly interpretation, represented by ‘deep’ hermeneutics, and data-driven, ‘shallow’ DH endeavours is indeed misconstrued, given that humanities have always been using ‘superficial’ representation of cultural items as their research objects:

Historically, modes of working in the Humanities emerged in close liaison with operations involving artificial flatness. What is represented on a surface always forms an empirical, and therefore countable, fact. Thus, the number has always played an essential role in the generation and representation of knowledge in the Humanities: Concordances, library signatures, catalogs of works, historical timelines and dating tables, and diagrammatic inscriptions of all kinds bear witness to this. (Krämer, 2023a, 15–16)

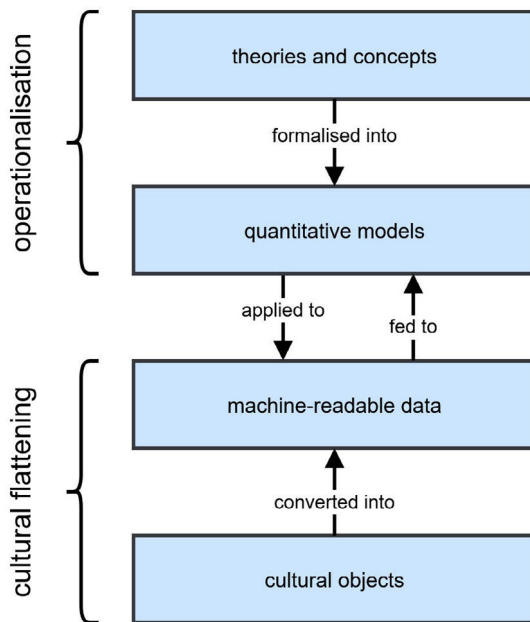
Consequently, computational methods differ from traditional methods only in the extension of their ‘flattening’ procedure, i.e. in the way in which they transform cultural objects into machine-actionable data. Following the model of the natural sciences, however, they also require thorough documentation of modelling choices as a standard practice, and this attitude helps the critical re-evaluation of assumptions that often remain hidden in non-digital hermeneutics.²⁵

24 I use here the term in its broader sense, referring to “the process by which objects are represented as vectors” (Dobson, 2021, 82). This should not be confused with its more common meaning in machine learning, where it refers to accelerating code execution by performing computations on a set of values simultaneously rather than iterating through individual elements.

25 “The Digital Humanities can be understood as an extrapolation and radicalization of the productive dimensions of the ‘cultural technique of flattening’. Its critical role towards the traditional Humanities may consist in making explicit

On an abstract level, one could even argue that the cultural flattening described by Krämer is a necessary methodological counterpart to another core DH practice, operationalisation: as tentatively schematised in Figure 4, these two practices converge in providing computational scholars respectively with the data and the interpretative models they need.

Figure 4: A minimal flowchart outlining two interlaced DH practices: cultural flattening and operationalisation.



just that kind of ‘surface orientation’ which – mostly concealed by the ‘depth rhetoric’ of hermeneutics – has always formed a genuine dimension in scholarly work (Kirschenbaum, *The Remaking of Reading*), but which has mostly remained a blind spot in Humanities’ self-understanding” (Krämer, 2023a, 8–9).

Within this perspective, drama vectorisation should be interpreted not as a reductionist attempt to overtly simplify the complexity of the-
atrical texts, but rather as a continuation of formalist morphological
thinking with modern computational methods. The ‘flattening’ of the
plays’ structural properties into vectors thus opens up new possibilities
of inquiry both in the exploratory and confirmatory research paradigms
(Gius and Jacke, 2022), without removing the need for ‘scalable’ readings
to make sense of the patterns the vectors show.

A typology of features

The feature vectors I built to represent EmDraCor plays include several
components from different sources. Most of them derive from the Dra-
Cor API,²⁶ whose endpoints allow one to access 96 features distributed
across four levels: corpus, play, segment, and character. At the play level,
it is possible to access 43 unique features²⁷ containing metadata and
structural information (cf. Börner and Trilcke, 2023, 37–41). Out of these
features, about half are non-numeric or not related to the text in itself
(e.g. composition/performance dates), and thus were excluded from my
vectors.

Additionally, I enriched the play embeddings with some further cen-
trality measures (computed on the basis of data from the API) and with
metrics developed in the context of three significant CLS studies I al-
ready mentioned (Algee-Hewitt, 2017, Szemes and Vida, 2024, Trilcke et
al., 2017). I ended up with 40 features; four additional, non-numeric de-
scriptors (slug,²⁸ language, genre, and year as string) were retained for

26 <https://dracor.org/doc/api>.

27 Those are the features one can see for each play in the corpus-wide metadata
table downloadable for each corpus. The API ontology actually names 59 play-
level features, but some of them are excluded from the metadata table, e.g. due
to redundancy (such as the English translations of play title and author name).

28 A slug is the last part of a clean URL, akin to a filename. In DraCor, it consti-
tutes a human-readable, unique identifier for a play and usually consists of the
author’s surname and the shortened title (e.g., *marlowe-doctor-faustus*).

identification purposes but excluded from the actual computations performed with the vectors.

Relying directly on the DraCor API to obtain most features was convenient but also entailed accepting its algorithmic choices, which are not always immediately understandable. A case in point is the computation of the ‘average path length’ and ‘diameter’ metrics.²⁹ When a graph has multiple independent components (i.e. is disconnected), the functions used by the metrics service return averages over within-component paths only. The resulting values are computable but semantically ambiguous: an average path length of 2.3 for a three-component graph is not comparable to the same value for a fully connected one. Given this uncertainty, I removed these two features to avoid interpretive confusion, bringing the final tally to 38.

The initial choice of so many features was motivated by an admittedly inclusive approach, in the belief that collecting a higher number of features would help give the vector more ‘depth’, i.e. more informational power. Such a strategy closely mirrored that adopted by Krautter et al. (2023, §39): on one side, the “features [chosen] cover as broad a field as possible, in order to exploratively determine which information is actually relevant” for the research question, and thus “complementary features can be found, each of which contributes new information that is not covered by other features”. On the other side, the features picked are not novel, and their usage is grounded in previous theoretical and applied research, as illustrated in the following pages.

Nonetheless, this implementation deviates from standard machine learning procedures in dispensing with the feature selection phase, i.e. not removing features which are correlated (such as density and average number of characters per scene). This can be motivated by two reasons. On one hand, as in Szemes and Vida (2024, 173), most correlations between pairs of drama features do not necessarily mean the relationship between them is exclusive, i.e. “they cannot be said to describe the same phenomenon or phenomena [...] in a deterministic relationship”. On the other hand, some techniques I will later employ natively deal with

29 Cf. also <https://dlina.github.io/Subgraphs/>.

feature redundancy and multicollinearity, while others focus on the independent interpretation of individual features, where multicollinearity becomes less relevant. Accordingly, I chose to maintain such a “multidimensional approach when it comes to the quantitative analysis” of plays (Fischer and Skorinkin, 2021, 528).

Furthermore, given the rich tradition of previous computational research on drama, I do not believe, as Fischer et al. (2017) do, that one must develop additional measures in order to enhance computational understanding of drama. Rather, I advocate for a rational and well-informed reuse of metrics already established, and consider developing new metrics only if the informational power of the existing ones proves insufficient. Nonetheless, I concur with Fischer and colleagues in considering “decisive to hold interpretations of [any] measur[e] against the backdrop of historical drama poetics” (4) – i.e. to employ metrics which are meaningful for one’s specific research object (in this case, early modern drama).

The EmDraCor vectors contain a diverse selection of features, which can be roughly grouped into four thematic areas: network data, cast features and speech distribution data, size-related data, and data linked to plot dynamics. Altogether, these metrics contribute to an understanding of dramatic form, insofar as they concretely operationalise aspects of the three ‘elementary components’ of drama previously introduced. Network metrics and cast features, for example, address character relations and configurations; speech distribution data quantifies the structure of dialogue; plot dynamics approximate aspects of *mythos*.

As already mentioned, most metrics were obtained directly from the DraCor metrics service, while some supplementary measures are also calculated on data from the API itself. Furthermore, a significant amount of work was dedicated to re-computing metrics from third-party studies, which in most cases meant “developing reverse-engineered code that would functionally approximate the analysis that the author[s] must have performed” (Schöch, 2023, 392, *italics mine*). Indeed, the original scripts were either written in other programming

languages or not available at all,³⁰ and thus some approximation could not be avoided.³¹

Detailed ‘model cards’ explaining each feature’s source, meaning, and computation process can be found in Appendix C. For the reader’s convenience, a list of the vector features is also provided in Table 1.

Beyond classifying them according to their category (network, cast, size or plot), vector metrics can also be systematised according to other categories. More specifically, they can be arranged on two axes – one which goes from *static* to *dynamic* features, and another one that goes from *global* to *local* features. Not all features, of course, fit neatly on these axes; this taxonomy is proposed here as an interpretive aid rather than a rigid schema.

First, a distinction can be introduced between metrics which immediately express a property of the whole play (e.g. its number of female characters), and are thus static, and metrics which attempt instead to capture the work’s temporal dimension, usually by reflecting developments throughout the plot. While in principle all features can be sampled repeatedly to reflect, even if coarsely, variation across time (e.g. by computing the average number of any measurable component per scene), some metrics have been designed with this specific purpose in mind. A notable example in this sense is represented by the already mentioned

30 Plot measures developed in Trilcke et al. (2017) are not directly retrievable through API calls since they are dynamically generated via JavaScript to be displayed in the ‘Speech distribution’ tab for each play. Szemes and Vida’s (2024) code is written in R and available on *GitHub*: https://github.com/SzemesBotton/d/drama_cluster_genre. Algee-Hewitt’s (2017) script was written in R but is not publicly deposited.

31 For example, in a few plays the computation of some metrics failed, mainly due to API call issues resulting in null values. Since the play embeddings needed to have the same dimensions, I had to fill these missing values. Although there is no optimal solution for this, I adopted a common approach in data science by using the mean value for that metric within the play’s own subcorpus. To make an example: if the metric ‘mediumSpeech’ could not be computed for Baron’s *Mirza*, I filled it with the average ‘mediumSpeech’ value from the English subcorpus. All these changes are thoroughly documented in the code used for processing the data.

studies by Trilcke et al. (2017) and Fischer et al. (2017), which exploited network theory measures in order to investigate plot development, modelling the “[t]he structure of a literary text [...] as a changing sequence of network states” (Fischer et al., 2017, 1).

Table 1: Vector components, grouped by type.

Type	Features
Network	Size; Average clustering coefficient; Density; Average degree; Maximum degree; Average betweenness; Average closeness; Average eigenvector centrality; Number of edges; Number of connected components; Ratio, average degree to maximum degree; Ratio, maximum degree to number of characters; Weighted degree distribution (high/medium/low); Protagonism; Mediatedness
Cast and speech	Average characters per scene; Average length of character speech; Speech intensity distribution (high/medium/low); Number of speakers; Number of speakers by gender (male/female/unknown); Number of collective speakers
Size	Number of acts; Number of segments; Word count, whole text; Word count, spoken text; Word count, stage directions; Number of prose lines; Number of verse lines
Plot	All-in index; Final scene size; Drama change rate

Secondly, metrics can be classified based on the levels of scale they encompass. While some of them describe a ‘local’ component in the play, such as a character or a scene, others directly relate to the play as a whole. Network metrics, for instance, include both types: centrality measures mostly profile individuals in their relationship with the rest of the cast, while clustering measures, by contrast, offer a broader picture by describing all-encompassing features like group cohesion, information dispersal, etc.

Ultimately, the taxonomies outlined in the previous paragraphs underscore the heterogeneity of the metrics collected. This is a key feature in the design of the EmDraCor embeddings and finds its inspiration, once again, in the work of Yarkho – who argued that quantifying literary form requires “deriv[ing] consolidated (mean) indicators for all features under analysis, regardless of their heterogeneity (as it is impossible to prove, for example, that the percentage of dialogues is a feature more noticeable or, alternatively, less noticeable than the number of scenes) ” (Gasparov, 2017, 144). Accordingly, the vectors I created aim to capture dramatic form as broadly as possible, combining metrics of different types and scales precisely because no feature alone can serve as a sufficient proxy for a whole play.

Experiments

The operationalisation of drama proposed in the previous chapter, consisting in the vectorisation of plays according to their formal features, aims at enabling meaningful and systematic comparison across diverse corpora. This serves two complementary goals – providing an empirical framework for testing Moretti’s thesis of progressive diversification in European early modern drama while addressing the broader question of “how dramatic forms change and are preserved within a framework of cultural evolution” (Szemes and Vida, 2024, 18).

To reach these aims, different strategies can be employed. In my view, the simplest method would involve measuring the degree of similarity or dissimilarity between play embeddings, since it can be argued that two texts whose vectors are highly different from each other will also be different in terms of dramatic form. One implementation in this sense entails directly computing distances between coeval plays from different linguistic traditions, and then assessing whether such values are increasing across time. Alternatively, one could also use dimensionality reduction techniques to project the play embedding on a Cartesian plane and then look for clustering by comparing snapshots from consecutive time frames.

In the following pages, I experiment with both approaches, testing their hermeneutical power while pointing out their limitations, before eventually focusing on a more traditional analysis of the development of individual metrics. While exploratory by design, these experiments yield some first empirical insights into data, thus revealing whether the branching of dramatic traditions theorised by Moretti is mirrored

through a ‘progressive distancing’ of the plays’ vectors, or what other dynamics are at play.

Distances

Data scientists and, by derivation, digital humanists and cultural analytics scholars have plenty of options at their disposal for measuring similarity or distance among any pair of vectors.¹ The choice of the right method, however, depends first and foremost on the properties of the vectors themselves. Since, as Nan Z. Da polemically argued, “it is good practice to ask users of CLS to show their vectors”, because “it demystifies much of the process and often reveals conceptual weaknesses” (Da, 2019, 622), let us observe the composition of a sample vector from the EmDraCor corpus, representing Ruiz de Alarcón’s play *La verdad sospechosa* (1619):

$V = \{41, 3, 14, 11, 3, 0, 0, 1, 3306, 16919, 16461, 473, 6.571428571428571, 0.505494505494505, 0.7699649556792411, 14, 11, 1, 46, 0.0494505494505494, 0.6608466660481829, 0.2431203018946239, 0.926829268292683, 0.5714285714285714, 0.6391666666666669, 3.0, 0.5974025974025974, 0.8461538461538461, 0.3571428571428571, 0.2857142857142857, 0.3571428571428571, 4.99, 0.3571428571428571, 0.2857142857142857, 0.3571428571428571, 0.0175417232479634, 0.5493558858726955\}$

Within this ‘bag-of-numbers’, one immediately notices a sizeable difference in the values’ magnitudes, ranging from small (0, 1, 2) to very large (16919), with many of them being long decimals. As statistical practice prescribes, I applied min-max scaling² to normalise them in the range

-
- 1 In their survey, Choi et al. (2010) list 76 such measures employed across different scientific fields.
 - 2 Min-max scaling was preferred over z-score standardisation for three reasons. First, the play vectors combine heterogeneous metrics with different distributions (some skewed, some bounded), while z-score assumes each feature is roughly normally distributed. Second, rescaling to a fixed [0,1] range en-

from 0 to 1. At the same time, some play embeddings contained null values, indicating that not all features were present. However, since the number of zero values across all metrics in the corpus is relatively low (less than 5%), one can safely consider the EmDraCor vectors as substantially populated across all dimensions.

The vectors' characteristics, such as their relative denseness, guided the choice of the appropriate measure for computing the distances between them. Cosine distance was excluded because, by measuring only the angle between vectors in a scale-invariant fashion, it disregards magnitude differences that could be meaningful. Manhattan distance, for its part, is more suited to high-dimensional spaces, but the EmDraCor vectors only contain 38 dimensions; furthermore, it is less responsive to sharp differences on individual features, which are crucial for assessing the diversification hypothesis. Given the density, low dimensionality, and normalised scale of the play embeddings, I decided to adopt Euclidean distance³ as the most appropriate metric, insofar as it requires no assumptions about distribution and remains reliable on this scale.

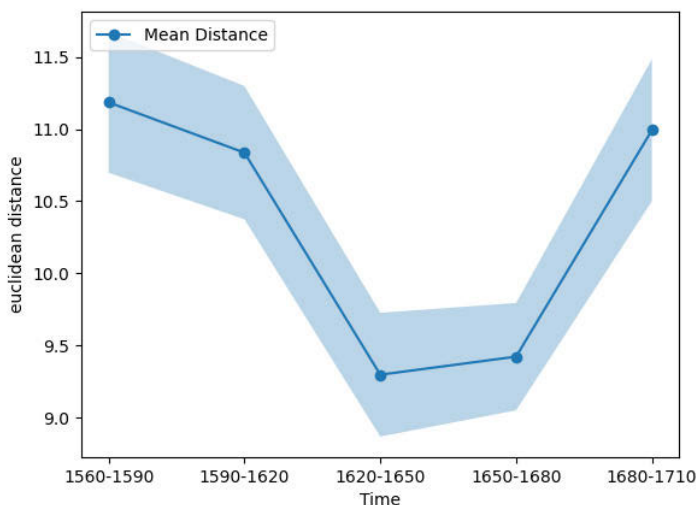
To test whether a general trend towards formal diversification could be recognised across the whole corpus, I first performed an experiment based on the repeated measurement of vector distances. To do so, I divided the 150 vector-plays into groups according to their normalised year; within each group, I calculated the Euclidean distances between each pair of plays. Then, I used the average of these distances to represent the 'formal difference' within each time frame, and plotted these averages to visually observe how they changed over time.

tures that no single feature dominates Euclidean distance calculations, while also yielding more interpretable values for a non-technical readership. Third, z-score standardisation tends to compress outlying values, which in this context may represent meaningful formal features rather than noise. On the other hand, however, min-max scaling is very sensitive to extreme values, as a single anomalous data point can compress the rest of the distribution.

- 3 This type of distance can be described as the length of a segment connecting two points on a Cartesian plane. It can be computed with the formula $d = \sqrt{(v_2 - v_1) \cdot (v_2 - v_1)}$ – where v_1 and v_2 represent any given vector (cf. Grootendorst, 2021).

Identifying the most suitable time sequences for visualising evolution would be worth another book in itself. While many DH studies use the century “as the typical yardstick of quantitative cultural history” (Moretti and Sobchuk, 2019, 90), there is no definitive answer to what Beausang (2022, 646) calls a “deceptively straightforward question”, i.e. “at what scale does literature change?”. In this context, and given the corpus’ relatively short temporal arc, I employed fixed thirty-year-long time spans – a frame which corresponded to that of the ‘generational cycles of literary evolution’ postulated by Moretti himself in *Graphs, Maps, and Trees* (Moretti, 2007, 20–21).

Figure 5: Mean Euclidean distances between groups of coeval plays (30-year time frame, fixed windows), with confidence interval.



The results of this implementation, shown in Figure 5, describe a different narrative from that put forward in ‘Modern European Literature’. Instead of a steady increase in diversification from the beginning of the seventeenth century, the data suggest a different trend, where a clear

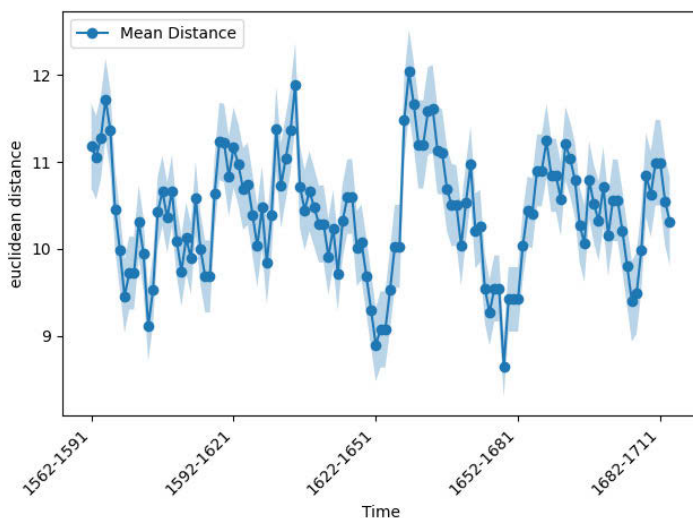
drop in formal differences in the first half of the century – a period of formal convergence, one might say – is followed by a sharp increase in diversity toward the beginning of the next one.⁴

Interpreting time series, however, is a notoriously difficult task, especially when they emerge from limited datasets. A different implementation, using instead a sliding rolling window of thirty years (Figure 6), returns a considerably more complex picture, where periods of formal convergence and divergence keep alternating. This greater volatility reflects both the sensitivity of the technique to small changes – with relatively few data points per window, the addition or removal of a single distance can noticeably shift the mean – and the finer temporal resolution it provides. A fixed frame analysis, based on larger and non-overlapping groups, trades granularity for stability, and smooths such oscillations into a single U-shaped pattern; however, as previously noted, a *t*-test shows that the downward and upward slopes are statistically significant. Furthermore, the same pattern still emerges (even if less forcefully) when the frame is scaled down to a 15-year window.

Computing pairwise distance between play embeddings within the same time frames, however, is not the only way of measuring distances inside the corpus. A complementary approach involves calculating an average vector for each group of coeval works from the same linguistic area and considering it a proxy for the ‘position’ of a given dramatic culture within a period. Then, fluctuations in the distance between such ‘centroids’ can be measured to assess diachronic divergence in formal features. This approach is not unprecedented in computational drama analysis: Ilseman (1995, 17) followed an analogous path in “establish[ing], if not invent[ing], the shape of the average play on the basis of [values from] the plays already examined”, so that “deviations from the model could be made visible”. By collapsing plays into averages, however, this method trades individual variation for a clearer view of aggregate trajectories.

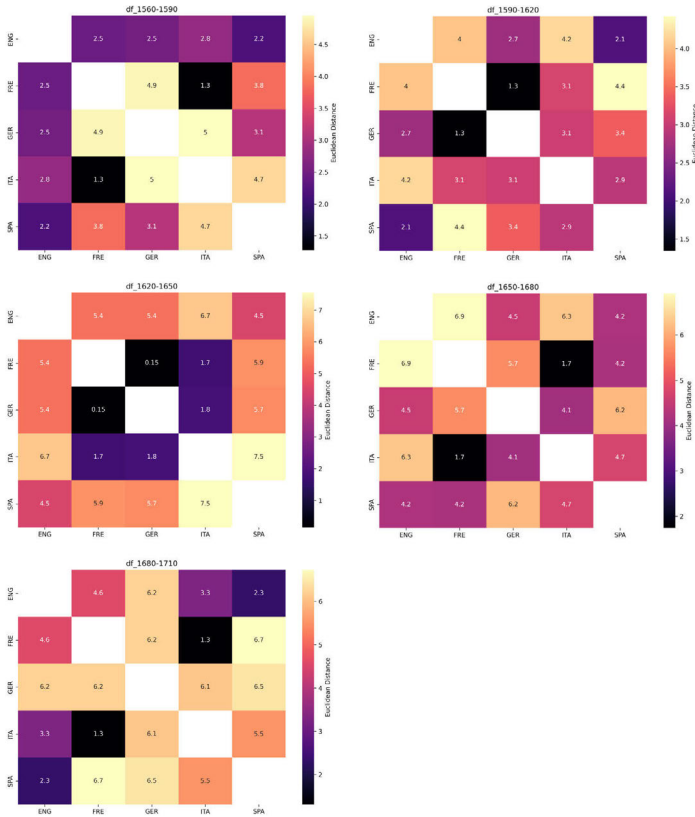
4 The reliability of the two slopes (but not of the plateaus) is confirmed by Student’s *t*-test (which determines whether the difference between the means of two groups is statistically significant) and visually displayed by the relatively small confidence intervals around them.

Figure 6: Mean Euclidean distances between groups of coeval plays (30-year time frame, rolling window), with confidence interval.



If Moretti's hypothesis holds, one would expect centroid distances between traditions to increase over time, and the heat maps presented in Figure 7 partially support this interpretation. Unlike the pairwise analysis, which revealed a U-shaped pattern, centroid distances suggest a more consistent divergence. Looking at the matrices, a general shift from darker to lighter tones, reflecting a movement from similarity to dissimilarity, can be noticed. At the same time, the graphs highlight known relations between theatrical cultures, such as the continued formal closeness between the Italian and French traditions (the 'regular' theatres in Lochert's reconstruction).

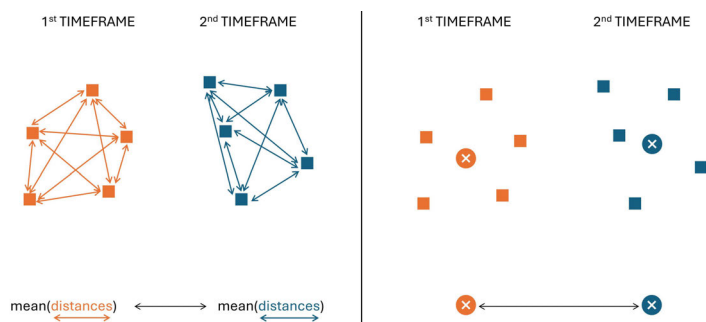
Figure 7: Heat maps showing distances between ‘centroids’ for different linguistic groups at different chronological stages (30-year time frames).



The contrasting results from the previous experiments can be explained by considering the different logic underlying each, which are summarised in Figure 8. In the first case, the iterative computation of pairwise distance between any two vectors returns a list of distances, whose mean is then compared across time frames. Being based on the repeated comparison of individual works, this procedure could be seen

as more granular but also more sensitive to outliers. In the second case, one identifies average values for groups, creating synthetic vectors for any spatio-temporal context, and then measures their distance. Each method captures a different aspect of the same data, but neither alone is decisive in supporting or refuting Moretti's claims, suggesting the need for another approach.

Figure 8: Visualising the logic behind two options for computing vector distances: pairwise (left) and centroid-based (right). Each dot represents a play embedding.



Clusters

As shown, preliminary results from the exploration of distances between plays paint a more complex picture of drama evolution than the progressive diversification postulated by Moretti. To examine the question from another perspective, I decided to switch to a graphical representation of the data, plotting the vectors on a Cartesian plane and tracing their clustering in successive temporal snapshots. The reasoning behind this strategy was straightforward: if a diversification of European drama had indeed taken place across the early modern era, this should be signalled by the progressive emergence of clusters organised by linguistic tradi-

To plot multi-feature vectors on a two-dimensional space, however, one first needs to apply an appropriate algorithm for scaling down the data. Machine learning research has developed several algorithms for dimensionality reduction (cf. Ayesha et al., 2020; Van der Maaten et al., 2009), which can be classified according to various criteria.⁵ More generally, an overarching distinction is often formulated between methods trying to preserve the general structure of the data and those focusing on preserving its local geometry. While any compression of high-dimensional vectors invariably leads to a loss of informational accuracy, it also allows for an intuitive, visual grasp of how data points relate to each other.

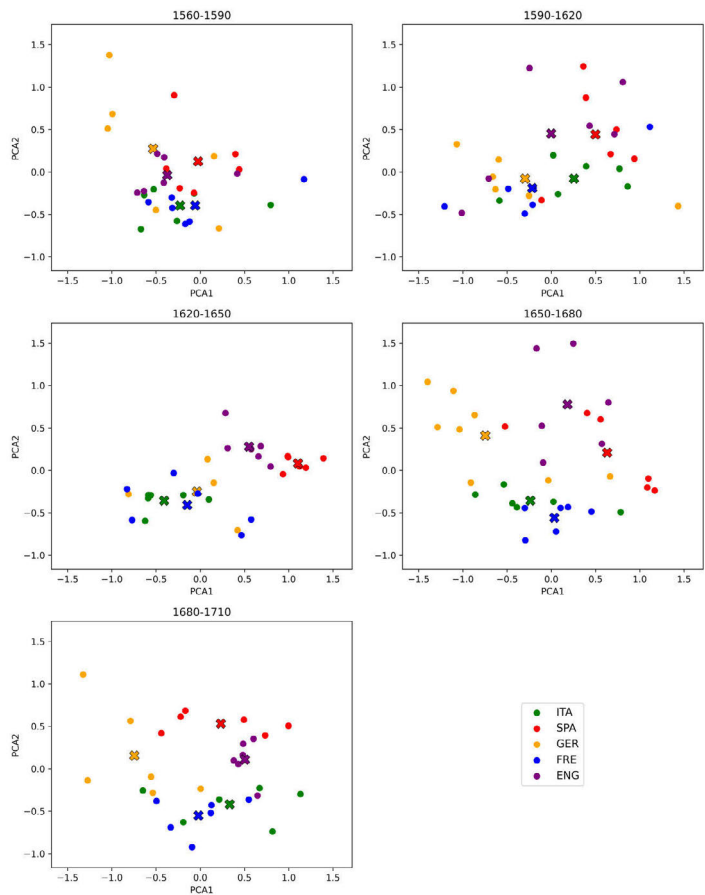
In the context of computational literary studies, Principal Components Analysis (PCA, see Lever et al., 2017) is often employed as the standard technique for dimension reduction. This preference likely stems from its early adoption in the subfield of computational stylistics, as testified already in work by John Burrows and others (Beausang, 2020, 185 ff.), from which it later entered the CLS mainstream. The relative ease in implementing PCA makes it arguably the most popular option for scholars looking to build “a low-dimensional representation of the data that describes as much of the variance in the data as possible” (Van der Maaten et al., 2009, 3), even though issues in its correct application by DH scholars have been noted (see Leufkens, 2020).

Despite also running some tests with other unsupervised techniques such as t-SNE (van der Maaten and Hinton, 2008) and UMAP (McInnes et al., 2018), I ultimately chose to employ PCA not only because of its straightforward implementation and interpretability of results, but also because of its better performance in preserving global structures – and thus unveiling patterns of divergence or convergence. In my implementation, I used 30-year time frames as observation windows, thus plotting five different snapshots from my corpus: 1560–90, 1590–1620,

5 Anowar et al. (2021) individuates three main criteria: the linearity of the scaling process (linear/non-linear), the existence of predetermined class labels (supervised/non-supervised), and the type of mapping (random projection-/manifold-based).

1620–1650, 1650–1680, 1680–1710. Data points were marked according to the dramatic tradition they belong to, while centroids for each group were added as well.

Figure 9: *EmDraCor* play embeddings projected on a Cartesian plane through PCA and grouped into 30-year timeframes, with centroids.



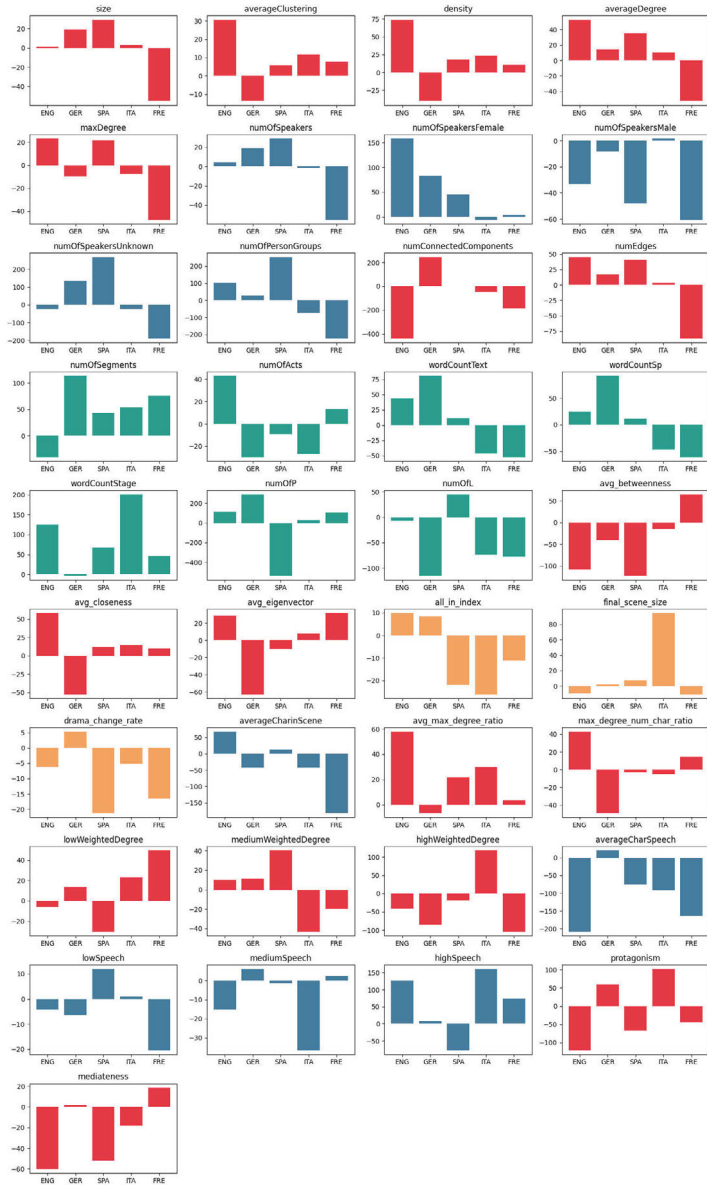
Results presented in Figure 9 partially support, once again, a narrative of diversification. While the extent of the vectors' dispersion across the PCA space remains relatively stable in the five plots, rough clusters of plays begin to appear, especially from 1650 onwards. This stands out especially when comparing the first graph, where plays from all regions cluster in the central area, and the last one, where the zones occupied by different dramatic traditions almost do not overlap – with the notable exception of French and Italian, whose formal proximity, shaped by a common adherence to neo-Aristotelian poetics, has already been noted in the heat maps.

Furthermore, inspecting the feature loadings of the two components plotted (which together account for almost 50% of the variance) shows which dimensions mostly contribute to positioning the play embeddings across the PCA space. While the first one (PCA1) loads mostly on network *density* features, such as density (19.2%), closeness (13.2%), and the ratio between average and maximum degree (13.7%), the second one (PCA2) relies on network *size* features, such as number of edges (11.2%), average of characters per scene (11.5%), and maximum degree (10.2%). This suggests, with some approximation, that the two main axes of formal variation within the corpus are network concentration and network scale.

Patterns

The evidence presented so far captures some variation of early modern drama at a formal level, but it still does not reveal a trend as clear as the one hypothesised in 'Modern European Literature'. While vector distances calculated using different approaches (pairwise or centroid-based) tell divergent stories, examining the clustering of the play embeddings via PCA seems to confirm a tendency towards diversification, but the pattern is not pronounced enough to be considered conclusive.

Figure 10: Percentual variation (from 1561 to 1710) for the value of each vector component across the five linguistic traditions.



Another approach, however, is possible. Instead of performing operations on the vectors as a whole, one could decompose them again into their constituent metrics and then examine their individual diachronic evolution. In other words, even if it is not possible to pinpoint a global trajectory for the whole corpus, it might be possible to single out which features are distinctive of a given literary tradition, i.e. which formal aspects are most decisive in distinguishing it from the other coeval ones.

An intuitive approach would involve plotting the values for a given metric after arranging the texts in each subcorpus in chronological order. Applying this method systematically, one can then compute shifts in absolute value for each metric across the time span investigated, and thus verify whether the variation of a given feature has actually become highly distinctive of a national tradition against the others. While I could have calculated these shifts directly, measuring the change between each pair of adjacent values, a few metrics had null values, thus creating spurious slopes.⁶ Therefore, I fitted a linear regression to each metric's values over time and plotted the resulting rate of change, expressed as a percentage, for each tradition (Figure 10).

The risks of such a simplifying approach are well known and have been eloquently outlined by Moretti and Sobchuk (2019): trends obscure the complexity of data insofar as they “make data easy to read” and “give them a meaning”, while reducing “multiple trajectories to a single big story” (93–94). Even when trends are “unquestionably there”, and “quite uniform and gradual [...] this doesn't necessarily mean that the forces behind them are equally uniform” (97).

As far as the EmDraCor corpus is concerned, the best antidote to any overinterpretation of trendlines lies in a thorough analysis of the data behind any of them. Some slopes, for example, appear skewed by the presence of outliers: the values for mediatedness in the French subset remain mostly stable during the 1561–1710 time frame, but a couple of

6 E.g. a play without any stage direction, and thus with ‘wordCountStage’ = 0.0. This happens in about 6% of the overall values.

peculiar early works⁷ creates the appearance of a slight upward trend; a similar argument could be made also for the plot-related measure ‘all-in index’ in the Italian subcorpus.

Similarly, the distribution of the texts in terms of genre can lead to some red herrings. The Italian subcorpus, for example, seemingly displays a notable growth in the final scene size values, but by observing its composition one notices how this is likely due to the random concentration of comedies (known to feature large ensemble casts in the last scenes) in the later decades.

Instances like this suggest that each percent variation needs to be double-checked with the actual trend lines and the corpus composition before it can be used to support broader arguments about literary history. Furthermore, not all the shifts in the features’ evolution can be taken as equally significant: changes in the number of prose lines (‘numOfP’) and verse lines (‘numOfL’), for example, are not very telling, since the two modes are most often mutually exclusive (i.e. a play would usually be either fully in prose or fully in verse).

While keeping these caveats in mind, however, I believe tracking the evolution of individual metrics across EmDraCor still represents a useful first step towards a quantitative profiling of dramatic traditions. Such findings are preliminary, constrained by the size and composition of the corpus; once the data landscape improves and bigger multilingual corpora become available, it will be possible to re-run these experiments on them, following standard reproducibility practices. In the meantime, the following pages collect and briefly discuss relevant features marking the autonomous development of each tradition – in a necessarily rhapsodical manner, moving freely across the most relevant metrics as the data warrants.

7 Both Bienvenu’s *Comédie du monde malade et mal pansé* (1568) and Virey’s *La Machabée* (1596) have fully-connected character networks, where every node is equally central to all connections, and because every shortest path is direct, no node serves as an intermediary on any shortest path. Thus, each node has a betweenness centrality of 0, and this translates into a null mediatedness value.

English drama

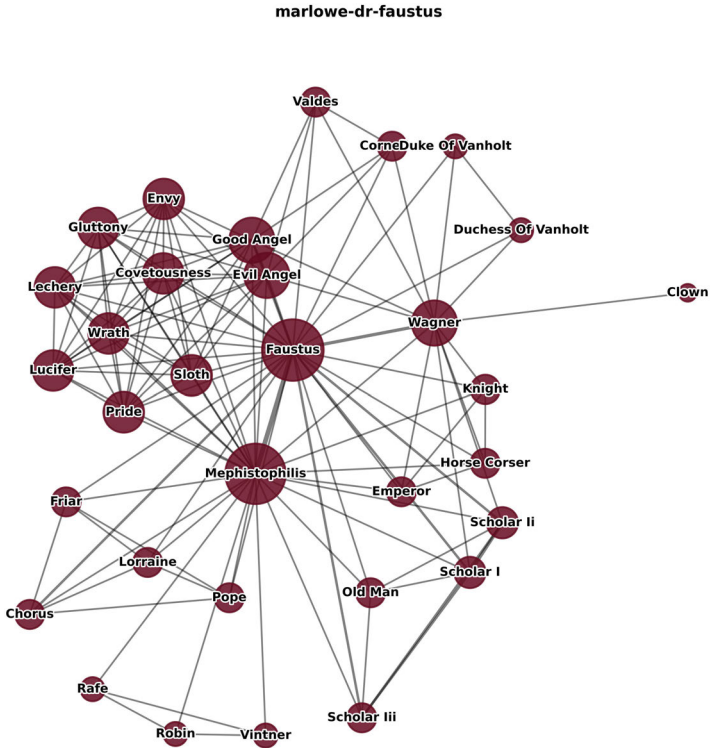
The English tradition displays the most significant metric variations across EmDraCor, suggesting that its theatrical culture was perhaps the one to develop the strongest formal autonomy during the time frame considered. This plays quite aptly into the narrative of the ‘exceptionalism’ of English drama, which has often been tied to the country’s peculiar geography:

Would there be Shakespeare, had England not been an island? Who knows? But that the greatest novelties of tragic form should arise away from the mainland, and from someone with ‘small Latin and less Greek’, is quite a sign of what European literature had to gain from losing its unity, and forgetting its past. (Moretti, 2013a, 13)

In terms of network structures, one notices immediately the parallel spike in average clustering and density, suggesting an enhanced connectivity both at a local (clusters) and general (network) level. The tightening of relations is also confirmed by the fact that the plays’ size remains broadly unchanged. Furthermore, in keeping with the findings of Algee-Hewitt (2017), protagonism and mediatedness also drop across the time span.

Such shifts in network structure can be easily grasped by comparing play networks from the two chronological extremes of the corpus. Marlowe’s *Doctor Faustus* (1588, Figure 11) is a typical example of diffuse Elizabethan tragedy, where the central character ties together different groups that would otherwise not interact. In this case, the eponymous scholar – embedded in a high-density clique with his demonic aide Mephistophilis and his servant Wagner – acts as the bridge between clearly distinct subnetworks, such as the courts of the Pope and the Holy Roman Emperor, the Vanholt ducal household, or the magicians Cornelius and Valdes.

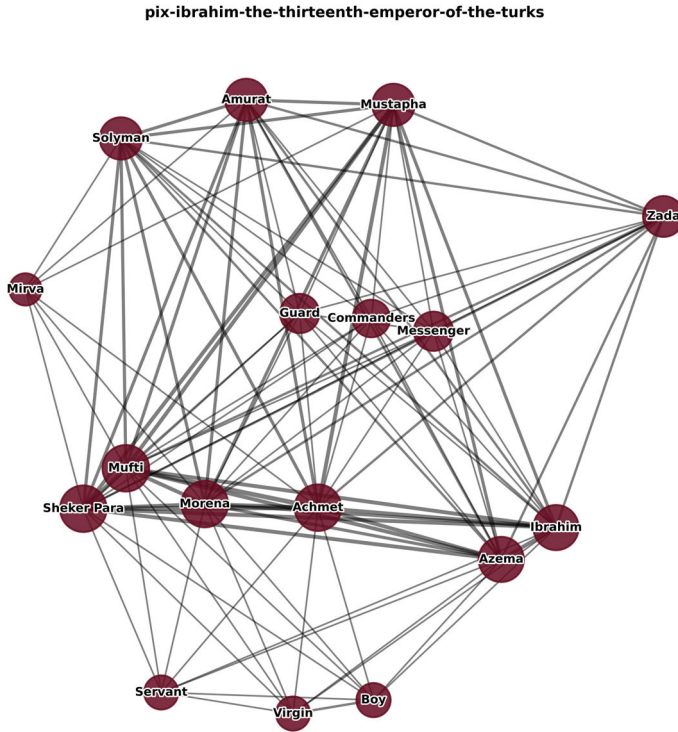
Figure 11: Co-presence character network of Marlowe's *Doctor Faustus*.



More than one century later, Pix's *Ibrahim* (1696, Figure 12) shows instead the 'distributed' features whose increase Algee-Hewitt tracked in this essay. While no single character stands out as a towering protagonist, a closely-knit group of central figures assumes such function. These include the titular character (the sultan Ibrahim), the villains (Sheker Para, Azema, and Achmet), their opponents (Solyman, Mustapha, and

Amurat), and the heroine Morena, whose scholarly-validated centrality in the plot economy⁸ is not fully reflected by her network position.

Figure 12: Co-presence character network of Pix's Ibrahim.



- 8 As Lechner (2016) argues, the treatment of the character of Morena (first violated by Ibrahim at the instigation of Sheker, then inciting a rebellion against the sultan, and ultimately emulating Lucretia in killing herself) signals the transition from the traditional heroic drama *à la* Dryden to the pathetic *she-tragedy* of the late seventeenth and early eighteenth century.

Furthermore, cast composition shifts significantly across the corpus' time frame: while the total number of characters remains roughly the same, the quota of females increases by a notable 80%, while that of males drops by 10%. This trend is strongly influenced by the plays of 1660–1680, where the number of female characters increases rapidly, peaking in 1675 with Thomas Shadwell's graphic rewriting of the *Don Juan* myth, *The Libertine*. In this play, the stage is indeed almost always occupied by some of the rake's female victims, who are brought forth both in groups (e.g. the nuns and shepherdesses he attempts to assault, or the "whole Batalion of couragious Women" – to quote his servant Jacomo – he tricked into believing they were his wives) and as individuals (Leonora, with her lady-in-waiting Maria and maid Flora; the sisters Clara and Flavia).

Such an increase can likely be explained by the groundbreaking introduction of actresses on stage starting in the 1660s (see Howe, 2001, ch. 1). After all, the choice of the *dramatis personae* was determined not only by poetics, but also by pragmatic reasons: since "most playwrights wrote their plays with a particular company in mind and fashioned roles to fit the specialties in that company" (Howe, 2001, 11), the sudden availability of female performers might have prompted playwrights to insert more female roles in the pieces.

At the same time, the little drop in the overall drama change rate suggests that only small cast changes kept taking place between scenes. One may speculate that this trend reflects the growing, but not uncontested influence of neoclassical French theatre, where the practice of the *liaison des scènes* required at least one character to always remain on stage and link different narrative segments. As noted already by Friedland (1911), the attempt to keep 'scenes unbroken' – alongside with the adoption of the three main unities of time, space, and action – was an important component in the reception of French theatrical models and was explicitly discussed by dramatists who tried to adhere – not without reservation – to such prescriptions.⁹

9 John Dryden, in the first prologue to *Secret Love; or, the Maiden Queen* (1688), assures he followed "the exactest rule, by which a play is wrought", including "the

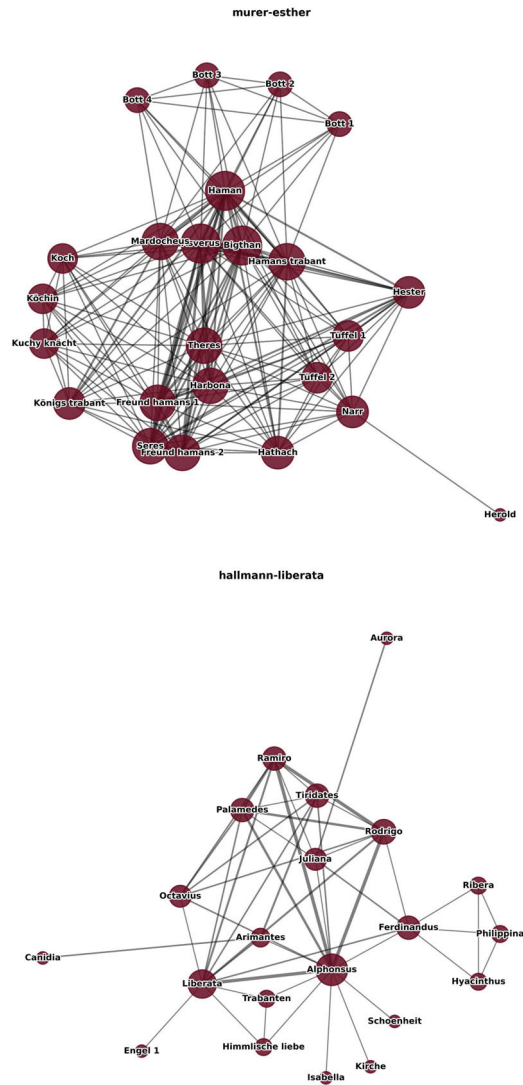
German drama

By contrast, German plays gradually shift towards sparser networks, as slight size increases are accompanied by decreases in density and average clustering — characteristics unique to this subcorpus. Other indicators of this trend include a rise in the number of independent subnetworks, which might also explain a reduction in centrality measures such as betweenness, closeness, and eigenvector centrality. A visual comparison between a play from the late sixteenth century and one from a century later may be useful to illustrate this point; Figure 13 contrasts Jos Murer's *Esther* (1567) and Johann Christian Hallmann's *Liberata* (1699).

While the former network features a compact structure, where high density likely reflects the intricate power dynamics within Ahasuerus' court, the latter exhibits a star-shaped pattern, where a core group of characters largely acts as the sole link between peripheral individuals. Furthermore, density is further reduced by the presence of individual characters who occupy distinct segments by themselves, frequently interacting with the real-world audience through humorous asides (e.g. by Pollio Asinius) or moralising tirades (e.g. by the personified Justice).

scenes unbroken", but also contrasts these arid prescriptions ("dead colours") to his own contribution, in terms of "wit" and "plot", which he characterised as "the living beauties of a play". By his part, Thomas Shadwell prefaces *The Sullen Lovers: or, The Impertinents* (1688) by saying he has "as often as [he] could naturally, kept the Scenes unbroken, which (though it be not so much practised, or so well understood, by the English) yet among the French-Poets is accompted a great Beauty". He later warns, however, "not to expect a very Correct Play", since "[i]t is so difficult a thing to write well in this kind". Both examples are mentioned by Friedland (1911, 286; 291).

Figure 13: Co-presence character networks of Murer's *Esther* (above) and Hallmann's *Liberata* (below).



At the same time, works become longer, mostly through an increase in lines spoken by characters, while the share of stage directions remains constant. This fact, however, cannot alone explain the strong rise in the number of segments, i.e. scenes, which could instead be attributed to the influence of French theatrical poetics. Following neo-classical precepts, it seems indeed that playwrights were prompted to create a new scene each time the character configuration changed, thus significantly increasing the number of scenes.¹⁰ This tendency reaches its apex in authors such as Christian Weise, whose *Ungleich und gleich gepaarte Liebes-Alliance* [*Unequally and equally matched love-alliance*] (1708) features 53 segments. Reading the play, one notices how every slight variation in character configuration comes to constitute a new *Auftritt*, with most scenes featuring only a couple of characters engaging in short, stichomythic exchanges.

Spanish drama

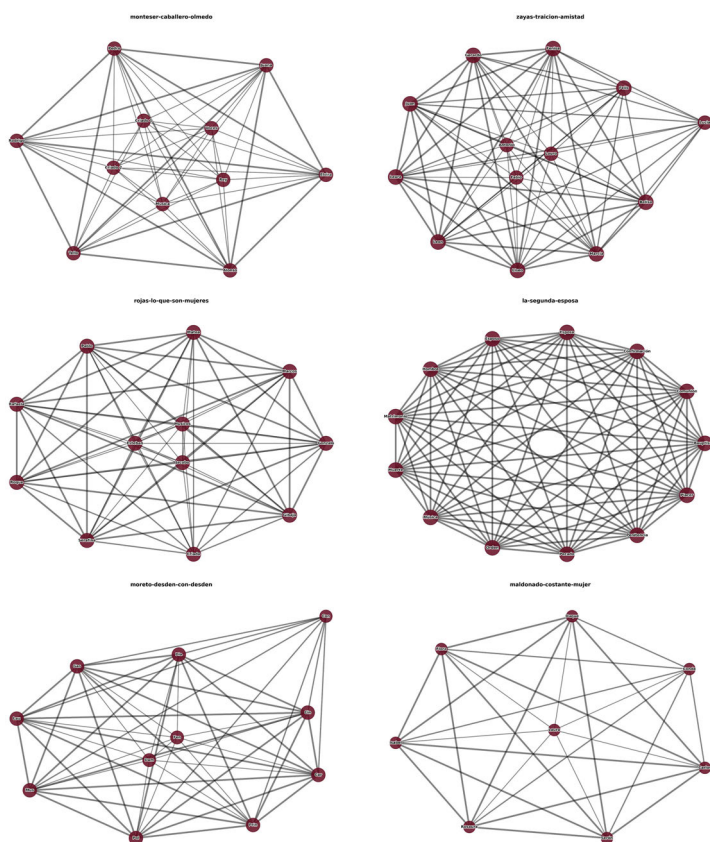
The Spanish subcorpus displays growth in all basic network measures (size, average clustering, density, average and maximum degree), indicating expanding and more densely connected dramatic networks. Furthermore, it shows an interesting pattern in terms of centrality values, with a drop in betweenness centrality and mediatedness. This is clearly visible in the structure of many Spanish character systems, especially from the 1620–1660 period, which take the shape of stable, often fully connected networks (Figure 14). In these examples, no character seems to act as a clear ‘mediator’ among groups; conversely, the different characters associate freely, forming complex constellations.

The absence of a central mediating figure is also underscored by the distribution of weighted degree as classed by Szemes and Vida (2024). While the number of characters with high and low weighted degree tends to diminish, characters in the middle tertile, i.e. having a

10 For a detailed reconstruction of the development of the concept of ‘scene’ in German drama and its related terminology (from *scena* to *Teil*, *Aufzug*, and finally *Auftritt*) see Jordan (1939, §2–3).

moderate number of connections, increase, pointing to networks where character centrality is more equally distributed.

Figure 14: Network structures of six 17th century Spanish plays. From the upper left: Monteser's *El caballero de Olmedo* (1621), Zayas' *La traición de la amistad* (1630), Rojas Zorrilla's *Lo que son mujeres* (1644), Calderón's *La segunda esposa* (1648), Moreto's *El desdén con el desdén* (1654), and Maldonado's, Dueña's and Cifuentes' *La más constante mujer* (1658).



Another factor setting the Spanish corpus apart from the others is the massive increase in ungendered speakers ('numOfSpeakersUnknown'). According to the corpus' encoding conventions, the 'unknown' tag marks two different categories: abstract or allegorical entities without a clear gender (e.g. 'Virtue'¹¹) and individuals/groups whose gender is not possible to determine. The first type features prominently, for example, in Juan Montenegro y Neira's *La toma de Buda* [*The Conquest of Buda*] (1687), where the stage is populated both by historical characters (Turks and Christians involved in the eponymous siege) and allegorical figures like Unity, Discord, and the Seven Deadly Vices.

The second case is exemplified by the many unspecified voices that populate plays like those by Antonio de Zamora, often characterised as collective speakers – thus contributing to the notable increase in the 'numOfSpeakersGroups' parameter as well. In his two works, *Ser fino y no parecerlo* [*Being refined and not looking like it*] (1701) and *Cada uno es linaje aparte y los Mazas de Aragón* [*Each one comes from a separate lineage, and the maces of Aragón*] (1708), Zamora makes ample use of non-specified individual or plural voices for dramatic effect, employing them to better convey the hectic atmosphere of battles and sieges.¹²

Finally, the decrease in the drama change rate, while moderate, could be explained by a progressive implementation of the *liaison des scènes*, as seen in the English corpus as well. However, one does not necessarily need to explain such practice through the exogenous influence of French classicism. Lope de Vega himself was indeed a proponent of this compositional strategy, as he recommended that "[v]ery seldom should the stage remain without someone speaking, because the crowd becomes restless in these intervals and the story spins itself out at great length;

11 While allegorical figures in early modern texts are often depicted under feminine appearance (see Warner, 1985), they still embody abstract concepts, and thus their gender has been consistently annotated throughout EmDraCor as 'unknown', unless it was culturally acquired (e.g. the Devil as a male character).

12 See, e.g., the third act of *Ser fino y no parecerlo*, where several unnamed Phoenician and Achaean warriors clash.

for, besides its being a great defect, the avoidance of it increases grace and artifice" (Vega, 1914; cf. Ehrlicher et al., 2020, 7).

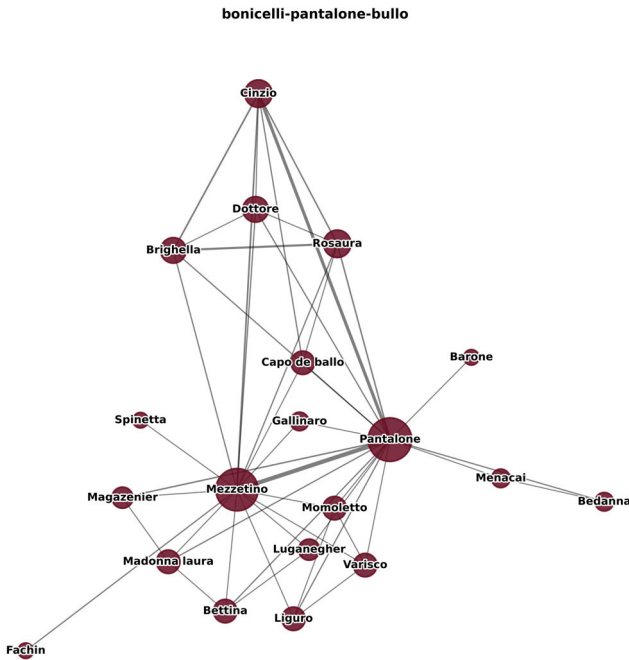
Italian drama

The Italian subcorpus, for its part, displays mostly moderate shifts in terms of network values, and generally exhibits less distinctive features compared to the others. Towards the end of the corpus, however, two metrics in particular show a sharp increase: the number of characters with a high weighted degree, i.e. with sustained interactions with other characters, and protagonism. Together, they suggest a pattern of progressive concentration of the protagonist role in one or few characters, in contrast with Algee-Hewitt (2017)'s findings on English early modern drama. Speech statistics support this reading, with an increase in the number of prolific speakers (reflected by the 'highSpeech' metric) at the expense of quieter figures.

A good example of this tendency is Giovanni Bonicelli's comedy, *Pantalone Bullo* [*Pantalone, the Bully*] (1688), where the relevance of the titular character – a vibrant rewriting of the stock type from the *commedia dell'arte* repertoire – cannot be overstated. Roaming across Venice with his gangs of thugs and causing mischief in taverns and shops, Pantalone is constantly in the foreground of the action together with his sidekick Mezzettino, as the network (Figure 15) shows. By contrast, the conventional, trite love subplot between Pantalone's daughter Rosaura and Cinzio is clearly confined to the periphery of the graph, and their interactions never become relevant in the network's overall economy.

Another interesting feature of Italian drama is the notable increase, especially towards the end of the seventeenth century, of stage directions. This might be explained with the presence of several works which, while not strictly belonging into the improvisation-based *commedia dell'arte*, still share its emphasis on the corporeal and performative component, and clearly foreground it in the written text. A clear example in this sense is Niccolò Amenta's *La somiglianza* [*The Resemblance*] (1706), where the author continuously includes stage directions (almost 200!) on how characters should come forth or interact with each other.

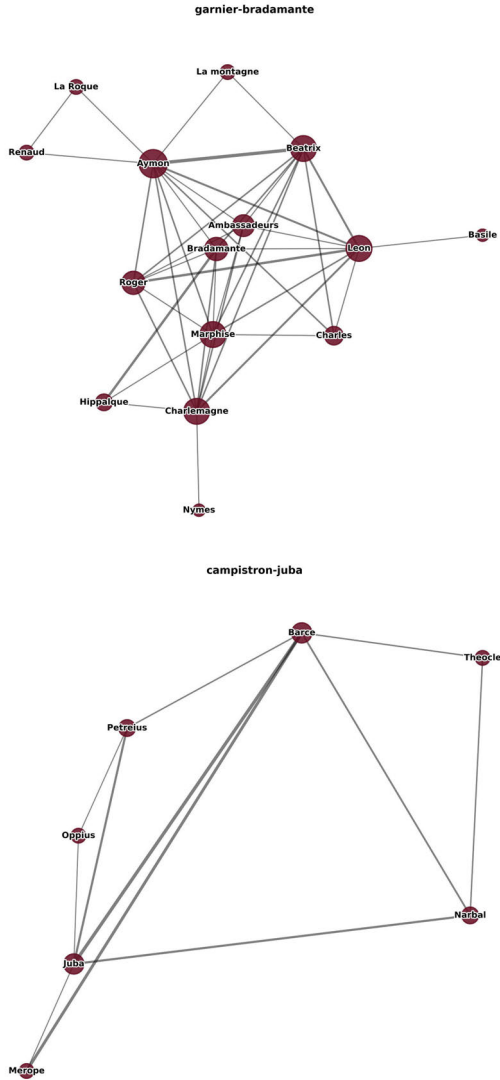
Figure 15: Co-presence character networks of Bonicelli's *Pantalone bullo*.



French drama

The most notable pattern found in the French subcorpus is a significant decrease in cast size, which does not affect all genders equally: while the percentage of female characters remains quite stable, the share of male and group characters decreases more sharply. This reduction produces leaner networks where, with fewer characters sharing the stage, all basic centrality measures (betweenness, closeness, eigenvector) rise mechanically: each figure is closer to all others, mediates more interactions, and is more directly connected to other central characters – even as the absolute number of connections and the number of edges decrease by virtue of the smaller casts.

Figure 16: Co-presence character networks of Garnier's Bradamante (above) and Campistron's Juba (below).



Although this trend can partly be explained by the presence of several one-act plays in the final decades of the corpus, it also reflects the extent to which the poetics of *classicisme* had progressively become ingrained in contemporary dramatic production, leading to more densely interconnected network structures. This shift can be seen clearly when comparing, for instance, the character network of an early play, such as Robert Garnier's *Bradamante* (1582), with that of the later tragedy by Jean-Galbert de Campistron, *Juba, Roi de Mauritanie* [*Juba, King of Mauritania*] (1685), as shown in Figure 16.

The evolution of French dramatic structures towards streamlined network models also affects the properties of the individual characters: since the cast is reduced, the proportion of prolific speakers increases (as testified by the increase in the 'highSpeech' value). At the same time, the decreasing drama change rate, suggesting more stability in stage configurations, is clearly linked to the authors' adherence to the *liaison des scènes* practice. The two phenomena can be observed again in the play by Campistron, a disciple of Racine, where all characters (and not only the central pair, represented by the tragic figures of King Juba and his spouse Barcé) deliver substantial speeches and serve as connectors between successive scenes, maintaining the *liaison*.

Coda: A reproduction experiment

Starting from the early 2000s, the debate on the 'replication crisis' in the sciences (see e.g. Ioannidis, 2005; Open Science Collaboration, 2015) has also reached other data-driven fields, including the digital humanities, where it prompted sustained reflection on the reliability of computational findings (Peels and Bouter, 2018; Herrmann et al., 2023; Schöch, 2023; Joyeux-Prunel, 2024). Within computational literary studies, a milestone in the discussion was Nancy Da's polemical attempt to build 'a computational case' against the entire subfield on the grounds that some results from influential studies could not be replicated (Da, 2019). Despite its significant conceptual and methodological shortcom-

ings,¹³ Da's intervention was instrumental in consolidating the view that replicability and reproducibility cannot be taken for granted and need instead to be built into any sound study design.

Following this principle, all experiments presented in the previous pages are fully replicable through the data (TEI-XML files and CSV meta-data) and code (Python scripts) provided in the *GitHub* repository. From a cultural analytics-oriented point of view, however, these options may seem insufficient: even if the rationale for the corpus building has been described in detail, one is still left wanting to re-implement the same methods on broader text collections to reinforce its empirical results.

The scenario I am describing here could be defined, according to the taxonomy developed by Schöch (2023), as a 'reproduction (of results)': "research question and method of analysis remain identical", but "the dataset, instead of also being identical, [is] more or less similar, but not unrelated" (389). The goal of such a procedure would be to assess whether findings obtained hold true across more than one dataset, and thus are valid more generally, and whether methods developed are robust enough for further reuse.

In my case, performing a reproduction experiment is a necessary step towards assessing the trustworthiness of my findings. While gaining insights into the EmDraCor corpus, following the practice of 'scalable reading', is a valuable goal in itself, it is also necessary to assess their significance in a broader context. Accordingly, a reproduction experiment would allow one to measure how many of the results presented in the previous pages are due to the idiosyncrasies of the text selection, and how many are actually generalisable and can be treated as reliable contributions towards a quantitative history of early modern drama.

Unfortunately, most of the plays featured in EmDraCor have not been sampled from larger machine-actionable collections; to a large extent, the collection itself has been assembled to address the lack of extensive

13 See e.g. the responses in the *Critical Inquiry* online forum (<https://critinq.wordpress.com/2019/03/31/computational-literary-studies-a-critical-inquiry-online-forum>) and the contributions by Jannidis (2020) and Piper (2020).

early modern ‘programmable corpora’. Nonetheless, even if the experiment cannot be fully reproduced on wider datasets, it is still possible to do it partially by leveraging the only two corresponding DraCor corpora that, at the time of writing, contained a significant number of early modern texts, i.e. the French (FreDraCor) and English (EngDraCor) ones.

Together, these two corpora contain more than one thousand texts (602 French, 529 English) from the 1561–1710 time frame, and thus represent a significant benchmark collection for reproducing the experiments previously presented. Furthermore, deepening the analysis of these two dramatic traditions also holds some inherent theoretical interest, since French and English drama can be considered to embody, in Lotmanian terms, the two main configurations of European drama of the age, i.e. a ‘regular’ (French) and an ‘irregular’ (English) model. A reproduction experiment would thus enable further exploration of this dichotomy from a quantitative-formalist angle.

In practical terms, the procedure for obtaining the ‘play embeddings’ for French and English texts was the same as that developed for EmDraCor: I downloaded the metadata tables and enriched them through the usual Python scripts, excluding all texts outside the chronological boundaries of the early modern period as previously defined (1561–1710). Once the vectors were constructed, I reran all scripts to investigate again the three dimensions of interest (distances, clusters, patterns), and compared the results with those obtained from EmDraCor.

Before presenting the findings, however, a caveat is in order. The data available from the English and French drama corpora at the time of writing mostly overlapped only between the 1620s and the 1650s, preventing one from tracking the parallel evolution of these two dramatic traditions in a more comprehensive way. Furthermore, in the case of English drama, data availability meant that I could only compare the English subset of EmDraCor with early works from the main corpus (up to ca. 1650).

One should therefore tread lightly in analysing the results presented in Figure 17, which compares the evolution in the values of the 38 vector metrics across the four collections under investigation (EngDraCor, EmDraCor-Eng, FreDraCor, EmDraCor-Fre). Once again, some trends can

be misleading and require scalable readings in order to be appropriately contextualised, as two examples from the French tradition illustrate.

To begin with, the increase (ca. +25%) of the metric 'size' in FreDraCor seems to contradict not only the results from the French subset in EmDraCor but also the scholarly intuition about the shrinking of casts in the *théâtre classique*. A closer inspection of the actual values for each play reveals, however, that this last observation still holds for most tragedies, while the outliers pushing the trend upwards are comedies and hybrid forms featuring music. This last category was excluded from the EmDraCor sample and includes, for example, many *tragédies lyriques* by Philippe Quinault, such as *Isis* (1677) – the period's largest play with a total of 63 speaking characters.

In any case, the increase in the number of characters in FreDraCor represents the biggest departure from EmDraCor-Fre, and has a ripple effect on all network-related metrics, which therefore show slight growth. Data scarcity explains instead the other notable counter-trend, i.e. the evolution of the 'protagonism' metric: an initially higher value, determined by only about 15 works from 1580 to 1630, starts dropping as soon as the corpus size dramatically increases.

Nonetheless, many results from the larger corpora confirm the trends seen in their smaller counterparts. For example, all four collections show a noticeable growth of the number of stage directions ('wordCountStage'). While it seems premature to interpret such a trend as an early sign of the 'tendency towards epification' that Trilcke et al. (2020) recognised in later German drama, it still points to an interesting development: the fact that the playwrights started providing more and more cues to the performers suggests indeed a gradual shift in the relationship between written text and stage practice.

Figure 17: Evolution of vector components in the English and French subsets of EmDraCor (lighter tones) and the respective main DraCor corpora (darker tones).

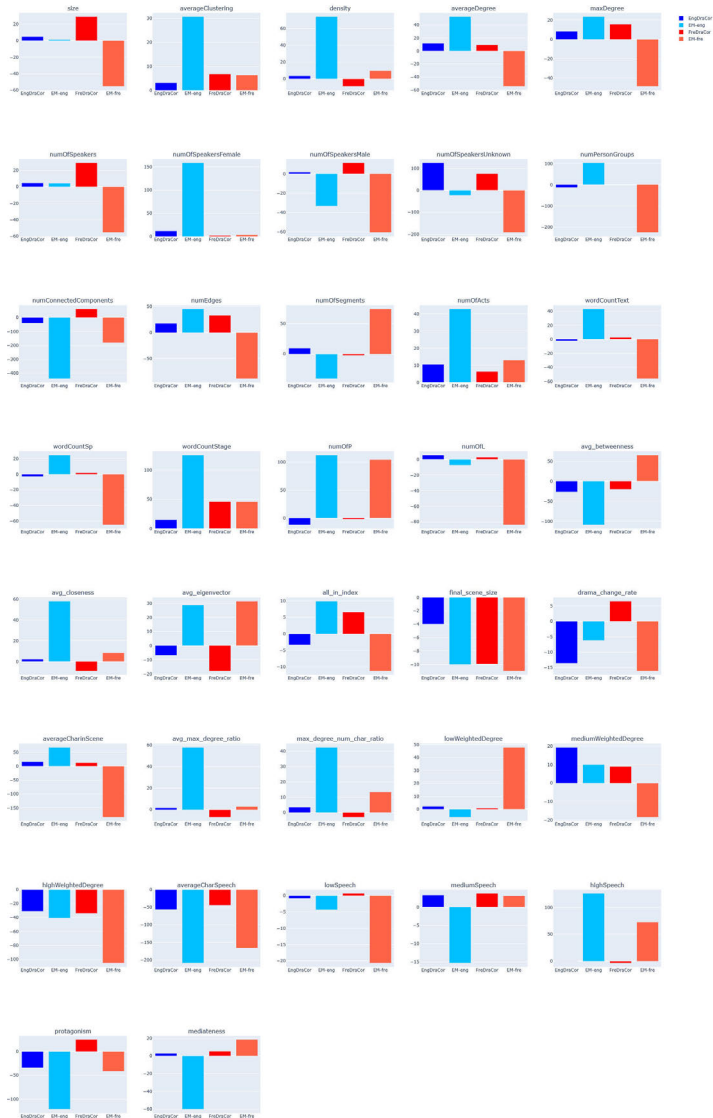


Figure 18: Evolution of the metric ‘averageCharSpeech’ in the English (1570–1660) and French (1561–1710) DraCor corpora (trendline with 95% confidence interval; extreme outliers removed using the IQR method).



Another phenomenon magnified in both EmDraCor subsets, but also finding confirmation in FreDraCor and EngDraCor, is the decrease in average character speech (Figure 18). The reason for this change becomes clearer if one takes into account the distribution of genres within the corpora. From the first half of the seventeenth century, both of them feature an increasing number of comedies – a form that naturally favours shorter, wittier exchanges over the lengthy tirades typical of high

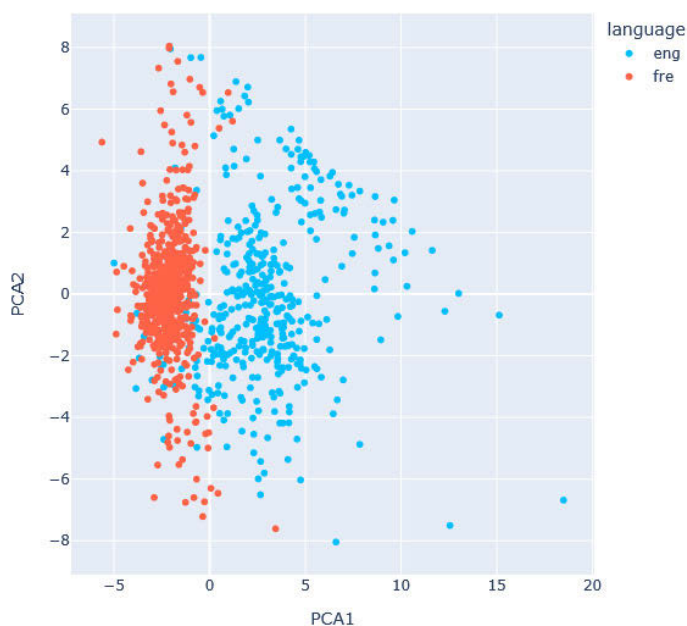
tragedy. As comedies come to represent a larger share of the dataset, it is not unlikely that their characteristically brief speeches progressively pull down the overall average for the metric.

In sum, results from this reproduction experiment return a mixed picture. To begin with, data show that EmDraCor-Eng is actually closer to EngDraCor, as the general trends of the metrics (positive/negative) correspond in 62% of the cases. By contrast, less than 30% of the EmDraCor-Fre metrics exhibit the same trend as in FreDraCor. This discrepancy can be explained by two factors: some known issues in the markup quality of FreDraCor, which influence various network-related metrics and whose impact cannot at this stage be quantified more clearly, and the idiosyncratic composition of EmDraCor-Fre.¹⁴ When trends diverge, the larger corpora nearly always exhibit moderate shifts, indicating that more pronounced increases or decreases in the smaller samples could be reclassified as typical small-sample bias. In only two instances ('size' and 'protagonism' in the two French corpora) are trends significantly opposed; the divergence can be traced to the factors discussed above.

In conclusion, the results from EmDraCor are not fully replicated in the larger corpora, and the picture of dramatic evolution emerging from the corpus cannot be considered statistically robust in absolute terms. The trends it captures are, however, best understood as amplified echoes of those present in the wider dramatic production of the period – exaggerated by the small sample, but not invented by it. In this sense, EmDraCor acts as a sort of 'weather vane', suggesting the main lines of formal development in European theatre even if it cannot prove them beyond doubt. Future improvements in corpus size and markup quality, especially for FreDraCor, will allow these findings to be tested more rigorously.

14 It should be noted that EmDraCor-Fre was not entirely sampled from FreDraCor; one fifth of the texts – mostly early ones – was added through manual encoding.

Figure 19: Principal Component Analysis, English and French DraCor data, 1561–1710.



Ultimately, this reproduction experiment also underlines the viability of vectorisation as a computational method for exploring dramatic forms. As seen in EmDraCor, the compression of plays into meaningful vectors of formal features appears a useful tool for assessing differences between extensive dramatic collections. A contrastive PCA of EngDraCor and FreDraCor (Figure 19) demonstrates this, since it not only displays the clear divide between the two traditions, but also underscores the ‘regularity’ of French theatre (as testified by a more pronounced clustering) as against the ‘irregularity’ of the English one (marked by sparse structure with many outliers).

Conclusion

In 2021, a group of American scholars published in *Cultural Analytics* a study attempting to computationally model ‘minor’ European literatures in light of Pascale Casanova’s theories. Their essay, they argued, “address[ed] the current lack of comparative computational literary studies, with most research still overwhelmingly focused on single national literary frames” (Erlin et al., 2021, 85). Three years later, the newly formed Research Committee on Digital Comparative Literature within the International Comparative Literature Association (ICLA) still acknowledged that “while [...] the development of techniques like computer-assisted distant reading and the digitization of large textual archives have opened new perspectives for the study of literature, the terrain of Comparative Literary Studies remains one of the roughest for digital research”.¹

Treading on such a ‘rough terrain’, this book has tried to employ computational methods to investigate the evolution of European drama during the early modern period. The bid to offer an empirical, data-driven – or better, to use Escobar Varela’s (2021, 7 ff.) more precise terminology, ‘data-assisted’ – perspective on the matter situated it not only within the computational literary studies milieu, but also within the framework of cultural analytics. Accordingly, the analysis followed, albeit not systematically, the blueprint thus laid out by Lev Manovich:

1 See https://icla-dcl.quarto.pub/info/files/ICLA_DCL_Description.pdf.

[...] to observe and analyze culture means to be able to map and measure five fundamental characteristics. The first four are diversity, uniqueness, dynamics (temporal changes), and structure. The last term here means clusters, networks, and other types of relations between many objects — that is, structure as it is understood in exploratory data analysis and unsupervised machine learning, as opposed to in 1960s structuralism. [...] In situations in which artifacts were created using a prescriptive aesthetic or template [...] I can also consider a fifth characteristic: variability. (Manovich, 2020, 215)

Diversity, uniqueness, dynamics, structure, and variability: the previous pages tried, through various proxies and methodologies, to touch upon all five components mentioned by Manovich. In doing so, I have striven to bring together two different approaches: one more goal-oriented, consisting in the re-assessment of the argument put forth in 'Modern European Literature', and another based on the paradigm of exploratory data analysis (Tukey, 1977). I shall now conclude by briefly recapitulating my findings and critically reflecting on the methodology adopted and its shortcomings.

Judging from the results of experiments performed in Chapter Four, Moretti's thesis about the evolution of early modern drama – intended as a consistent movement towards progressive diversification of formal features – stands partially confirmed. On the one hand, the two implementations I adopted for measuring vector distances did not completely agree: the pairwise method showed a convergence-divergence movement, while the centroid-based one rather supported a broad increase in diversification. On the other hand, the inspection of the PCA projections supported more forcefully the narrative of a branching of dramatic traditions; even if the sample per snapshot (30 works) is limited, some clustering does become recognisable.

Finally, the most interesting results come from profiling dramatic spaces according to the evolution of specific metrics. By examining their variation across the different subcorpora, I found traces of persisting shifts in plot structures, characters' features, and network arrangements, suggesting that it is possible to discriminate between dramatic

cultures on the basis of form, and that some of these differences have actually been increasing throughout time. The reproduction experiment on the English and French DraCor collections, while not statistically validating the evidence from EmDraCor, still suggested that it retains value in pointing out the general direction in which the different traditions were developing at the time.

Together, the three approaches proposed – based on the observation of distances, clusters, and patterns – yielded some insights into European drama which go beyond the assessment of Moretti's thesis. The PCA, for example, shows a process which is better described as specialisation than as diversification: dramatic traditions are not expanding into new formal territory, but progressively sorting themselves within a broadly stable space of possibilities, each consolidating a distinct formal identity. Furthermore, computational evidence gives substance to previously qualitative observations, such as the persistent proximity of the French and Italian traditions based on shared poetics. Moreover, several metric shifts correlate with documented historical events, such as the introduction of actresses on the English stage from 1660, showing how formal features meaningfully reflect developments in theatrical practice.

The results of this investigation, however, reveal the limits of applying a single rigid framework, such as speciation, to literary history. While Moretti's 1994 introduction of the 'branching' model within literary studies was pioneering, its emphasis on a straightforward and monodirectional development of drama history has been widely criticised. Tom Eyers, for example, has spoken both of a New Positivist posture and an "implicit Hegelianism" latent in Moretti's theories – recognising in his work a strong teleological drive, based on "the tacit, humanistic assumption that history progresses toward a goal, cultural form perspiring on cue to [...] changes that are to be charted from the perspective of their already having happened, and thus, one may say, from the purview of the victors" (Eyers, 2017, 46–47).

From a disciplinary point of view, it should also be emphasised that, in the thirty years following 'Modern European Literature', many more models of cultural evolution, going far beyond basic speciation, have been proposed. Any attempt at reconstructing the evolution of Euro-

pean early modern drama would thus have to take into account several other models – from general hedonistic selection to drift, cumulation, or co-evolution, and possibly even cultural attraction (cf. Mesoudi, 2016; Morin, 2016; Sobchuk, 2023).

In general terms, the evolutionist template advanced by Moretti might be substituted with a more tempered vision of the process of change, based on ‘horizontal’ movements through the system, and emphasising the notion of variation over evolution, as in Küpper’s cultural net theory. In this perspective, the local branching of forms, while still present, would not necessarily be tied to geographical and linguistic spaces, but could also be explained by other poetological (e.g. the reception or rejection of Aristotle) or sociological factors (e.g. the material conditions of theatre).

Beyond these theoretical considerations, the reader should necessarily take into account the limitations of this research while assessing its contribution. The first one, as already mentioned, concerns the composition and the size of the EmDraCor corpus: while the theoretical and pragmatic choices informing its building process have been extensively discussed in Chapter Two, I am aware that a more accurate evaluation of evolutive dynamics would require retrieving a larger number of plays – and specifically of tragedies, if one wanted to restrict oneself to the assessment of Moretti’s original argument on ‘Baroque tragedies’.

More generally, a larger number of texts would reinforce the statistical reliability of my experiments and also allow one to test other modes of explanation for literary evolution (such as the ‘cohort succession’ hypothesis advanced by Underwood et al., 2022). An expansion to other linguistic domains beyond the five so far investigated would be desirable as well, especially if including traditions defined by transnational circulation, such as the Dutch one – situated at the crossroads between the French ‘regular’ and the English ‘irregular’ models.

From a methodological point of view, moreover, the operationalisation of the concept of drama I proposed in Chapter Three stands open to criticism. First, it could be argued that the specific choice of structural features I employed to represent the plays was inadequate to verify Moretti’s thesis, as it did not cover all elements he listed as

marking the divergence of early modern drama (such as those linked to the plays' language or performance). Secondly, it should be kept in mind that, while most scholars naturally focused on "the surface features of dramatic texts" because of their "easy computational accessibility", they are still "less reliable than they might seem at first glance" (Andresen and Reiter, 2024b, 2). A glaring example of this, which I previously discussed, is the comparability of segmentation practices across corpora and its effect on metric computation.

Considering the project's original conception, the first objection can be addressed on practical grounds: after all, the majority of the elements mentioned by Moretti are still form-related and have thus found concrete expression in the array of features employed. It does, however, underline once again an aspect the reader might want to keep in mind: the research behind this book focused exclusively on the written aspect of theatre, i.e. the dramatic text, and did not address questions related to its performative aspects or the social implications, which are central to other digital humanities investigations, such as Escobar Varela's (2021), and played an important role in early modern culture. This restriction of scope, however, is justified by the research object itself, since, as Lochert (2009, 29) notes, the European theatre of the sixteenth and seventeenth centuries is still characterized by a primacy of the text.

The second objection, regarding the broader validity of surface textual features, could instead be approached by pointing out the inherent limits of the practice of modelling. "Built into the modeling process", Richard Jean So (2017, 671) argues, "is a self-reflexive account of what the model has sought to measure and the limitations of its ability to produce such a measurement". Accordingly, the model itself should not be considered "as a potential antagonist, an entity that has brought forth intractable empirical truths that one must accept or reject", but rather a mediating instance which allows the scholar to confront data in a more efficient way, and which should be continuously updated and refined according to its performance. In this case, the operationalisation of dramatic form I proposed, and thus the formal features I selected to represent the plays, are justified by their usefulness in recognising patterns in

drama history, and can thus be used as a point of departure for further research.

Ultimately, the global significance of this work is twofold. On the one hand, considering the still unsatisfying data landscape for computational comparative drama analysis, the creation of EmDraCor – a multilingual, open-access, machine-actionable corpus of 150 TEI/XML-encoded plays – represents in itself a significant contribution to future digital research in early modern studies.

On the other hand, this project – like many others in digital humanities research – began by trying to provide an empirical reassessment (and, to some extent, a validation) of a previous qualitative theory. In doing so, it placed itself within the “triangulation” process outlined by Fotis Jannidis:

Characteristic of literary research is precisely the abundance of interpretative theses, especially on the canonical texts, an abundance of statements that contradict each other. So what does it mean when, in this context, such a thesis is taken up again through quantitative research and substantiated with quantitative arguments? In this way, the well-known thesis is supported anew, from a methodologically different perspective. In the social sciences, this phenomenon is referred to as ‘triangulation’ in the context of the discussion about ‘mixed methods’ and is seen by many advocates of a pragmatic approach to qualitative and quantitative methods as a gain in knowledge. (Jannidis, 2022, 10–11)

As shown in the previous pages, the experiments’ outcomes lent some support to Moretti’s global hypothesis, but also suggested that ‘horizontal’ variation could be a better interpretative key for the evolution of early modern drama than ‘vertical’ speciation – thus partly moving beyond the mere triangulation paradigm, and delivering some new insights. The book’s main contribution, however, amounted to the further development of a largely ignored method, i.e. the vectorisation of texts based on their structural properties. Accordingly, the main hermeneutical gain it provides is not found in the actual scrutiny of ‘Modern European Liter-

ature', but rather in the development and refinement of a methodology which might represent an efficient approach for dealing computationally with issues of literary form. This technique, I argue, might be particularly productive for comparative literary studies, insofar as it allows scholars to operationalise the fuzzy concepts of 'comparison' and 'similarity' with regard to formal aspects. Operations such as the comparison of vectors according to different implementations, or the inspection of clustering, might thus offer quantitative avenues for exploring humanistic models of comparison such as Wittgenstein's 'family resemblances' or Rosch's 'prototype theory'.

In this perspective, the work situates itself within what Anne-Sophie Bories and colleagues have called the "polite revolution of computational literary studies", i.e. "a moment of acute methodological creativity, with researchers branching out toward novel and often heterogeneous tools, experimenting with a number of possibilities, some of which, in turn, inform and sometimes change what they are striving to grasp" (Bories et al., 2022, 13). The methodological contribution this book offers should thus represent the most interesting takeaway for the reader, going beyond the results of specific experiments conducted; after all, "[t]he goal of experimental inquiry is not to get everything right the first time[,] but to advance knowledge in a way that invites further testing and refinement" (Underwood, 2019, 183).

Almost one hundred years ago, Boris Yarkho closed his highly experimental essay on speech distribution in five-act tragedies by bluntly stating: "[w]e have demonstrated the technique; the path has been cleared for those more fortunate than us" (Yarkho, 2019, 72). In a similar vein, but perhaps with a less laden rhetoric, I hope that other scholars will improve the methods I began prototyping here and fully explore the potential of formal vectorisation for the comparative study of literary form across traditions.

Appendix A

This appendix synthetically documents the sources, processing pipelines, and editorial decisions behind each of the five EmDraCor subcorpora. It complements the discussion of corpus design in Chapter Two by providing granular information on the provenance of the texts, the strategies used to fill temporal gaps, any exclusion criteria applied, and known issues that may affect downstream analysis.

French subcorpus (EmDraCor-Fre)

Sources	24 plays from FreDraCor (originally from <i>Théâtre Classique</i> , cf. Fièvre, 2007); 2 plays from Wikisource; 4 plays encoded from scans (<i>Google Books</i> , <i>Internet Archive</i> , <i>Gallica</i>).
Data landscape	Wide availability of texts thanks to FreDraCor (containing 552 plays from 1561 to 1710); some gaps before 1600.
Issues	FreDraCor's sheer size has delayed fine-grained curation and fixing of markup inconsistencies, mostly regarding wrong speaker attribution.
Gap-filling strategy	Missing texts retrieved through the <i>Répertoire des pièces de la Renaissance française, 1537–1615</i> (Bourassa, 2009) and the CÉSAR (<i>Calendrier électronique des spectacles sous l'Ancien Régime et la Révolution</i>) database launched by Barry Russell (cf. Bannister, 2009).

Processing	Plays from FreDraCor (XML): checked and corrected for speaker attribution errors, missing/wrong sex attributes, dating issues. Remaining plays (TXT, sometimes from OCR): onboarded through the EasyDrama pipeline
-------------------	---

German subcorpus (EmDraCor-Ger)

Sources	22 plays from <i>TextGrid</i> ; 5 plays encoded from scans (<i>Google Books</i>); 1 play from GerDraCor (originally from <i>TextGrid</i>); 1 play from the <i>Deutsches Textarchiv</i> ; 1 play from a critical edition (Brauneck, 1975).
Data landscape	Very limited initial availability: only one early modern text (by Gryphius) in GerDraCor at the start of the project. Retrieval further complicated by the wide diffusion of Jesuit Neo-Latin drama in the German-speaking space during the seventeenth century (e.g. Jakob Bidermann's <i>Cenodoxus</i> , 1602), which reduced the pool of vernacular candidates.
Issues	/
Gap-filling strategy	Missing titles identified through Alexander (1984) and Meyer (1993); corresponding scanned copies retrieved from Google Books.
Processing	Plays from TextGrid (XML): markup transformation pipeline via Python scripts. Play from the Deutsches Textarchiv (XML): converted to TXT and encoded through the EasyDrama pipeline. Remaining plays (from scans): encoded through the EasyDrama pipeline.

English subcorpus (EmDraCor-Eng)

Sources	28 plays from the newly-created EngDraCor (originally from <i>EarlyPrint</i> 's EEBO (<i>Early English Books Online</i>) subset; 2 plays directly from <i>EarlyPrint</i> 's ECCO (<i>Eighteenth Century Collections Online</i>) subset.
Data landscape	Massive text availability through <i>EarlyPrint</i> , covering the entire time frame. Stratified sampling was performed on the basis of a master list of all plays in the corpus, which however stopped at 1700 (i.e. included only files from EEBO and not from ECCO). The existing DraCor Shakespeare corpus (ShakeDraCor) was deliberately not included in favour of a broader, non-hypercanonical selection.
Issues	Before being used for sampling, the <i>EarlyPrint</i> master list of plays required manual cleaning due to the presence of foreign-language texts and translations.
Gap-filling strategy	Two plays for 1701–1710 randomly sampled from titles listed in the relevant Wikipedia category pages, then manually downloaded from the ECCO collection.
Processing	All <i>EarlyPrint</i> texts (XML): markup transformation to DraCor standards, additional curation.

Spanish subcorpus (EmDraCor-Spa)

Sources	17 plays from EMOTHE; 3 plays from CalDraCor; 4 plays from open-access scholarly editions; ¹ 6 plays encoded from scans (<i>Google Books</i> , <i>Internet Archive</i>).
Data landscape	DraCor's Spanish-language collections are limited to nineteenth-century drama (SpanDraCor) and Calderón's full works (CalDraCor). The otherwise useful <i>Biblioteca Virtual Miguel de Cervantes</i> offers no direct XML download and its metadata tagging is often imprecise.
Issues	Dating is particularly complex in Spanish early modern drama (e.g. the chronology of Lope de Vega's opus ²); all dates were cross-referenced against multiple scholarly sources, prioritising the earliest attested year (usually the composition year). Scene segmentation is often unclear in these plays and required particular attention.
Gap-filling strategy	Missing titles identified through the CATCOM (<i>Catálogo de autores teatrales del siglo XVII</i>) database (Ferrer Valls, 2013) and other bibliographic resources.
Processing	EMOTHE (XML): markup transformation pipeline via Python scripts. Plays from CalDraCor (XML): additional curation. Scholarly editions (PDF): converted to TXT and onboarded through the EasyDrama pipeline. Remaining plays (TXT, from scans): onboarded through the EasyDrama pipeline.

-
- 1 More specifically, from the BIADIG (*Biblioteca Áurea Digital*) collection of the *Grupo de Investigación Siglo de Oro* (GRISO): <https://hdl.handle.net/10171/34258>.
 - 2 For an attempt to tackle this problem via stylometric methods see Cuéllar (2023).

Italian subcorpus (EmDraCor-Ita)

Sources	12 plays from ItaDraCor; 14 plays encoded from scans (<i>Google Books</i> , <i>Internet Archive</i>); 3 plays from open-access scholarly editions; ³ 1 play from <i>Wikisource</i> .
Data landscape	Many decades were completely unrepresented; in general, the scarce availability of texts in TEI-XML format required substantial encoding efforts from scratch.
Issues	Exclusion criteria: libretti and texts entirely written in local dialects (plays mixing different languages for stylistic reasons, such as Bonicelli's <i>Pantalone bullo</i> , were accepted).
Gap-filling strategy	Missing titles identified through historical bibliographies, such as Leone Allacci's <i>Drammaturgia</i> (in its revised 1755 edition), and modern ones, such as Herrick (1964).
Processing	Plays from ItaDraCor (XML): additional curation. Remaining plays (TXT, often from scans): onboarded through the EasyDrama pipeline.

3 More specifically, from the *Archivio del Teatro Pregoldoniano* project (<https://www.usc.gal/goldoni/biblio/>).

Appendix B

The following tables provide an overview on the plays included in the Early Modern Drama Corpus (EmDraCor). The year indicated is normalised according to DraCor dating conventions.¹ DraCor-compatible TEI-XML files for each play are available in the project's *GitHub* repository.

German-language plays

Author	Title	Genre	Year
Jos Murer	<i>Absolom</i>	drama	1565
Jos Murer	<i>Hester</i>	drama	1567
Nicodemus Frischlin	<i>Frau Wendelgard</i>	comedy	1579
Bartholomäus Krüger	<i>Von dem Anfang und Ende der Welt</i>	drama	1580
Martin Hayneccius	<i>Hans Pfriem</i>	comedy	1582
Georg Rollenhagen	<i>Vom reichen Mann und armen Lazaro</i>	drama	1590
Heinrich Julius von Braunschweig-Wolfenbüttel	<i>Der Fleischhawer</i>	drama	1593

1 <https://dracor.org/doc/faq>.

Author	Title	Genre	Year
Heinrich Julius von Braunschweig-Wolfenbüttel	<i>Von einem ungeratenen Sohn</i>	tragedy	1594
Ludwig Hollonius	<i>Somnium Vitae Humanae</i>	drama	1605
Joachim Schlue	<i>Isaac</i>	comedy	1606
Martin Rinckart	<i>Der Eislebische Christliche Ritter</i>	comedy	1613
Anonym	<i>Ein ander lustig Pickelherings Spiel</i>	comedy	1620
Jakob Balde	<i>locus serius theatralis</i>	drama	1629
Johann Rist	<i>Irenaromachia</i>	tragi-comedy	1630
Johann Rist	<i>Perseus</i>	tragedy	1634
Martin Opitz	<i>Judith</i>	tragedy	1635
Andreas Gryphius	<i>Absurda Comica oder Herr Peter Squenz</i>	comedy	1648
Daniel Casper von Lohenstein	<i>Ibrahim</i>	tragedy	1649
Andreas Gryphius	<i>Cardenio und Celine</i>	tragedy	1650
Johann Georg Schoch	<i>Comoedia Vom Studenten-Leben</i>	comedy	1657
Johann Sebastian Mitternacht	<i>Der unglückselige Soldat</i>	drama	1662
Daniel Casper von Lohenstein	<i>Epicharis</i>	tragedy	1665
Johann Christian Hallmann	<i>Sophia</i>	tragedy	1671
Christian Weise	<i>Baurischer Machiavellus</i>	comedy	1679
Christian Weise	<i>Masaniello</i>	tragedy	1682
Joseph Anton Stranitzky	<i>Türkisch-bestraffter Hochmuth</i>	tragi-comedy	1685
Christian Reuter	<i>Der ehrlichen Frau Schlampampe Krankheit und Tod</i>	tragi-comedy	1696

Author	Title	Genre	Year
Johann Christian Hallmann	<i>Liberata</i>	tragedy	1699
Barthold Feind	<i>Das verwirrte Haus Jacob</i>	drama	1703
Christian Weise	<i>Ungleich und gleich gepaarte Liebes-Alliance</i>	comedy	1708

French-language plays

Author	Title	Genre	Year
Jean De La Taille	<i>Les Corrivaus</i>	comedy	1562
Jacques Bienvenu	<i>Comédie du monde malade et mal pansé</i>	comedy	1568
Robert Garnier	<i>Hippolyte</i>	tragedy	1573
Odet de Turnèbe	<i>Les Contents</i>	comedy	1581
Robert Garnier	<i>Bradamante</i>	tragedy	1582
Pierre Matthieu	<i>Aman</i>	tragedy	1588
Simon Belyard	<i>Le Guisien</i>	drama	1592
Jean de Virey	<i>La Machabée</i>	tragedy	1596
Antoine de Montchrestien	<i>La Carthaginoise</i>	tragedy	1601
Nicolas Soret	<i>La Céciliade</i>	drama	1606
Théophile de Viau	<i>Pyrame et Thisbé</i>	tragedy	1617
Honoré d'Urfé	<i>Sylvanire</i>	pastoral	1617
Théophile de Viau	<i>La Tragédie de Pasiphaé</i>	tragedy	1627
Pierre Du Ryer	<i>Clitophon</i>	tragicomedy	1629
Balthazar Baro	<i>La Clorise</i>	pastoral	1632

Author	Title	Genre	Year
Charles Dalibray	<i>Soliman</i>	tragicomedy	1637
Pierre Corneille	<i>Rodogune</i>	tragedy	1644
Gabriel Gilbert	<i>Téléphonte</i>	tragicomedy	1643
François Tristan l'Hermitte	<i>Le Parasite</i>	comedy	1654
Molière	<i>Sganarelle ou Le Cocu Imaginaire</i>	comedy	1660
Jean Racine	<i>La Thébaïde</i>	tragedy	1664
Adrien-Thomas Perdou de Subligny	<i>La Folle Querelle ou La Critique d'Andromaque</i>	comedy	1668
Thomas Corneille	<i>La Mort d'Achille</i>	tragedy	1673
Gaspard Abeille	<i>Argélie</i>	tragedy	1673
Pierre Saint-Glas	<i>Les Bouts Rimés</i>	comedy	1682
Dancourt	<i>Le Chevalier à la Mode</i>	comedy	1687
Dancourt	<i>La Parisienne</i>	comedy	1691
Jean Galbert de Campistron	<i>Juba Roi de Mauritanie</i>	tragedy	1685
François Joseph La-grange-Chancel	<i>Amasis</i>	tragedy	1702
Joseph de Lafont	<i>Le Naufrage ou La Pompe Funèbre de Crispin</i>	comedy	1710

Italian-language plays

Author	Title	Genre	Year
Ludovico Dolce	<i>Marianna</i>	tragedy	1565
Lionardo Salviati	<i>Il granchio</i>	comedy	1566

Author	Title	Genre	Year
Torquato Tasso	<i>Aminta</i>	pastoral	1573
Vincenzo Giusti	<i>Irene</i>	tragedy	1579
Battista Guarini	<i>La idropica</i>	comedy	1584
Pomponio Torelli	<i>La Merope</i>	tragedy	1587
Antonio Decio	<i>Acripanda</i>	tragedy	1592
Torquato Tasso	<i>Gli intrichi d'amore</i>	comedy	1598
Giambattista Della Porta	<i>La carbonaria</i>	comedy	1601
Silvio Fiorillo	<i>La Lucilla costante</i>	comedy	1609
Giovan Battista Andreini	<i>Lo schiavetto</i>	comedy	1612
Flaminio Scala	<i>Il finto marito</i>	comedy	1618
Bartolomeo Tortoletti	<i>Gionata</i>	tragedy	1624
Nicolò Barbieri	<i>L'inavvertito</i>	comedy	1629
Pier Maria Cecchini	<i>L'amico tradito</i>	comedy	1633
Decio Memmolo	<i>La Susanna</i>	tragedy	1632
Vincenzo Nolfi	<i>Romilda</i>	tragedy	1643
Fulvio Testi	<i>L'isola di Alcina</i>	tragedy	1648
Giovanni Oransi	<i>Rosina</i>	comedy	1652
Carlo De' Dottori	<i>Aristodemo</i>	tragedy	1657
Giacinto Andrea Cicognini	<i>Il Mustafâ</i>	drama	1663
Mattias Maria Bartolommei	<i>Amore opera a caso</i>	comedy	1668
Girolamo Graziani	<i>Il Cromuele</i>	tragedy	1671
Giovanni Domenico Bregholini	<i>Pausania</i>	tragedy	1679

Author	Title	Genre	Year
Girolamo Gigli	<i>Un pazzo guarisce l'altro</i>	comedy	1687
Giovanni Bonicelli	<i>Pantalone bullo</i>	comedy	1688
Alessandro Guidi	<i>Endimione</i>	pastoral	1692
Girolamo Gigli	<i>I litiganti ovvero Il giudice impazzato</i>	comedy	1695
Niccolò Amenta	<i>La somiglianza</i>	comedy	1706
Luigi Lodovico Riccoboni	<i>Tito Manlio</i>	tragedy	1707

Spanish-language plays

Author	Title	Genre	Year
Alonso de La Vega	<i>Seraphina</i>	tragedy	1566
Alonso de La Vega	<i>La Duquesa de la Rosa</i>	comedy	1566
Andrés Rey de Artieda	<i>Los amantes</i>	tragedy	1577
Cristóbal de Virués	<i>La gran Semíramis</i>	tragedy	1579
Miguel de Cervantes	<i>El cerco de Numancia</i>	tragedy	1585
Félix Lope de Vega y Carpio	<i>El perseguido</i>	comedy	1590
Félix Lope de Vega y Carpio	<i>La viuda valenciana</i>	comedy	1595

Author	Title	Genre	Year
Félix Lope de Vega y Carpio	<i>El galán castrucho</i>	comedy	1598
Antonio Mira de Amescua	<i>El esclavo del demonio</i>	drama	1605
Guillén de Castro	<i>Las mocedades del Cid</i>	drama	1605
Luis Vélez de Guevara	<i>La serrana de la vera</i>	tragedy	1613
Juan Ruiz de Alarcón y Mendoza	<i>La verdad sospechosa</i>	comedy	1619
Francisco de Monteser	<i>El caballero de Olmedo</i>	comedy	1621
María de Zayas Sotomayor	<i>La traición en la amistad</i>	comedy	1630
Pedro Calderón de la Barca	<i>La vida es sueño</i>	drama	1635
Francisco de Rojas Zorrilla	<i>Entre bobos anda el juego</i>	comedy	1638
Francisco de Rojas Zorrilla	<i>Lo que son mujeres</i>	comedy	1644
Pedro Calderón de la Barca	<i>La segunda esposa</i>	auto	1648
Juan Maldonado, Diego de la Dueña, Jerónimo de Cifuentes	<i>La más constante mujer</i>	comedy	1658
Agustín Moreto y Cavana	<i>El desdén con el desdén</i>	comedy	1654
Agustín Moreto y Cavana	<i>El lindo don Diego</i>	comedy	1662
Pedro Calderón de la Barca	<i>Sueños hay que verdad son</i>	auto	1670

Author	Title	Genre	Year
Juan de Matos Fragoso, Juan Bautista Dia- mante, Juan Vélez de Guevara	<i>El hidalgo de la Mancha</i>	comedy	1673
Juan Bautista Dia-mante	<i>El Hércules de Ocaña</i>	comedy	1673
Sor Juana Inés de la Cruz	<i>Los empeños de una casa</i>	comedy	1683
Juan de Mon- tenegro y Neira	<i>La toma de Buda</i>	auto	1687
Francisco Bances y Candamo	<i>El esclavo en grillos de oro</i>	comedy	1692
Antonio de Zamora	<i>Ser fino y no pare- cerlo</i>	comedy	1692
Ferdinando Romero	<i>Aunque las razones basten nunca la justicia sobra</i>	comedy	1701
Antonio de Zamora	<i>Duendes son los alcahuetes y el espíritu Foletto</i>	comedy	1709

English-language plays

Author	Title	Genre	Year
Thomas Norton, Thomas Sackville	<i>Gorboduc</i>	tragedy	1561
Ulpian Fulwell	<i>Like Will to Like</i>	interlude	1568
Nathaniel Woodes	<i>The Conflict of Conscience</i>	comedy	1572

Author	Title	Genre	Year
George Gascoigne	<i>The Glass of Government</i>	tragicomedy	1575
Christopher Marlowe	<i>Dr. Faustus</i>	tragedy	1587
George Peele	<i>The Love of David and Fair Bathsheba</i>	tragedy	1590
Samuel Daniel	<i>Cleopatra</i>	tragedy	1594
Ben Jonson	<i>Every Man Out of His Humour</i>	comedy	1599
Samuel Daniel	<i>Philotas</i>	tragedy	1605
Lording Barry	<i>Ram Alley</i>	comedy	1608
Thomas Dekker	<i>If It Be Not Good the Devil Is in It</i>	comedy	1611
Henry Shirley	<i>The Martyred Soldier</i>	tragedy	1622
John Fletcher	<i>A Wife for a Month</i>	tragicomedy	1624
James Shirley	<i>The Example</i>	comedy	1634
Thomas Dekker	<i>The Noble Spanish Soldier</i>	tragedy	1634
Thomas Nabbes	<i>The Unfortunate Mother</i>	tragedy	1639
James Shirley	<i>The Court Secret</i>	tragicomedy	1642
John Denham	<i>The Sophy</i>	tragedy	1642
Robert Baron	<i>Mirza</i>	tragedy	1655
Thomas Tomkis	<i>Lingua</i>	comedy	1657
Thomas Porter	<i>The Carnival</i>	comedy	1664
Abraham Bailey	<i>The Spiteful Sister</i>	comedy	1667
Thomas Duffet	<i>The Mock Tempest</i>	comedy	1675
Thomas Shadwell	<i>The Libertine</i>	tragedy	1676
Edward Ravenscroft	<i>Dame Dobson</i>	comedy	1684

Author	Title	Genre	Year
Francis Fane	<i>The Sacrifice</i>	tragedy	1687
Mary Pix	<i>Ibrahim</i>	tragedy	1696
Peter Motteux	<i>Beauty in Distress</i>	tragedy	1698
Nicholas Rowe	<i>The Fair Penitent</i>	tragedy	1702
Susannah Centlivre	<i>The Man's Be-witched</i>	comedy	1709

Appendix C

I present here a series of ‘model cards’ describing the metrics included in the EmDraCor vectors. Each card includes a descriptive name of the feature, the related variable as found in the Python scripts, its source, and an explanation on how it was calculated – especially important when the value was not directly computed by the DraCor metrics service and called via API, but rather calculated autonomously on the basis of other scholars’ work. Metrics are grouped according to the four textual aspects they cover: network, cast and speech, size, plot.

Network metrics

Metric Name	Variable Name	Source	Explanation
Size	Size	DraCor API	Represents the total number of nodes in the network.
Average Clustering Coefficient	averageClustering	DraCor API	Calculated as the mean of the local clustering coefficients, which indicates the extent to which nodes cluster or form tightly inter-connected groups.

Metric Name	Variable Name	Source	Explanation
Density	density	DraCor API	Represents the ratio of actual edges to possible edges, reflecting the overall interconnectivity of the network.
Average Degree	averageDegree	DraCor API	Represents the average number of edges any node has, measuring the overall connectivity.
Maximum Degree	maxDegree	DraCor API	Represents the degree of the character with the highest number of connections.
Average Betweenness	Betweenness	Computed on data from the DraCor API	Betweenness measures the extent to which a node in a network lies on the shortest paths between other nodes, indicating its importance in facilitating communication or information flow.
Average Closeness	closeness	Computed on data from the DraCor API	Closeness quantifies how close a node is to all other nodes in a graph, with nodes having higher closeness centrality being more central in terms of overall network accessibility. It is calculated as the reciprocal of the sum of the shortest path distances from that node to all other nodes in the network.

Metric Name	Variable Name	Source	Explanation
Average Eigenvector Centrality	eigenvector	Computed on data from the DraCor API	Eigenvector centrality assesses the importance of a node based on both its direct connections and the importance of the nodes it is connected to, with nodes being more central if they are connected to other highly central nodes.
Number Of Edges	numEdges	DraCor API	Represents the total number of edges in the network.
Connected Components	numConnected-Components	DraCor API	Represents the number of independent subgraphs (connected components) within the network.
Ratio, Average Degree To Maximum Degree	avg_max_degree_ratio	Derived from Szemes and Vida (2024), computed on DraCor API data	"This measure therefore gives the proportion of the average degree number relative to the maximum degree number, and in a sense resembles density" (Szemes and Vida, 2024, 172).
Ratio, Maximum Degree To Number Of Characters	max_degree_num_characters_ratio	Derived from Szemes and Vida (2024), computed on DraCor API data	"The ratio shows what percentage of the entire cast the node with the highest degree interacts with" (Szemes and Vida, 2024, 172).

Metric Name	Variable Name	Source	Explanation
Weighted Degree Distribution	low/medium/high-WeightedDegree	Derived from Szemes and Vida (2024), computed on DraCor API data	Weighted degree quantifies the importance or strength of the connections a node has by assigning a weight to each edge. While Szemes and Vida (2024) used <i>k</i> -means clustering to identify different groups in the distribution, I preferred a simpler solution, i.e. counting how many characters have a weighted degree falling in the first, second, or third tertile (= low, medium, high WD) and then converting it to a share.
Protagonism	Protagonism	Derived from Algee-Hewitt (2017), computed on DraCor API data	A variation of eigenvector centrality, computed as follows: the ratio between the Gini score (a common measure for inequality, here describing the relationship between the core and the peripheral cast in a play) and the percentage of characters in the top quartile of the eigenvector distribution.

Metric Name	Variable Name	Source	Explanation
Mediatedness	Mediatedness	Derived from Algee-Hewitt (2017), computed on DraCor API data	A variation of betweenness, refers to the maximum normalised betweenness among the characters.

Cast and speech

Metric Name	Variable Name	Source	Explanation
Average Characters Per Scene	averageCharin-Scene	Derived from Szemes and Vida (2024), computed on DraCor API data	Computed by dividing the number of characters in one scene by the total number of scenes.
Average Length Of Character Speech	averageChar-Speech	Derived from Szemes and Vida (2024), computed on DraCor API data	Average number of words spoken by a character per line and/or sentence. Only characters who speak more than 10 times are considered.

Metric Name	Variable Name	Source	Explanation
Speech Intensity Distribution	lowSpeech, medium-Speech, high-Speech	Derived from Szemes and Vida (2024), computed on DraCor API data	Computed following the same logic of weighted degree distribution (word count + division in tertiles), shows the share of characters which speak 'much', 'little', or 'a moderate amount' in the text. Szemes and Vida excluded collective characters from the count, but I reintroduced them since I believe them, in most instances in my corpus, to be fully fledged dramatic characters (i.e. not only 'background noise').
Number Of Speakers	numOfSpeakers	DraCor API	Number of speaking characters in a play (marked with a <i>@who</i> attribute inside any <code><sp></code> element).
Number Of Speakers By Gender	numOfSpeakersFemale, numOfSpeakersMale, numOfSpeakerUnknown	DraCor API	Number of speaking characters which have a given <i>@sex</i> attribute in the <code><castList></code> element.
Number Of Collective Speakers	numPersonGroups	DraCor API	Number of speaking characters marked by the element <code><personGrp></code> in the <code><particDesc></code> .

Size

Metric Name	Variable Name	Source	Explanation
Number Of Acts	numOfActs	DraCor API	Number of <div> elements whose attribute <i>@type</i> is “act” (within the <body> element in the TEI markup).
Number Of Segments	numOfSegments	DraCor API	Number of <div> elements of all types, within the <body> element in the TEI markup, including acts, scenes, configurations, and other types of textual segments. Segments inside the <front> or <back> elements, such as prefaces or epilogues, are not counted.
Word Count, Whole Text	wordCountText	DraCor API	Total number of words (tokens) in the entire text.
Word Count, Spoken Text	wordCountSp	DraCor API	Total number of words (tokens) within any <sp> element (i.e. within any speech act).
Word Count, Stage Directions	word-CountStage	DraCor API	Total number of words (tokens) within any <stage> element (i.e. within any stage direction).

Metric Name	Variable Name	Source	Explanation
Number Of Prose Lines	numOfP	DraCor API	Total number of paragraphs in prose (<p>) in the play.
Number Of Verse Lines	numOfL	DraCor API	Total number of lines of verse (<l>) in the play.

Plot dynamics

Metric Name	Variable Name	Source	Explanation
All-In Index	all_in_index	Derived from Trilcke et al. (2017), computed on DraCor API data	Identifies the point in time at which all characters have occurred at least once in a drama, expressed as a percentage (position of the segment where it happens / total number of segments, e.g.: all-in at segment 41/42: 98%).
Final Scene Size	final_scene_size	Derived from Trilcke et al. (2017), computed on DraCor API data	Shows the percentage of all characters of a drama which appear on stage in the last scene.

Metric Name	Variable Name	Source	Explanation
Drama Change Rate	drama_change_rate	Derived from Trilcke et al. (2017), computed on DraCor API data	Based on a modified Levenshtein distance, it takes into account only insertions (+1) and deletions (-1) of characters between scenes (see Fischer et al., 2017). I first compute the change rate between segments through the ratio between the number of edits and the number of individual speakers in any pair of scenes. The mean of all segment change rates is then assumed as the overall drama change rate.

Bibliography

- Ädel, Annelie. "Corpus Compilation." In *A Practical Handbook of Corpus Linguistics*, edited by Magali Paquot and Stefan Gries, 3–24. Cham: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-46216-1_1.
- Ahnert, Ruth, Sebastian E. Ahnert, Catherine Nicole Coleman, and Scott B. Weingart. *The Network Turn: Changing Perspectives in the Humanities*. Cambridge: Cambridge University Press, 2021. <https://doi.org/10.1017/9781108866804>.
- Alexander, Robert J. "Balthasar Crusius (1550?–1630). Zu seinem Leben und einem unbekannten Drama." *Wolfenbütteler Barock-Nachrichten* 7 (1980): 106–112.
- Alexander, Robert J. *Das deutsche Barockdrama*. Stuttgart: J.B. Metzler, 1984. <https://doi.org/10.1007/978-3-476-03913-2>.
- Algee-Hewitt, Mark. "Distributed Character: Quantitative Models of the English Stage, 1550–1900." *New Literary History* 48, no. 4 (2017): 751–782. <https://doi.org/10.1353/nlh.2017.0038>.
- Allegri, Luigi. "Teoria e poetica del teatro moderno." In *Storia del teatro moderno e contemporaneo*, vol. 1, edited by Roberto Alonge and Guido Davico Bonino, 1183–1205. Torino: Einaudi, 2000.
- Allison, Sarah, Ryan Heuser, Matthew Jockers, Franco Moretti, and Michael Witmore. "Quantitative Formalism: An Experiment." Stanford Literary Lab Pamphlet 1, January 2011. <https://litlab.stanford.edu/LiteraryLabPamphlet1.pdf>.

- Andresen, Melanie, and Nils Reiter, eds. *Computational Drama Analysis: Reflecting on Methods and Interpretations*. Berlin: De Gruyter, 2024a. <https://doi.org/10.1515/9783111071824>.
- Andresen, Melanie, and Nils Reiter. "Introduction." In *Computational Drama Analysis: Reflecting on Methods and Interpretations*, edited by Melanie Andresen and Nils Reiter, 1–6. Berlin: De Gruyter, 2024b. <https://doi.org/10.1515/9783111071824-001>.
- Andresen, Melanie, Benjamin Krautter, Janis Pagel, and Nils Reiter. "Who Knows What in German Drama? A Composite Annotation Scheme for Knowledge Transfer. Annotation, Evaluation, and Analysis." *Journal of Computational Literary Studies* 1, no. 1 (2022). <https://doi.org/10.48694/jcls.107>.
- Angelini, Franca. "Barocco italiano." In *Storia del teatro moderno e contemporaneo*, vol. 1, edited by Roberto Alonge and Guido Davico Bonino, 193–276. Torino: Einaudi, 2000.
- Anowar, Farzana, Samira Sadaoui, and Bassant Selim. "Conceptual and Empirical Comparison of Dimensionality Reduction Algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE)." *Computer Science Review* 40 (May 2021): 100378. <https://doi.org/10.1016/j.csrev.2021.100378>.
- Arnaud, Charles. *Les théories dramatiques au 17^e siècle: étude sur la vie et les œuvres de l'abbé d'Aubignac*. Genève: Slatkine, 1970. First published 1888.
- Artico, Tancredi. "Giason Denores, Poetica." *Studi giralddiani. Letteratura e teatro* (October 2020): 249–272. <https://doi.org/10.6092/2421-4191/2020.6.249-272>.
- Asmuth, Bernhard. *Einführung in die Dramenanalyse*. Stuttgart: Metzler, 1980.
- Ayesha, Shaeela, Muhammad Kashif Hanif, and Ramzan Talib. "Overview and Comparative Study of Dimensionality Reduction Techniques for High Dimensional Data." *Information Fusion* 59 (July 2020): 44–58. <https://doi.org/10.1016/j.inffus.2020.01.005>.
- Bamman, David. "Born Literary Natural Language Processing." In *Computational Humanities*, edited by Jessica Marie Johnson, David Mimno, and Lauren Tilton. Minneapolis: University of Minnesota

- Press, 2024. <https://dhdebates.gc.cuny.edu/read/computational-humanities-5c64bbab-d7ca-41be-8f87-f26117a9a20f/section/7c98d16f-ef21-4201-ad00-24b4e3ba6be1#cho3>.
- Bannister, Mark H. "La base de données CESAR et l'avenir des ressources électroniques." *Revue d'histoire du théâtre*, no. 3 (2009).
- Bartsch, Sabine, Evelyn Gius, Marcus Müller, Andrea Rapp, and Thomas Weitin. "Sinn und Segment. Wie die digitale Analysepraxis unsere Begriffe schärft." *Zeitschrift für digitale Geisteswissenschaften* 8 (2023). https://doi.org/10.17175/2023_003.
- Bassnett, Susan. *Translation Studies*. 3rd ed. London: Routledge, 2002.
- Beausang, Chris. "A Brief History of the Theory and Practice of Computational Literary Criticism (1963–2020)." *magazén*, no. 2 (December 2020). <https://doi.org/10.30687/mag/2724-3923/2020/02/002>.
- Beausang, Chris. "Diachronic Delta: A Computational Method for Analysing Periods of Accelerated Change in Literary Datasets." *Digital Scholarship in the Humanities* 37, no. 3 (August 2022): 644–659. <https://doi.org/10.1093/llc/fqabo41>.
- Beine, J. J. "Introducing NeoLatDraCor, a Research-Oriented Corpus for the Study of Neo-Latin Drama." Poster presented at the 19th Congress of the International Association for Neo-Latin Studies, Aix-en-Provence, 2025. <https://doi.org/10.5281/zenodo.15551942>.
- Benjamin, Walter. *Ursprung des deutschen Trauerspiels*. Revised ed., edited by Rolf Tiedemann. Frankfurt am Main: Suhrkamp, 1963.
- Biber, Douglas. "Representativeness in Corpus Design." *Literary and Linguistic Computing* 8, no. 4 (1993): 243–257. <https://doi.org/10.1093/llc/8.4.243>.
- Biet, Christian. "French Tragedy during the Seventeenth Century: From Cruelty on a Scaffold to Poetic Distance on Stage." In *Politics and Aesthetics in European Baroque and Classicist Tragedy*, edited by Jan Bloemendal and Nigel Smith, 294–316. Leiden: Brill, 2016. https://doi.org/10.1163/97890004323421_012.
- Bloemendal, Jan, and Nigel Smith. "Introduction." In *Politics and Aesthetics in European Baroque and Classicist Tragedy*, edited by Jan Bloemendal and Nigel Smith, 1–39. Leiden: Brill, 2016. https://doi.org/10.1163/97890004323421_002.

- Bod, Rens. *A New History of the Humanities: The Search for Principles and Patterns from Antiquity to the Present*. Oxford: Oxford University Press, 2013.
- Bod, Rens. "Has There Ever Been a Divide? A Longue Durée Perspective." *History of Humanities* 3, no. 1 (2018): 15–25. <https://doi.org/10.1086/696299>.
- Bonfatti, Emilio. "La scena tedesca tra particolarismo e gran teatro del mondo." In *Storia del teatro moderno e contemporaneo*, vol. 1, edited by Roberto Alonge and Guido Davico Bonino, 805–865. Torino: Einaudi, 2000.
- Bories, Anne-Sophie, Pablo Ruiz Fabo, and Petr Plecháč. "The Polite Revolution of Computational Literary Studies." In *Computational Stylistics in Poetry, Prose, and Drama*, edited by Anne-Sophie Bories, Petr Plecháč, and Pablo Ruiz Fabo, 1–18. Berlin: De Gruyter, 2022. <https://doi.org/10.1515/9783110781502-001>.
- Börner, Ingo, and Peer Trilcke. "CLS INFRA D7.1 On Programmable Corpora." February 2023. <https://zenodo.org/record/7664964>.
- Börner, Ingo, and Peer Trilcke. "CLS INFRA D7.3 On Versioning Living and Programmable Corpora." May 2024. <https://doi.org/10.5281/zenodo.11081934>.
- Börner, Ingo, Frank Fischer, Luca Giovannini, Carsten Milling, Christopher Lu, Henny Sluyter-Gäthje, Daniil Skorinkin, and Peer Trilcke. "Onboard onto DraCor: Prototyping Workflows to Homogenize Drama Corpora." In *DHd2023 Book of Abstracts*. Belval/Trier: University of Trier, 2023a. <https://doi.org/10.5281/zenodo.7715333>.
- Börner, Ingo, Peer Trilcke, Carsten Milling, Frank Fischer, and Henny Sluyter-Gäthje. "Dockerizing DraCor: A Container-based Approach to Reproducibility in Computational Literary Studies." In *DH2023 Book of Abstracts*. Graz: University of Graz, 2023b. <https://doi.org/10.5281/zenodo.8107836>.
- Bourassa, André G. "Répertoire des pièces de la Renaissance française, 1537–1615." February 2009. <https://www.theatrales.uqam.ca/soufflebaroque.html>.

- Brainerd, Barron, and Victoria Neufeldt. "On Marcus' Methods for the Analysis of the Strategy of a Play." *Poetics* 3, no. 2 (January 1974): 31–74. [https://doi.org/10.1016/0304-422X\(74\)90013-8](https://doi.org/10.1016/0304-422X(74)90013-8).
- Brauneck, Manfred, ed. *Spieltexte der Wanderbühne. Band 2: Liebeskampff* (1630). Berlin: Walter de Gruyter, 1975.
- Brockett, Oscar G., and Franklin J. Hildy. *History of the Theatre*. 10th ed. London: Pearson, 2013.
- Buringh, Eltjo, and Jan Luiten van Zanden. "Charting the 'Rise of the West': Manuscripts and Printed Books in Europe, A Long-Term Perspective from the Sixth through Eighteenth Centuries." *The Journal of Economic History* 69, no. 2 (June 2009): 409–445. <https://doi.org/10.1017/S0022050709000837>.
- Calder, Andrew. *Molière: The Theory and Practice of Comedy*. London: Athlone, 1993.
- Capra, Daniela. "I Secoli d'Oro del teatro spagnolo." In *Storia del teatro moderno e contemporaneo*, vol. 1, edited by Roberto Alongo and Guido Davico Bonino, 669–803. Torino: Einaudi, 2000.
- Cardinali, Giacomo. "George Buchanan's Tragedies and Contemporary Dramatic Theory." In *George Buchanan: Poet and Dramatist*, edited by Philip Ford and Roger P. H. Green, 163–182. Swansea: Classical Press of Wales, 2009.
- Carroll, Joseph. *Evolution and Literary Theory*. Columbia: University of Missouri Press, 1995.
- Carroll, Joseph. "Evolutionary Literary Theory." In *A Companion to Literary Theory*, edited by David Richter, 425–438. Hoboken: Wiley-Blackwell, 2018.
- Casanova, Pascale. *La république mondiale des lettres*. Paris: Éditions du Seuil, 1999.
- Choi, Seung-Seok, Sung-Hyuk Cha, and Charles C. Tappert. "A Survey of Binary Similarity and Distance Measures." *Journal of Systemics, Cybernetics and Informatics* 8 (2010): 43–48. <https://www.iiisci.org/journal/pdv/sci/pdfs/GS315JG.pdf>.
- Ciotti, Fabio. "Computational Approaches to Literary Periodization: An Experiment in Italian Narrative of 19th and 20th Century." In *DH2022 Book of Abstracts*, 181–183. Tokyo: University of Tokyo, 2022.

- Civardi, Jean-Marc, ed. *La querelle du 'Cid' (1637–38)*. Paris: Champion, 2004.
- Clubb, Louise George. *Italian Drama in Shakespeare's Time*. New Haven: Yale University Press, 1990. <https://doi.org/10.2307/j.ctt211qz0b>.
- Cohen, Margaret. *The Sentimental Education of the Novel*. Princeton: Princeton University Press, 2002.
- Conte, Antonio. "La rinascita della «Poetica» nel Cinquecento italiano." In *La «Poetica» di Aristotele e la sua storia*, edited by Diego Lanza. Pisa: ETS, 2003.
- Cordner, Michael. "Playwright versus Priest: Profanity and the Wit of Restoration Comedy." In *The Cambridge Companion to English Restoration Theatre*, edited by Deborah Payne Fisk, 209–225. Cambridge: Cambridge University Press, 2000. <https://doi.org/10.1017/CCOLO521582156.013>.
- Cuéllar, Álvaro. "Cronología y estilometría: datación automática de comedias de Lope de Vega." *Anuario Lope de Vega. Texto, literatura, cultura* XXIX (2023): 97–130. <https://doi.org/10.5565/rev/anuariolopedevga.483>.
- Da, Nan Z. "The Computational Case against Computational Literary Studies." *Critical Inquiry* 45, no. 3 (March 2019): 601–639. <https://doi.org/10.1086/702594>.
- Dennerlein, Katrin, Thomas Schmidt, and Christian Wolff. "EmoDrama. Ein Korpus mit Emotionsinformationen in Dramen von 1650–1815." *Zeitschrift für digitale Geisteswissenschaften* 8 (2023). https://doi.org/10.17175/2023_010.
- Dennerlein, Katrin, Martin Rupnig, and Christian Reul. "Zum Aufbau digitaler Dramenkorpora. OCR4alltoDraCorTEI als Baustein für die Edition von maschinenlesbaren Versionen historischer Dramendrucke." Paper presented at DHD 2025 Under Construction, Bielefeld, February 26, 2025. <https://doi.org/10.5281/zenodo.14942992>.
- Desagulier, Guillaume. "Digital Corpora." In *Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics*, 51–67. Cham: Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-64572-8_3.

- Dobson, James E. "Vector Hermeneutics: On the Interpretation of Vector Space Models of Text." *Digital Scholarship in the Humanities* 37, no. 1 (2021): 81–93. <https://doi.org/10.1093/llc/fqab079>.
- Douguet, Marc. *Composition dramatique et liaison des scènes dans le théâtre français du XVIIe siècle*. Genève: Droz, 2023.
- Ehrlicher, Hanno, Jörg Lehmann, Nils Reiter, and Marcus Willand. "La poética dramática desde una perspectiva cuantitativa: la obra de Calderón de la Barca." *Revista de Humanidades Digitales* 5 (November 2020): 1–25. <https://doi.org/10.5944/rhd.vol.5.2020.27716>.
- Erlin, Matt, Andrew Piper, Douglas Knox, Stephen Pentecost, Michaela Drouillard, Brian Powell, and Cienna Townson. "Cultural Capitals: Modeling Minor European Literature." *Journal of Cultural Analytics* 6, no. 1 (2021). <https://doi.org/10.22148/001c.21182>.
- Escobar, Miguel. *Theater as Data: Computational Journeys into Theater Research*. Ann Arbor: University of Michigan Press, 2021.
- Estill, Laura. "Digital Humanities' Shakespeare Problem." *Humanities* 8, no. 1 (March 2019): 45. <https://doi.org/10.3390/h8010045>.
- Eyers, Tom. *Speculative Formalism: Literature, Theory, and the Critical*. Evanston, IL: Northwestern University Press, 2017.
- Faisant, Claude. *Mort et résurrection de la Pléiade*. Paris: Champion, 1998.
- Ferrer Valls, Teresa. "CATCOM: A Database on Performances of Spanish Classical Theater." *Journal for Early Modern Cultural Studies* 13, no. 4 (2013): 160–163. <https://doi.org/10.1353/jem.2013.0059>.
- Fièvre, Paul, ed. *Théâtre classique*, 2007. <http://www.theatre-classique.fr>.
- Firth, John R. "A Synopsis of Linguistic Theory, 1930–55." *Studies in Linguistic Analysis* 23 (1957): 1–32.
- Fischer, Frank. "Von dramalys.exe zu DraCor: Eine Geschichte der digitalen Dramenanalyse von 1984 bis heute." Talk delivered at the #RosDH Ringvorlesung, University of Rostock, May 2024. <https://le.hkost.github.io/slides/2024-05-02-rosdh/index.html>.
- Fischer, Frank, and Daniil Skorinkin. "Social Network Analysis in Russian Literary Studies." In *The Palgrave Handbook of Digital Russia Studies*, edited by Daria Gritsenko, Mariëlle Wijermars, and Mikhail Kopotev, 517–536. Cham: Springer International Publishing, 2021. https://doi.org/10.1007/978-3-030-42855-6_29.

- Fischer, Frank, Marina Akimova, and Boris Orekhov. "Preface: Data-Driven Formalism." *Journal of Literary Theory* 13, no. 1 (March 2019a): 1–12. <https://doi.org/10.1515/jlt-2019-0001>.
- Fischer, Frank, Mathias Göbel, Dario Kampkaspar, Christopher Kittel, and Peer Trilcke. "Network Dynamics, Plot Analysis: Approaching the Progressive Structuration of Literary Texts." In *DH2017 Book of Abstracts*. Montréal: McGill University, 2017. <https://dh2017.adho.org/abstracts/071/071.pdf>.
- Fischer, Frank, Carsten Milling, and Mark Schwindt. "Streamlining the DraCor API." 2024. <https://weltliteratur.net/streamlining-the-DraCor-api/>.
- Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtel, Christopher Kittel, Carsten Milling, and Peer Trilcke. "Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama." In *DH2019 Book of Abstracts*. Utrecht: University of Utrecht, 2019b. <https://doi.org/10.5281/zenodo.4284002>.
- Fischer, Frank, Peer Trilcke, Christopher Kittel, Carsten Milling, and Daniil Skorinkin. "To Catch a Protagonist: Quantitative Dominance Relations in German-Language Drama (1730–1930)". In *DH2018 Book of Abstracts*. Mexico City: Red de Humanidades Digitales A. C., 2018. <https://dh2018.adho.org/en/to-catch-a-protagonist-quantitative-dominance-relations-in-german-language-drama-1730-1930/>.
- Forestier, Georges. "De la modernité anti-classique au classicisme moderne. Le modèle théâtral (1628–1634)." *Littératures classiques* 19 (1993): 87–128.
- Friedland, Louis Sigmund. "The Dramatic Unities in England (Continued)." *The Journal of English and Germanic Philology* 10, no. 2 (1911): 280–299. <https://www.jstor.org/stable/27700090>.
- Fucks, Wilhelm. *Nach allen Regeln der Kunst*. Stuttgart: Deutsche Verlags-Anstalt, 1968.
- Fucks, Wilhelm. "Über den Gesetzesbegriff einer exakten Literaturwissenschaft, erläutert an Sätzen und Satzfolgen." *Zeitschrift für Literaturwissenschaft und Linguistik* I/II (1971): 113–137.

- Gallucci, Giorgia, and Luca Giovannini. “‘Un mare senza lidi’: digitalizzazione e modellazione della Drammaturgia di Leone Allacci”. *Umanistica Digitale* 20 (2025): 583–599.
- García Berrio, Antonio, and Javier Huerta Calvo. *Los géneros literarios: sistema e historia (una introducción)*. Madrid: Cátedra, 1992.
- Gasparov, Mikhail. “Boris Yarkho’s Works on Literary Theory.” *Studia Metrica et Poetica* 3, no. 2 (January 2017): 130–150. <https://doi.org/10.12697/smp.2016.3.2.05>.
- Genette, Gérard. *Figures III*. Paris: Seuil, 1972.
- Gilodi, Roberto. “Morphology and Literature.” In *Bilddenken und Morphologie*, edited by Laura Follesa and Federico Vercellone, 115–132. Berlin: De Gruyter, 2021. <https://doi.org/10.1515/9783110674194-008>.
- Giovannini, Luca, Ingo Börner, Frank Fischer, Carsten Milling, Daniil Skorinkin, and Peer Trilcke. “Quali prospettive per ItaDraCor? Risorse e strumenti per la codifica di testi drammatici in lingua italiana.” In *AIUCD 2023 Book of Abstracts*, 108–111. Siena: University of Siena, 2023. <https://doi.org/10.5281/zenodo.10009101>.
- Giovannini, Luca, and Giorgia Gallucci. “Allacci Digitale: A Historical Dataset for Early Modern Italian Drama”. *Journal of Open Humanities Data* 10 (2024): 55. <https://doi.org/10.5334/johd.250>.
- Giovannini, Luca, and Daniil Skorinkin. “Computational Approaches to Opera Libretti.” *Journal of Computational Literary Studies* 3 (2024): 1–23. <https://doi.org/10.48694/jcls.3595>.
- Giovannini, Luca, and Daniil Skorinkin. “Towards a LLM-powered Encoding Workflow for Plays / Hacia un flujo de trabajo de codificación para obras de teatro impulsado por LLM.” In *TEI2024 Book of Abstracts*, 94–95. Buenos Aires: Universidad del Salvador, 2024. <https://zenodo.org/records/13883242>.
- Giovannini, Luca, and Daniil Skorinkin. “Computational Modelling of Literary Characters.” Talk delivered at the Digital Humanities colloquium, Friedrich-Alexander University of Erlangen-Nuremberg, January 28, 2025. <https://lucagiovannini7.github.io/slides/dh-erlangen-2025.html>.

- Gius, Evelyn, and Janina Jacke. "Are Computational Literary Studies Structuralist?" *Journal of Cultural Analytics* 7, no. 4 (2022). <https://doi.org/10.22148/001c.46662>.
- Gius, Evelyn, Fotis Jannidis, Markus Krug, Albin Zehe, Andreas Hotho, Frank Puppe, Jonathan Krebs, Nils Reiter, Nathalie Wiedmer, and Leonard Konle. "Detection of scenes in fiction". In *DH2019 Book of Abstracts*. Utrecht: University of Utrecht, 2019. <https://web.archive.org/web/20200811094038/https://dev.clariah.nl/files/dh2019/boa/0608.html>. Grootendorst, Maarten. "9 Distance Measures in Data Science." February 2021. <https://towardsdatascience.com/9-distance-measures-in-data-science-918109d069fa>.
- Hatzfeld, Helmut. "The Baroque from the Viewpoint of the Literary Historian." *The Journal of Aesthetics and Art Criticism* 14, no. 2 (1955): 156. <https://doi.org/10.2307/425853>.
- Henke, Robert, and Eric Nicholson, eds. *Transnational Exchange in Early Modern Theater*. Aldershot: Routledge, 2008.
- Henke, Robert, and Eric Nicholson. "Introduction." In *Transnational Mobilities in Early Modern Theater*, edited by Robert Henke and Eric Nicholson. Aldershot: Routledge, 2014a. <https://doi.org/10.4324/9781315549811>.
- Henke, Robert, and Eric Nicholson, eds. *Transnational Mobilities in Early Modern Theater*. Farnham: Ashgate, 2014b. <https://doi.org/10.4324/9781315549811>.
- Herrick, Marvin T., ed. *Italian Plays, 1500–1700, in the University of Illinois Library*. Urbana: University of Illinois Press, 1964.
- Herrmann, Berenike, and Gerhard Lauer. "Korpusliteraturwissenschaft. Zur Konzeption und Praxis am Beispiel eines Korpus zur literarischen Moderne." *Osnabrücker Beiträge zur Sprachtheorie*, no. 92 (2018). <https://pub.uni-bielefeld.de/record/2955632>.
- Herrmann, Berenike, Anne-Sophie Bories, Francesca Frontini, Clémence Jacquot, Steffen Pielström, Simone Rebora, Geoffrey Rockwell, and Stéfan Sinclair. "Tool Criticism in Practice. On Methods, Tools and Aims of Computational Literary Studies." *DHQ: Digital Humanities Quarterly* 17, no. 2 (2023). <http://www.digitalhumanities.org/dhq/vol/17/2/000687/000687.html>.

- Hirsh, James E. *The Structure of Shakespearean Scenes*. New Haven: Yale University Press, 1981. <https://archive.org/details/structureofshake0000hirs>.
- Hoftijzer, Paul G. "Dutch Printing and Bookselling in the Golden Age." In *Two Faces of the Early Modern World: The Netherlands and Japan in the 17th and 18th Centuries*, 59–67. Kyoto: International Research Center for Japanese Studies, 2001. <https://doi.org/10.15055/00001557>.
- Howard-Hill, Trevor H. "The Evolution of the Form of Plays in English during the Renaissance." *Renaissance Quarterly* 43, no. 1 (1990): 112–145. <https://doi.org/10.2307/2861794>.
- Howe, Elizabeth. *The First English Actresses: Women and Drama, 1660–1700*. Cambridge: Cambridge University Press, 2001.
- Ilseman, Hartmut. "Computerized Drama Analysis." *Literary and Linguistic Computing* 10, no. 1 (January 1995): 11–21. <https://doi.org/10.1093/lc/10.1.11>.
- Ilseman, Hartmut. *Handbuch Internet Drama Analyse Programm — Programmbeschreibung*. 2006. <https://web.archive.org/web/20150905080328/http://www.shakstat.engsem.uni-hannover.de/p.pdf>.
- Ioannidis, John P. A. "Why Most Published Research Findings Are False." *PLoS Medicine* 2, no. 8 (August 2005): e124. <https://doi.org/10.1371/journal.pmed.0020124>.
- Jannidis, Fotis. "On the Perceived Complexity of Literature. A Response to Nan Z. Da." *Journal of Cultural Analytics* 5, no. 1 (January 2020). <https://doi.org/10.22148/001c.11829>.
- Jannidis, Fotis. "Digitale Literaturwissenschaft. Zur Einführung." In *Digitale Literaturwissenschaft*, edited by Fotis Jannidis, 1–16. Stuttgart: J.B. Metzler, 2022. https://doi.org/10.1007/978-3-476-05886-7_1.
- Jansen, Steen. "Esquisse d'une théorie de la forme dramatique." *Lan-gages*, no. 12 (1968): 71–93. <http://www.jstor.org/stable/41680691>.
- Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. Champaign, IL: University of Illinois Press, 2013.
- Jokisch, Jan Niklas, and Antonio Rojas Castro. "Revisiting Network Analysis in Drama: Operational Challenges and Methodological Insights." Paper presented at *Digital Humanities 2025*, Lisbon, July 16, 2025.

- Jordan, Gilbert J. "Die Akt- und Szeneneinteilung im deutschen Drama." *The Journal of English and Germanic Philology* 38, no. 3 (1939): 396–415.
- Joyeux-Prunel, Béatrice. "Digital Humanities in the Era of Digital Reproducibility: Towards a Fairest and Post-Computational Framework." *International Journal of Digital Humanities* 6 (2024): 23–43. <https://doi.org/10.1007/s42803-023-00079-6>.
- Kafitz, Dieter. *Grundzüge einer Geschichte des deutschen Dramas von Lessing bis zum Naturalismus*. 2nd ed. Königstein: Athenäum, 1989.
- Katritzky, M. A., and Pavel Drábek, eds. *Transnational Connections in Early Modern Theatre*. Manchester: Manchester University Press, 2019.
- Krautter, Benjamin. "Die Operationalisierung als interdisziplinäre Schnittstelle der Digital Humanities." *Scientia Poetica* 26, no. 1 (November 2022): 215–244. <https://doi.org/10.1515/scipo-2022-009>.
- Krautter, Benjamin. "Skalierungspraktiken in der computergestützten Analyse von literarischen Texten." In *DHd2023 Book of Abstracts*, edited by Anna Busch and Peer Trilcke, 291–295. Belval/Trier: Zenodo, March 2023. <https://doi.org/10.5281/ZENODO.7688632>.
- Krautter, Benjamin. "The Scales of (Computational) Literary Studies: Martin Mueller's Concept of Scalable Reading in Theory and Practice." In *Zoomland*, edited by Florentina Armaselu and Andreas Fickers, 261–286. Berlin: De Gruyter, 2024. <https://doi.org/10.1515/978311317779-011>.
- Kretz, Nicolette. "Grundelemente (1): Bausteine des Dramas (Figur, Handlung, Dialog)." In *Handbuch Drama*, edited by Peter W. Marx, 105–121. Stuttgart: J.B. Metzler, 2012. https://doi.org/10.1007/978-3-476-00512-0_9.
- Krämer, Sybille. "Should I Really 'Hermeneutise' the Digital Humanities? A Plea for the Epistemic Productivity of a 'Cultural Technique of Flattening' in the Humanities." *Journal of Cultural Analytics* 7, no. 4 (January 2023a). <https://doi.org/10.22148/001c.55592>.
- Krämer, Sybille. "The Cultural Technique of Flattening." *Metode* 1 (2023b). <https://metode.r-o-m.no/en/articles/essay/the-cultural-technique-of-flattening>.
- Küpper, Joachim. *The Cultural Net: Early Modern Drama as a Paradigm*. Berlin: De Gruyter, 2018. <https://doi.org/10.1515/9783110536638>.

- Labatut, Vincent, and Xavier Bost. "Extraction and Analysis of Fictional Character Networks: A Survey." *ACM Computing Surveys* 52, no. 5 (September 2020): 1–40. <https://doi.org/10.1145/3344548>.
- Lande, Joel B. "German Trauerspiel and Its International Nexus: On the Migration of Poetic Forms." In *Politics and Aesthetics in European Baroque and Classicist Tragedy*, edited by Jan Bloemendal and Nigel Smith, 319–343. Leiden: Brill, 2016. https://doi.org/10.1163/9789004323421_013.
- Lechner, Doris. "And Am I Thus Rewarded? The Rejected Hero and the Raped Heroine in Mary Pix's Ibrahim." *Zeitschrift für Anglistik und Amerikanistik* 64, no. 4 (2016): 367–383. <https://doi.org/10.1515/zaa-2016-0035>.
- Leopold, Silke. "Die Anfänge von Oper und die Probleme der Gattung." *Journal of Seventeenth-Century Music* 9, no. 1 (2003). <https://sscm-jscm.org/v9/no1/leopold.html>.
- Leufkens, Kirsten. "Data Visualization Using Principal Component Analysis within the Digital Humanities." Master's thesis, University of Utrecht, 2020. <https://studenttheses.uu.nl/handle/20.500.12932/35459>.
- Lever, Jake, Martin Krzywinski, and Naomi Altman. "Principal Component Analysis." *Nature Methods* 14, no. 7 (July 2017): 641–642. <https://doi.org/10.1038/nmeth.4346>.
- Levi, Anthony. "L'Europe au XVIIe siècle: unité ou pluralité des cultures?" In *Horizons européens de la littérature française au XVIIe siècle: l'Europe: lieu d'échanges culturels? La circulation des œuvres et des jugements au XVIIe siècle*, edited by Wolfgang Leiner. Tübingen: Narr, 1988.
- Lochert, Véronique. *L'écriture du spectacle: les didascalies dans le théâtre européen aux XVIe et XVIIe siècles*. Genève: Droz, 2009.
- Lochert, Véronique. "A. Acte." In *Le Théâtre au miroir des langues*, edited by Véronique Lochert, Marc Vuillermoz, and Enrica Zanin, 191–200. Genève: Droz, 2018a.
- Lochert, Véronique. "B. Scène." In *Le Théâtre au miroir des langues*, edited by Véronique Lochert, Marc Vuillermoz, and Enrica Zanin, 201–208. Genève: Droz, 2018b.

- Lotman, Jurij M. "Problema 'obucheniya culture' kak ee tipologicheskaya kharakteristik [The Problem of 'Learning Culture' as Its Typological Characteristic]." *Trudy po znakovym sistemam* 5 (1971): 167–176. <http://hdl.handle.net/10062/47714>.
- Lotman, Jurij M. "La cultura e il suo 'insegnamento' come caratteristica tipologica." In *Tipologia della cultura*, by Boris A. Uspenskij and Jurij M. Lotman, 69–81. Milano: Bompiani, 1985.
- Lotman, Jurij M., and Boris A. Uspenskij. "On the Semiotic Mechanism of Culture." *New Literary History* 9, no. 2 (1978): 211–232. <http://www.jstor.org/stable/468571>.
- Lvoff, Basil. "Distant Reading in Russian Formalism and Russian Formalism in Distant Reading." *Russian Literature* 122–123 (May 2021): 29–65. <https://doi.org/10.1016/j.ruslit.2021.07.003>.
- Maaten, Laurens van der, and Geoffrey Hinton. "Visualizing Data Using t-SNE." *Journal of Machine Learning Research* 9 (2008): 2579–2605. <https://www.jmlr.org/papers/volume9/vandermaaten08a/vandermaaten08a.pdf>.
- Maaten, Laurens van der, Eric Postma, and Jaap van den Herik. "Dimensionality Reduction: A Comparative Review." Technical Report TiCC-TR 2009–005, Tilburg Centre for Creative Computing, Tilburg University, 2009. https://lvdmaaten.github.io/publications/papers/TR_Dimensionality_Reduction_Review_2009.pdf.
- Manjavacas, Enrique, and Lauren Fonteyn. "Adapting vs. Pre-training Language Models for Historical Languages." *Journal of Data Mining & Digital Humanities* (2022). <https://doi.org/10.46298/jdmdh.9152>.
- Manovich, Lev. *Cultural Analytics*. Cambridge, MA: MIT Press, 2020.
- Marcus, Leah S. "Antitheatricality: The Theater as Scourge." In *A New Companion to Renaissance Drama*, edited by Arthur F. Kinney and Thomas W. Hopper, 182–192. Hoboken: Wiley Blackwell, 2017. <https://doi.org/10.1002/9781118824016.ch14>.
- Marcus, Solomon. *Poetica matematică*. București: Academiei, 1970.
- Marcus, Solomon. "Ein mathematisch-linguistisches Dramenmodell." *Zeitschrift für Literaturwissenschaft und Linguistik* 1, no. 1 (1971): 139. <https://www.proquest.com/scholarly-journals/ein-mathematisch-l>

- inguistisches-dramenmodell/docview/1300203364/se-2?accountid=14491.
- Marcus, Solomon. "The Theater of Mathematics and the Mathematics of the Theatre." In *BRIDGES 1999: Conference Proceedings*, edited by Reza Sarhangi, 293–294. Winfield, KS: Southwestern College, 1999. <http://t.archive.bridgesmathart.org/1999/bridges1999-293.pdf>.
- Marenco, Franco. "Gli inizi del teatro moderno." In *Storia del teatro moderno e contemporaneo*, vol. 1, edited by Roberto Alonge and Guido Davico Bonino, 277–472. Torino: Einaudi, 2000.
- Margoni, Thomas, and Mark Perry. "Scientific and Critical Editions of Public Domain Works: An Example of European Copyright Law (Dis)Harmonization." *Canadian Intellectual Property Review* 27 (2011): 157–161. <https://ssrn.com/abstract=1961535>.
- Marx, William. "Brève histoire de la forme en littérature." *Les Temps Modernes* 676, no. 5 (December 2013): 35–47. <https://doi.org/10.3917/lm.676.0035>.
- McInnes, Leland, John Healy, and James Melville. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction." arXiv preprint, 2018. <https://doi.org/10.48550/arXiv.1802.03426>.
- Meneghelli, Donata. "Periodization, Comparative Literature, and Italian Modernism." *CLCWeb: Comparative Literature and Culture* 15, no. 7 (December 2013). <https://doi.org/10.7771/1481-4374.2386>.
- Mesoudi, Alex. "Cultural Evolution: A Review of Theory, Findings and Controversies." *Evolutionary Biology* 43, no. 4 (2016): 481–497. <https://doi.org/10.1007/s11692-015-9320-0>.
- Meyer, Reinhart, ed. *Bibliographia dramatice et dramaticorum*. Vol. 2, pt. 2, *Einzeltitel 1701–1708*. Tübingen: Niemeyer, 1993.
- Moretti, Franco. "La letteratura europea." In *Storia d'Europa. L'Europa oggi*, vol. 1, edited by Perry Anderson, Walter Barberis, and Carlo Ginzburg, 837–866. Torino: Einaudi, 1993.
- Moretti, Franco. "Modern European Literature: A Geographical Sketch." *New Left Review*, no. 206 (1994): 86–109.
- Moretti, Franco. "The Slaughterhouse of Literature." *Modern Language Quarterly* 61, no. 1 (2000): 207–227.

- Moretti, Franco. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso, 2007.
- Moretti, Franco. *Distant Reading*. London: Verso, 2013a.
- Moretti, Franco. "Operationalizing: or, the Function of Measurement in Modern Literary Theory." Stanford Literary Lab Pamphlet 6, December 2013b. <https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf>.
- Moretti, Franco. "La critica quantitativa come progetto e come problema." *Intersezioni*, no. 3 (2021): 381–390. <https://doi.org/10.1404/102239>.
- Moretti, Franco. *Falso movimento. La svolta quantitativa nello studio della letteratura*. Milano: nottetempo, 2022.
- Moretti, Franco. "Network Semantics. A Beginning." *Transilvania*, no. 5 (2024): 5–11. <https://doi.org/10.51391/trva.2024.05.01>.
- Moretti, Franco, and Oleg Sobchuk. "Hidden in Plain Sight: Data Visualization in the Humanities." *New Left Review* 118 (2019): 86–115. <https://newleftreview.org/issues/iii18/articles/franco-moretti-oleg-sobchuk-hidden-in-plain-sight>.
- Morin, Olivier. *How Traditions Live and Die*. Oxford: Oxford University Press, 2016.
- Mrugalski, Michał, Carolin Odebrecht, Vera Charvat, Ingo Börner, and Matej Ďurčo. "CLS INFRA D5.1. Review of the Data Landscape." July 2022. <https://doi.org/10.5281/zenodo.6861022>.
- Mueller, Martin. "Scalable Reading." 2020. <https://sites.northwestern.edu/scalablereading/2020/04/26/scalable-reading/>.
- Newels, Margarete. *Die dramatischen Gattungen in den Poetiken des Siglo de Oro: eine einleitende Studie zum Thema der Dramentheorie im goldenen Zeitalter*. Wiesbaden: Steiner, 1959.
- Niefanger, Dirk. "Barock." In *Handbuch Drama: Theorie, Analyse, Geschichte*, edited by Peter W. Marx, 230–241. Stuttgart: Metzler, 2012.
- Norman, Buford. *Touched by the Graces: The Libretti of Philippe Quinault in the Context of French Classicism*. Birmingham, AL: Summa, 2001.
- Oleza, Joan, and Jesús Tronch. "Emothe: un proyecto digital para el teatro europeo de los siglos XVI y XVII." *Viceversa*, no. 109 (2020): 56–59.

- Open Science Collaboration. "Estimating the Reproducibility of Psychological Science." *Science* 349, no. 6251 (August 2015). <https://doi.org/10.1126/science.aac4716>.
- O'Sullivan, James. "The Humanities Have a 'Reproducibility' Problem." 2019. <https://talkinghumanities.blogs.sas.ac.uk/2019/07/09/the-humanities-have-a-reproducibility-problem/>.
- Pagel, Janis, Benjamin Krautter, Melanie Andresen, Marcus Willand, and Nils Reiter. "On Designing Collaboration in a Mixed-Methods Scenario. Reflecting Quantitative Drama Analytics." In *Mixing Methods: Practical Insights from the Humanities in the Digital Age*, edited by Birgit Schneider, Beate Löffler, Tino Mager, and Carola Hein, 81–102. Bielefeld, 2023. <https://doi.org/10.1515/9783839469132-010>.
- Peels, Rik, and Lex Bouter. "The Possibility and Desirability of Replication in the Humanities." *Palgrave Communications* 4, no. 1 (2018). <https://doi.org/10.1057/s41599-018-0149-x>.
- Perella, Nicolas J. "The Autonomy of Poetry in Battista Guarini's Polemical Tracts." *Forum Italicum: A Journal of Italian Studies* 7, no. 2 (June 1973): 338–352. <https://doi.org/10.1177/001458587300700209>.
- Pfister, Manfred. *The Theory and Analysis of Drama*. Cambridge: Cambridge University Press, 1988.
- Pichler, Axel, and Nils Reiter. "From Concepts to Texts and Back: Operationalization as a Core Activity of Digital Humanities." *Journal of Cultural Analytics* 7, no. 4 (December 2022). <https://doi.org/10.22148/001c.57195>.
- Pierazzo, Elena. *Digital Scholarly Editing: Theories, Models and Methods*. London: Routledge, 2020.
- Pilshchikov, Igor. "Apoética quantitativa do Formalismo Russo." *RUS (São Paulo)* 11, no. 16 (September 2020): 9–42. <https://doi.org/10.11606/issn.2317-4765.rus.2020.172896>.
- Pilshchikov, Igor. "Formalismo cuantitativo 'viejo' y 'nuevo' (El Círculo Lingüístico de Moscú y el Laboratorio Literario de Stanford)." *Literatura: teoría, historia, crítica* 24, no. 1 (January 2022): 265–295. <https://doi.org/10.15446/lthc.v24n1.98440>.
- Piotrowski, Michael. *Natural Language Processing for Historical Texts*. Cham: Springer, 2012.

- Piper, Andrew. "Do We Know What We Are Doing?" *Journal of Cultural Analytics* 5, no. 1 (2020). <https://doi.org/10.22148/001c.11826>.
- Propp, Vladimir. *Morphology of the Folktale*. 2nd ed. Austin: University of Texas Press, 1968.
- Ranke, Wolfgang. "Schillers Shakespeare: Von den 'Räubern' zum Weimarer 'Macbeth.'" *Jahrbuch der Deutschen Schillergesellschaft* 54 (2010): 706–724. <https://doi.org/10.46500/83530771-031>.
- Refini, Eugenio. "«Profiter de quelque chose à ma république». Poetica e politica nell'epistola 'De l'art de la tragédie' di Jean de La Taille." *Studi Francesi*, no. 158 (July 2009): 234–250. <https://doi.org/10.4000/studifrancesi.7787>.
- Reppen, Randi. "Building a Corpus: What Are Key Considerations?" In *The Routledge Handbook of Corpus Linguistics*, 13–20. London: Routledge, 2022.
- Reul, Christian, Dennis Christ, Alexander Hartelt, Nico Balbach, Maximilian Wehner, Uwe Springmann, Christoph Wick, Christine Grundig, Andreas Büttner, and Frank Puppe. "OCR4all—An Open-Source Tool Providing a (Semi-)Automatic OCR Workflow for Historical Printings." *Applied Sciences* 9, no. 22 (November 2019): 4853. <https://doi.org/10.3390/app9224853>.
- Revzina, Olga G., and Isaak I. Revzin. "Nekotorye matematicheskie metody analiza dramaturgicheskogo postroeniia [Some Mathematical Methods for Analyzing Dramatic Constructions]." In *Tochnye metody v issledovaniakh kul'tury i iskusstva (materialy k simpoziumy)*, 291–300. 1971.
- Romanska, Magda. "Drametrics: What Dramaturgs Should Learn from Mathematicians." In *The Routledge Companion to Dramaturgy*. London: Routledge, 2014. <https://doi.org/10.4324/9780203075944-88>.
- Ruggerio, Michael J. "Some Approaches to Structure in the Spanish Golden Age 'Comedia.'" *Orbis Litterarum* 28, no. 3 (September 1973): 173–191. <https://doi.org/10.1111/j.1600-0730.1973.tb01866.x>.
- Russo, Luigi. "La tragedia nel cinque e seicento." *Belfagor* 14 (1959): 14–21.
- Sánchez Laílla, Luis. "«Dice Aristóteles»: la reescritura de la Poética en los Siglos de Oro." *Críticón* 79 (2000): 9–36. https://cvc.cervantes.es/literatura/criticon/pdf/079/079_011.pdf.

- Sapogov, Vyacheslav A. "Nekotorye kharakteristiki dramaturgicheskogo postroeniia komedii A. N. Ostrovskogo 'Les' [Some Characteristics of the Dramatic Construction of A. N. Ostrovsky's Comedy 'The Forest']." In *A. N. Ostrovskij i russkaia literatura* [A. N. Ostrovsky and Russian Literature], edited by Vyacheslav Alexandrovich Sapogov, 60–69. Kostroma: Iaroslavskij pedagogicheskij institut, 1974.
- Savoy, Jacques. "Stylometric Analysis of Characters in Shakespeare's Plays." *Digital Scholarship in the Humanities* 38, no. 3 (September 2023): 1238–1246. <https://doi.org/10.1093/llc/fqac092>.
- Schings, Hans-Jürgen. "Consolatio Tragoediae: Zur Theorie des barocken Trauerspiels." In *Deutsche Dramentheorien*, edited by Reinhold Grimm, 19–55. 3rd ed. Wiesbaden: Akademische Verlagsgesellschaft Athenaion, 1980.
- Schlegel, August Wilhelm. "Erste Vorlesung." In *Vorlesungen über dramatische Kunst und Litteratur* [1809–1811]. Leiden: Brill Schöningh, 2018. https://doi.org/10.30965/9783657778546_004.
- Schmidt, Thomas, Manuel Burghardt, Katrin Dennerlein, and Christian Wolff. "Katharsis—A Tool for Computational Drametrics." In *DH2019 Book of Abstracts*. Utrecht: University of Utrecht, 2019. <https://epub.uni-regensburg.de/43579/1/Katharsis%20%E2%80%93%20A%20Tool%20for%20Computational%20Drametrics.pdf>.
- Schöch, Christof. "Aufbau von Datensammlungen." In *Digital Humanities: Eine Einführung*, edited by Fotis Jannidis, Hubertus Kohle, and Malte Rehbein, 223–233. Stuttgart: J.B. Metzler, 2017a. https://doi.org/10.1007/978-3-476-05446-3_16.
- Schöch, Christof. "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama." *Digital Humanities Quarterly* 11, no. 2 (2017b). <https://dhq.digitalhumanities.org/vol/11/2/000291/000291.html>.
- Schöch, Christof, Roxana Patras, Tomaž Erjavec, and Diana Santos. "Creating the European Literary Text Collection (ELTeC): Challenges and Perspectives." *Modern Languages Open*, no. 1 (December 2021): 25. <https://doi.org/10.3828/mlo.voio.364>.
- Schöch, Christof. "Quantitative Semantik. Word Embedding Models für literaturwissenschaftliche Fragestellungen." In *Digitale Literaturwis-*

- senschaft*, edited by Fotis Jannidis, 535–562. Stuttgart: J.B. Metzler, 2022. https://doi.org/10.1007/978-3-476-05886-7_22.
- Schöch, Christof. “Repetitive Research: A Conceptual Space and Terminology of Replication, Reproduction, Revision, Reanalysis, Reinvestigation and Reuse in Digital Humanities.” *International Journal of Digital Humanities* 5, no. 2 (2023): 373–403.
- Shaw, Bernard. *Back to Methuselah: A Metabiological Pentateuch*. Vol. 16. London: Constable, 1921.
- Šeĵa, Artjoms. “Boris Yarkho’s Data on Literary History.” 2020. https://github.com/perechen/data_archaeology/tree/master/yarkho.
- Šeĵa, Artjoms, Ben Nagy, Joanna Byszek, Laura Hernández-Lorenzo, Botond Szemes, and Maciej Eder. “From Stage to Page: Stylistic Variation in Fictional Speech.” In *Computational Drama Analysis: Reflecting on Methods and Interpretations*, edited by Melanie Andresen and Nils Reiter, 149–166. Berlin: De Gruyter, 2024. <https://doi.org/10.1515/9783111071824-008>.
- Skorinkin, Daniil. “EasyDrama: A Lightweight Solution for Encoding Plays in TEI/XML.” In *TEI2024 Book of Abstracts*, 38–39. 2024. <https://doi.org/10.5281/zenodo.13883242>.
- Snow, C. P. *The Two Cultures*. Cambridge: Cambridge University Press, 2012.
- So, Richard Jean. “All Models Are Wrong.” *PMLA* 132, no. 3 (2017): 668–673. <https://doi.org/10.1632/pmla.2017.132.3.668>.
- Sobchuk, Oleg. “Charting Artistic Evolution: An Essay in Theory.” PhD dissertation, University of Tartu, 2018. <https://dspace.ut.ee/items/8fbbf6b8-d5e0-4e50-a05d-cd7f73f44208>.
- Sobchuk, Oleg. “Evolution of Modern Literature and Film.” In *The Oxford Handbook of Cultural Evolution*, edited by Jamshid J. Tehrani, Jeremy Kendal, and Rachel Kendal. Oxford: Oxford University Press, 2023. <https://doi.org/10.1093/oxfordhb/9780198869252.013.45>.
- Sonkin, Mikhail. “Digital Adaptation of Boris Yarkho’s Research on the Works of Pierre Corneille.” 2022. <https://github.com/Migabaj/yarkho>.
- Sperantov, Vladimir V. “Poetika remarki v russkoy tragedii XVIII–nachala XIX vv. (K tipologii literaturnykh napravleniy)

- [The Poetics of Stage Directions in Russian Tragedy of the Eighteenth and Early Nineteenth Centuries (On the Typology of Literary Movements)]." *Philologica* 5, no. 11/13 (1998): 9–48. <https://rvb.ru/philologica/05/05sperantov.htm>.
- Sullivan, Henry W. *Tragic Drama in the Golden Age of Spain: Seven Essays on the Definition of a Genre*. Kassel: Reichenberger, 2018.
- Szemes, Botond, and Bence Vida. "Tragic and Comical Networks: Clustering Dramatic Genres According to Structural Properties." In *Computational Drama Analysis: Reflecting on Methods and Interpretations*, edited by Melanie Andresen and Nils Reiter. Berlin: De Gruyter, 2024. <https://doi.org/10.1515/978311071824-009>.
- Trilcke, Peer. "Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft." In *Empirie in der Literaturwissenschaft*, edited by Philip Ajouri, Katja Mellmann, and Christoph Rauen, 201–247. Münster: Brill, 2013. https://doi.org/10.30965/9783957439710_012.
- Trilcke, Peer, Frank Fischer, Mathias Göbel, Dario Kampkaspar, and Christopher Kittel. "Netzwerkdynamik, Plotanalyse – Zur Visualisierung und Berechnung der progressiven Strukturierung literarischer Texte." In *DHd2017 Book of Abstracts*. Bern, 2017. <https://doi.org/10.5281/ZENODO.4622799>.
- Trilcke, Peer, Christopher Kittel, Nils Reiter, Daria Maximova, and Frank Fischer. "Opening the Stage: A Quantitative Look at Stage Directions in German Drama." In *DH2020 Book of Abstracts*. Ottawa: University of Ottawa, 2020. https://dh2020.adho.org/wp-content/uploads/2020/07/337_OpeningtheStageAQuantitativeLookatStageDirectionsinGermanDrama.html.
- Trilcke, Peer, Evgeniya Ustinova, Frank Fischer, Carsten Milling, and Ingo Börner. "Detecting Small Worlds in a Corpus of Thousands of Theater Plays." In *Computational Drama Analysis: Reflecting on Methods and Interpretations*, edited by Melanie Andresen and Nils Reiter, 7–34. Berlin: De Gruyter, 2024. <https://doi.org/10.1515/978311071824-002>.
- Trivero, Paola. "Il cinquecento italiano: soggetti e varianti del canone tragico." In *La tragédie et son modèle à l'époque de la Renaissance en-*

- tre France, Italie et Espagne*, edited by Michele Mastroianni. Torino: Rosenberg & Sellier, 2015.
- Tukey, John W. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, 1977.
- Underwood, Ted. *Distant Horizons: Digital Evidence and Literary Change*. Chicago: University of Chicago Press, 2019.
- Underwood, Ted, Kevin Kiley, Wenyi Shang, and Stephen Vaisey. "Co-hort Succession Explains Most Change in Literary Culture." *Sociological Science* 9 (2022): 184–205. <https://doi.org/10.15195/v9.a8>.
- Vega y Carpio, Félix Lope de. *The New Art of Writing Plays*. New York: Dramatic Museum of Columbia University, 1914.
- Veltruský, Jiří. "Dramatic Text as a Component of Theatre." In *Theatre Theory Reader: Prague School Writings*, edited by David Drozd, 247–268. Prague: Charles University, Karolinum Press, 2017.
- Vergeer, Tim. *Spanish Drama on the Dutch Stage: Transgressive Emotions in the Seventeenth Century*. Leiden: Brill, 2026.
- Vince, Ronald W. "Act." In *The Oxford Encyclopedia of Theatre and Performance*, vol. 2. Oxford: Oxford University Press, 2003.
- Vitse, Marc. *Éléments pour une théorie du théâtre espagnol du XVIIe siècle*. Toulouse: University of Toulouse-Le Mirail, 1988.
- Warner, Marina. *Monuments and Maidens: The Allegory of the Female Form*. New York: Atheneum, 1985.
- Waumans, Michaël C., Thibaut Nicodème, and Hugues Bersini. "Topology Analysis of Social Networks Extracted from Literature." *PLOS ONE* 10, no. 6 (June 2015): e0126470. <https://doi.org/10.1371/journal.pone.0126470>.
- Weinberg, Bernard. *A History of Literary Criticism in the Italian Renaissance*. Chicago: University of Chicago Press, 1961.
- Weingart, Scott B., Nickoal Eichmann-Kalwara, and Matthew Lincoln. *Index of Digital Humanities Conferences*. 2024. <https://dh-abstracts.library.cmu.edu>.
- Wellek, René. "The Concept of Baroque in Literary Scholarship." *The Journal of Aesthetics and Art Criticism* 5, no. 2 (December 1946): 77–109. <https://doi.org/10.2307/425797>.

- Wellek, René, and Austin Warren. *Theory of Literature*. New York: Harcourt, Brace and Jovanovich, 1949.
- Wendell, Inna A. "A Statistical Analysis of Genre Dynamics: Evolution of the Russian Five-Act Comedy in Verse in the Eighteenth and Nineteenth Centuries." PhD dissertation, University of California, Los Angeles, 2021. <https://escholarship.org/uc/item/9rr5k9p7>.
- West, William N. "Introduction: Italy in the Drama of Europe." *Renaissance Drama* 36/37 (2010): ix–xiv. <http://www.jstor.org/stable/41917450>.
- Wiedemann, Conrad. *Der galante Stil*. Tübingen: Niemeyer, 1969.
- Wilkinson, Mark, Michel Dumontier, IJsbrand Jan Aalbersberg, et al. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3 (2016): 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Willand, Marcus, Peer Trilcke, Christof Schöch, Nanette Rißler-Pipka, Nils Reiter, and Frank Fischer. "Aktuelle Herausforderungen der digitalen Dramenanalyse." In *DHd2017 Book of Abstracts*. Bern, 2017. <https://doi.org/10.5281/ZENODO.4622643>.
- Windelband, Wilhelm. "History and Natural Science." *Theory & Psychology* 8, no. 1 (1998): 5–22. <https://doi.org/10.1177/0959354398081001>.
- Winston, Jessica. "Gorboduc Now! The First English Tragedy in Modern Print and Performance." *English: Journal of the English Association* 68, no. 261 (June 2019): 184–203. <https://doi.org/10.1093/english/efz019>.
- Xanthos, Aris, Isaac Pante, Yannick Rochat, and Martin Grandjean. "Visualising the Dynamics of Character Networks." In *DH2016 Book of Abstracts*. 2016. <http://dh2016.adho.org/abstracts/407>.
- Yarkho, Boris. *Metodologiya tochnogo literaturovedeniia: izbrannye trudy po teorii literatury* [Methodology for the Exact Study of Literature: Selected Works in Literary Theory]. Moskva: Jazyki slavjanskikh kul'tur, 2006.
- Yarkho, Boris I. "Speech Distribution in Five-Act Tragedies (A Question of Classicism and Romanticism)." *Journal of Literary Theory* 13, no. 1 (March 2019): 13–76. <https://doi.org/10.1515/jlt-2019-0002>.
- Zehe, Albin, Leonard Konle, Lea Katharina Dümpelmann, Evelyn Gius, Andreas Hotho, Fotis Jannidis, Lucas Kaufmann, Markus Krug, Frank Puppe, Nils Reiter, Annekea Schreiber, and Nathalie Wied-

mer. “Detecting Scenes in Fiction: A New Segmentation Task.” In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 3167–3177. Association for Computational Linguistics, 2021. <https://doi.org/10.18653/v1/2021.eacl-main.276>.