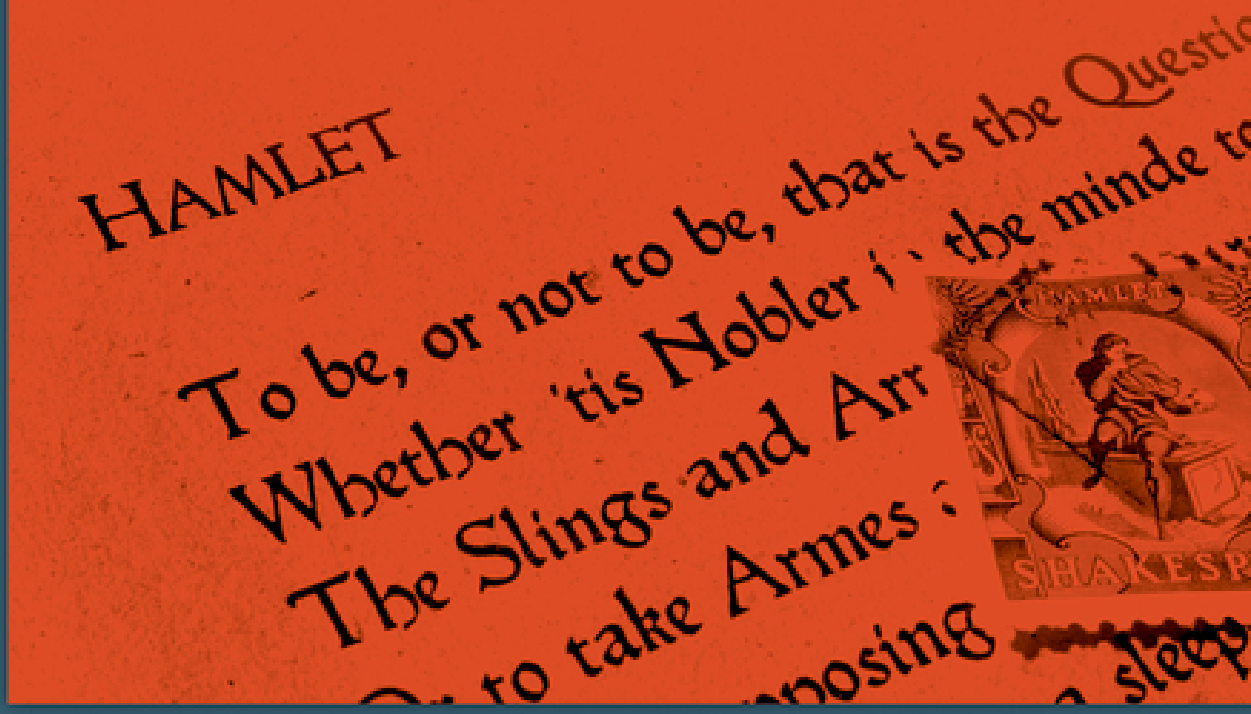
DE GRUYTER

COMPUTATIONAL DRAMA ANALYSIS

REFLECTING ON METHODS AND INTERPRETATIONS

Edited by Melanie Andresen and Nils Reiter



DE G

Computational Drama Analysis

Computational Drama Analysis

Reflecting on Methods and Interpretations

Edited by
Melanie Andresen and Nils Reiter

DE GRUYTER

ISBN 978-3-11-107176-3
e-ISBN (PDF) 978-3-11-107182-4
e-ISBN (EPUB) 978-3-11-107191-6
DOI https://doi.org/10.1515/9783111071824

Library of Congress Control Number: 2024930373

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.dnb.de.

Cover image: William Shakespeare's *Hamlet* (original text from the First Folio, 1623) with a stamp. claudiodivizia / istockphoto
Printing and binding: CPI books GmbH, Leck

www.degruyter.com

Contents

Melanie Andresen and Nils Reiter

Introduction

Computational drama analysis is the field of research that attempts to model, analyze, and interpret dramatic texts using computational methods. It is part of the larger field of computational literary studies and, even more broadly, the digital humanities. Computational drama analysis is part of the current boom in quantitative data analysis methods including machine learning and artificial intelligence. More and more work is done on structuring, analyzing, understanding, or generating literary texts with the help of algorithms.

The papers in this volume are the result of the two-day workshop on "Computational Drama Analysis: Achievements and Opportunities," held at the Theaterwissenschaftliche Sammlung of the University of Cologne[1] on September 14–15, 2022. Following an open call for papers published in March 2022, nine papers were presented at the workshop, each accompanied by an in-depth discussion. All papers were extensively revised on the basis of the discussions and additional reviewer comments and are now available in this volume.

1 Challenges in Computational Drama Analysis

Compared to other literary genres, dramatic texts have unique characteristics that make them both well suited and challenging for quantitative and formal methods. On the one hand, they have a very clear structure, manifested in the segmentation into acts and scenes, a clear separation between stage directions and character speech, and mostly unambiguous speaker designations. These elements can be automatically recognized and formally modeled relatively easily. Importantly for quantitative analysis, these elements allow for an unambiguous and descriptive approach: no interpretation is required to know which character speaks a particular utterance and when. With the DraCor corpora,[2] the community of computational drama analysis also has an exemplary resource at hand that provides large corpora of TEI-encoded plays for many languages.

1 https://tws.phil-fak.uni-koeln.de.

2 https://dracor.org.

Melanie Andresen, Department of Natural Language Processing, University of Stuttgart
Nils Reiter, Department of Digital Humanities, University of Cologne

https://doi.org/10.1515/9783111071824-001

On the other hand, there are also a number of challenges which require our attention. The multiple internal structures of plays can also be a challenge in the application of computational methods, many of which were developed for continuous text. Computational drama analysis therefore needs to adapt methods to account for this structure and the multiperspectivality that it represents.

Due to the easy computational accessibility of the surface features of dramatic texts, research in computational drama analysis has had a strong focus on these formal features. While this already leads to interesting insights, the majority of non-computational scholarly work on drama does not focus on these formal features, making it difficult to integrate computational and non-computational research in literature. In addition, formal features are less reliable than they might seem at first glance. While most – but not even all – plays feature a segmentation into acts and scenes, the way these segments have been used by playwrights differ significantly. Interpreting results based on these formal features must take this development into account.

Dramatic texts are (often, not always) intended to be performed on stage. This adds an additional layer to the text that is very difficult to access computationally, as a reader's experience may drastically differ from that of an audience in a live theater. Investigating such differences, for example, would require the incorporation of multi-modal data, such as recordings of theater performances. Such recordings are not available on the same scale as text corpora. They would have to be extensively processed before the essence of the plays could be analyzed in relation to the text. Here, a close exchange and collaboration with the field of theater studies would be beneficial and one possible avenue for computational drama analysis would be to foster this interdisciplinary endeavor.

Resources such as DraCor, which provide a standardized format for dramatic texts and comparable corpora for several languages, allow for another strand of analysis that has not yet been widely explored: the comparative analysis of drama across languages and cultures. This can be achieved by focusing on formal, language-independent features of drama or by analyzing the speech contents more closely, but in a multilingual setting. Another possibility is to rely on translations, either of full texts or of the resources used for analysis. With the current, stunning developments in natural language processing, we can expect this to be a viable path of increasing quality.

While it is possible to automate many types of analysis today, there are still tasks that require manual labor by researchers. This is especially true as we move away from the text surface and attempt to account for more complex phenomena that promise easier interpretation of the findings in terms of the traditional literary studies. This means that many analyses will not scale well to large corpora, making the goal of e.g. analyzing the historical development difficult to achieve. By

closely following the technological development in natural language processing, computational drama analysis could be able to propose more powerful methods as well.

In summary, computational drama analysis is a field of research that has established a solid foundation of shared research questions, data, and methods and is now ready to tackle the next level of complexity. The papers in this volume hopefully give a sense of how this might be put into practice.

2 Chapters in this Volume

In the following we will give a short overview of the papers in this volume and how they address the issues and desiderata of computational drama analysis.

One crucial limitation for any digital analysis is the availability of data. This has led to an imbalance in representation, as efforts at digitizing plays have been mainly made for large language communities like English, German, and French. **Ruiz Fabo, Bernhard, Briand, and Werner** tackle this issue by creating a drama corpus for Alsatian, which is already partly available via the DraCor platform (pp. 53–85). Their paper gives us background on the Alsatian dramatic tradition that is influenced by both German and French drama and centered around the comic genre. The authors share valuable insights in some of the challenges of such an endeavor, like the semi-automatic creation of TEI versions of the plays, and dealing with spelling variation. They also present a manual annotation of character properties like gender, profession, and social class. Preliminary analyses show, for instance, that the plays mirror urbanization in a decline of characters involved in agriculture and an increase in the crafts.

Trilcke, Ustinova, Börner, Fischer, and Milling observe on pp. 6–33 that comparative work on drama that crosses language boundaries is still rare, even though data is available. Network analysis is particularly suited for comparative analyses as network metrics abstract away from concrete words. They work towards a typology of dramatic networks starting with the "Small World" concept. This concept describes networks that are highly clustered, but these clusters are also connected by rather short paths. The authors explore and compare different operationalizations of the concept with regard to their potential for typological descriptions of drama. A diachronic case study exemplifies this potential by revealing distinct distributions of small world plays across time and language.

Szemes and Vida also contribute to a language-independent typology of drama based on character networks (pp. 166–188), this time framed as a classification task about the dramatic genres of tragedies vs. comedies. In contrast

to most previous work, they strive to find measures that are independent of the networks' size and contribute to the plays' interpretation. For this purpose, they define 16 measures based on network properties, some new, some commonly used. Results from a principal components analysis and a classification based on a support-vector machine show that there is a lot of overlap between the groups, but instructive tendencies can be derived. For instance, comedies have a denser character network and more characters with a lot of speech, while in tragedies, characters tend to have longer utterances. The authors also test the hypothesis that it is mainly the last act that determines the structure of the network (as the protagonist typically ends up isolated in tragedies vs. celebrating their wedding with everybody in comedies). Especially for the Shakespeare corpus, such a tendency can be confirmed.

On pages 85–105, **Hicke and Mimno** provide us with a feminist view on Shakespeare's comedies, which feature many prominent female characters. In order to analyze more closely how female characters enact power in the comedies, they approximate a character's influence in a play by measuring how much they speak and how many other characters they speak to. They provide interactive visualizations of how this influence measure changes in the course of the play. In a number of examples that productively complement the visualizations with close readings, the authors show that women influence the plays' development in many different ways, for instance by continuously gaining in influence throughout the play, appearing as an influential character right away, or by avoiding societal limitations by cross-dressing.

Krautter and Pagel (pp. 122–148) also focus on characters, albeit one specific group of characters: the schemer, i.e., characters that somehow interfere with the main plot. They identify schemers in the plays by reviewing secondary literature and attempt to train an automatic classifier based on machine learning that predicts which characters belong to the class of schemers. Drawing on a broad range of features like stage presence, speech, sentiment scores, topic modeling, and network metrics, they arrive at a model with high precision, but low recall, meaning that it identifies almost all schemers as schemers, but also predicts many characters to be schemers that do not belong to this group. The authors assume that the group of schemers is very diverse and therefore hard to pin down by generalizable features.

Character speech is at the core of the paper on pages 148–166, written by **Šeļa, Nagy, Byszuk, Hernández-Lorenzo, Szemes, and Eder**. More specifically, they analyze the distinctiveness of a character: is character speech dominated by author style or are characters identifiable by their speech? The authors present and compare two methods for measuring character distinctiveness that build on measures in authorship verification. The first is comparing the probability distribu-

tions of character trigrams based on bootstrapping. The second is measuring the area under a word keyness curve, which is smaller if a character is only characterized well by few keywords and larger if many keywords reach high scores for the character. Both compare the speech of a target character to all other characters' speech in a play. The two measures complement each other as the one based on trigram probabilities has more statistical power while the measure based on keywords informs interpretation by these words.

One important aim that points into the future of computational drama analysis is going beyond structural features and tackling the content of plays. **Mareček, Nováková, Vosecká, Doležal, Musil, and Rosa** (pp. 105–122) attempt this by focusing on situations as building blocks of plays. They update dramatic situations described in the literature and use the new categories to annotate 52 mostly contemporary Czech plays. It is a promising avenue to see how plays are composed of sequences of situations like "Interrogation," "Parting," and "Seduction" and Mareček et al. hope to at some point use it for the generation of plays. However, their annotation effort shows that this is a difficult task, as the agreement between annotators is moderate and also the attempt of automating the annotations presents many challenges. More data is required to make this feasible.

While most approaches to drama are based on spoken text, **Gavazza, Glaser, Mastellari, and Novokhatko** enrich the volume with a theater studies perspective on drama that focuses on material objects and their role in the play (pp. 33–53). Based on material engagement theory, they argue for the importance of objects in ancient Greek comedy and show how the interaction of characters and objects can create humorous effects, among others. The authors present the Lexicon of Objects from Greek Comedy, which collects information on all material objects in a mySQL database. This comprehensive resource can be used, for instance, to explore the quantitative distribution of object categories in ancient Greek comedy, showing that food is the most popular category overall, which corresponds to what scholars know about the genre.

On pp. 188–215, **Dennerlein, Schmidt, and Wolff** go beyond character speech by focusing on emotions in stage directions. They observe that emotions can be directly or indirectly expressed in character speech, but can also appear in stage directions, which we might consider a more reliable source than character speech. Based on the German Drama Corpus and manual as well as automatic annotations, they analyze the differences in emotions mentioned in character speech vs. stage directions and how their relation develops diachronically. They find an increase in emotion depiction in stage directions, which corresponds to known trends towards more narrative patterns in drama. Their paper provides a fine-grained analysis of developments for individual emotions like abhorrence, friendship, and despair.

Tab. 1: Summarizing metadata about the papers in this volume

Authors	Data	Methods	Target of Analysis
Trilcke et al.	Very Big Drama Corpus (VeBiDraCor)	Network analysis	Typology of character networks
Gavazza et al.	Ancient Greek drama	Manual annotation, database creation	Use of objects
Ruiz Fabo et al.	MeThAL Corpus (Alsatian)	Corpus creation, machine learning, manual annotation	Spelling variation, character properties
Hicke and Mimno	Shakespeare's plays	Manual annotation, speech patterns	Female speech distribution
Mareček et al.	Modern Czech plays	Manual annotation, classification	Dramatic situations
Krautter and Pagel	German Drama Corpus (GerDraCor)	Classification	Character type 'schemer'
Šeļa et al.	French Drama Corpus (FreDraCor), German Drama Corpus (GerDraCor), Russian Drama Corpus (RusDraCor), Shakespeare Drama Corpus (ShakeDraCor)	Distinctiveness measures	Character speech
Szemes and Vida	Shakespeare Drama Corpus (ShakeDraCor), German Drama Corpus (GerDraCor)	Classification, network analysis	Network typology, genre differences
Dennerlein et al.	German Drama Corpus (GerDraCor)	Manual and automatic annotation	Emotions

Table 1 provides a concise overview of the papers in this volume with a focus on their target of analysis as well as the data and methods used.

Peer Trilcke, Evgeniya Ustinova, Ingo Börner, Frank Fischer, and Carsten Milling

Detecting Small Worlds in a Corpus of Thousands of Theater Plays

A DraCor Study in Comparative Literary Network Analysis

Abstract: Although homogenized TEI corpora of theater plays from different languages are becoming more and more available, research on plays with a comparative angle is still rare in the field of Computational Literary Studies (CLS). At the same time, approaches of formal network analysis in particular bear huge potential for comparative research due to their modeling of texts as asemantic structures. An attempt to integrate the paradigm of such a formal analysis with general network research on the one hand and literary history on the other hand is the approach of a typification of networks with respect to structural properties such as the the "Small World" concept. However, studies have so far remained limited to smaller and monolingual corpora. In this study, we report on the implementation of different operationalizations of the "Small World" concept and their applications to a corpus of almost 3000 plays. Looking at the results of these analyses, we examine how the different operationalizations of the "Small World" concept relate to each other and discuss how they could be used for a network-based typology of dramatic forms. We finally develop initial ideas for a network-grounded history of dramatic forms in a transnational perspective.

1 Introduction and Research Agenda

1.1 Literary Network Studies

Modeling and analyzing fictional artifacts – such as epics, novels, movies, or plays – as networks has become a widespread procedure in computational humanities (Labatut and Bost 2019). Accordingly, the extraction of network structures from those artifacts and their analysis is not only a challenge for computer scientists (Elson et al. 2010; Lee and Yeung 2012; Agarwal et al. 2013; Krug 2020), but is also being conducted within the scope of decidedly humanities-related research. Thus,

Peer Trilcke, Ingo Börner, Carsten Milling, University of Potsdam
Evgeniya Ustinova, Charles University Prague
Frank Fischer, Free University of Berlin

https://doi.org/10.1515/9783111071824-002

for example, the importance of the Horatio character for the structural stability of the network in Shakespeare's *Hamlet* was described (Moretti 2011, pp. 4–5) or the role of the Mouse character in *Alice's Adventures in Wonderland* as a connector between many characters (Agarwal et al. 2012, pp. 93–94). The networks of Homer's *Illiad* and of his *Odyssey* have been (frequently) investigated (Kydros et al. 2015; Miranda et al. 2018), and there are studies of the amalgamated network of the *Íslendinga sögur* and its subcomponents (Mac Carron and Kenna 2013) as well as of community structures in *Les Misérables* (Newman and Girvan 2004, pp. 12–13). The network-based distinctions between rural and urban novels have been discussed (Elson et al. 2010; Jayannavar et al. 2015), and network measures have been tested as a possible input for genre classification tasks (Coll Ardanuy and Sporleder 2015; Hettinger et al. 2015) and as measures of similarity (Reger 2016). Several papers have also proposed network-based concepts of 'protagonism' as well as quantitative classifications of the characters of fictional texts (Park et al. 2013; Algee-Hewitt 2017; Fischer et al. 2018; Krautter et al. 2018).

So, while network analysis has already proven to be insightful for literary studies, some research has also shown that the "structural intuition" (Freeman 2004, p. 4) of network-based approaches and the related high level of formalization hold particularly high potential for comparative analyses. This is because the 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.

Three systematic approaches can be distinguished here: first, a comparative approach is possible in the sense of *general network sciences*, in which a universal morphology of all possible forms of networks is elaborated. Second, there are what might be called *cultural* or, third (and more narrowly), *literary network studies*, which focus on the analysis of cultural products, fictional artifacts or even literary works, though against the backdrop of general network science, whose research especially on universal network morphology provides an important theoretical frame of reference. In this sense, networks of literary works and other fictional worlds are regularly discussed with respect to universal topological structures. Examples include studies on mythological networks (Carron and Kenna 2012), on the Marvel universe (Alberich et al. 2002) or on generative approaches to literary network structures (Moretti 2020).

1.2 Studying Small Worlds (in Literature)

One type of network structure that has driven and inspired comparative research is the so-called "small world network" (Watts and Strogatz 1998; Watts 1999a,

1999b), which is defined by a twofold difference: it differs from regular network types as well as from random network types by certain topological characteristics and is suspected "to be widespread in biological, social and man-made systems" (Watts and Strogatz 1998, p. 442). Small world networks seem to be a perhaps universal form that empirical networks tend to take (in contrast to strictly mathematically constructed ones).

It does not come as a surprise that the small world concept has also been used in the computational analysis of literary texts. However, these studies are usually limited to stating that the network constructed for the analyzed work has topological characteristics similar to small world networks – and is therefore structured like a real-world network (Mac Carron and Kenna 2013; Kydros and Anastasiadis 2014; Miranda et al. 2018). In contrast, Stiller et al. (2003), guided by an elaborated anthropological theory, used the small world concept in their analysis of ten Shakespearean plays, in which they not only proved that the plays can be described as small worlds, but also offered evolutionary hypotheses as an explanation for this specific structure. However, even in this approach, which is further developed in Stiller and Hudson (2005), the focus of the analysis (following the paradigm of evolutionary history) lies on the similarity between real-world networks and small world-structured plays.

As interesting as such analyses may be from the perspective of general network science or in light of evolutionary anthropology, from a literary studies point of view they lack at least the reflection of literary form, which has to be considered a major determinant of the networks we extract from literary texts. If, in contrast, one takes into account that the fictional worlds of literature only exist in their mediation through form, there are strong arguments for conceptualizing networks extracted from literary texts first as phenomena of form and not so much as representations of real-world social structures. In this sense, Trilcke (2013, pp. 223–226) has suggested that methods of network science could be applied in literary studies in the context of analyzing form, referring in his own analyses to Klotz's morphological typology of drama (Klotz 1969).

With their corpus-based analysis of more than 450 German-language plays, Trilcke et al. (2016) have taken up this form-oriented approach of literary network analysis and related it to the small world concept. The *small world test* (SWT) proposed by them, a procedure for the algorithmic classification of small world characteristics in the plays of a corpus, aims at analyzing the structural composition of plays and thus could be understood as a network-based contribution to the typology of dramatic forms. For the following re-implementation of this calculation method and the comparison with other implementations, it is crucial that Trilcke et al. (2016) designed their small world test as an outlier test with the aim of identifying particularly exceptional structures in the history of theater plays.

In the present study, we are building on this typology-oriented, form-based approach to small world phenomena in literature. However, we do not want to introduce an additional operationalization of the small world concept for literary studies. Instead, we will, on the one hand, re-implement the outlier-based approach of Trilcke et al. (2016) and, on the other hand, relate it to a much broader operationalization of the small world concept, which Humphries and Gurney (2008) designed to describe the so called "small-world-ness" (SWN) of networks. In this context, we will show that while the two conceptualizations have some overlap, they also open up two different possibilities for typification: while the *SWT small world type* of Trilcke et al. (2016) describes a historic extreme type of dramatic form, the *SWN small world type* of Humphries and Gurney (2008) identifies a systematic standard type of dramatic form.

Overall, with our study we aim to explore different operationalizations of the small world concept as applied to dramatic texts to better understand the conceptual logic behind these different operationalizations. In doing so, we are less concerned with deciding which of the two conceptualizations is the "more correct" one – this would be, in our view, a theoretically fruitless question. Rather, our aim is to contribute to a network-based, structure-oriented literary history of the dramatic form by uncovering and discussing the quite different descriptive potentials of the operationalizations in question.

To be concrete, in the following we will re-implement the above-mentioned small world test (Trilcke et al. 2016) and apply it to a very big drama corpus (VeBiDraCor) with nearly 3,000 structurally fairly homogeneous plays, written in more than ten languages. In addition, we will also implement the scale free test performed by Trilcke et al. (2016) (with reference to Albert and Barabási (2002)), which describes a more strictly defined variant of small world networks. Simultaneously, we will implement the measure of "small-world-ness" (SWN) proposed by Humphries and Gurney (2008), which follows a different conceptual logic than the small world test. Our DraCor-based research corpora (Fischer et al. 2019) as well as both concepts and their corresponding implementations are described in chapter 2. In chapter 3, we will discuss the results of our analysis with regard to the potentials that both small world concepts offer for a comparative typology of dramatic forms. Instead of an outlook, we will eventually sketch some ideas about the history of dramatic form from a network science point of view in chapter 4.

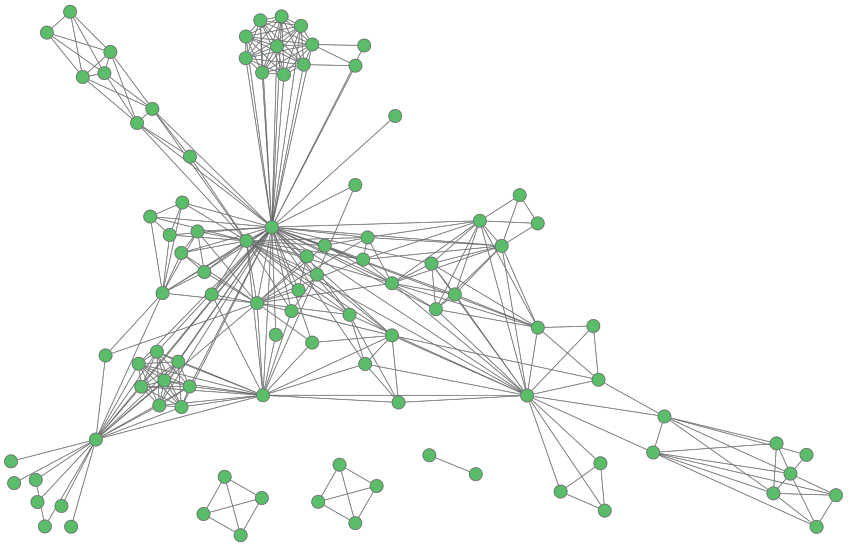


Fig. 1: Johann Wolfgang Goethe, *Götz von Berlichingen mit der eisernen Hand*, 1773.

2 Operationalization and Corpora

2.1 Operationalizing Small Worlds

Small worlds, theorized by Watts as a "deep feature of the social world" that constitutes "a family of graphs" (Watts 1999b, pp. 493–494, 502), are usually characterized in research by two properties, each of which is defined relatively to a "regular" and a "random" network type. Phenomenologically described, small world networks (or more precisely: small world graphs) are, like regular networks, "highly clustered," whereas random graphs are "poorly clustered"; at the same time, small worlds have "small characteristic path lengths, like random graphs" (Watts and Strogatz 1998, p. 440). The description already specifies the two network measures used for the formalization: the network average clustering coefficient C and the average shortest path length L. We will come back to these two measures in a moment.

To begin with a visual representation of the structural properties of small world networks, it is worth taking a look at the network of Goethe's play *Götz von Berlichingen mit der eisernen Hand* (1773) in Figure 1.[1]

1 See https://dracor.org/id/ger000123.

While, on the one side, we see some densely connected areas ("highly clustered"), some very central nodes (such as the character Götz, situated in the upper center) ensure, on the other side, that the paths between these different clusters remain quite short ("small characteristic path lengths").

This specific structure can be measured. However, before we can apply network measures to dramatic texts, these texts must be *modeled as networks*, where G is a network with n nodes and m edges. This information is extracted from plays encoded in TEI as the following pseudo-structure exemplifies:

```
<div type="segment">
    <sp who="#speaker_A">
        <p>text</p>
    </sp>
    <sp who="#speaker_B">
        <p>text</p>
    </sp>
</div>
<div type="segment">
    <sp who="#speaker_B">
        <p>text</p>
    </sp>
    <sp who="#speaker_C">
        <p>text</p>
    </sp>
</div>
```

While each distinct speaker represents a node n, a relation m is established if the speeches `<sp>` of two or more speakers, assigned via the `@who` attribute (`"#speaker_A"`, `"#speaker_B"`), appear in the same segment `<div>`. The algorithms used for extracting our network data are open-source and can be viewed on our GitHub repositories.[2] It is important to note that this method of extracting networks from dramatic texts is only one possibility among many others and that the scope for interpretation is determined by this mechanism of extraction. At the

2 The xQuery function of the *DraCor API* (https://github.com/dracor-org/dracor-api) *metrics:get-network-metrics* (https://bit.ly/3wDIJ5Y) extracts the segments of a given TEI file using the function *dutil:get-segments* (https://bit.ly/3AWraAK) and for each of these segments gets the distinct speakers with the function *dutil:distinct-speakers* (https://bit.ly/3CIvDYT). The network metrics are calculated based on these extracted features with the *DraCor metrics service* (https://github.com/dracor-org/dracor-metrics) using the Python package *networkx*.

same time, this extraction mechanism has two merits: first, it can be implemented with a manageable effort, and thus can be applied quite quickly to large sets of texts; second, in line with our research interest, it strongly relies on the specific form of the dramatic text. Last but not least, it is of particular importance for our approach that all dramatic networks are constructed in the same way to ensure comparability, which is possible thanks to the homogeneous structure of plays on the DraCor platform (Fischer et al. 2019).

Having extracted dramatic networks from plays in this way, we run a series of small world-related analyses on the data.

First, the small world test SWT proposed by Trilcke et al. (2016): Following the conceptualization by Watts and Strogatz (1998), this test relies on the two above-mentioned measures for network topology, the average clustering coefficient C and the average shortest path length L. We calculate C referring to Watts and Strogatz (1998), while L is implemented as a ratio of the sum of average path lengths for each node and the number of such paths.[3]

For SWT, we assume that each corpus Cor is a set of n plays P, to each of which corresponds a network G. We first calculate C and L for each G. For each G we then generate 1,000 random graphs (following the Erdős-Rényi model (E–R), which, simply put, distributes all edges according to constant probability between the nodes of a graph, see Erdős and Rényi (1959)), and calculate the mean for C and L, resulting in C_{rand} and L_{rand}. Next, for each P we calculate C^{dev} and L^{dev} with

$$C^{dev} = \frac{C_G}{C_{rand}} \tag{1}$$

$$L^{dev} = \frac{L_G}{L_{rand}} \tag{2}$$

As stated above, dramatic small worlds can in general be understood as a type of dramatic form that is shaped by a characteristic combination of *difference* from (i.e. higher clustering than random networks) and *similarity* to (i.e. average path length similar to random networks) another type. Although with C^{dev} and L^{dev} there are measures of similarity and difference, there is not yet a procedure for classifying small worlds. Watts and Strogatz (1998) provide no rule as to when C^{dev} is high enough and when L^{dev} is low enough to classify a network as a small world. Humphries and Gurney (2008, p. 2) therefore propose a "continuously graded notion of small-world-ness" and base the categorical concept of small

3 For networks with more than one component, only the paths between connected nodes are counted.

world derived from this small-world-ness on a postulated threshold value (see remarks below).

The approach to solving this problem suggested by Trilcke et al. (2016) is rooted in the domain focus of their study design. They assume that threshold values for the difference (or similarity) between C_G and C_{rand} (or L_G and L_{rand}), as indicated by C^{dev} (or L^{dev}), must be obtained by comparing them to the whole domain, i.e. in the present case to the particular corpus studied. So, for all *P* in *Cor*, they calculate C^{dev} and L^{dev} and average them. Correspondingly, for SWT we classify G as a dramatic small world if the following two criteria are met:

- **Criterion I**: C^{dev} of a single *P* has to be significantly higher than $\overline{C}^{dev}_{Cor}$ and
- **Criterion II**: L^{dev} of a single *P* must not differ significantly from $\overline{L}^{dev}_{Cor}$.

To decide if the values are significantly higher resp. do not significantly differ, following Trilcke et al. (2016), we calculate a simple deviation test, where we consider anything above or below *mean* $\pm 2\times$ *standard deviation* to be a significant deviation from the norm. For **criterion I**, this means:

$$C^{dev} > \overline{C}^{dev}_{Cor} + 2\delta \tag{3}$$

And for **criterion II** this means:

$$\overline{L}^{dev}_{Cor} - 2\delta \quad < \quad L^{dev} \quad < \quad \overline{L}^{dev}_{Cor} + 2\delta \tag{4}$$

It is, at this point, important to note again the outlier-oriented approach to the small world concept proposed by Trilcke et al. (2016). Assuming a normal distribution of the plays in the corpus (resp. of the values for their structural properties), according to Equation 3 and Equation 4 about 5% of the plays should be typed as SWT type small worlds, whereas 95% should be within the range *mean* $\pm 2\times$ *standard deviation*.

Second, in addition to SWT, following Trilcke et al. (2016), we carry out a scale-free test SFT, in which we check whether dramatic networks can be typified as scale-free networks following the conceptualization by Albert and Barabási (2002). In the interpretation of Trilcke et al. (2016), scale-free networks fulfill criterion I and criterion II *and* are characterized by having a node degree distribution that follows a power law (**criterion III**). Since strict power law distributions are rare (Broido and Clauset 2019), especially in networks as small as the plays studied, in line with we use the following operationalization in line with Trilcke et al. (2016): for the node degree distribution of each P that meets criterion I and criterion II, we calculate the coefficient of determination R^2 for a) a linear, b) a quadratic, c) an exponential, and d) a power law fit. When R^2 of the power law fit is highest, we consider criterion III to be fulfilled and the play has passed the SFT.

Both tests, SWT as well as SFT, follow, as we have already stated, on the one hand a *categorical* logic: the network of a play is either a SWT small world (or a SFT scale-free network) or it is not. On the other hand, the categorical attribution, which both tests provide, is not an absolute one, but always only possible *in relation* to a specific corpus. These two aspects of our operationalization distinguish them fundamentally from the approach of Humphries and Gurney (2008). The measure for small-world-ness S proposed by them operates, in its conceptual basis, with a "continuously graded notion of small-world-ness" (Humphries and Gurney 2008, p. 2), so that a graph, in fact, can be more or less small-world-ish. Only in a second step do the authors introduce a threshold value $S = 1$ with which their operationalization can also be used for categorical attributions. More significant, however, is the difference between our operationalization and the one proposed by Humphries and Gurney (2008) when it comes to the relational aspect of the term. In Humphries and Gurney's conceptualization, small-world-ness S is an absolute measure that can be calculated without referring to a reference group (a corpus in our case). To be more specific, S – with G as any given graph – is calculated as follows (Humphries and Gurney 2008, p. 2):

$$\gamma_G = \frac{C_G}{C_{rand}} \tag{5}$$

and

$$\lambda_G = \frac{L_G}{L_{rand}} \tag{6}$$

so that

$$S = \frac{\gamma_G}{\lambda_G} \tag{7}$$

Thus, while the operationalization of SWT by Trilcke et al. (2016) proposes a *categorial-relative* term, the measure of small-world-ness S by Humphries and Gurney (2008) proposes a *continuous-absolute* term (for which they at the same time indicate options to be applied in a *categorial-absolute* way, resulting in a small world type we will call SWN). It must be noted, in addition, that Humphries and Gurney's threshold $S = 1$ for categorical attribution aims to show small worlds as a *widespread, general phenomenon*; the operationalization by Trilcke et al. (2016), on the other hand, conceptualizes small worlds as a *rare, structurally exceptional phenomenon*. Correspondingly, we expect a high number of dramatic small worlds of the type SWN, but relatively few of the type SWT (and, consequently, even fewer of the type SFT).

Tab. 1: Research corpora overview.

	Number of		Year				Mean Number of	
Corpus Name	**Plays**	**Authors**	**Min**	**Max**	**Mean**	**SD**	**Speakers**	**Segments**
VeBiDraCor	2978	797	–472	2017	1719	332.2	14.3	19.9
VeBiDraCor_Struc	2327	702	–472	2017	1701	367.1	15.3	24.9
VeBiDraCor_Struc_Hist	2246	690	1508	2017	1766	91.7	15.4	25.3

2.2 Research Corpora

To understand how the different operationalizations behave and what typological and historical conclusions can be drawn from their application, we conduct a series of analyses on different corpora of plays (see Table 1). At the heart of our analyses is *VeBiDraCor* – our very big drama corpus, which we created by aggregating all individual corpora currently available through DraCor.[4] For a more detailed breakdown of the corpora aggregated for *VeBiDraCor* and their basic metadata, see Table 2.

From *VeBiDraCor*, we made two derivations, each of which we used to restrict the heterogeneity of dramatic forms somewhat. *VeBiDraCor_Struc* includes only plays whose number of segments is ≥ 5, so that we exclude plays which, due to their shape (i.e. few segments), tend to have rather exceptional networks. With *VeBiDraCor_Struc_Hist* we add a second step of homogenization by restricting ourselves to plays published or first performed after 1500 ("yearNormalized"[5]).

2.3 Analyses

For all research corpora listed above, we

- calculate the small-world-ness measure S;
- perform a categorical application of the small-world-ness measure S, using the threshold = 1 $\rightarrow$ SWN (Humphries and Gurney 2008, p. 2);
- perform the small world test $\rightarrow$ SWT;
- perform the scale-free test $\rightarrow$ SFT.

4 *VeBiDraCor* was created on August 09, 2022 using a dedicated, fully functional Docker image of DraCor (incl. metrics services and API functions), which we also use to version the state of the corpora at a given point in time to provide the identical data and API functionalities for use in replication studies. For more info, see at the end of the chapter.

5 On calculating the "Normalized Year," cf. https://dracor.org/doc/faq.

Tab. 2: Composition of VeBiDraCor.

Corpus			Number of		Year				Mean Number of	
Acronym	Name	Title	Plays	Authors	Min	Max	Mean	SD	Speakers	Segments
AlsDraCor	als	Alsatian Drama Corpus	25	13	1816	1937	1914	24.8	12	16.2
BashDraCor	bash	Bashkir Drama Corpus	3	2	1917	1975	1952	30.8	18.7	3.7
CalDraCor	cal	Calderón Drama Corpus	205	1	1623	1760	1663	32.6	18	2.3
FreDraCor	fre	French Drama Corpus	1556	395	1170	2017	1747	85.6	10	19.3
GerDraCor	ger	German Drama Corpus	589	220	1650	1947	1837	57.3	22.3	24.8
GreekDraCor	greek	Greek Drama Corpus	39	4	−472	−388	−425	21.9	11.2	12.7
HunDraCor	hun	Hungarian Drama Corpus	41	25	1558	1970	1860	77.9	24.3	22.9
ItaDraCor	ita	Italian Drama Corpus	139	49	1449	1933	1657	123.2	10.9	26.1
RomDraCor	rom	Roman Drama Corpus	36	3	−215	96	−121	111	11.2	16.3
RusDraCor	rus	Russian Drama Corpus	212	56	1747	1947	1848	51.1	17.5	25.1
ShakeDraCor	shake	Shakespeare Drama Corpus	37	1	1591	1613	1600	6.5	38.7	20.8
SpanDraCor	span	Spanish Drama Corpus	25	9	1878	1945	1913	18.8	22.1	33.8
SweDraCor	swe	Swedish Drama Corpus	68	20	1880	1900	1889	5.7	11.6	14.4
TatDraCor	tat	Tatar Drama Corpus	3	1	1908	1912	1910	2.1	10	12

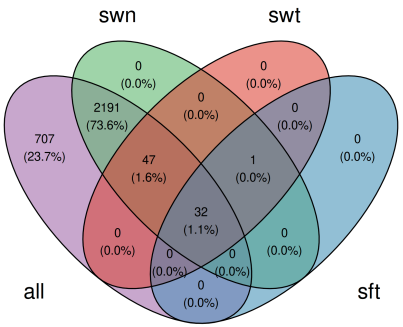


Fig. 2: Relations of the sets of small worlds, with "all" = *VeBiDraCor* and "swn," "swt" and "sft" as the corresponding subsets.

Analyses were carried out based on the output of the DraCor API using a purpose-developed R script.[6]

3 Discussing Results from a Typological Point of View

In the following, we will present and discuss the results of our analyses and further inspect the received data. In our discussions and inspections, we will focus on the following guiding questions: first, how do, in particular, the two small world operationalizations SWN and SWT– the first from general network science, the second from literary network studies – relate to each other in quantitative and phenomenological respects? Second, if we assume that a larger or smaller group of the dramatic networks in our corpora are typified as small worlds: can certain properties be specified for the dramatic networks that do not fall into this group? In other words: can other network-based types of dramatic form be described relative to the group of small worlds?

3.1 Overview of the Results

Table 3 reports both the mean S as well as the absolute number and the share of dramatic networks typified as small world (or scale-free) networks in the cor-

6 The code for the analyses is online, see note at the end of this article.

Tab. 3: Small world attributions by measure and corpus.

			Amount (share) of		
Corpus Name	**N**	**Mean S**	**SWN plays**	**SWT plays**	**SFT plays**
VeBiDraCor	2978	1.68	2,270 (76.2%)	79 (2.7%)	32 (1.1%)
VeBiDraCor_Struc	2327	1.8	2,071 (89.0%)	70 (3.0%)	31 (1.3%)
VeBiDraCor_Struc_Hist	2246	1.8	1,990 (88.6%)	71 (3.3%)	31 (1.4%)

responding corpora according to the different measures. Figure 2 visualizes the relations of the sets of detected small world networks with respect to the different measures.

We first keep three observations:

- A. As expected, SWN typifies a large part of the corpora as small worlds (between 76.2% and 89%), whereas SWT seems to describe only a very specific type of dramatic network (between 2.7% and 3.3%). SFT is even more rare. We will discuss the striking dominance of type SWN small worlds in the subsequent chapter 3.2.
- B. As can be seen in Figure 2 (and as can partly also be mathematically stated a priori), we do have a subset-inclusion order (i.e. a nested set collection), so that *VeBiDraCor* $\supseteq$ *VeBiDraCor_swn* $\supseteq$ *VeBiDraCor_swt* $\supseteq$ *VeBiDraCor_sft*. In other words, the operationalization SWT proposed by Trilcke et al. (2016) turns out to typify not so much a different type, but rather something like an extreme type of a dramatic small world. We will discuss in chapter 3.3 why this extreme type is nevertheless not (necessarily) identical with plays that have the highest S value.
- C. Homogenizing the corpora in structural terms (see *VeBiDraCor_Struc*) leads to a slight increase in S and in the proportion of plays typified as small world resp. scale-free networks. Combining structural and historical homogenization (see *VeBiDraCor_Struc_Hist*) also has only a small effect. We will come back to this observation in chapter 4.

Before discussing the results in more detail from a typological point of view, let us take a brief and exemplary look at how some of the features of dramatic networks influence the small world typification. Table 4 lists the minimum number of speakers (nodes) as well as the minimum number of segments that were sufficient for a play to be typified as SWN, SWT or SFT. Both speakers and segments are important factors in the construction of the networks. As can be seen, SWN can also be assigned to very small, low-segment dramatic networks (see also our discussion of SWN type as a standard type below). In contrast, SWT and even more

Tab. 4: Lowest number of speakers/segments in a play that passed the test.

	Minimal Number of	
Type	**Speakers**	**Segments**
SWN	4 (39 plays, e.g. ger000237)	2 (23 plays, e.g. fre001424)
SWT	17 (greek000027)	9 (rus000091)
SFT	20 (ita000098)	13 (ger000279)

so SFT are sensitive in particular to the size of a network (number of speakers in the play): while the mean of the plays in VeBiDraCor is 14.3 speakers (see Table 1), the SWT play with the lowest size has a speaker count of 17 (for SFT it is even 20). The number of segments, on the other hand, does not need to be above average: while the mean of the plays in VeBiDraCor is 19.9 segments (see Table 1), the SWT play with the lowest size has a segment count of 9 (for SFT it is 13).

These findings are supported by tests of correlation. There is a positive correlation between the value for small-world-ness S and the number of speakers;[7] and there is a slightly lower positive correlation between the value for small-world-ness S and the number of segments.[8] Since both factors are essential for the specific form of a dramatic network, and since small worlds are a type more likely to be found in larger networks, these correlations are unsurprising. Of greater interest are those cases where there is no direct correlation between, say, the number of speakers and the small world status of a dramatic network. More on this later.

3.2 SWN Small Worlds as a Standard Type of Dramatic Form

For our discussion of the typological potentials of the different small world concepts, we return to the three observations A, B, C above – and begin with A, of which two aspects are noteworthy. First, according to SWN, the small-world-ness of dramatic networks is nothing special, but the rule: described from a network analytical point of view, SWN small worlds are a (or even: the) standard type of dramatic form. Second, however, it is important to note that not all dramatic networks can be classified as SWN small worlds in the sense of Humphries and Gurney (2008), which raises the question of what other characteristics these networks have.

7 Pearson's r = .612 (p < .001); Spearman's r = .578 (p < .001).

8 Pearson's r = .225 (p < .001); Spearman's r = .454 (p < .001).

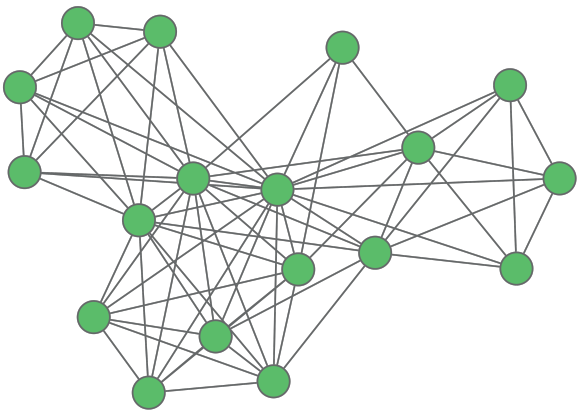


Fig. 3: Georg Kaiser, *Die Koralle*, 1917.

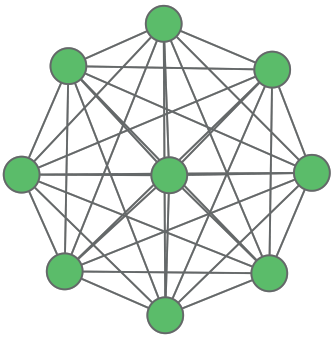


Fig. 4: Calderón, *No hay burlas con el amor*, 1635.

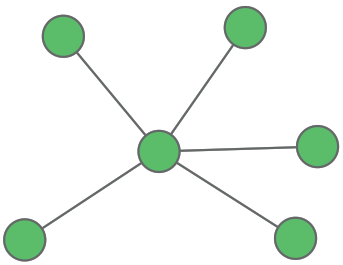


Fig. 5: Louis de Boissy, *L'Apologie du Siècle*, 1734.

To recap: small worlds are a type of network that occupies "a middle ground between regular and random networks," with "high local clustering of elements [...], but also short path lengths between elements" (Humphries and Gurney 2008, p. 1). Let us calculate the average of S for all 2,270 dramatic SWN-small worlds in *VeBiDraCor* and pick the play whose S is closest to this average (which is 1.85), a play called *Die Koralle*, written by Georg Kaiser,[9] which for now shall be taken as our average dramatic SWN small world.

If you look at the respective graph (cf. Figure 3), it actually proves nothing spectacular. *Die Koralle* is a rather average dramatic network, as you get to see it again and again when browsing through DraCor. In this case, we can spot three

9 https://dracor.org/id/ger000545.

somewhat 'more' highly clustered areas (top left, top right, and bottom), and also some nodes in the center of the network that connect these areas, ensuring a rather short path length between them. High clustering, short path length: taken with Humphries and Gurney (2008), this is the core of the small world concept, not more – but not less, either.

Now it might be rather obvious to conclude that the ubiquity of the small world phenomenon, as postulated by general network science, is thus also evident in the dramatic networks that can be typified by SWN. And indeed, there is nothing to argue against this interpretation. From the point of view of literary network studies, however, there is something essential to add. What manifests itself in an average SWN small world network like *Die Koralle* is first of all an effect of *form*, that is: highly clustering areas usually emerge as an effect of the form elements *scene* or *act*. The connection of these areas by central nodes has something to do with the concept of "protagonism," understood as the idea of someone or something carrying the plot that (according to Aristotle) in a play regularly follows the "law of the single strand." In addition, the connections between the areas can also be understood as an effect of the traditional form principle of *liaison des scènes* (even if this principle is not applied in a completely strict sense).

Understanding the structure of dramatic networks (also) as effects of form does not mean to principally reject the views of general network science. It does mean: to speak about dramatic small worlds not as a direct replica of real-world networks, but to always reflect on the literary form as an indispensable mediator of structure, world, society.

And sometimes, form dominates (social) structure. After having discussed small worlds as a form-grounded standard type of dramatic networks, we will now look briefly at the 708 plays from *VeBiDraCor* that were not typified as SWN small worlds (see Table 3). Do the non-small world plays in *VeBiDraCor* share any characteristics, or is every non-small world un-small-world-ish in its own way?

A total of 149 plays from *VeBiDraCor*, for which our analyses did not come to any result at all ("NA") for various reasons,[10] cannot be adequately treated in the context of this question, however. Hence, we set them aside, so 559 plays remain.

Of these, 536 have (at least) two shared characteristics: for their networks, it is both $C = 1$ and $L = 1$, which, third, implies that the density D of the network = 1, which in turn indicates the structural property underlying all of this, namely that the graph is fully connected. So, these fully connected (so to say "full den-

10 Almost all reasons have to do with the fact that we have to divide by 0 at some step of the calculation, mostly because the dramatic network is just too small, and as a result L or $C = 0$. Watts also notes that applying the small world concept usually presupposes that a network is "numerically large" (Watts 1999a, p. 495).

sity") graphs are a second type of dramatic form; Figure 4 shows a randomly selected example, *No hay burlas con el amor*, written by Pedro Calderón de la Barca in 1635.[11] Rarely can such dramatic *full density networks* be understood as a kind of reflection of real-world networks; rather, their maximum clustering is primarily an effect of form, typically caused by specific conventions for structuring the scenes in a play (and often determined by the practical conditions of historical types of theater stage).

In the end, 23 plays still remain. Most of them are variants of full density networks, in which only one or very few edges are missing for being fully connected (we call them *high density networks*). Yet there is another small group that can be typified based on a common characteristic. This group of dramatic non-small worlds is the antithesis of the full density networks in that their clustering coefficient is $C = 0$. These are the *dramatic star networks*, as shown in Figure 5 using the example of *L'Apologie du Siècle* by Louis de Boissy, written in 1734.[12] All in all, we have identified eleven dramatic star networks (or slight variants of them) in *VeBiDraCor*.[13]

3.3 SWT Small Worlds as an Extreme Type of Dramatic Form

While the SWN small world type was found to be the quantitatively dominant standard type of dramatic network, the results of the analysis in Table 3 show that the SWT small world type has a very low prevalence. The construction of the SWT measure using an outlier test may suggest that the SWT measure simply typifies those dramatic networks for which the highest small-world-ness value S was calculated in the SWN analysis. This is not the case. In fact, the plays typified by SWT are extreme types of dramatic small worlds, but their identification follows a somewhat different logic than the assignment of very high S values to dramatic networks. In other words, the extreme types typed by SWT are not necessarily the small-worldiest plays in the sense of S. As is shown in Table 5, among the top 30 dramatic networks ranked by S, there are only eight networks (column SWT = TRUE) that are also categorized by SWT as dramatic small worlds.

We, in a first step, choose a phenomenological approach to the different conceptual logics of S and SWT revealed in the ranking and take a look at two

11 https://dracor.org/id/cal000142.

12 https://dracor.org/id/fre000152.

13 These plays deserve to be studied in more detail. Here is a list of their DraCor IDs: fre000029, fre000036, fre000098, fre000152, fre000223, fre000248, fre000317, fre000574, fre001117, fre001511, tat000001.

Tab. 5: *VeBiDraCor* plays ranked by S, top 30.

Rank	Id	Author	Title	Year	S	SWN	SWT	C^{dev}	L^{dev}
1	rus000053	Gogol	Театральный разъезд после представления новой комедии	1842	26.95	TRUE	FALSE	10.5	0.39
2	ger000348	Gleich	Der Eheteufel auf Reisen	1821	14.63	TRUE	FALSE	8.2	0.56
3	fre000709	La Tessonerie	L'Art de Régner	1645	12.58	TRUE	FALSE	6.7	0.53
4	ger000258	Voß	Faust	1823	11.37	TRUE	FALSE	9.6	0.84
5	fre001169	Quinault	La Comédie sans Comédie	1655	11.34	TRUE	FALSE	5.8	0.51
6	fre000518	Dorimond	La Comédie de la Comédie	1662	10.98	TRUE	FALSE	4.9	0.45
7	fre001014	Molière	Le Malade Imaginaire	1673	10.74	TRUE	FALSE	6.4	0.59
8	ger000393	Grabbe	Hannibal	1835	10.64	TRUE	TRUE	10.3	0.97
9	fre000612	Fagan	Momus à Paris	1770	10.29	TRUE	FALSE	4.7	0.45
10	fre001036	Monselet	L'Enfer des Gens de Lettres	1859	10.28	TRUE	FALSE	6.4	0.62
11	ger000123	Goethe	Götz von Berlichingen mit der eisernen Hand	1773	9.79	TRUE	TRUE	9.9	1.01
12	fre001206	Regnard	Le Carnaval de Venise	1699	9.08	TRUE	FALSE	6.5	0.72
13	fre000427	Dancourt	Nouveau Prologue, et Nouveaux Divertissements	1704	8.81	TRUE	FALSE	4.6	0.53
14	ger000302	Schaefer	Faustine, der weibliche Faust	1898	8.52	TRUE	FALSE	6.9	0.81
15	fre000868	Legrand	La Nouveauté	1727	8.2	TRUE	FALSE	5.9	0.71
16	fre000997	Molière	Le Bourgeois Gentilhomme	1671	8.14	TRUE	FALSE	5.9	0.72
17	ger000085	Büchner	Dantons Tod	1835	7.93	TRUE	TRUE	8.8	1.11
18	rus000191	Bulgakov	Война и мир	1932	7.7	TRUE	TRUE	8.2	1.07
19	ger000378	Grabbe	Napoleon oder Die hundert Tage	1831	7.67	TRUE	FALSE	10.7	1.4
20	fre000993	Molière	Les Amants Magnifiques	1670	7.46	TRUE	FALSE	5.4	0.72
21	fre000068	d'Aure	Geneviève ou L'Innocence Reconnue Tragédie Chrétienne	1669	7.45	TRUE	FALSE	3.9	0.52
22	shake000037	Shakespeare	Pericles	1609	7.18	TRUE	TRUE	6.9	0.97
23	swe000047	Strindberg	Lycko-Pers resa	1882	7.17	TRUE	FALSE	5.8	0.81
24	ger000201	Goethe	Faust. Der Tragödie zweiter Teil	1832	6.97	TRUE	FALSE	8.3	1.19
25	ger000278	Avenarius	Faust	1919	6.92	TRUE	TRUE	6.4	0.93
26	ger000564	Büchner	Woyzeck	1837	6.86	TRUE	FALSE	5.4	0.79
27	ger000555	Baggesen	Der vollendete Faust oder Romanien in Jauer	1808	6.86	TRUE	FALSE	8.7	1.27
28	ger000532	Wolfram	Faust	1839	6.83	TRUE	TRUE	6.2	0.91
29	ger000149	Sorge	Der Sieg des Christos	1924	6.76	TRUE	FALSE	5.3	0.79
30	fre001180	Quinault	Roland	1685	6.68	TRUE	TRUE	6.2	0.93

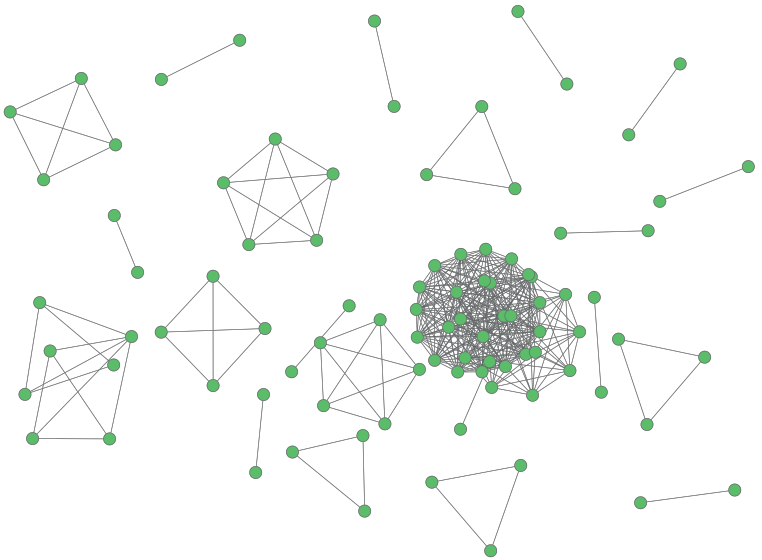


Fig. 6: Nikolaj Vasilevich Gogol, *Театральный разъезд после представления новой комедии*, 1842.

network visualizations. Figure 6 shows the play *Театральный разъезд после представления новой комедии* (*Leaving the Theater after the Presentation of a New Comedy*) by Nikolai Gogol, published in 1842;[14] Figure 7 shows the highest-ranked play also typified according to SWT as a dramatic small world, Christian Dietrich Grabbe's *Hannibal* from 1835.[15] Both dramatic networks are complex, and both dramatic networks are notable at first glance for their numerous semi-autonomous clusters. However, while in Gogol's play the majority of these clusters are unconnected components, in Grabbe's play there are only four separate components in total. Gogol's play thus almost prototypically exemplifies what Watts has called the "caveman graph" (Watts 1999a, p. 500). Yet a caveman network is not a small world.

In contrast, the dramatic network for Grabbe's play (Figure 7) also breaks down into several components. Overall, however, its combination of high clustering and short path length (generated by the bridging edges between the components) makes it much more small-world-ish in the SWT sense.

14 https://dracor.org/id/rus000053.
15 https://dracor.org/id/ger000393.

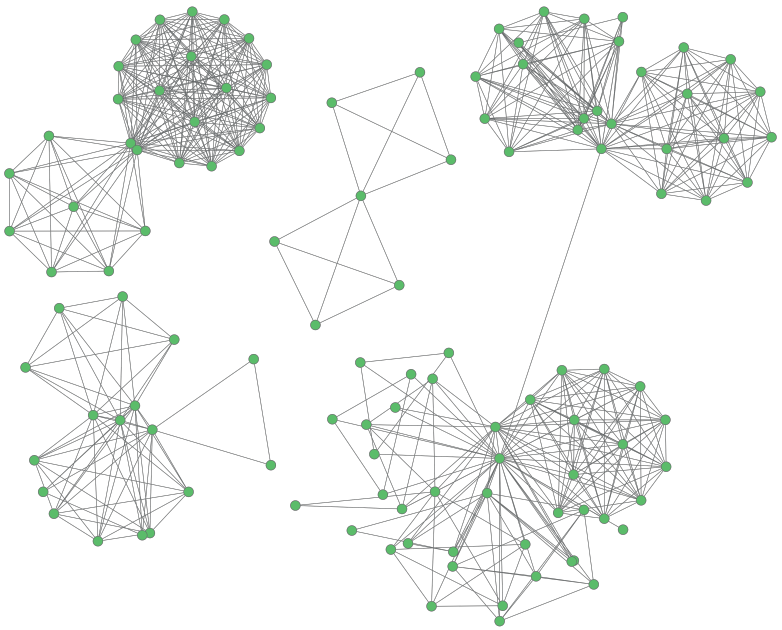


Fig. 7: Christian Dietrich Grabbe, *Hannibal*, 1835.

So, we can note from phenomenological inspection already that our implementation of Humphries and Gurney (2008)'s small-world-ness measure S seems to be quite insensitive to the disintegration of small world candidates into e.g. caveman graphs. But what is the computational reason that Gogol's play has extremely high small-world-ness according to SWN resp. based on S, but is not typified as dramatic small world according to SWT?

The answer lies in the different way the two small world characteristics (high clustering; short average path length) are implemented in the measures. In SWT, which was constructed as a test where both characteristics are checked independently, both characteristics must be met (see equation 3 and equation 4). Humphries and Gurney (2008), on the other hand, relate the two measures to each other by division to create the integrative measure S (see equation 7). However, this makes the small-world-ness measure S susceptible to peaks of one characteristic, which may allow the other characteristic behaving inconspicuously at the same time. To give an example: on the basis of S, a dramatic network may be classified as small world if it has an extremely high clustering (one characteristic with extreme value fulfilled), but at the same time only a medium path length (the other characteristic not fulfilled). This exactly is the case with Gogol's

play (as well as with many other top-ranked plays according to S, which SWT does not typify as small world): here, the L^{dev} deviates too much to pass the criterion II test (see equation 4) and thus cannot be typified as a dramatic small world according to SWT.

What follows from this high sensitivity of S to the usefulness of the measure from the point of view of a general network science cannot be discussed here. From a literary network studies perspective, we suggest taking the two measures as a starting point for quite different directions of research. On the one hand, the idea of a measure of small-world-ness may be an interesting starting point for developing a *general morphology of dramatic networks* in which – as outlined above – network structures are described and interpreted as effects of form. In this direction, small-world-ness could turn out to be something like *a general, at least transhistorical and transnational form property of dramatic networks*. On the other hand, dramatic networks of the SWT small world type offer an approach for *a network-based account to dramatic genres*, with genres understood as*historical forms* that, accordingly, emerge under certain historical conditions – and may disappear again.

4 Instead of an Outlook: Some Thoughts on a Network-Grounded History of Dramatic Form

In a last step, we want to at least briefly discuss the potentials of the considerations just presented for research questions on literary history, namely with reference to a set of diagrams showing historical distributions. We first look at how the different network types that we typified in our analysis (i.e. SWN, SWT, SFT) are distributed historically. Afterwards, we further look into the historical distribution, taking into account that the small worlds come from different DraCor corpora (which are merged in *VeBiDraCor*), usually differentiated by national language

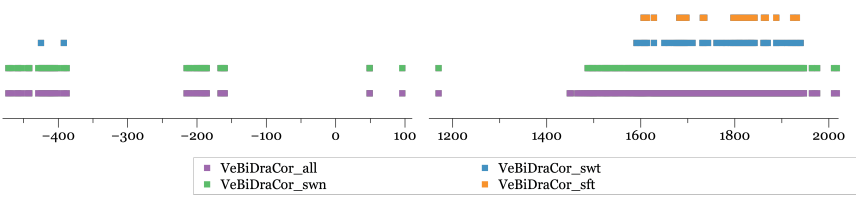


Fig. 8: Historical distribution of dramatic small worlds by different measures.

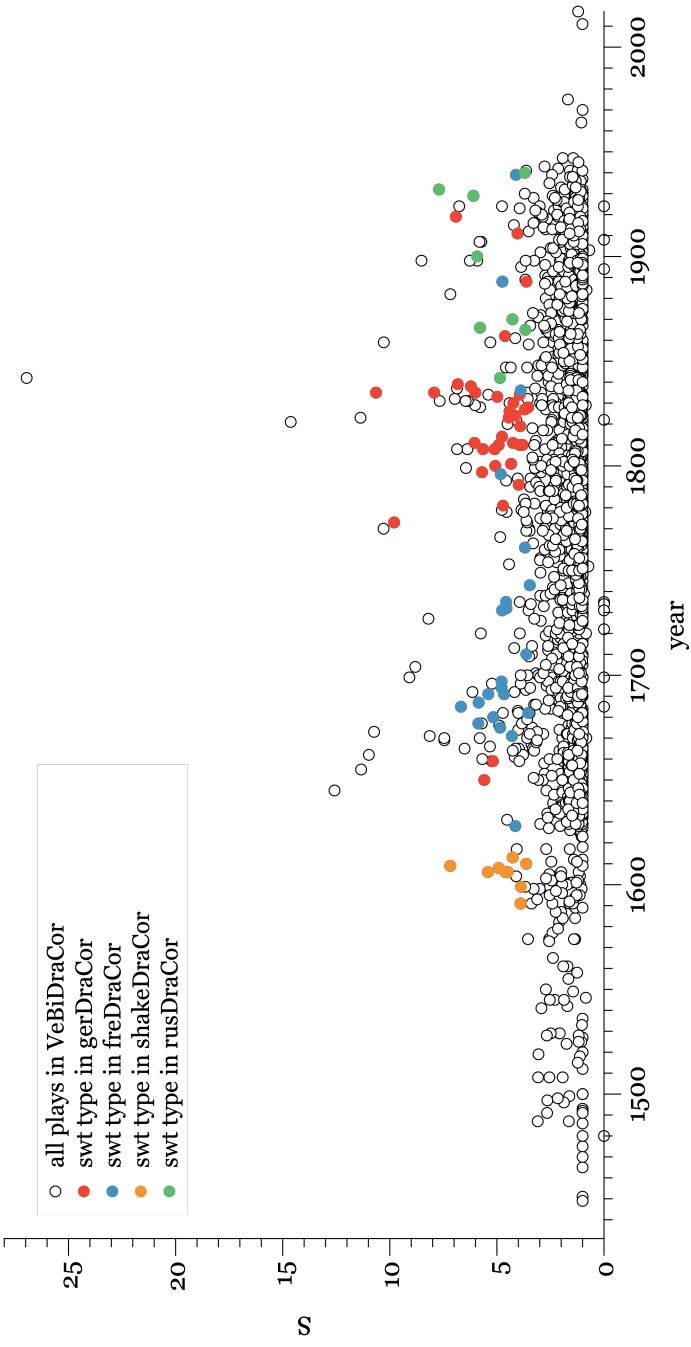


Fig. 9: Historical distribution of all modern SWT type dramatic small worlds differentiated by sub-corpora.

Figure 8 allows a comparison of the historical occurrence of the dramatic networks typified according to the three different concepts. While the purple "baseline" shows all plays (each represented as small square) from VeBiDraCor, the green line represents dramatic small worlds of the SWN type, the blue line represents dramatic small worlds of the SWT type, and the orange line represents dramatic networks of the SFT type. As we have already noted, dramatic small worlds of the SWN type can be understood as a historically more or less indifferent standard type; they describe a transhistorical form option of dramatic networks rather than a specific historical phenomenon.[16] In contrast, dramatic small worlds of the SWT type turn out to be a primarily modern phenomenon. With the exception of two plays by Aristophanes (which would be worth a separate study),[17] the first dramatic small worlds of the SWT type do not appear until the end of the 16th century. Not surprisingly, these first dramatic SWT small worlds of the modern era are plays by Shakespeare, in specific, the two parts of *Henry VI*.[18] Furthermore, the dramatic scale-free networks of the SFT type, additionally characterized by a power-law distribution of the node degree distribution (see criterion III, defined in chapter 2.1), prove to be a genuinely modern phenomenon, at least on the basis of *VeBiDraCor*. Here, too, Shakespeare is the first: the networks of plays like *Timon of Athens*[19] and *Antony and Cleopatra*[20] turn out to be the first examples of dramatic scale-free networks in history. The extraordinary impact of Shakespeare on the history of dramatic form, which is repeatedly attested to him (just to mention studies on the open form in Klotz (1969)), seems to be confirmed by network-grounded analyses.

This, however, is where further research would now have to start, looking much more closely at the structures of these networks, at their realization through dramatic forms and at their relation to dramatic worlds. At this point, we would like to add just one more observation.

Figure 9 shows all dramatic networks from *VeBiDraCor* since 1450, with the small-world-ness value S plotted on the y-axis and the year on the x-axis. The diagram suggests one historical hypothesis: could it be that the form of dramatic net-

16 At this point, at the latest, we however must emphasize that due to the composition of our corpus, we remain limited to a Western perspective in our analyses and thus in our findings. We would be very interested in breaking this Western perspective and including other, non-Western traditions of theatrical texts and so begin to argue from a genuinely global point of view.

17 See https://dracor.org/id/greek000027 and https://dracor.org/id/greek000032.

18 See https://dracor.org/id/shake000020 and https://dracor.org/id/shake000021.

19 See https://dracor.org/id/shake000029.

20 See https://dracor.org/id/shake000035.

works founded by Shakespeare spread in a wave-like fashion through European literatures?

Our analyses, of course, do not provide an adequate answer to this question. In fact, *VeBiDraCor* is far from balanced enough to do so. Moreover, many of the corpora we have brought together in *VeBiDraCor* assemble canonic plays. For a truly comprehensive transnational research on a network-based history of dramatic forms, it turns out, even the already quite large *VeBiDraCor* is still far too small.

Acknowledgment: We warmly thank all the members of and contributors to the DraCor community whose work on the corpora made *VeBiDraCor* – and thus this study – possible in the first place. An overview of all participants can be found at https://dracor.org/doc/credits and, with regard to the TEI files, in the corresponding repositories on GitHub, see https://github.com/dracor-org.

DraCor has, in the context of the project "Computational Literary Studies Infrastructure. CLS INFRA," recently received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 101004984.

Data and Code: Data and code of this study are published on GitHub: https://github.com/dracor-org/small-world-paper. The *Very Big Drama Corpus (VeBiDraCor)* was compiled from data published on DraCor.

We conducted the analysis using Docker containers based on an image of *VeBiDraCor* (*vebidracor-api:3.0.0*) (see https://github.com/dracor-org/vebidracor) and *RStudio* as research environment. Instructions on how to setup the Docker containers to reproduce the results using the provided Docker images published on Dockerhub can be found in the README file in the `publication-version` branch.

Calculations were made using a script written in R, which consists of three parts: reading all necessary data and combining it into one dataframe, intermediate calculations of values for the criteria, and checking of conditions of different tests.

References

Agarwal, Apoorv, Augusto Corvalan, Jacob Jensen, and Owen Rambow (2012). "Social Network Analysis of Alice in Wonderland." In: *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, pp. 88–96. URL: https://aclanthology.org/W12-2513.

Agarwal, Apoorv, Anup Kotalwar, and Owen Rambow (2013). "Automatic Extraction of Social Networks from Literary Text: A Case Study on Alice in Wonderland." In: *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pp. 1202–1208. URL: https://aclanthology.org/I13-1171.

Alberich, Ricardo, Joe Miro-Julia, and Francesc Rosselló (2002). *Marvel Universe Looks Almost Like a Real Social Network*. DOI: 10.48550/arxiv.cond-mat/0202174.

Albert, Réka and Albert-László Barabási (2002). "Statistical Mechanics of Complex Networks." In: *Reviews of Modern Physics* 74.1, pp. 47–97. DOI: 10.1103/RevModPhys.74.47.

Algee-Hewitt, Mark (2017). "Distributed Character: Quantitative Models of the English Stage, 1500–1920." In: *Digital Humanities 2017. Conference Abstracts*, pp. 119–121. URL: https://dh2017.adho.org/abstracts/103/103.pdf.

Broido, Anna D. and Aaron Clauset (2019). "Scale-free Networks Are Rare." In: *Nature Communications* 10.1. DOI: 10.1038/s41467-019-08746-5.

Carron, Pádraig Mac and Ralph Kenna (2012). "Universal Properties of Mythological Networks." In: *EPL (Europhysics Letters)* 99.2. DOI: 10.1209/0295-5075/99/28002.

Coll Ardanuy, Mariona and Caroline Sporleder (2015). "Clustering of Novels Represented as Social Networks." In: *Linguistic Issues in Language Technology* 12. DOI: 10.33011/lilt.v12i.1379.

Elson, David, Nicholas Dames, and Kathleen McKeown (2010). "Extracting Social Networks from Literary Fiction." In: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 138–147. URL: https://aclanthology.org/P10-1015.

Erdős, Paul and Alfréd Rényi (1959). "On Random Graphs I." In: *Publicationes Mathematicae Debrecen* 6.3–4, pp. 290–297. DOI: 10.5486/PMD.1959.6.3-4.12.

Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling, and Peer Trilcke (2019). "Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama." In: *DH2019: »Complexities«. 9–12 July 2019. Book of Abstracts*. DOI: 10.5281/zenodo.4284002.

Fischer, Frank, Peer Trilcke, Christopher Kittel, Carsten Milling, and Daniil Skorinkin (2018). "To Catch a Protagonist: Quantitative Dominance Relations in German-Language Drama (1730–1930)." In: *DH2018: »Puentes/Bridges«. 26–29 June 2018. Book of Abstracts / Libro de resúmenes*. URL: https://dh2018.adho.org/to-catch-a-protagonist-quantitative-dominance-relations-in-german-language-drama-1730-1930/.

Freeman, Linton C. (2004). *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press.

Hettinger, Lena, Martin Becker, Isabella Reger, Fotis Jannidis, and Andreas Hotho (2015). "Genre Classification on German Novels." In: *2015 26th International Workshop on Database and Expert Systems Applications (DEXA)*, pp. 249–253. DOI: 10.1109/DEXA.2015.62.

Humphries, Mark D. and Kevin Gurney (2008). "Network 'Small-World-Ness': A Quantitative Method for Determining Canonical Network Equivalence." In: *PLoS ONE* 3.4. DOI: 10.1371/journal.pone.0002051.

Jayannavar, Prashant, Apoorv Agarwal, Melody Ju, and Owen Rambow (2015). "Validating Literary Theories Using Automatic Social Network Extraction." In: *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, pp. 32–41. DOI: 10.3115/v1/W15-0704.

Klotz, Volker (1969). *Geschlossene und offene Form im Drama*. 4th edition. Hanser.

Krautter, Benjamin, Janis Pagel, Nils Reiter, and Marcus Willand (2018). *Eponymous Heroes and Protagonists – Character Classification in German-Language Dramas*. URL: https://www.digitalhumanitiescooperation.de/wp-content/uploads/2019/06/p07_krautter_et_al_eng-1.pdf.

Krug, Markus (2020). "Techniques for the Automatic Extraction of Character Networks in German Historic Novels." PhD thesis. Universität Würzburg. DOI: 10.25972/OPUS-20918.

Kydros, Dimitrios and Anastasios Anastasiadis (2014). "Social Network Analysis in Literature. The Case of 'The Great Eastern' by A. Embirikos." In: *Proceedings of the 5th European Congress of Modern Greek Studies*, pp. 681–702.

Kydros, Dimitrios, Panagiotis Notopoulos, and Georgios Exarchos (2015). "Homer's Iliad – A Social Network Analytic Approach." In: *International Journal of Humanities and Arts Computing* 9.1, pp. 115–132. DOI: 10.3366/ijhac.2015.0141.

Labatut, Vincent and Xavier Bost (2019). "Extraction and Analysis of Fictional Character Networks: A Survey." In: *ACM Computing Surveys* 52.5, 89:1–89:40. DOI: 10.1145/3344548.

Lee, John and Chak Yan Yeung (2012). "Extracting Networks of People and Places from Literary Texts." In: *Proceedings of the 26th Pacific Asia Conference on Language, Information, and Computation*, pp. 209–218. URL: https://aclanthology.org/Y12-1022.

Mac Carron, Pádraig and Ralph Kenna (2013). "Network Analysis of the Íslendinga Sögur – the Sagas of Icelanders." In: *The European Physical Journal B* 86.10. DOI: 10.1140/epjb/e2013-40583-3.

Miranda, Pedro Jeferson, Murilo Silva Baptista, and Sandro Ely de Souza Pinto (2018). "The Odyssey's mythological network." In: *PLOS ONE* 13.7. DOI: 10.1371/journal.pone.0200703.

Moretti, Franco (2011). "Network Theory, Plot Analysis." In: *Pamphlets of the Stanford Literary Lab* 2. URL: https://litlab.stanford.edu/LiteraryLabPamphlet2.pdf.

Moretti, Franco (2020). "Simulating Dramatic Networks: Morphology, History, Literary Study." In: *Journal of World Literature* 6.1, pp. 24–44. DOI: 10.1163/24056480-20201001.

Newman, Mark E. J. and Michelle Girvan (2004). "Finding and Evaluating Community Structure in Networks." In: *Physical Review E* 69.2. DOI: 10.1103/PhysRevE.69.026113.

Park, Gyeong-Mi, Sung-Hwan Kim, Hye-Ryeon Hwang, and Hwan-Gue Cho (2013). "Complex System Analysis of Social Networks Extracted from Literary Fictions." In: *International Journal of Machine Learning and Computing*, pp. 107–111. DOI: 10.7763/IJMLC.2013.V3.282.

Reger, Isabella (2016). "Figurennetzwerke als Ähnlichkeitsmaß." MA thesis. Universität Würzburg. URL: https://nbn-resolving.org/urn:nbn:de:bvb:20-opus-149106.

Stiller, James and Mathew Hudson (2005). "Weak Links and Scene Cliques Within the Small World of Shakespeare." In: *Journal of Cultural and Evolutionary Psychology* 3.1, pp. 57–73. DOI: 10.1556/jcep.3.2005.1.4.

Stiller, James, Daniel Nettle, and Robin I. M. Dunbar (2003). "The Small World of Shakespeare's Plays." In: *Human Nature* 14.4, pp. 397–408. DOI: 10.1007/s12110-003-1013-1.

Trilcke, Peer (2013). "Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft." In: *Empirie in der Literaturwissenschaft*. Ed. by Philip Ajouri, Katja Mellmann, and Christoph Rauen. Mentis, pp. 201–247.

Trilcke, Peer, Frank Fischer, Mathias Göbel, and Dario Kampkaspar (2016). "Theatre Plays as 'Small Worlds'? Network Data on the History and Typology." In: *DH2016. Digital Identities: the Past and the Future*. Jagiellonian University, Pedagogical University of Krakow, pp. 385–387. DOI: 10.5281/zenodo.6974706.

Watts, Duncan J. (1999a). "Networks, Dynamics, and the Small-World Phenomenon." In: *American Journal of Sociology* 105, pp. 493–527. DOI: 10.1086/210318.

Watts, Duncan J. (1999b). *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton Univ. Press.

Watts, Duncan J. and Steven H. Strogatz (1998). "Collective Dynamics of 'Small-World' Networks." In: *Nature* 393.6684, pp. 440–442. DOI: 10.1038/30918.

Beatrice Maria Vittoria Gavazza, Leon A. Glaser,
Virginia Mastellari, and Anna Novokhatko

Exploring Ancient Greek Comedy: Material Engagement Theory and Computational Tools

Abstract: Material objects are of outstanding importance for theater studies and the analysis of drama, the genre in which the imaginary world is performed and embodied on stage. Material Engagement Theory (the world of things, artifacts and material signs added to the cognitive equation) proves to be an insightful tool when applied to ancient Greek comedy, a genre which preserves a significant number of objects from antiquity. Material Engagement Theory is the theoretical basis of the database we have been building since 2019, namely the *Lexicon of Objects from Greek Comedy* (https://www.lggk.uni-freiburg.de/wordpressNew/), the first database collecting and visualizing mentions of artifacts from the entire field of ancient Greek comedy.

This paper explains the principles of applied Material Engagement Theory (section 1) and gives examples of how objects in drama influence human perceptions (section 2). It then discusses the database and its functions (section 3), and finally presents the research results from the database with detailed explanations of several case studies (section 4). The examples in the last section show that both quantitative analysis and the computer-assisted processing of the dramatic corpus are necessary for contemporary research on ancient drama. They open up fresh avenues for interpretation and raise philological and scholarly work to a new level of research.

1 Materialities and Our Engagement With Them

To be human is to be closely connected to the non-human, material things that surround us. This is the main area and goal of new materialism: an interdisciplinary, theoretical and politically engaged field of research that emerged at the dawn of the new millennium as part of what might be called the post-constructivist, ontological or material turn (Coole and Frost 2010). The revival of materialist ontologies has been stimulated by a productive friction with the linguistic turn and the

Beatrice Maria Vittoria Gavazza, Leon A. Glaser, Virginia Mastellari, Anna Novokhatko, University of Freiburg

https://doi.org/10.1515/9783111071824-003

social constructivist framework, critically questioning their limits, which derive from the meaning of language, culture and representation. Accordingly, the polycentric investigations that have consolidated the scientific corpus of new materialism focus on the primacy of matter or "things" as an unexplored question and seek a renewed substantive engagement with the dynamics of materialization, whether in relation to bodily life or material phenomena. The new materialism revises the earlier notion of things as a unitary, inert substance or a socially constructed fact and foregrounds novel accounts of their active force, processual character, formative impulse and self-organizing capacities.

Things, then, are capable of influencing our thinking because their material components attract and appeal to our senses, and the values we attach to such experiences can influence our behavior in profound and diverse ways. Recently, much work has been done on approaches to material culture that express the role of things in shaping human practices and values. Literary studies have also taken a "material turn," due to Bill Brown's "thing theory" (Brown 2001). The things of the past, the present and the (imaginary) future – both in their material reality and in their literary and artistic renderings – are experiencing a rebirth.

The original aim of the database *Lexicon of Objects from Greek Comedy* is to explore how the mind and things interact by focusing on everyday objects in classical Greece, such as weapons, shoes, clothes, vases, textiles and utensils (among many others). These objects are part of a fluid economy of exchange between mind, body, and things. There are clear advantages to focusing on concepts that span multiple material registers (metal, fabric, wood, clay, etc.) in an interdisciplinary way rather than looking at just one type of object or material. In this way, we gain a better understanding of (1) how, by applying the same terminology to a variety of things, the ancient Greek mind conceived of the relationship between different kinds of artifacts and (2) how the perception and effect of objects remained constant or changed across different media and eras. Through its interest in experience as a form of interaction between people and things, the database project joins the growing body of works in the humanities that address the question of the ways in which people make sense of things and, conversely, how things make sense of people – how things make us who we are as individuals, societies and epochs.

This paper shows how the use of computer data in the analysis of dramatic texts can contribute to exploring (1) how matter influences or even constitutes human thought and action and (2) how human cognitive processes develop and unfold via interaction with things. Literary representations of material objects and human experiences with them can reveal much about the contours of the human mind and body as perceived by authors across different eras and genres. Conversely, the textual approach adopted here ensures that equal attention is paid

to the ways in which texts give meaning to material substances and bring them into effect.

Furthermore, a recurrent feature of "new materialisms" is their concern with the ontological distinctions between people, persons, things and materials. In the context of objects from ancient Greek comedy – which encompasses all kinds of things, both living and non-living, material and immaterial, organic and inorganic – the interstices and transitions between these categories have proven to be a pervasive theme. Jane Bennett's 2010 concept of vital or "vibrating" materialism is particularly relevant to the comic stage. By the term "vibrant" Bennett refers to the affective vitality of material things – the ability of non-human substances to color and shape the human experience of the world. According to Bennett, drawing attention to material things in this way means highlighting the extent to which human beinghood and thinghood intersect, the extent to which the "we" and the "it" slide into one another. One moral of the story is that we are also non-human and that things are also important actors in the world (Bennett 2010, p. 4).

In this sense, the interpretations that emerge during the textual analysis will not only address the effect of things on people and thus the power of materiality, but also the mixing of human and non-human elements in the representation of cognition and perception. In analyzing the data, we come across objects and materials that are both vital and alive. By this, we are specifically referring to the acting and affective capacities of things as manifested in the evocation of affects, emotions and other kinds of human responses. Vital refers to the appearance or impression of vitality in objects, i.e. the way in which the sensorial qualities of the materials mentioned in dramatic texts signal their own acting capacities.

Whereas the new materialism approach aims to establish a balance between persons and things and between human and non-human beings, the embodied approach to cognition challenges the notion of thought as a process tied to the brain. A number of theories of embedded, enactive extended and/or embodied ("4E") cognition have emerged in recent decades, but they all share the attempt to decipher how the brain responds to and engages with forces outside the skull, be they physical senses or features of the external world (Newen et al. 2018). New materialisms and the 4E cognition theory challenge Cartesian dualism, which juxtaposes material things and mental processes. More generally, embodied cognition emphasizes that mental activity is linked to the functioning of the body. This means that cognition is also dependent on the body's interactions with its physical environment, including material entities.

Some manifestations of embodied cognition prove particularly relevant to the cognitive facets of the objects referred to in dramatic texts: enactive perception and the extended mind on the one hand and material engagement on the other. The enactive perception and extended mind theories – as formulated by Alva Noë

(2004) and Andy Clark (2008), respectively – assume that sensory perception and higher order thinking (e.g. mathematical calculation) emerge from our dynamic, embodied engagement with the things that populate our everyday world. While Noë's enactive approach looks at the input of our sensorimotor abilities with respect to our qualitative experience of the phenomenal world, Clark's work focuses on the tools and external "props" (including language itself) that guide and structure thought:

> The proper response is to see mind and intelligence themselves as mechanically realized by complex, shifting mixtures of energetic and dynamic coupling, internal and external forms of representation and computation, epistemically potent forms of bodily action, and the canny exploitation of a variety of extrabodily props, aids, and scaffolding. Minds like ours emerge from this colourful flux as surprisingly seamless wholes: adaptively potent mashups extruded from a dizzying motley of heterogeneous elements and processes. (Clark 2008, p. 219)

The findings of 4E cognition shed light on the particular ways in which our physical environment can direct and promote human thought processes and thus provide useful patterns for the interplay between cognitive activity and material properties so often mentioned in Greek comedies.

Lambros Malafouris' theory of material engagement instead emphasizes that the objects in our environment are not only cognitive tools, but also contribute to constituting thought (Malafouris 2013, 2018, 2019; Poulsgaard and Malafouris 2020). He argues, for example, that Linear B tablets were not simply passive vehicles for the transmission of thought in the form of linguistic signs, but that these tablets, in their sensory and spatial dimensions, actively shaped and transformed the task of recording and remembering data:

> Thinking is not simply a matter of reading a series of meaningful linguistic signs inscribed on the surface of a tablet, but also a matter of meaningfully engaging with the tablet itself as a material sign. (Malafouris 2013, p. 238)

More particularly, Malafouris' theory of material engagement explains how the specific sensory qualities of certain materials (e.g. textile, clay, wood, metal) can alter and shape the course of human engagement with that material.

The objects of the literary and material tradition provide information about the media in which this quality can materialize, as well as the effects of this materialization on human beings. At the same time, the descriptions of the objects themselves, as found especially in literary representations, reveal the ways in which the ancient sources conceptualized the relationships between the human body, the mind and material things.

In a previous publication we emphasized the significance of objects in comic texts and argued that access to these objects through our database enables a vivid perception of them not as objects mentioned in ancient Greek texts, but as real haptic and visual artifacts that are present in our consciousness (Novokhatko et al. 2021). This will also be made clear through some examples in the next section. In the following pages we will briefly present the structure of the database and show the recent changes it has undergone to meet the needs of our ongoing work. We will then explore the interpretive tools/possibilities in relation to the material objects found in comedy and demonstrate how one object can indicate a great deal about the overall dramatic context.

2 The Relevance of Objects in Greek Comedy

The interaction between objects and characters is crucial for ancient comedy, not least in terms of its effect on the audience. Through their visual impact, objects determine and enrich the set design, dictating the timing, space and humorousness of the scenes.[1] Characters interact with objects, manipulate them, use their names for verbal puns or create absurd or ridiculous situations by misusing them.

Complete surviving comedies offer enlightening examples of the effects produced on the audience's perception by the interaction of objects and characters. In Aristophanes' *Acharnians* one of the characters, the Athenian general Lamachus, enters the scene armed with a plumed helmet and a shield decorated with the head of the Gorgon (ll. 574–575).[2] Given the context of this scene, these weapons have an intrinsically comic effect: the protagonist of the play, the Athenian Dikaiopolis, has just signed a personal peace treaty with the Spartan enemies, and Lamachus shows up in arms with very bad timing. He embodies the image of war, but is ridiculously out of place, since the peace has already been established. Lamachus' weapons do not exhaust their comic effect with the plain contrast between war and peace. On the contrary, the hilarity of the scene grows to a climax through the characters' physical interactions with these objects. While Lamachus and Dikaiopolis are talking, the weapons are displaced and manipulated, and gradually lose their dignity. Dikaiopolis makes Lamachus remove the

1 For humor studies, especially the application of humor theories to classical texts, see e.g. Raskin (1985), Attardo (2001), Bolens (2001), Morreall (2009), Davies (2011), Dennett et al. (2011), Double (2014), Attardo (2017), and Lowe (2020).

2 On the symbolic value of these objects, cf. Castellaneta and Maffione (2016, pp. 459–460). On the plot of the *Acharnians*, cf. Olson (2002, pp. xl–lii).

helmet from his head and place it on the ground (so that the plume – the noblest part – enters in contact with dirt, l. 583); he then asks him for a feather from the plume (l. 584), which in a sign of utmost irreverence Dikaiopolis uses to induce vomiting (l. 587). Later in the play, a similar scene makes fun of Lamachus' weapons by contrasting them with foods (ll. 1095–1149). As Lamachus assembles his military equipment, everything he mentions is ridiculed by Dikaiopolis, who is getting his dinner ready. They ask their slaves to bring out of the house weapons and food (respectively), and the slaves run back and forth collecting what has been called for. The result is the materialization of the objects onstage. The strong contrast between the two characters is reflected not only by the hostile language they employ in this scene, but also by the display of the objects. Food abundance, such as that resulting from Dikaiopolis' requests, jars with a war context, such as the one Lamachus is getting ready to enter, since in war time food is anything but abundant. Funnily enough, Dikaiopolis echoes Lamachus' requests by changing either a single word or very few ones, but the final result of his requests, assembled onstage, is antithetical to the general's orders.

A similar comic technique, whereby objects first create a visual contrast and then, through their interaction with the characters, achieve dynamic comic effects, is exploited by Aristophanes in the prologue to another comedy, the *Thesmophoriazusae*.[3] The famous tragedian Euripides is featured here as a character and appears on stage together with his relative Mnesilochus. Euripides is convinced that the Athenian women want to sentence him to death because of his habit of depicting female vices and wants to persuade one of his colleagues, the tragic poet Agathon – a young, handsome and effeminate man – to infiltrate the celebrations of the Thesmophoria festival, which is restricted to women, to avert the danger.[4] Agathon's character is brought on stage, probably by means of a sliding or rotating stage machine, certainly together with a bed (l. 261). In this way the setting (Agathon's furnished dwelling and accessories) enters the scene together with the character. Faced with the confusion of Agathon's wardrobe, Mnesilochus launches into a long interrogation (ll. 136–140). The presence on stage of a male figure dressed as a woman must have instantly achieved a visual comic effect. Verbal mockery ensues: Mnesilochus focuses on Agathon's accessories, which are listed in order to emphasize the absurdity of their juxtaposition. The list is interesting because it also highlights a series of gender stereotypes: typical feminine objects include the saffron gown (l. 138), the hairnet (l. 138), the breast-band (l.

3 On the prologue of the *Thesmophoriazusae*, see Austin and Olson (2004, p. lv).

4 On the plot of the *Thesmophoriazusae*, see Austin and Olson (2004, pp. lii–lxviii).

139) and the mirror (l. 140).[5] The Greek term used here for "garb" (*stolē*, l. 136) could indicate either a male or female garment; however, the context of the joke determines the identification of the *stolē* as a feminine item (cf. Austin and Olson 2004, pp. 82, 101): as the actor playing the character enters the scene in what is clearly a woman's dress, one should not expect a more precise designation. At the same time, Agathon possesses objects that are exclusive to men: the *barbitos* ("bass") and the lyre, stringed musical instruments used by poets, the oil-flask used in gymnasiums (l. 139) and a soldier's sword (l. 140).[6]

In response to Agathon's refusal to infiltrate the Thesmophoria celebrations, Mnesilochus offers to dress as a woman and plead Euripides' cause before the Athenians. Mnesilochus' disguise scene opens with him shaving: the old man must be completely smooth, so Euripides asks Agathon for a razor (l. 219). Agathon tells Euripides to take one from the case (l. 220): we can imagine Euripides' movements on the stage as he grabs the razor case and draws out a blade. After the razor, Euripides asks for a torch or a lamp (l. 238) to singe the hair on the rest of the man's body, particularly his buttocks. Since the sources report that women used a lamp for this operation,[7] it is probable that the object brought on stage and used to prepare the playwright's relative was a large torch, which would have amplified the ridiculousness of the scene when brought close to the reluctant character's bottom (Sommerstein 1994, p. 173). After the shaving comes the dressing. Mnesilochus must wear a *krokōtos*, a yellow-dyed garment typically worn by women at festivals and religious celebrations.[8] In an evidently awkward manner, Mnesilochus first wears the robe and, only afterwards, a girdle to hold up the breasts (ll. 255–253). It is unclear whether the fact of first wearing the robe and then the bra was a way of emphasizing the characters' awkwardness or whether, as a result of this action, he then wore the girdle over the robe, instantly achieving a visual comic effect. The preparation of Euripides' relative is completed by hair accessories (l. 257) and shoes (l. 262, with the generic term *hypodēmata*, but obvious recourse to women's shoes on stage). Agathon invites Euripides to take everything from his bed (l. 261): the props construct the setting and their location dictates the space and time of the stage action.

5 On these objects, see Austin and Olson (2004, pp. 101–103). On their symbolic value in the scene, see Castellaneta and Maffione (2016, pp. 320–321).

6 On these objects, see Austin and Olson (2004, pp. 102–103).

7 Aristoph. *Lys*. 825–828, *Eccl*. 12–13.

8 In a theatrical context, this was a costume of Dionysus, the effeminate god par excellence, e.g. in Aeschylus' *Aedonians* or Aristophanes' *Frogs*; cf. Stone (1981, pp. 174–175); Dover (1993, p. 40); Mastromarco and Totaro (2006, p. 312 n. 9, p. 452 n. 23, p. 564 n. 6.).

Not only complete dramas, but also comic fragments testify to the centrality of the interaction between characters and objects for the staging of dramatic scenes. In a fragment from the Sicilian comic repertoire (Epicharmus' fr. 147 K.-A.), two characters talk about an object that lies in front of them by playing on its name. This is a tripod – plainly, says character (B.) (l. 1). Tripods were composed of three-footed supports of bronze, or a similar material, and a fixed or movable vessel supported by the three feet and by a fourth, more slender element at the center. They were used as tables in houses, but also had a votive function: they were typically displayed in sanctuaries as votive offerings.[9] This tripod, however, has four legs, which would actually make it a 'tetrapod,' as both characters (A.) and (B.) state (ll. 2–3). The final stanza alludes once again to the name of the tripod with a reference to the myth of Oedipus, who solved the Sphinx's riddle about something that has four, two or three feet (i.e. man). In Epicharmus' fragment, at a micro-level the object provides a cue for the gag about the contrast between the name and the identity of the four-footed tripod; at a macro-level, it suggests what the setting of the action might be, namely a sanctuary (a usual place for votive tripods) or a household.

The examples just discussed show that in dramas, too, people (i.e. the characters) interact with their environment in material ways, as Malafouris (2013) pointed out. This interaction extends beyond the stage: through their visual impact, objects determine the setting and ensure the immediate involvement of the spectators. On a more complex level, the interaction of props and characters allows the playwright to develop entire scenes by playing on dynamic comic effects. Finally, on an abstract level, they can allow the characters to verbally play upon the contradictions between objects' forms and their denominations.

The examples provided once again confirm the value of a materialistic approach to ancient Greek comedy. An essential step to bringing materiality into focus – and to exploiting the information it can provide for drama analysis – is the cataloguing of the objects found in comic plays and fragments.

3 The *LGgK* Database

The *LGgK* (from the German *Lexikon der Gegenstände aus der griechischen Komödie*) is a database which not only collects data, but proves to be a useful tool for the quantitative analysis of ancient drama. While this second point will be discussed

9 Novokhatko (2020) identifies it as a votive tripod. On this object, see Reisch (1905, coll. 1669–1696); Suhr (1971).

in the next section, in the present one we shall present the contents and structure of the database.[10]

The *LGgK* is an interdisciplinary tool designed to collect all objects mentioned in ancient Greek comedy,[11] a genre notoriously rich in information on artifacts.[12]. Information is derived from already existing commentaries on ancient texts.[13] The material is catalogued as follows. For each object the database provides: a translation; references to author(s) and fragment(s) where the object is found; potential additional information on the object, such as dialectal variants attested in the texts or different spellings for the same object; modern literature on the object; and an image.[14] To facilitate not only the cataloging, but also the research, we have divided all the objects on the database into categories (see the next section). The "search-by-category" function allows the user to get an overview of groups of objects belonging to the same field and proves useful to understand the diffusion of certain artifacts compared to others, as we shall demonstrate in the next section.

The structure of the *LGgK* is a mySQL database supported by WordPress and hosted by the server of the University of Freiburg (https://www.lggk.uni-freiburg.de/wordpressNew/). We originally decided to develop the database according to the relational database model:[15] this seemed like the most suitable for organizing and searching a large amount of rather simply structured data, as it had proven to serve as a valuable basis for the quantitative and qualitative analysis of large corpora.

However, by expanding and refining the categories used to catalogue the objects, it became clear that the database needed a more complex system of links. The first problem was that we had objects that could be assigned to more than one category. The second problem was the need for a more complex division into

10 We offered a thorough description of the database and its functions in Novokhatko et al. (2021, pp. 262–267). The account presented here is functional to introduce those points that are most relevant for the discussion in the next section.

11 The aim is to make the data accessible to researchers from different disciplines and with different backgrounds. On this requirement, see Poulsgaard and Malafouris (2020).

12 For the reasons behind the choice of Greek comedy, see Novokhatko et al. (2021, pp. 256–262)

13 We started from the commentaries on comedic fragments published as part of the project *Kommentierung der Fragmente der griechischen Komödie* (https://www.komfrag.uni-freiburg.de/). The series is still ongoing. Upon completion, the corpus will also include the complete comedies by Aristophanes and Menander.

14 The "image" section is still a work in progress and is designed to collect available pictures and reproductions of the artifacts or hyperlinks to archaeological databases.

15 See https://cloud.google.com/learn/what-is-a-relational-database. For a description of the relational database model, see, moreover, Codd (1970).

subcategories: on the one hand, to split categories into further subcategories, and on the other hand, to combine different existing categories into macro-categories.

Although the relational model has long been a popular one for databases, it suffers from the limitations imposed by the structure of the data. In this model, values attributed to the objects (in the *LGgK*: term, annotation, translations, references, literature, image and category) are fixed according to a defined scheme. What we need is a model in which an object's attribute can evolve over time. Therefore, the relational data model is to be replaced by a Resource Description Framework (RDF), also known as "Triplestore." Rather than assigning a single value to a given key item, the *triplestore* consists of "subject-predicate-object" entities that allow any number of "objects" (according to RDF terminology) to be assigned to a given "subject." While relational data models associate items with attributes (e.g. categories, authors etc.) that do not communicate with each other, the RDF model allows a single "subject" to be associated with multiple "objects." Therefore, RDF will enable more complex attributions, guarantee more precision and allow for more options in the retrieval process. At the same time, it will also provide additional access to the data via standard query languages (e.g. SPARQL) for more experienced users.

4 The *LGgK* for Quantitative Drama Analysis

Notwithstanding future improvements, the *LGgK* already shows great potential as a starting point for quantitative drama analysis. As anticipated in the previous section, a key aspect of the database is the cataloging of objects according to categories, which also proves to be an insightful tool for material analysis. So far, the process of identifying categories and cataloging objects has been carried out manually. The search for objects starts from the index of the commentaries on comic fragments and the selection process is subject to vetting by the research team. The choice of categories has been made empirically on the basis of the data collected. The reason for this is threefold: first, research on fragments is still largely ongoing. The full corpus has not yet been studied in detail[16] and thus has yet to be coded in

16 Comic fragments are still the focus of commentaries and editions today, and some of them have not yet been studied in depth. The last volume of R. Kassel and C. Austin's edition of comic fragments (*Poetae Comici Graeci* (PCG), De Gruyter: Berlin and New York 1983–2001), containing Menander's papyrus fragments (edited by R. Kassel and S. Schröder), has only recently been published (2022). Concerning the 73 planned volumes of commentaries on comic fragments, 37 have been published to date (October 2022).

such a way that it might be used in machine learning processes. The second reason lies in the characteristics of the corpus, which mostly consists of fragments that are sometimes in a poor textual condition. Gaps and variants occur, and human intervention is still crucial to overcome such textual challenges. The third reason is the risk of coming across cases that do not conform to the established standard, which would preclude a machine-based approach to identifying and categorizing objects. This is the case, for instance, with the lemma *kanabeuma*, "carver's sketch," attested in an Aristophanes fragment (fr. 719.2 K.-A.).[17] This word only occurs in this passage as an alternative form of *kinabeuma*, translated by lexica as "knavish trick." The standard translation cannot be applied to this word: one needs to consider not only the text, but also the context and the general meaning.

On the other hand, computational analysis may be an effective and easily applicable method to study the data collected in the database (see below). At the moment, the *LGgK* database includes objects from the 21 commentaries scrutinized so far.[18] This involves 75 ancient authors chronologically divided into three groups – 6th- and 5th-century BC poets, 5th–4th-century poets, and 4th-century poets (see Tables 1a–1c)[19] – and a total of 526 objects. The chronological distribution of the various categories (Table 2) reflects the fact that so far, we have included more material from the 5th–4th centuries and from 6th and 5th centuries than from the 4th.[20] A greater variety is attested for the 5th–4th centuries than for the 6th and 5th centuries, as Tables 1a–2 show. Although the number of 6th- and 5th-century poets in the *LGgK* significantly exceeds that of 5th–4th-century poets, the number of commentaries covering the two periods is 9 for the 6th- and 5th-centuries, and 8 for the 5th–4th centuries, thus constituting very similar percentages of the total volumes contained in the database (42.9 % and 38.1 % respectively). The chronological distribution of the objects can also be interpreted on the basis of qualitative explanations, namely the characteristics of the comic genre, both in general and in a given period (see below), and the characteristics of the fragments (for instance, their state of preservation or the sources' interest in them).

17 See Bagordo (2017, pp. 122–127). All comic fragments are quoted according to K(assel)–A(ustin)'s edition mentioned in the previous footnote.

18 At the present stage, items from the last 16 published commentaries on comic fragments are yet to be included in the database, as are the data from the remaining 36 volumes that have not yet been published.

19 Depending on the number of fragments preserved for each author, the commentaries may collect a single playwright or several ones.

20 This depended on the publication of the commentaries, which began with the most ancient texts.

Although the results are not yet to be regarded as definitive, they offer interesting insights that can fruitfully be combined with approaches inspired by Material Engagement Theory. The most widely represented category is "food" (*Essen*, 91 entries, approximately 16 % of the total of 559 categorizations),[21] followed by the category encompassing all "vessels" associated with dining (*Gefäß*, 51 entries, approximately 9 % of the total – as distinct from *Behälter*, applied to "containers" used outside the culinary context) and the category "clothing" (*Kleidung*, 41 entries, approximately 8 % of the total). The last statistics concern categories of objects from more specific fields, such as "medicine" or "fishing" (*Medizin* and *Fischfang*, both with 2 headwords, 0.358 % of the total). For more details, see Figure 1.

Individual authors can also provide insights on these statistics. One example is Eupolis (5th century BC), a poet who accounts for a total of 494 fragments (Olson 2014, 2016, 2017), 348 of which have already been included in the database. Given this material, Eupolis' statistics slightly differ from the general ones (chart 3). Several categories are not represented at all; out of the 37 total categories, only 15 are found in his extant production (in order of frequency: *Gefäß*, *Essen*, *Mobiliar*, *Konstruktion*, *Getränk*, *Kleidung*, *Landwirtschaft*, *Werkzeug*, *Musik*, *Geld*, *Gewürz*, *Küchengerät*, *Kult*, *Material* and *Textilie*). The categories most frequently associated with this author are "tableware" (*Gefäß*, 7 records, approximately 20 % of the 33 categorizations), followed by "food" and "furniture" (*Essen* and *Mobiliar*, 4 records each, approximately 12 %) (Figure 2).

In the analysis of quantitative data from the database a few aspects should be considered in order to avoid reaching hasty conclusions. First, the characteristics of the ancient comic genre must be taken into account. As mentioned, the most widely represented category in the database is "food," followed by "tableware." The above-mentioned case of Eupolis presents a slight inversion – "tableware" as the first category, with "food" and "furniture" sharing second place – but the podium is often occupied by the same winners. This is perfectly consistent with something we have long known about the comic genre, namely that it assigned a prominent role to corporeality (cf. Davidson 1993; Gilula 1995; Davidson 1997; Wilkins 2000). Indeed, this genre is rich in all those elements that shape the physicality of bodies, specifically food and drink, and objects useful to consume them (Wilkins 2000, particularly pp. 1–51). Even things that seem distant from the culinary sphere, such as politics, are often discussed onstage through food images

21 We are considering the overall number of categorizations (559), which is higher than that of the headwords (526), because a lemma can be assigned to more than one category.

(cf. Beta 2004, pp. 144–147). This tendency increases the recurrence of terms pertaining to the semantic field of food.

A second point to consider concerns the sources of the fragments. Many fragments have not been transmitted "directly" (on papyrus scraps, for instance), but rather "indirectly" as quotations in later authors. As a consequence, fragments are the result of a selection depending on the needs of the authors quoting them. If one considers Eupolis again, the majority of his fragments have been handed down by the lexicographical tradition (351 out of 494, 71.05 %).[22] This figure is also mirrored by the database: out of 348 Eupolis fragments already catalogued, 250 of them are transmitted by lexicographical sources (71.84 %). It is possible that the sources play a role in the recurrence of certain types of objects. On the other hand, indirect transmission could provide a reliable sample of the objects mentioned in the complete plays. To ascertain this, a comparison between the data from comedic fragments and that from complete comedies is needed, ideally works by contemporary authors (for Eupolis one can take Aristophanes, whose 11 complete comedies, with a total of 15 291 lines,[23] would provide a good point of comparison).[24] This will be possible once the database has been completed.

Besides the analytical data it provides, the database also offers a number of insights and interpretative possibilities. As far as the data are concerned, we mentioned the fact that the most widely represented category is *Essen* ("food") and we have laid out the reasons why this is the case (i.e. the characteristics of the genre, the sources of the fragments). Comedy allows us not only to learn about a vast range of everyday ancient foods (like bread, olives and cheese), but also to acquire information about more specific types of food, recipes and cooking methods. For instance, besides providing repeated occurrences of the simple word for "bread," *artos* (in the database: Telekleides fr. 1.4; Aristophanes fr. 955.2; Nicophon fr. 6.1; Mnesimachos fr. 4.11), the *LGgK* lists numerous specific types of bread and baking techniques: "bread of whole wheaten meal" (*autopyritēs*) and "baked bread" (*dipyritēs*) in Phrynichus fr. 40; "roll or loaf baked or toasted on a spit" (*obelias*) in Nicophon fr. 6.2; "bread baked in a baking dish" (*klibanitēs*) in Ameipsias fr. 5; "bread baked over the fire" (*escharitēs*) in Antidotos fr. 3.1 and Krobylos fr. 2.2; "white bread" (*katharos*) in Krobylos fr. 2.2; "oven bread" (*ipnitēs*) in Timokles fr. 3.2.

Everyday items also provide instances of objects that are either attested by authors who are chronologically very distant from one another or span the whole

22 These data are drawn from Nesselrath (2010, p. 430).

23 According to the edition by Wilson (2007).

24 Menander is the only other author for whom we still have a complete comedy, running to a total of 969 lines (according to the Arnott 1979 edition).

evolution of the genre. The most common ones are again foods, household objects and everyday clothing items. Concerning foods, the above-mentioned term for "bread" (*artos*) provides an example of a staple food found throughout the history of comedy. In this case, we can safely state that comedy mirrors everyday life. By the same token, we also register a good number of occurrences for objects linked to the household: *chytra*, for instance, defining a common cooking "pot," is found in 6th- and 5th-century BC comedy (Krates frr. 16.8; 32.1; 47; Eupolis fr. 218.3), as well as in the 5th and 4th century BC (Alkaios fr. 24; Diokles fr. 9; Polyzelos fr. 6; Aristophanes frr. 606.1; 693) and in the 4th (Timokles frr. 7.4; 23.3). The same can be said for ordinary objects like the "table" (*trapeza*), listed as *Mobiliar* ("furniture") and found across all periods (5th cent. BC: Crates fr. 16.5; Telekleides fr. 1.7; 5th–4th cent. BC: Apollophanes fr. 5.2; Aristophanes fr. 903; 4th cent. BC: Klearchos fr. 4.3; Mnesimachos fr. 8.2; Timokles fr. 13.4) and common clothing items like the "cloak" (*himation*), listed as *Kleidung* (clothing) and found in the 5th century BC (Crates fr. 35), the 5th–4th (Aristophanes fr. 92) and the 4th (Timokles fr. 10.3).

Another revealing case is represented by objects that are only attested from a certain point in time onwards and in relation to which the comic sources provide important pieces of information. Two examples might be the musical instrument called *giggras* and the catapult. The *giggras* is mentioned in a 4th-century BC fragment by Amphis (fr. 14).[25] A speaker, probably Dionysus, boasts about his latest invention, a musical instrument his interlocutor has never heard about: a new kind of pipe called *giggras*. This small instrument (a hand's span in length), with a high-pitched and plaintive tone,[26] was imported from Phoenicia or Caria (West 1992, p. 92). It has never appeared on stage, says the first speaker, but it is already[27] fashionable at symposia in Athens. The surprise shown by the second speaker, who apparently does not know what a *giggras* is, is probably intended to stress the exotic provenance of the object, which is not an Athenian product. Although we have no precise information on when this instrument was imported

25 The fragment reads: "And I, for one, the most cleverly devised *giggras*. (B.) But what is a '*giggras*'? (A.) A new invention of mine, which I have never yet presented at the theatre, although in Athens it has already become fashionable at banquets. (B.) Why don't you bring it forth then to the crowd? (A.) Because I am waiting to be allotted a tribe that is really fond of victory. For I know that it will shake with applause everything as with a trident" (Trans. Papachrysostomou 2016, p. 97).

26 Athenaeus 4.174f (the source of Amphis' fragment), who defines this instrument as "mournful," adds that the Carians used it in their laments: see Barker (1984, pp. 261–263).

27 Amphis' dates are very vague: Papachrysostomou (2016, p. 11) places his *floruit* in first half of the 4th century BC.

to Greece, the fragment provides a *terminus post quem* not only for the object – namely one of the earliest dates at which the item is attested – but also for its uses (as an instrument played first at symposia and later at the theater).

The second example too sheds light on the interpretative possibilities offered by the database. It regards the mention of the "catapult" in two fragments from the 4th century BC, namely Mnesimachos fr. 7.9 and Timokles fr. 12.5.[28] The invention of the catapult is due to Dionysius I of Syracuse, who in 399 BC gathered a large group of technicians to conduct research on new weaponry, including – precisely – the tension catapult. It is not known with certainty when this weapon was introduced into Greece. A likely guess is in the 370s, with the auxiliaries that Dionysius I dispatched to the Spartans following the Battle of Leuctra. At about the same time catapults would appear to have reached Athens. Nevertheless, both fragments mention the weapon in relation to Philip II of Macedon, who was probably responsible for introducing a new catapult model with a twisting mechanism. This type of catapult can be traced back to the Macedonian king's innovations in siege warfare.[29] Again, the comic mention of an object from a specific point in time onwards can prove revealing with respect to the chronological diffusion of that object and provide hints that most often are not found in other sources.

Lastly, dialectal variants – noted in the Annotation column of the database – can prove insightful for interpretative purposes, for instance to identify a speaker's origin, even in fragmentary contexts. Thus the Doric (presumably Laconian) word for "broth" in Epilykos fr. 4 (*dōmos*, the Attic equivalent being *zōmos*: see Metagenes fr. 18.2), together with the rest of the fragment's lexicon (other words are also in Doric dialect) and the context of the source (a discussion on Spartan symposia), suggests that the speaker is a Spartan (cf. Orth 2014, pp. 263–273).

5 Conclusions

This paper has outlined the theoretical background of the *LGgK* database and the analytical possibilities it affords. In the first section, we introduced MET theory as a cognitive reading for objects in ancient drama. In the second section, we showed to what extent objects brought on stage, or even just mentioned, come alive through the characters' interaction with them. Objects are therefore inde-

28 Concerning these fragments, see Mastellari (2020, pp. 453–467) (on Mnesimachus) and Apostolakis (2019, pp. 115–123) (on Timocles).

29 See Mastellari (2020, pp. 465–467) with further bibliography.

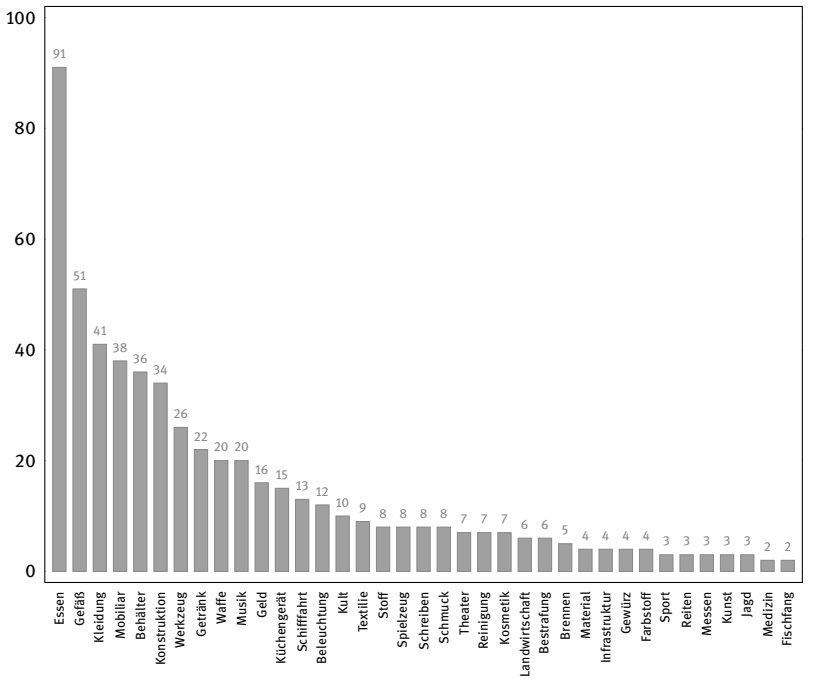


Fig. 1: Number of records per category, main table.

pendent and central research targets. In the third section, we presented the structure of the database for all objects mentioned in comedy and the improvements that are currently being made to it. Finally, the fourth section discussed the possibilities which *LGgK* offers in the field of computer-aided drama analysis. As the database is an ongoing project, we were able to present only provisional results, which nevertheless show its potential.

Specifically, we considered some categories which include different objects, showing their chronological distribution. Secondly, we presented the most widely attested categories in relation to the production of individual authors (taking the 5th-century BC comic poet Eupolis as an example). We issued a specific caveat with respect to the results obtained, in light of the characteristics of the genre and the sources of the fragments. We then listed a number of other interpretative opportunities: the variety of items and detailed information associated with specific categories; the possibility of identifying staple items, attested in Greek comedy across all periods, or objects introduced from a certain point onwards; and, lastly, dialectal variants as a means to identify characters.

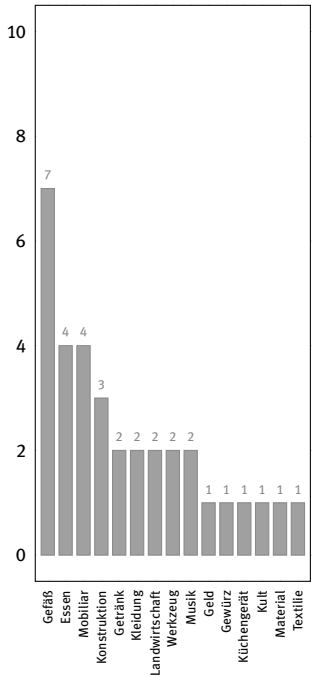


Fig. 2: Number of records per category. Author: Eupolis.

This paper has illustrated the usefulness of having a database of objects mentioned in dramatic texts, arguing that computer-based tools can contribute to the analysis of such textual material. This entails a shift from looking at individual objects to looking at the relationships between objects, characters and settings, on both a synchronic and diachronic level. Database analysis can and should be based on a critical awareness of the contextual research that shapes the compilations, a reflective practice for ongoing material documentation and analysis, and a responsive, sensitive and community-oriented approach to interpretation and access.

References

Apostolakis, Kostas (2019). *Timokles. Introduction, Translation, Commentary (FrC21)*. Vandenhoeck & Ruprecht.

Arnott, W. Geoffrey (1979). *Menander (ed. with an Engl. transl.)* Harvard University Press.

Attardo, Salvatore (2001). *Humorous Texts: A Semantic and Pragmatic Analysis*. De Gruyter.

Attardo, Salvatore (2017). "The General Theory of Verbal Humor." In: *The Routledge Handbook of Language and Humor*. Ed. by Salvore Attardo. New York and London, pp. 126–142.
Austin, Colin and S. Douglas Olson (2004). *Aristophanes. Thesmophoriazusae. Edited with Introduction and Commentary.* Oxford University Press.
Bagordo, Andreas (2017). *Aristophanes fr. 675–820: Übersetzung und Kommentar (FrC10.10)*. Verlag Antike.
Barker, Andrew (1984). *Greek Musical Writings*. Cambridge University Press.
Bennett, Jane (2010). *Vibrant Matter: A Political Ecology of Things*. Duke University Press.
Beta, Simone (2004). *Il linguaggio nelle commedie di Aristofane. Parola positiva e parola negativa nella commedia antica.* Accademia Nazionale dei Lincei.
Bolens, Guillemette (2001). *Kinesic Humor: Literature, Embodied Cognition, and the Dynamics of Gesture*. Oxford University Press.
Brown, Bill (2001). "Thing Theory." In: *Critical Inquiry* 28.1, pp. 1–22.
Castellaneta, Sabina and Anna Lisa Maffione (2016). "Gli oggetti nelle commedie di Aristofane." In: *Gli oggetti sulla scena teatrale ateniese. Funzione, rappresentazione, comunicazione. Giornate di studio Università degli Studi di Padova 1-2 dicembre 2015*. Ed. by Alessandra Coppola, Caterina Barone, and Monica Salvador. CLEUP, pp. 447–550.
Clark, Andy (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Oxford University Press.
Codd, Edgar F. (1970). "A Relational Model of Data for Large Shared Data Banks." In: *Communications of the ACM* 13, pp. 377–387.
Coole, Diana and Samantha Frost, eds. (2010). *New Materialisms: Ontology, Agency, and Politics*. Duke University Press. DOI: 10.1515/9780822392996.
Davidson, James (1993). "Fish, Sex and Revolution in Athens." In: *CQ* 43, pp. 53–66.
Davidson, James (1997). *Courtesans and Fishcakes*. Harper Collins.
Davies, Christie (2011). *Jokes and Targets*. Indiana University Press.
Dennett, Daniel C., Matthew M. Hurley, and Reginald B. Adamas (2011). *Inside Jokes: Using Humor to Reverse-Engineer the Mind*. MIT Press.
Double, Oliver (2014). *Getting the Joke: The Inner Workings of Stand-Up Comedy*. 2nd ed. Bloomsbury Publishing.
Dover, Kenneth. J. (1993). *Aristophanes. Frogs.* Clarendon Press.
Gilula, Dwora (1995). "Comic Food and Food for Comedy." In: *Food in Antiquity*. Ed. by John Wilkins et al. University of Exeter Press, pp. 75–90.
Lowe, Nick J. (2020). "Beyond a joke: making humour theory work with Aristophanes." In: *Aristophanic Humour: Theory and Practice*. Ed. by Peter Swallow and Edith Hall. Bloomsbury Academic, pp. 13–22. DOI: 10.5040/9781350101555.ch-02.
Malafouris, Lambros (2013). *How Things Shape the Mind: a Theory of Material Engagement.* MIT Press.
Malafouris, Lambros (2018). "Bringing Things to Mind: 4Es and Material Engagement." In: *The Oxford Handbook of 4E Cognition*. Ed. by Albert Newen, Leon De Bruin, and Shaun Gallagher. Oxford University Press, pp. 755–772.
Malafouris, Lambros (2019). "Mind and Material Engagement." In: *Phenomenology and the Cognitive Sciences* 18.1, pp. 1–17.
Mastellari, Virginia (2020). *Calliade–Mnesimaco. Introduzione, traduzione, commento (FrC16.5)*. Vandenhoeck & Ruprecht.
Mastromarco, Giuseppe and Piero Totaro (2006). *Commedie di Aristofane*. UTET.
Morreall, John (2009). *Comic Relief: A Comprehensive Philosophy of Humor*. Blackwell's.

Nesselrath, Heinz-Günther (2010). "Comic Fragments: Transmission and Textual Criticism." In: *Brill's Companion to the Study of Greek Comedy*. Ed. by Gregory W. Dobrov. Brill, pp. 423–453.

Newen, Albert, Leon De Bruin, and Shaun Gallagher, eds. (2018). *The Oxford Handbook of 4E Cognition*. Oxford University Press.

Noë, Alva (2004). *Action in Perception*. MIT Press.

Novokhatko, Anna (2020). "δηλαδὴ τρίπους: on Epicharmus fr. 147 PCG." In: *Fragmentation in Ancient Greek Drama*. Ed. by Anna Lamari, Franco Montanari, and Anna Novokhatko. De Gruyter, pp. 337–352.

Novokhatko, Anna, Virginia Mastellari, Beatrice Gavazza, and Leon Glaser (2021). "LGgK. Lexical database of the realia in ancient Greek comedy." In: *FuturoClassico* 7, pp. 1–18.

Olson, S. Douglas (2002). *Aristophanes Acharnians. Ed. with Introduction and Commentary*. Oxford University Press.

Olson, S. Douglas (2014). *Eupolis frr. 326–497: Translation and Commentary*. Verlag Antike.

Olson, S. Douglas (2016). *Eupolis: Heilotes–Chrysoun Genos (frr. 147–325): Translation and Commentary*. Verlag Antike.

Olson, S. Douglas (2017). *Eupolis: Testimonia and Aiges–Demoi (frr. 1–146)*. Verlag Antike.

Orth, Christian (2014). *Aristomenes–Metagenes. Einleitung, Übersetzung, Kommentar (FrC9.2)*. Verlag Antike.

Papachrysostomou, Athina (2016). *Amphis. Introduction, Translation, Commentary (FrC20)*. Verlag Antike.

Poulsgaard, Kåre Stokholm and Lambros Malafouris (2020). "Understanding the hermeneutics of digital materiality in contemporary architectural modelling: a material engagement perspective." In: *AI & Society*. DOI: 10.1007/s00146-020-01044-5.

Raskin, Victor (1985). *Semantic Mechanisms of Humor*. Reidel.

Reisch, Emil (1905). "Dreifuß." In: *Realencyclopädie der classischen Altertumswissenschaft*, coll. 1669–1696.

Sommerstein, Alan H. (1994). *Aristophanes. Thesmophoriazusae*. Aris & Philips.

Stone, Laura M. (1981). "Costume in Aristophanic Comedy." PhD thesis. University Microfilms.

Suhr, Elmer G. (1971). "The Tripod." In: *Folklore* 82.3, pp. 216–232.

West, Martin L. (1992). *Ancient Greek Music*. Oxford University Press.

Wilkins, John (2000). *The Boastful Chef. The Discourse of Food in Ancient Greek Comedy*. Oxford University Press.

Wilson, Nigel G. (2007). *Aristophanis Fabulae. Rec. brevique adnotatione critica instruxit N. G. Wilson, Acharnenses, Equites, Nubes, Vespae, Pax, Aves, Vol. II: Lysistrata, Thesmophoriazusae, Ranae, Ecclesiazusae, Plutus*. Clarendon Press.

Tab. 1: Percentage presence of authors from a given period of time in the database

(a) 6th and 5th century BC

Author	Volume
Alkimenes	FrC 1.1
Arkesilaos	FrC 1.1
Aristagoras	FrC 1.1
Chionides	FrC 1.1
Diop]eithes	FrC 1.1
Ekphantides	FrC 1.1
Euphronios	FrC 1.1
Euthykrates	FrC 1.1
Euxenides	FrC 1.1
Hegemon	FrC 1.1
Ion	FrC 1.1
Kallias	FrC 1.1
Kallistratos	FrC 1.1
Kantharos	FrC 1.1
Leukon	FrC 1.2
Lykis	FrC 1.2
Lysippos	FrC 1.2
Magnes	FrC 1.2
Menandros II	FrC 1.2
Menekrates	FrC 1.2
Myllos	FrC 1.2
Myrtilos	FrC 1.2
Philonides	FrC 1.2
Susarion	FrC 1.2
Thugenides	FrC 1.2
Xenophilos	FrC 1.2
Krates	FrC 2
Kratinos	FrC 3.2
Telekleides	FrC 4
Hermippos	FrC 6
Phrynichos	FrC 7
Eupolis	FrC 8.2, 8.3
32 (42.7%)	9 (42.9%)

(b) 5th and 4th century BC

Author	Volume
Alkaios	FrC 9.1
Ameipsias	FrC 9.1
Apollophanes	FrC 9.1
Aristomenes	FrC 9.2
Aristonymos	FrC 9.2
Autokrates	FrC 9.2
Demetrios I	FrC 9.2
Diokles	FrC 9.2
Epilykos	FrC 9.2
Eunikos	FrC 9.2
Kephisodoros	FrC 9.2
Krates II	FrC 9.2
Metagenes	FrC 9.2
Nikochares	FrC 9.3
Philonikos	FrC 9.3
Philyllios	FrC 9.3
Poliochos	FrC 9.3
Polyzelos	FrC 9.3
Sannyrion	FrC 9.3
Xenophon	FrC 9.3
Aristophanes	FrC 10.3, 10.9, 10.10, 10.11
Nikophon	FrC 15
22 (29.3%)	8 (38.1%)

(c) 4th-century BC

Author	Volume
Asklepiodoros	FrC 16.1
Dionysodoros	FrC 16.1
Alkenor	FrC 16.1
Anaxilas	FrC 16.1
Antidotos	FrC 16.1
Apollinaris	FrC 16.1
Araros	FrC 16.1
Euboulides	FrC 16.5
Euphanes	FrC 16.5
Euthias	FrC 16.5
Euthykles	FrC 16.5
Heniochos	FrC 16.5
Heraklides	FrC 16.5
Kalliades	FrC 16.5
Kallikrates	FrC 16.5
Klearchos	FrC 16.5
Kratinos iunior	FrC 16.5
Krobylos	FrC 16.5
Mnesimachos	FrC 16.5
Anaxandrides	FrC 17
Timokles	FrC 21
21 (28%)	4 (19%)

Tab. 2: Chronological distribution of the objects according to category

Category		Number of records	Century		
German	**English**		**6th, 5th BC**	**5th–4th BC**	**4th BC**
Medizin	Medicine	2		✔	
Fischfang	Fishing	2		✔	
Sport	Sport	3	✔	✔	✔
Reiten	Horse riding	3		✔	
Messen	Measuring	3		✔	
Kunst	Art	3	✔	✔	
Jagd	Hunting 3		✔		
Material	Material	4	✔	✔	
Infrastruktur	Infrastructure	4	✔	✔	
Gewürz	Spice	4	✔	✔	✔
Farbstoff	Dye	4	✔	✔	
Brennen	Burning	5		✔	✔
Landwirtschaft	Agriculture	6	✔	✔	
Bestrafung	Punishment	6	✔	✔	✔
Theater	Theater	7	✔	✔	✔
Reinigung	Cleaning	7	✔	✔	
Kosmetik	Cosmetics	7	✔	✔	✔
Stoff	Fabric	8	✔	✔	✔
Spielzeug	Toys	8	✔	✔	
Schreiben	Writing	8	✔	✔	✔
Schmuck	Jewellery	8	✔	✔	✔
Textilie	Textile	9	✔	✔	
Kult	Cult	10	✔	✔	✔
Beleuchtung	Lighting	12	✔	✔	✔
Schifffahrt	Navigation	13	✔	✔	✔
Küchengerät	Kitchenware	15	✔	✔	✔
Geld	Money	16	✔	✔	✔
Waffe	Weapon	20	✔	✔	✔
Musik	Music	20	✔	✔	✔
Getränk	Drink	22	✔	✔	✔
Werkzeug	Tool	26	✔	✔	✔
Konstruktion	Construction	34	✔	✔	✔
Behälter	Container	36	✔	✔	✔
Mobiliar	Furniture	38	✔	✔	✔
Kleidung	Clothing	41	✔	✔	✔
Gefäß	Tableware	51	✔	✔	✔
Essen	Food	91	✔	✔	✔
Total		559			

Pablo Ruiz Fabo, Delphine Bernhard, Andrew Briand, and Carole Werner

Computational Drama Analysis from Almost Zero Electronic Text: the Case of Alsatian Theater

Abstract: At the MeThAl project, we are creating the first large TEI corpus of Alsatian theater; Alsatian refers to Germanic varieties spoken in Alsace (Eastern France). The corpus, covering mainly the 1870-1940 period, will have above 500 000 tokens (51 plays) for which no previous electronic text existed. We present our automatic TEI encoding workflow assisted by a Conditional Random Fields model based on OCR sources, followed by manual correction. As the corpus shows large orthographic variation (there is no standard spelling) and NLP resources for Alsatian are scarce, several text analyses are challenging; we discuss our approach to tackle this. We developed detailed character metadata using TEI feature structures, encoding characters' social variables like their socio-professional group and social class. This provided an overview of the evolution of social groups in the plays across time, complementing earlier, smaller-sample studies of the tradition. Besides metadata-based analyses, with a view to performing emotion detection, we present an emotion lexicon which handles orthographic variation in the corpus, and how it can be used to train automatic variant detection. We also outline how we are publicly sharing the resources, on open data repositories, on DraCor, and via a corpus navigation interface. An overall goal is starting off computational literary research (CLS) on Alsatian theater and helping compare Alsatian theater with other traditions for which rich CLS results already exist, including the two dominant ones which influenced Alsatian theater most: German and French.

1 Introduction

Quantitative drama analysis, assisted in recent years by automatic computational means, has delivered new insights on many issues. Some of the analyses performed pose certain preconditions: first, the availability of electronic text in adequate quantity, ideally with the relevant structural markup (e.g., TEI). Second,

Pablo Ruiz Fabo, Delphine Bernhard, Carole Werner, Université de Strasbourg
Andrew Briand, University of Washington

https://doi.org/10.1515/9783111071824-004

(NLP-based) tools to perform automatic annotations which can help operationalize a drama analysis research question.

Alsatian theater is a tradition for which these conditions were not fulfilled until very recently, even if dramatic works in Alsatian (i.e., Germanic varieties spoken in Alsace, Eastern France) have been steadily produced for over two centuries. A large-scale quantitative study of this tradition was so far impossible, given the absence of an electronic corpus. Within the MeThAl project,[1] we are taking the first steps towards enabling such analyses, focusing on the 1870–1940 period. We present the challenges encountered for corpus development and analysis, and the solutions adopted, paying attention to how this work may generalize to other low-resource settings. One of the challenges we discuss is the huge orthographic variation in the corpus, given the lack of a standard variety. This hinders statistical analyses typical of Computational Literary Studies (CLS); we will present the first resources developed to tackle this problem. An overall goal is to help start a dialogue between CLS research on Alsatian theater and existing CLS research on the dominant traditions that have influenced Alsatian theater (German and French), for which major results have already been achieved. Given the limited range of electronic text and NLP resources for Alsatian varieties, there is a gap between the CLS questions that can be addressed in this tradition and in larger ones, and our project seeks to start bridging this gap.

The paper is structured as follows: section 2 introduces the specificities of Alsatian theater and our corpus selection criteria. Section 3 outlines our encoding workflow. Since part of the corpus is still being encoded, and given orthographic variation in the text, we focused on detailed metadata as a way to reach generalizations about the material. Within this effort, section 4 describes our encoding of character social variables into a TEI personography, which gives an overview of the evolution of social groups present in the corpus. Deeper, text-based analyses require dealing with the corpus's large scriptural variation. In this respect, section 5 presents a lexicon for emotion analysis which is able to handle this task, besides methods for spelling variant detection that can be developed based on the lexicon. In Section 6, we outline how we are sharing the project resources, targeting both scholars and the general public. Section 7 concludes.

1 https://methal.pages.unistra.fr/en.

2 Alsatian Theater Particularities and Corpus Selection

A continuous dramatic production in Alsatian has been existing since the early nineteenth century (Gall 1974, pp. 5–6). Comedic subgenres predominate (from refined satires to farces), without excluding other genres like the *Volksstück* (popular drama) and *Weihnachtsmärchen* (Christmas tale). Alsatian theater is also interesting due to its ties to the two dominant traditions surrounding it, German and French, such that a polysystem analysis (Even-Zohar 1990) might be attempted.

Alsatian theater's foundational play is *Der Pfingstmontag* by Johan Georg Daniel Arnold, published in 1816. The tradition, however, experienced its golden age around 1900; Alsace was part of the German Empire between 1871 and 1918, and writing dialect theater was one means whereby Alsatian authors sought to affirm their identity as apart from the rest of the German-speaking world (Huck et al. 2007, p. 12). Our corpus, stretching mainly from 1870 to 1940, covers this period.

A salient characteristic of the corpus is its vast scripto-linguistic variation. No standard orthography existed during the corpus period, and practices vary across authors and also for characterization purposes within the same author; representing sociophonetic variation also leads to different scripturalizations for the same lexeme. Besides this variation, code-switching between Alsatian, French, and standard German is also characteristic in the corpus, reflecting a situation of language contact (and diglossia, see Huck 2015, pp. 151–161). Besides, the paratext (e.g., stage directions) can be in either Alsatian or German.

2.1 Corpus Selection

We are creating the first electronic-text corpus for Alsatian theater. The overall goal is to provide digital text amenable to large-sample quantitative analyses that complement existing knowledge of the tradition (see section 4.3 for some findings). In this scenario, we are aiming for breadth, including minor authors and a variety of author origins and publisher locations, rather than attempting an exhaustive coverage of the best-known authors. Besides, earlier literature has covered major authors based on samples specific to them, using non-quantitative approaches (Cerf 1972, 1975; Huck 1998; Hülsen 2003; Huck 2005). Regarding genres, we focus on comedies since this is the tradition's main genre, but are also encoding important *Volksspiele* or dramas (including socially engaged dra-

mas from the 1890s by Julius Greber).[2] As a starting point for our encoding, we prefer digital sources if available, such as the collection of image-mode digitizations at the Numistral heritage portal by Strasbourg's National and University Library.[3] Other platforms offering public-domain content like the Internet Archive or Google Books also complement our sources, as does a Wikisource collection with the complete works of Alsatian dramatist August Lustig (transcribed by Mireille Libmann in 2013).[4] Besides, by consulting secondary literature on the history of Alsatian theater (Cerf 1972; Gall 1974; Hülsen 2003; Huck 2005), we identified a small number of plays unanimously considered important by those sources but not yet digitized, and obtained digitizations.

When selecting plays for encoding, we prefer those mentioned by the secondary literature just cited, and, more generally, we strive for representing all decades in the corpus period, different geographical origins for authors and publishers and different subgenres, with a majority of comedic works of various types. Even if we seek variety, a limitation of the corpus is that male authors predominate; increasing the share of women authors is a challenge, since they are underrepresented in digital sources and are minoritary in preserved sources. We intend to counter this bias with deeper searches in non-digitized collections (e.g., at city archives).

We have so far performed OCR on, and TEI-encoded, 51 plays (30 plays with 329 969 text tokens already published and 21 more with 188 275 text tokens awaiting validation). We also converted to TEI the 26 plays by August Lustig on Wikisource (136 675 text tokens); this involved no OCR and only wiki-markup conversion to TEI with rule-based methods, thus being a much simpler process than the workflow for plays for which no electronic text existed.

3 TEI Encoding Workflow

We adopted the Text Encoding Initiative's (TEI) recommendations (TEI Consortium 2022). The TEI very naturally matches our project goals: it allows for encoding structural divisions, which are crucial for classical quantitative drama analyses like configuration matrices (Pfister [1977] 2001, p. 236), and it has rich possi-

2 The distribution can be seen in a metadata table in our repository: https://git.unistra.fr/methal/methal-sources#plays-available-in-tei.

3 https://www.numistral.fr/fr/theatre-alsacien. Some of the plays have OCR, but it has not been corrected, and we produce our own OCR and its correction (see section 3).

4 https://als.wikipedia.org/wiki/Text:August_Lustig/A._Lustig_Sämtliche_Werke:_Band_2.

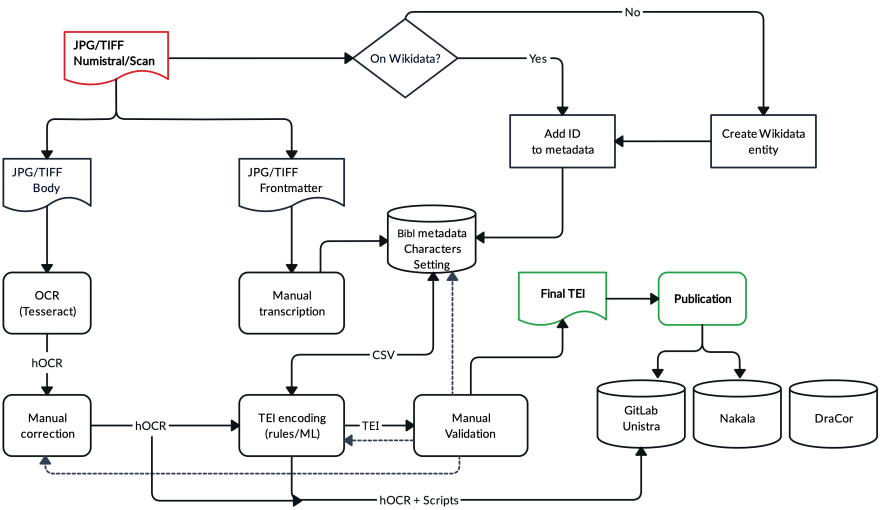


Fig. 1: TEI encoding workflow.

bilities for encoding author and character metadata. Besides, there is an ecosystem of corpora and tools for working with dramatic texts in TEI, e.g., the DraCor platform (Fischer and Börner 2019) or the DramaAnalysis R package (Reiter and Pagel 2020). It is thus an advantageous format to help start a dialogue between research on other traditions and on Alsatian theater.

Our workflow combines manual transcription of bibliographic and character metadata (3.1), automatic TEI encoding of the plays' body based on OCR output (3.2), and manual revision for both the OCR and the automatic TEI-encoding. We believe that our encoding approach could be used in similar projects where little or zero electronic text exists. The workflow is summarized in figure 1.

3.1 Transcription of Bibliographic and Character Metadata

Based on initial tests, we decided to treat differently the plays' front matter from their body. The front matter, i.e., the cover page with bibliographic metadata, and other material before a play's beginning, such as the cast list, takes many different formats depending on the publisher or series. Given this challenge, we manually transcribed bibliographic and character information into a delimited format and later generated the relevant TEI elements (e.g., `teiHeader`, `castList`, `listPerson`) based on this; we also annotated characters with social attributes describing them (section 4). As the volume of manual transcription amounts to one or two pages

per play, the approach scaled well to over three hundred plays and over 2000 characters.[5] The manual work also includes creating Wikidata entities for plays and authors to refer to them in the TEI versions.

3.2 Automatic OCR to TEI Conversion With Conditional Random Fields

The bodies of the plays (from the first act or scene to the final curtain) are quite regular and lend themselves to automated encoding, followed by manual corrections. As we start off from image-mode digitizations, the first step is optical character recognition (OCR), for which we use Tesseract (v4, Smith 2018). We found that for Alsatian text from the corpus period, best results are obtained by combining German and French models plus the "script" model matching each play's typeface, i.e., a language-independent model trained on either blackletter (Fraktur model) or roman typeface ("Latin" model). For plays typeset in blackletter, German Fraktur models were used in addition to the language-independent ones.[6] A possible reason why combining these models works best is that Alsatian scripturalization in the corpus includes characters and diacritics typical of German (eszett and umlaut), but also typical of French (acute and grave accent). Besides, there is code switching involving both languages, and the paratext is often in German. We chose hOCR (Baierer 2020) as our output format; this format encodes word-position information on the page (bounding boxes), which can be used to compute features for TEI-element prediction (details below). OCR output was manually corrected, generally by a single person.[7]

A first challenge is converting hOCR output into TEI. Several projects have performed automatic TEI generation from OCR sources, often implemented as a sequence labelling task: Khemakhem et al. (2017) present GROBID dictionaries, which encodes dictionary entries in TEI based on layout and typographic cues present in OCR outputs, using a cascade of Conditional Random Fields (CRF)

5 We have bibliographic references for over 350 plays, but digitizations (and thence character lists) for ca. 230.

6 In other words, we used a combination of models from https://github.com/tesseract-ocr/tessdata, language codes `deu` and `fra` in general, plus `deu_frak` and `frk` for blackletter plays, and, among models in the `script` directory, `Fraktur` or `Latin` depending on the play's typeface.

7 Double keying has been shown to handle most OCR and transcription errors in electronic corpus creation (Geyken et al. 2012, §3), reaching >99.99 % accuracy (Deutsche Forschungsgemeinschaft 2016, p. 32), however for this first Alsatian theater corpus, we preferred to obtain first versions of a larger number of plays rather than a smaller number of double-keyed versions.

models (Lafferty et al. 2001), each of which handles a different TEI element. Khemakhem et al. (2018) also use GROBID to encode encyclopedia entries and auction catalogues in French, as do Gabay et al. (2021) for exhibition catalogues. Still using CRFs (a single model rather than a cascade), Erjavec et al. (2021) TEI-encoded academic theses' structure (e.g., abstract vs. acknowledgements) using simple surface features (most frequent words per page, page and word length).

Neural methods have also been used for TEI generation. Pagel et al. (2021) predict the TEI structure for plays' body elements in German, assuming one label per sentence (5-way classification task predicting act and scene divisions, speaker names and their speech, and stage directions). The input is not OCR, but rather plain text. The annotated corpus contains ca. 1.4 million sentences, the smallest class is *Act* with 1458 sentences, *Scene* has over 11 000 and the rest over 175 000.[8] They obtain best results (>0.97 F1 for all classes but stage directions, with 0.84 F1) by fine-tuning the German `bert-base-uncased` model. The results are higher than their baseline's, a CRF with surface features (token string, case, presence of digits) and lexical ones (presence of triggers indicating act/scene divisions), which got >0.92 F1 for all classes but stage directions, with only 0.44 F1.

In our project, we considered that the structure of a play's body is less complex than the text-types for which GROBID has been used (dictionaries, encyclopedias or catalogues). Accordingly, we implemented a simpler architecture: a single CRF model was created to predict, based on OCR input in hOCR or ALTO formats,[9] six types of content: act and scene divisions, speaker names and their speech, stage directions, and verse lines. The *verse lines* class refers to verse within character speech (e.g., a song or poem), typographically distinct from "prose-based" character speech. The models were created with `sklearn-crfsuite` (Korobov [Nov. 26, 2015] 2019). The system is publicly available at https://git.unistra.fr/methal/FETE.

At project outset, we had no TEI-encoded plays at all, so we produced a first set of seven plays using dictionary- and rule-based methods; a manual correction of the TEI-encoded documents was also carried out. These seven plays and their OCR were used as the initial training set for a CRF model. The initial model was used to tag some more plays, which were then manually corrected, up to the 16 plays that make up the training data described in table 1, comprising ca. 150 000 tokens.

8 For the *Act* or *Scene* classes, a "sentence" would be the expression that indicates an act or scene onset, e.g., *Erster Akt* for *First Act*.

9 ALTO: Analyzed Layout and Text Objects, (Alto Editorial Board 2022); hOCR: Baierer (2020).

Tab. 1: CRF training-data distribution (number of examples per class).

Content type	TEI element predicted	Train	Test
Act	`<div type="act">`	25	9
Scene	`<div type="scene">`	408	68
Speaker	`<speaker>`	9,675	978
Speech	`<p>`	104,504	12,239
Verse line	`<l>`	1,390	236
Stage direction	`<stage>`	21,371	3,192

The model predicts a label for each corpus token. Three sets of features were implemented in the CRF and computed for the current, previous and following token:

- **Set 1 (Token-level features):** These include the token string (also lowercased), initial case, the presence of digits, the presence of punctuation often used to structure plays (period, colon, parentheses). Based on hOCR or ALTO output, font size and token horizontal position (normalized per play), and an indication of font-size difference between previous and current token.
- **Set 2 (Set 1 + Heading features):** Includes the features in Set 1, plus new features to detect act/scene headings. In view of the small number of examples to learn from, we favored lexical features, i.e., words that mean *act* or *scene* in either Alsatian, German or French (paratext is often not in Alsatian).
- **Set 3 (Sets 1 + 2 + Verse line features):** Includes the features in sets 1 and 2 plus features to detect verse lines. As layout or token-based cues are inconclusive, the following two features were implemented. (1) Based on a comparison of lines' final characters, a boolean to indicate whether the current line rhymes with the previous or next. (2) An estimation of the difference in syllable count between consecutive lines, obtained by counting sequences of consecutive vowels, as an approximation to the number of syllable nuclei.

After prediction, the TEI hierarchy is recreated from the predicted labels. The TEI header and cast list are generated automatically based on our manual transcription of the relevant text (3.1) and added to the document. Errors made by the automatic encoding are corrected manually.

CRF prediction results are in table 2. Even if for some categories (table 1) the number of test-items is small, overall the results suggest that a simple CRF model producing token level annotations is a viable means to speed up TEI encoding of dramatic texts when little training data are available, also helping with segmenting verse-lines among prose.

We see the following limitation: the model uses punctuation as a feature and overfits to the main punctuation cues used in the training data to indicate speak-

Tab. 2: Precision (P), Recall (R) and F1 for the three feature-sets. Best F1 per class bold, second-best italicized.

	Set 1			Set 2			Set 3		
Content	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
Act	0.889	0.889	0.889	1	1	**1**	1	1	**1**
Scene	1	0.941	0.97	1	1	**1**	1	0.956	*0.977*
Speaker	0.988	0.995	**0.991**	0.988	0.994	**0.991**	0.988	0.994	**0.991**
Speech	0.98	0.993	0.986	0.985	0.991	*0.988*	0.989	0.991	**0.99**
Verse-line	1	0.475	0.644	0.921	0.695	*0.792*	0.898	0.898	**0.898**
Stage direction	0.971	0.875	0.908	0.972	0.964	**0.968**	0.971	0.962	*0.967*
Weighted mean	0.979	0.979	0.978	0.982	0.982	*0.982*	0.984	0.984	**0.984**

er/speech divisions (a colon in an overwhelming majority of examples) or stage directions (almost invariably in parentheses). While encoding new plays, we saw that performance is poor when those cues are absent (e.g., with older plays who use a period after the speaker instead of a colon, or with a play delimiting stage directions via square brackets). A simple workaround is to preprocess plays so that those delimiters are the majority ones prior to prediction and postprocess them back into the original delimiters after it. However, this does not solve the more difficult (although rare) case where no delimiter is used to indicate stage directions; for this, Pagel et al.'s contextual embeddings approach, which classifies entire sentences based on content, should help, even if stage directions were the hardest category for their model. In our model, the hardest class is verse-lines, which suffer from a similar problem: their lexical content is difficult to distinguish from non-verse lines.

A model which uses plain-text as its input like the one developed by Pagel et al. (2021) has the advantage that it requires no prior OCR output with layout and word-position information. Such a model can thus be used to TEI-encode already available electronic transcriptions of plays (e.g., manually performed transcriptions or manually corrected plain-text OCR for which no hOCR or ALTO output was produced). We would like to attempt this approach in the future; our corpus has ca. 10 times fewer tokens than theirs and it would be interesting to examine the impact of training data volume.

4 Character Social Annotations: A Corpus Overview From Its *Dramatis Personæ*

The TEI-encoding of the complete corpus, which will help address CLS questions related to the plays' structural properties and character configurations per genre, is still ongoing. Besides, the corpus' huge scripto-linguistic variation poses challenges for many text-based analyses. In this situation, a feasible effort is to obtain an overview of the social groups portrayed in the corpus, thanks to creating detailed character metadata. Since earlier literature on Alsatian theater has also focused on social groups represented in the plays, based on smaller samples, we can compare our findings to what is known from earlier studies.

Available studies on Alsatian theater have examined how social status relates to plot, interacting with setting (rural or small town vs. urban) and character gender. Cerf (1972, pp. 339–348) speaks of rural plots in which parents (rich farmers) wish for their daughter to marry into the same social class, whereas the daughter prefers a less wealthy farmer whom she likes or an urban suitor; this is one of the plays' sources of conflict. In urban plots, parents wish for their daughters to marry someone with a better socio-economic status; this is punished in late nineteenth century plays (the young women are abandoned and become outcasts with a child to raise) but the daughters' ascent by marriage to a social class immediately above their own is possible in early twentieth-century plays. Sons' social ascent takes place via education. Hülsen (2003, p. 104) also documents marriage plots showing a clash between the parents' will and the younger generation's choices, besides tension between French and German cultures (used for comic purposes), as sources of dramatic conflict.

Cerf (loc. cit.) provides an overview of socio-professional groups in the plays. In small towns or villages, most characters are farmers, rich or poor, and domestic employees at the farms. In city settings, mostly the middle classes (small and middle bourgeoisie) are represented: craftspeople, office workers, liberal professions, rentiers. Domestic workers are also frequent characters. Less frequently, we find very poor characters (e.g., unemployed). Hülsen (2003, p. 140) reaches similar conclusions, noting also that both the lowest classes and French speaking nobility are absent from the plays.

4.1 Character Social Variable Annotations

The generalizations reported in the literature were arrived at by an examination of relevant text passages in samples focusing on the *Théâtre alsacien de Strasbourg*'s

Tab. 3: Character social variables annotated (besides *age*). Gender *both* is for group characters (not individuals).

Category	Values
Socio-professional groups	professionals, scientific, technical; intermediate professions; service and sales; crafts; industry and transportation; agriculture; elementary professions; *rentiers*; clergy; military; government officials; *volunteer positions*
Social class	upper class; upper middle class; lower middle class; lower class
Gender	both; female; male; unknown

repertoire. Cerf (1972) analyzed 39 plays between 1898 and 1939, and Hülsen (2003) covered 13 plays between 1898 and 1914. As a first step to complement such knowledge, we annotated socioprofessional groups, social class, age and gender for 2386 characters in 231 plays, creating a TEI personography. As a rule, only information that is explicit or can be deduced from the *dramatis personæ* of the plays was annotated, without recourse to the plays' text.

Regarding socio-professional groups, although classification schemes for historical professions exist, such as HISCO (Van Leeuwen et al. 2004) or an ontology by Moeller and Nasarek (2018) for German-speaking settings, these are intended for historical research rather than for describing literary characters, and their applicability to the corpus is not necessarily immediate. Fièvre (2017) presented a taxonomy for French classical drama, but its categories are obsolete for our period.

Given that existing taxonomies were not a good fit, we developed the one shown in table 3, which is inspired by the above but adapted to our corpus situation. In the table, groups in italics represent social rather than professional groups important for the corpus period. *Rentiers* are rich enough to not need to work for a living; in the corpus, this often happens after many years as a successful tradesperson. The *volunteer positions* group refers to occupations that are not professional, but that are important to define a character's social status, such as being an association leader (e.g., of a patriotic association or of a sports club) or in some cases being a member of such associations or societies.

Character professional groups were annotated by a single person (one of the authors or an intern) and were then revised by another author. Disagreement was rare (this was not measured more precisely). Social class annotations were carried out by one of the authors or by an intern and revised by another author. Based on

the sample discussed in 4.3, there was a disagreement in 8.2% of cases.[10] As future work, it could be considered whether adding an intermediate category *middle class* is a better option than a binary division of middle class characters into *upper* and *lower*.

There are some limitations to our social variable annotations: First, many female characters in the corpus are not described with a profession, but rather by reference to a male character (e.g., *wife of*); more detailed data about this are shown in in 4.3. Accordingly, quantitative generalizations about female characters using our socioprofessional taxonomy are limited, although when described as associated to another character, they could "inherit" class from it. A second limitation is that our annotations are static and refer to character state in the *dramatis personae*. Character end-state and social class changes (e.g., by education or marriage) are not directly captured. Third, given Alsace's status as alternating between French and German political rule (including during the corpus period), characters' origin is important as an indication of power relations, as is the language variety they use (Alsatian, other dialects, standard German, or French). At this point, we do not have these annotations.

4.2 TEI Personography

In order to share the social annotations in an interoperable format, we provide them as a TEI personography, using the *feature structures* (FS) formalism (Romary 2015).[11] FS represent feature hierarchies, using key-value pairs. Values can be either atomic, or a new feature structure; recursivity allows us to choose our description granularity. The data model is described in a feature system declaration `<fsdDecl>`, composed of feature declarations `<fDecl>` and feature structure declarations `<fsDecl>`.

Galleron (2017) created a feature library composed of *characterisemes* or characterization units, for describing the characters of French theater (1630 to 1810), including mimetic traits, which provide characters with human-like features, and synthetic traits related to characters' role in the play (e.g., an antagonist) (Phelan 1989).

10 The annotations for 2386 characters are now released in TEI format (see section 4.2). However, only a smaller sample with 1103 characters was available when we obtained the results discussed in 4.3 below. We have shared these initial annotations and analyses in spreadsheet format at https://page.hn/8ynxsp.

11 https://tei-c.org/release/doc/tei-p5-doc/fr/html/FS.html.

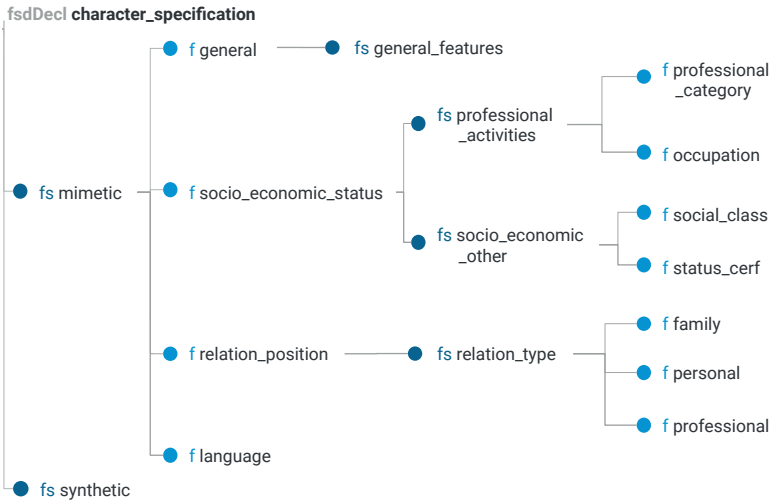


Fig. 2: Informal representation of the mimetic features in our feature library for character description. Node `general_features` refers to age, gender and origin. Dark nodes are feature structures, light nodes are features. The project repository gives the formal declaration. [13]

We believe that this characteriseme approach facilitates comparative studies across dramatic corpora. One of our goals is helping to compare Alsatian theater to the two hegemonic traditions surrounding it (German and French), so we adopted the approach.

In order to make the feature library applicable to new corpora, including Alsatian theater, we added new feature values (e.g., family relations such as *ex-husband* or *divorced wife*). When annotating, we abstracted away from variation in job titles by normalizing them into a list of ca. 350 professions; we added these as feature values as well. We also added intermediate levels to the FS for a more modular analysis of mimetic feature groups: (1) basic characteristics like gender or age, (2) features related to socio-economic status (including socioprofessional group and profession) and (3) characters' position in a relation (whether they are presented as e.g., a spouse, a relative or subordinate of another character).[12] Personography compliance was ensured with a schema automatically derived from the feature system declaration using the approach in (Bermúdez Sabel 2019).

Figure 2 shows an informal diagram of our FS declaration and figure 3 gives an example character annotated with its features. The complete personography

12 The relation itself (who is related to whom) is not part of the personography, although available in the plays' TEI via `<listRelation>`.

```
<person xml:id="mtl-per-0890">
 <bibl corresp="#mtl-090"/>
 <persName>Alice Sandel</persName>
 <note type="roleDesc">Dactylo</note>
 <occupation>Dactylo</occupation>
 <fs type="character_specification">
  <f name="specification_type">
   <fs type="mimetic_features">
    <f name="general">
     <fs type="general_features">
      <f name="sex">
       <symbol value="F"/>
      </f>
     </fs>
    </f>
    <f name="socio_economic_status">
     <vColl>
      <fs type="professional_activities">
       <f name="occupation">
        <symbol value="typist"/>
       </f>
       <f name="professional_category">
        <symbol value="intermediate_professionals"/>
       </f>
      </fs>
      <fs type="socio_economic_other">
       <f name="social_class">
        <symbol value="lower_class"/>
       </f>
      </fs>
     </vColl>
    </f>
    <f name="language">
     <default/>
    </f>
   </fs>
  </f>
 </fs>
</person>
```

Fig. 3: Personography entry for the character *Alice Sandel*.

is at the project repository.[13] Findings enabled by these annotations are presented in the next section.

4.3 Findings

Our character social variable annotations allowed us to both reproduce trends already described in the available literature and arrive at new generalizations. We

13 https://git.unistra.fr/methal/methal-sources/-/tree/master/personography.

Tab. 4: Character social class distribution.

Social class	Count	%
Upper	61	10.85
Upper middle	74	13.17
Lower middle	209	37.19
Lower	218	38.79
Total	562	100

present our results on character gender distribution, character social class distribution, and character professional group distribution across time. Note that the complete personography (231 plays, 2386 characters) presented in 4.2 was not yet available when we carried out the analyses in the present section and the sample we report about here includes 108 plays with 1103 characters.

Regarding character gender, 24.59 % of characters were female, 71.32 % were male, while for the rest, the gender cannot be determined from the *dramatis personæ* or we are dealing with mixed-gender group characters. One salient difference between female and male characters in the corpus is that most female characters are not described with a profession, while the opposite holds for male characters, as shown in figure 4. As we discuss at the end of the section, there is, however, a characterization difference in this respect depending on whether the author is female or male.

Regarding social class distribution, it was possible to annotate 562 characters with the classification introduced in table 4; we considered that no sufficient cues were present in the *dramatis personæ* to assign a class to the remaining characters. We found a majority of middle class characters, followed by lower class ones. "Lower class" in this corpus mostly refers to characters in roles such as domestic worker; marginalized characters are absent or very rare, as was pointed out in earlier literature (Cerf 1972, p. 344; Hülsen 2003, p. 140). Some high class characters with socially prestigious occupations appear. However, our corpus also confirms observations in earlier literature that the French-speaking high-bourgeoisie and nobility are rare or absent (Hülsen 2003, p. 140).

We examined the distribution of socioprofessional groups according to our taxonomy (see 4.1) and their temporal evolution (Ruiz Fabo and Werner 2021). We focused on the 1890 to 1939 period. Earlier decades show only 10 characters or less; given the small data volume, we do not report results about them. Several of our results are compatible with trends described in (Cerf 1972; Hülsen 2003). Figure 5 synthesizes our results.

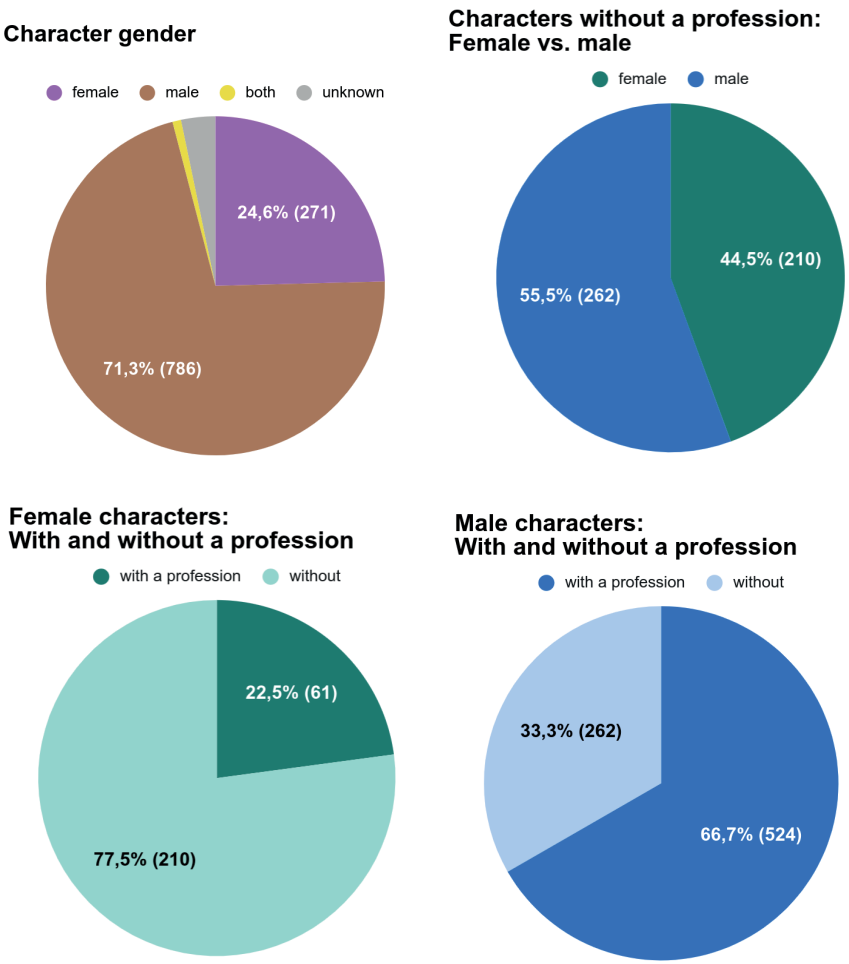


Fig. 4: Character gender and characters described with or without a profession.

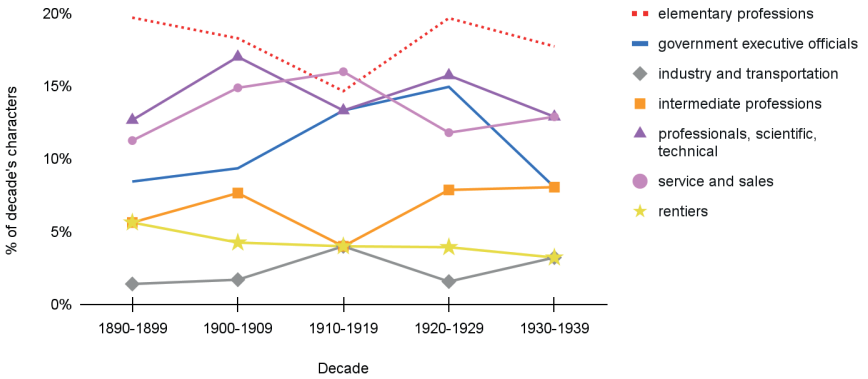
Subfigure 5a shows several groups whose representation is steady throughout the corpus period; the difference between the period minimum and maximum is no more than 5 percentage points, aggregating per decade. Most of these groups belong to the small and middle urban bourgeoisie. There are also two groups (*elementary professions* and *government*) whose interactions with the bourgeoisie groups can be a source of dramatic conflict in the plays. Characters in the elementary professions group are often domestic employees whose disagreements with their employers contribute to the plays' plots. Besides that, (comical) conflict between political authority figures and members of bourgeoisie groups also appears in the plays.

A further important bourgeoisie group, craftspeople, is shown in subfigure 5b. Unlike the groups displayed in 5a, this group's representation is not just steady across time, but rising; our sample shows a clear increase of professionals from the crafts group between the beginning and the end of the corpus period. This contrasts with the clear decline in character occupations related to agriculture between the beginning and the end of the period, a second trend also shown in 5b. As Cerf (1972, p. 344) notes, speaking of the Théâtre Alsacien de Strasbourg's (TAS) repertoire, it describes not only rural Alsace, but also the urban social groups that form its public. Our sample's increase in craftspeople during the corpus period points in the same direction.

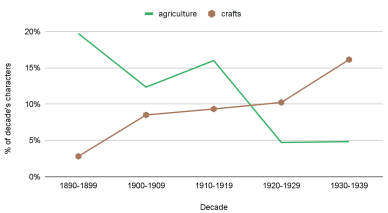
Cerf (1972, pp. 352–354) reports that Alsatian theater is careful not to promote controversy about religion and politics. This may be reflected in the fact that the clergy and the military only have a discontinuous presence in the plays; these and other character groups that have a minor presence in the plays, not appearing in all decades, are shown in subfigure 5c.

We also looked at the distribution of female characters described with a profession in the *dramatis personæ* (subfigure 5d); this character subgroup had not been addressed in earlier literature. It should be noted that data are scarce. As mentioned, only 24.6 % of characters are female and only 22.5 % of these are described via a profession in the plays' character lists. In other words, only about one character out of 20 in the corpus is a female character with a profession, which yields a total number of only 56 characters represented in subfigure 5d.[14] Even if the data are not abundant, a generalization that emerges is that most female characters with a profession are part of the elementary professions group (i.e., domestic workers in most cases). Several examples of female characters with professions

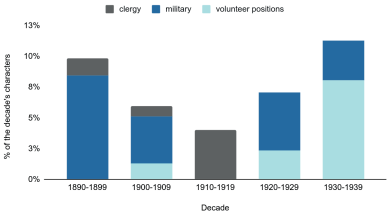
14 The count for female characters with a profession here (56) is slightly smaller than the one in figure 4 (61), because figure 4 includes all corpus decades, and we only consider plays from 1890 onwards when analyzing professional group distribution, given the small number of plays overall before 1890.



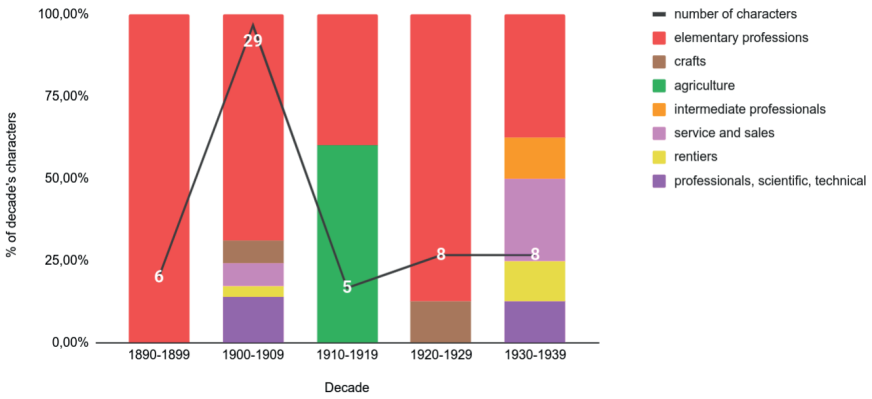
(a) Bourgeoisie groups (other than crafts), plus elementary professions and government.



(b) Crafts vs. agriculture professions.



(c) Minor groups.



(d) Female characters with a profession in the *dramatis personæ*: professional group distribution.

Fig. 5: Evolution of professional groups in the corpus, based on a sample of 108 plays and 1103 characters.

unrelated to domestic work exist in the 1900 and 1920 decades, e.g., a philosophy student, a textile merchant, a typist, an actress and a midwife. However, given the small amount of data, it is not possible to establish trends.

When crossing character-level metadata with play-level metadata we arrive at new generalizations. We annotated each play's scene for its urban or rural/small-town setting. Cerf (1972, p. 340), based on 39 TAS plays, states that craftspeople in small-town settings are absent from the repertoire. When looking at our larger sample this no longer holds, which attests to the usefulness of obtaining larger-scale annotations. We also annotated author gender and found a difference in the characterization of female characters in female vs. male authors. Among characters presented via a profession in the *dramatis personæ*, 46 % are female in plays written by female authors, vs. only 11 % in plays written by men. Such observations were not possible so far for this tradition, as the annotations that underlie them were not available.

5 Spelling Variants Matching

In Alsatian, all analysis methods that require a homogeneous representation of the corpus's vocabulary (such as corpus search tools or topic modelling) need to manage scriptural variation first, so as to detect and merge variants. For instance, we would like to perform automatic emotion analysis in the MeThAl corpus in the future. As a first step, we have built an emotion lexicon for the Alsatian dialects, named ELAL (Bernhard and Ruiz Fabo 2022).[15] In order to cater to the large amount of graphical variants observed in the MeThAl corpus, we have included graphical variants of Alsatian lexical items in this lexicon. Alsatian spelling variants (e.g., *Arbeewa* and *Ardbewa*) and closely related Alsatian and German word forms (e.g., Alsatian *Erdbewe* and German *Erdbeben*) have been identified automatically and corrected manually.[16]

As a by-product of the manual correction of the ELAL emotion lexicon, we obtained pairs of true, correct variants and false, incorrect variants. We have used this dataset to test and compare two different approaches to spelling variants matching. In the rest of this section, we first present these two approaches, then detail our training and test datasets and finally discuss the results obtained.

15 https://nakala.fr/10.34847/nkl.40cex998.

16 We refer the reader to (Bernhard and Ruiz Fabo 2022) for details about the process.

5.1 Methods for Spelling Variants Matching

The goal of spelling variants matching is to decide whether a given pair of forms are spelling variants or not. This can be considered as a binary classification task with two outcomes: true or false. The features traditionally used rely on string similarity measures or character n-grams, which are applicable out of context. In addition, features extracted from large corpora, e.g., word embeddings, make it possible to take the context of occurrence into account for validating variants. We have compared two methods using only word-internal features, since we operate on the vocabulary and not on the corpus level. The first method, proposed by Barteld et al. (2019), aims at filtering spelling variant candidates and was proposed for dealing with variation in texts from Middle Low German. The filter uses n-grams extracted from the pairs of word-forms given as input and word embeddings.[17] For example, for the pair (*beschtadiga*, *bschtadiga*), the following n-grams are extracted, based on the alignment of both words and the detection of the mismatches in the alignment; here, there is a single mismatch (insertion of 'e' between 'b' and 's'):

- 2-grams: [bb, -e], [-e, ss],
- 3-grams: [$$, bb, -e], [bb, -e, ss], [-e, ss, cc]
- 4-grams: [$$, bb, -e, ss], [bb, -e, ss, cc], [-e, ss, cc, hh]

These features are then used to train a Support Vector Machine (SVM).

The second method, DeezyMatch (Coll Ardanuy et al. 2020; Hosseini et al. 2020), uses a deep learning approach relying on a Siamese classifier consisting of two parallel recurrent layers and supporting the following architectures: Elman Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU). DeezyMatch was originally developed for toponym matching in the context of entity linking. The system converts input pairs to dense vectors after preprocessing steps which include tokenising the forms into smaller units (in our case, into characters and n-grams of characters).

5.2 Dataset

We have used the dataset which was manually annotated for building the ELAL emotion lexicon to generate training data. The development set contains 52 239 pairs, out of which 42 209 are positive instances (true variant pairs) and 10 030 are negative instances (non variant pairs). The held-out test set contains 22 389

17 We only used n-gram features, since we lack large corpora for building dense representations of word forms.

Tab. 5: Excerpts from the development dataset.

Form 1	Form 2	Matching
Gebortsdoe	Geburtsdäuj	TRUE
Pischtol	Pistol	TRUE
Gueter	Güeta	TRUE
unruehich	unruhig	TRUE

Form 1	Form 2	Matching
Schlöüi	schlag	FALSE
blute	blutt	FALSE
Grond	Grénd	FALSE
féjle	féle	FALSE

pairs, out of which 18 090 are positive instances and 4299 are negative instances. The dataset is imbalanced and contains about 4 times as many positive as negative instances. Figure 5 shows some examples from the development dataset,[18] which illustrate the difficulty of the task: very similar word forms can be negative instances (e.g., "féjle" – " féle") while more distant word forms can be positive instances (e.g., "Gebortsdoe" and "Geburtsdäuj").

5.3 Experimental Settings

In addition to the SVM evaluated in (Barteld et al. 2019), we tested two other classifiers: Ridge Classifier[19] and Logistic Regression.[20] We used the implementation provided by Barteld et al. (2019)[21]. for extracting the n-gram based features `ng` and the Python Scikit-learn library for the supervised classification experiments (Pedregosa et al. 2011). The best hyperparameters were chosen using 5-fold cross-validation on the development dataset.[22] The final models were trained on the entire development set. We also tested `other` features in addition to the n-grams: the frequency in the MeThAl corpus of each of the words in the pair, the number of French translations in the bilingual French-Alsatian lexicons for each of the words in the pair, the number of French translations shared by the two words in the pair.

We used the default parameters of DeezyMatch, except for the maximum n-gram length, which we changed from 3 to 4, and we did not lowercase the forms, so as

18 We considered inflectional variants, such as *Gueter* and *Güeta* to be valid pairs.

19 https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RidgeClassifier.html.

20 https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html.

21 Available at https://github.com/fab-bar/SpellVarDetection.

22 The best hyperparameters used in the experiments are as follows: Ridge Classifier alpha=1.0, SVM C=1.0, gamma=1.0, kernel='linear', Logistic Regression C=10000.

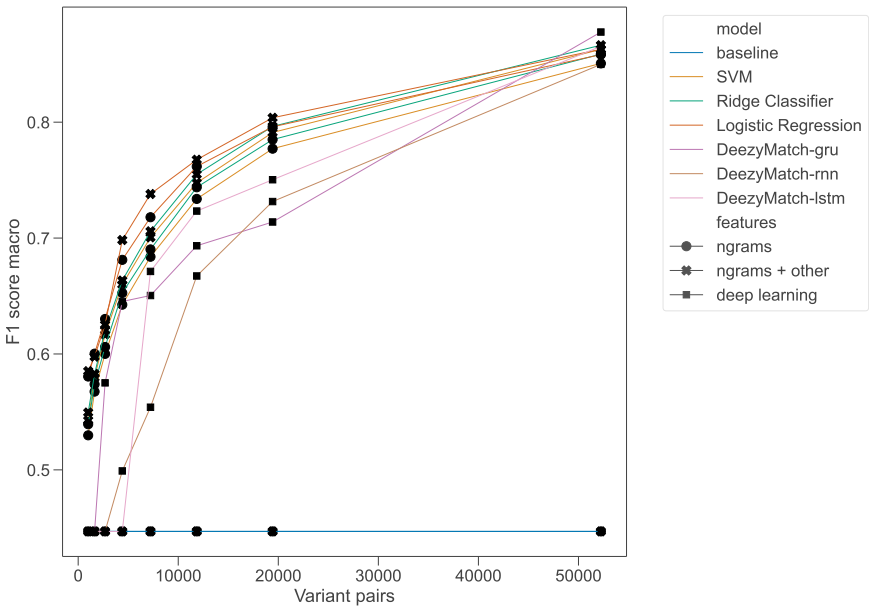


Fig. 6: F1-scores (macro average) on the test set for different amounts of training pairs.

to match the parameters used for Barteld et al.'s feature extractor. We used 85 % of the development dataset as training data and the rest for validation: the final model is the one with the lowest validation loss.

5.4 Results and Discussion

We compared the performance of the different models by varying the size of the training corpus (see Figure 6). The F1-scores increase logarithmically. The machine learning models SVM, Ridge Classifier and Logistic Regression overall perform better than DeezyMatch for the smallest training datasets, but obtain lower or equivalent results for the largest training dataset. The additional features bring some small improvements, which are more notable when less training data are available.

Table 6 shows the final results obtained on the held-out test set. The best model overall is DeezyMatch-gru. We compared the performances of the best performing methods on the macro F1 score using the paired bootstrap test (Berg-Kirkpatrick et al. 2012). The differences between DeezyMatch-gru and SVM, Ridge Classifier and Logistic Regression using the `ng+other` features are statistically sig-

Tab. 6: Evaluation results on the held-out test set. Scores are reported as macro-averages, since the dataset is imbalanced. We also include the F1 score for the minority class (non-variant NV). The baseline corresponds to always returning the majority class.

Model	Features	Precision	Recall	F1	F1 (NV)
Baseline		0.4	0.5	0.45	0
SVM	ng	0.88	0.83	0.85	0.75
Ridge Classifier	ng	**0.9**	0.83	0.86	0.77
Logistic Regression	ng	0.88	0.84	0.86	0.77
SVM	ng+other	0.89	0.84	0.86	0.77
Ridge Classifier	ng+other	**0.9**	0.84	0.87	0.78
Logistic Regression	ng+other	0.87	0.86	0.86	0.78
DeezyMatch-gru		0.88	**0.88**	**0.88**	**0.8**
DeezyMatch-rnn		0.85	0.85	0.85	0.76
DeezyMatch-lstm		0.87	0.86	0.87	0.78

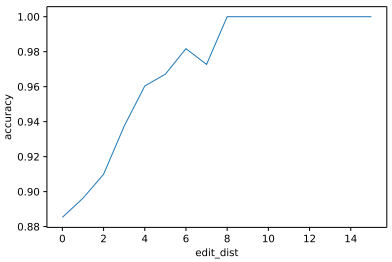


Fig. 7: Accuracy of DeezyMatch-gru as a function of the edit distance between the words in the pair.

nificant ($p<0.01$). The differences between DeezyMatch-gru on the one hand and DeezyMatch-rnn and -lstm on the other are also statistically significant ($p<0.01$). We can therefore conclude that DeezyMatch-gru is the best performing classifier for our dataset.

Finally, we analyze the accuracy of DeezyMatch-gru as a function of the edit distance between words (see Figure 7). As could be expected, word pairs with a larger edit distance (and thus more spelling differences) are easier to classify than word pairs with a lower edit distance (short words, or word pairs with few but distinctive differences).

The results obtained with these experiments are encouraging and we plan to utilize these methods to allow better access to the MeThAl corpus and to perform analyses based on the vocabulary and textual contents of the plays. Moreover, due

to limited workforce, we were only able to validate about 2/3 of the initial data used for building the ELAL emotion lexicon. We will try and use a semi-automatic method based on DeezyMatch to correct the data that have not been manually annotated yet.

6 Sharing the Project Resources

The resources are encoded in standard formats to facilitate reuse. They are also public, shared through several means: (1) our institutional GitLab repository;[23] (2) Nakala data repository, hosted at the French national infrastructure for digital humanites;[24] (3) the DraCor platform, which has accepted our first 25 plays, and where we plan on sharing the rest of the collection.[25] Thanks to a comprehensive TEI header and the DOI assigned by Nakala, our resources follow the FAIR principles.

We hope to promote interest in Alsatian in a wider public beyond the research community. For this reason, we developed a corpus navigation interface.[26] The usefulness of interfaces in literary research is disputed: whereas Schuwey (2019, pp. 12–15) has stressed their importance, Reiter et al. (2017, p. 1184) warn that their impact is not always positive, in the sense that they provide ready results rather than promoting a focus on the exact methodological choices used to arrive at them. Since our project targets not only specialists but also the general public, we believe that an interface can help make the tradition better known.

The interface allows for a structured navigation of the plays based on our character social annotations, besides on the bibliographic metadata (see figure 8). It is possible to filter the corpus based on the presence and co-occurrence of professional groups in the *dramatis personæ*, and filtering with other social variables like class will also be implemented. This can help find plays which show a conflict between certain groups and we hope it is an attractive way for the public to engage with the collection. A full-text search that handles variation in the corpus is under development.

23 https://git.unistra.fr/methal/methal-sources.
24 https://nakala.fr/collection/10.34847/nkl.feb4r8j9.
25 https://dracor.org/als.
26 https://methal.eu/ui.

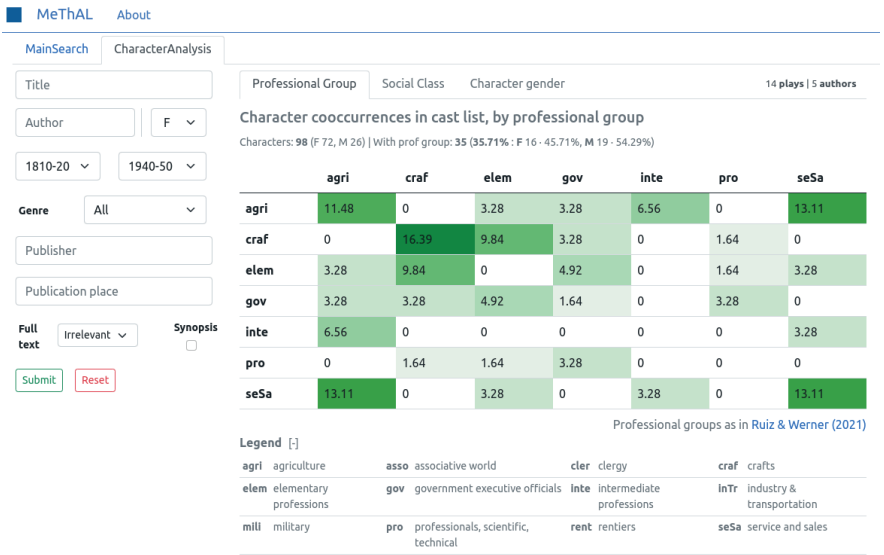


Fig. 8: Character view of the interface. The corpus is filtered to plays by female authors. It can be filtered further according to the character-group co-occurrences in the central pane.

7 Conclusion and Perspectives

We presented our work towards creating the first large-scale TEI corpus of Alsatian theater covering the 1870 to 1940 period. This is a dramatic tradition for which close to zero electronic text existed. NLP resources for Alsatian are also scarce, which limits the CLS questions that can be addressed. TEI encoding was aided by creating a simple sequence labelling model (Conditional Random Fields) which predicts elements based on OCR outputs. It was possible to train the model with our first seven plays only, which we had encoded using rules. This method can be applied to other dramatic corpora for which no electronic text existed and performing OCR was a requirement. Earlier literature on this tradition based on small samples had examined social groups in the plays. We performed an annotation of character social variables for all plays' *dramatis personæ*, later shared as a TEI personography implemented with feature structures which reflect characterisemes or characterization units. This gave an informative overview of the evolution of social groups in the plays which complements what was known about the tradition. It also attests to the usefulness of annotating character metadata systematically in a collection where the text is not yet encoded and where content analyses are hindered by large scripto-linguistic variation. In the future, characters from the

major traditions surrounding Alsatian theater (French and German) could be annotated with the feature library in order to compare the plays' character makeup. As a first step towards emotion analysis, we created an emotion lexicon that handles variation in the corpus. We presented variation detection methods that could be applied to similar scenarios and that can facilitate full-text search. As more linguistic resources get developed for Alsatian varieties (cf. DIVITAL project, Bernhard and Vergez-Couret (2022)), the range of possible NLP-based literary analyses of Alsatian theater will broaden.

Funding: This research was supported by Université de Strasbourg's IdEx program (Attractivité 2020 call).

Acknowledgment: We thank further project members, Dominique Huck and Pascale Erhart, for advice on corpus selection and other project aspects.

We also thank our interns: Nathanaël Beiner, Lena Camillone, Hoda Chouaib, Audrey Deck, Barbara Hoff, Valentine Jung, Salomé Klein, Audrey Li-Thiao-Té, Qinyue Liu, Kévin Michoud, Vedisha Toory, Heng Yang.

The authors would like to acknowledge the High Performance Computing Center of the University of Strasbourg for supporting this work by providing scientific support and access to computing resources. Part of the computing resources were funded by the Equipex Equip@Meso project (Programme Investissements d'Avenir) and the CPER Alsacalcul/Big Data.

References

Alto Editorial Board (2022). "ALTO: Technical Metadata for Layout and Text Objects." URL: https://www.loc.gov/standards/alto/.

Baierer, Konstantin, ed. (2020). *hOCR - OCR Workflow and Output Embedded in HTML*. URL: http://kba.cloud/hocr-spec/1.2/.

Barteld, Fabian, Chris Biemann, and Heike Zinsmeister (2019). "Token-based spelling variant detection in Middle Low German texts." In: *Language Resources and Evaluation*, pp. 1–30. DOI: 10.1007/s10579-018-09441-5.

Berg-Kirkpatrick, Taylor, David Burkett, and Dan Klein (2012). "An Empirical Investigation of Statistical Significance in NLP." In: *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pp. 995–1005.

Bermúdez Sabel, Helena (2019). "Encoding of Variant Taxonomies in TEI." In: *Journal of the Text Encoding Initiative* Issue 11. DOI: 10.4000/jtei.2676.

Bernhard, Delphine and Pablo Ruiz Fabo (2022). "ELAL: An Emotion Lexicon for the Analysis of Alsatian Theatre Plays." In: *Conference on Language Resources and Evaluation (LREC 2022)*, pp. 5001–5010. URL: https://hal.archives-ouvertes.fr/hal-03655148.

Bernhard, Delphine and Marianne Vergez-Couret (2022). *DiViTal: Increase the DIgital VITALity and Visibility of Languages of France: Linguistic Descriptions and Annotated Corpora*. URL: https://divital.gitpages.huma-num.fr/en/.

Cerf, Eve (1972). "Le théâtre alsacien de Strasbourg, miroir d'une société (1898-1939)." In: *Saisons d'Alsace* 43.

Cerf, Eve (1975). "Les contes merveilleux du théâtre alsacien de Strasbourg." In: *Revue des sciences sociales de la France de l'Est* 4, pp. 3–30.

Coll Ardanuy, Mariona, Kasra Hosseini, Katherine McDonough, Amrey Krause, Daniel van Strien, and Federico Nanni (2020). "A Deep Learning Approach to Geographical Candidate Selection through Toponym Matching." Tech. rep. arXiv:2009.08114. arXiv. DOI: 10.48550/arXiv.2009.08114.

Deutsche Forschungsgemeinschaft (2016). "DFG Practical Guidelines on Digitisation." Deutsche Forschungsgemeinschaft. URL: https://www.dfg.de/formulare/12_151/12_151_en.pdf.

Erjavec, Tomaž, Darja Fišer, and Nikola Ljubešić (2021). "The KAS Corpus of Slovenian Academic Writing." In: *Language Resources and Evaluation* 55.2, pp. 551–583. DOI: 10.1007/s10579-020-09506-4.

Even-Zohar, Itamar (1990). "Polysystem Theory." In: *Polysystem Studies [=Poetics Today]* 11.1, pp. 9–26. URL: https://www.tau.ac.il/~itamarez/works/books/Even-Zohar_1990--Polysystem%5C%20studies.pdf.

Fièvre, Paul (2017). *Rubrique Concernant Les Personnages*. Théâtre Classique. URL: http://www.theatre-classique.fr/pages/PagePersonnages.html.

Fischer, Frank and Ingo Börner (2019). "Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama." In: *Digital Humanities 2019*, p. 5. URL: https://dev.clariah.nl/files/dh2019/boa/0268.html.

Gabay, Simon, Barbara Topalov, Caroline Corbières, Lucie Rondeau Du Noyer, Béatrice Joyeux-Prunel, and Laurent Romary (2021). "Automating Artl@s – Extracting Data from Exhibition Catalogues." In: *EADH 2021 - Second International Conference of the European Association for Digital Humanities*.

Gall, Jean-Marie (1974). *Le théâtre populaire alsacien au XIXe siècle*. Istra.

Galleron, Ioana (2017). "Conceptualisation of Theatrical Characters in the Digital Paradigm: Needs, Problems and Foreseen Solutions." In: *Human and Social Studies* 6.1, pp. 88–108. DOI: 10.1515/hssr-2017-0007.

Geyken, Alexander, Susanne Haaf, Bryan Jurish, Matthias Schulz, Christian Thomas, and Frank Wiegand (2012). "TEI und Textkorpora: Fehlerklassifikation und Qualitätskontrolle vor, während und nach der Texterfassung im Deutschen Textarchiv." In: *Jahrbuch für Computerphilologie*. URL: http://computerphilologie.digital-humanities.de/jg09/geykenetal.html.

Hosseini, Kasra, Federico Nanni, and Mariona Coll Ardanuy (2020). "DeezyMatch: A Flexible Deep Learning Approach to Fuzzy String Matching." In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 62–69. DOI: 10.18653/v1/2020.emnlp-demos.9.

Huck, Dominique (1998). "D'r Herr Maire (1898) de Gustave Stoskopf. Entre ethnologie et littérature : les Alsaciens en auto-représentation." In: *Recherches Germaniques* 28, pp. 163–190.

Huck, Dominique (2005). "Le 'Théâtre Alsacien de Strasbourg' et la production dramaturgique de ses fondateurs (1898-1914)." In: *Culture et histoire des spectacles en Alsace et en Lorraine : De l'annexion à la décentralisation (1871-1946)*. Ed. by Jean-Marc Leveratto, Jeanne Benay, Olivier Thomas, and Séverine Wuttke. Peter Lang, pp. 198–222.

Huck, Dominique (2015). *Une Histoire Des Langues de l'Alsace*. La Nuée Bleue.

Huck, Dominique, Arlette Bothorel-Witz, and Anemone Geiger-Jallet (2007). "L'Alsace et ses langues. Eléments de description d'une situation sociolinguistique en zone frontalière." In: *Aspects of Multilingualism in European Border Regions: Insights and Views from Alsace, Eastern Macedonia and Thrace, the Lublin Voivodeship and South Tyrol*. Ed. by Andrea Abel, Mathias Stuflesser, and Leonhard Voltmer. EURAC Research (Europäische Akademie / Accademia Europea / European Academy), pp. 13–101.

Hülsen, Bernhard von (2003). *Szenenwechsel im Elsass: Theater und Gesellschaft in Strassburg zwischen Deutschland und Frankreich 1890-1944*. Leipziger Universitätsverlag.

Khemakhem, Mohamed, Luca Foppiano, and Laurent Romary (2017). "Automatic Extraction of TEI Structures in Digitized Lexical Resources using Conditional Random Fields." In: *electronic lexicography, eLex 2017*. URL: https://hal.archives-ouvertes.fr/hal-01508868.

Khemakhem, Mohamed, Laurent Romary, Simon Gabay, Hervé Bohbot, Francesca Frontini, and Giancarlo Luxardo (2018). "Automatically Encoding Encyclopedic-like Resources in TEI." In: *Proceedings of the annual TEI Conference and Members Meeting*. URL: https://hal.inria.fr/hal-01819505.

Korobov, Mikhail [Nov. 26, 2015] (2019). *sklearn-crfsuite*. TeamHG-Memex. URL: https://github.com/TeamHG-Memex/sklearn-crfsuite.

Lafferty, John D., Andrew McCallum, and Fernando C. N. Pereira (2001). "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data." In: *Proceedings of the Eighteenth International Conference on Machine Learning*, pp. 282–289.

Moeller, Katrin and Robert Nasarek (2018). "Die Ontologie historischer deutschprachiger Berufs- und Amtsbezeichnungen. Interoperationalität und Berufsklassifizierung durch semantisches Topic Modeling." In: *DHd 2018. Kritik der Digitalen Vernunft. Konferenzabstracts*. DHd 2018, pp. 173–177.

Pagel, Janis, Nidhi Sihag, and Nils Reiter (2021). "Predicting Structural Elements in German Drama." In: *Second Conference on Computational Humanities Research*, pp. 217–227.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay (2011). "Scikit-learn: Machine Learning in Python." In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Pfister, Manfred [1977] (2001). *Das Drama: Theorie und Analyse*. 11th ed. W. Fink.

Phelan, James (1989). *Reading People, Reading Plots: Character, Progression, and the Interpretation of Narrative*. University of Chicago Press.

Reiter, Nils, Jonas Kuhn, and Marcus Willand (2017). "To GUI or Not to GUI?" In: *INFORMATIK 2017*. URL: https://dl.gi.de/bitstream/handle/20.500.12116/3880/B14-8.pdf?sequence=1&isAllowed=y.

Reiter, Nils and Janis Pagel (2020). *quadrama/DramaAnalysis: 3.0.2*. Version v3.0.2. DOI: 10.5281/zenodo.4051004.

Romary, Laurent (2015). "Standards for Language Resources in ISO – Looking Back at 13 Fruitful Years." In: *edition - Die Fachzeitschrift für Terminologie* 11.2, pp. 13–19.

Ruiz Fabo, Pablo and Carole Werner (2021). "Exploration du théâtre alsacien à travers ses listes de personnages pendant la période 1870-1940." In: *Humanistica 2021*. DOI: 10.5281/zenodo.4762733.

Schuwey, Christophe (2019). *Interfaces. L'apport des humanités numériques à la littérature*. Alphil éditions.

Smith, Ray (main developer) (2018). *Tesseract (v4.0)*. URL: https://github.com/tesseract-ocr/tesseract.

TEI Consortium (2022). *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version v4.4.0. DOI: 10.5281/zenodo.6482461.

Van Leeuwen, Marco H. D., Ineke Maas, and Andrew Miles (2004). "Creating a Historical International Standard Classification of Occupations An Exercise in Multinational Interdisciplinary Cooperation." In: *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 37.4, pp. 186–197. DOI: 10.3200/HMTS.37.4.186-197.

Rebecca M. M. Hicke and David Mimno

"Let every word weigh heavy of her worth": Examining How Women Enact Power in Shakespeare's Comedies through Interactive Speech Pattern Visualizations

Abstract: Critics from a variety of disciplines have noted that Shakespeare's comedies are in some aspect unique, and that their female characters are particularly prominent. In order to better understand the impact of women on the comedies' narratives, in this paper we create and analyze interactive visualizations of their speech patterns. These visualizations illustrate that the conventions of Shakespeare's comedies allow women to emerge as prominent characters who impact the plays' trajectories and influence their own narratives. In particular, they show that the comedies' focus on ensembles of characters and localized conflicts, instead of political or military dramas, creates an environment in which women may become guiding members of their social groups. We see that many of these women address and impact more characters throughout a play as its social clusters increasingly interact. This increase in impact means that, although the comedies may frequently end with a return to patriarchal normativity, women are able to enact power throughout the course of the plays and ultimately influence their fate within the patriarchy.

1 Introduction

Scholars from several disciplines have identified the comedies as unique amongst Shakespeare's plays. Many feminist critics have argued that they are the plays that most prominently feature female characters and allow them the most influence (Dash 1981; Pitt 1981; Gay 1994; Rackin 2005; Callaghan 2016). In addition, digital humanists have found that the structure of the comedies is unique; they feature smaller social networks, more localized conflicts, and are more likely to focus on ensembles of characters instead of individuals or pairs (Stiller and Hudson 2005; Algee-Hewitt 2017; Masías et al. 2017; Evalyn et al. 2018; Murphy et al. 2020). In this work, we examine what makes the comedies' female characters unique by creating a specific, measurable definition of a character's *influence* within a play; namely, we work on the assumption that an influential character should say more

Rebecca M. M. Hicke, David Mimno, Cornell University

https://doi.org/10.1515/9783111071824-005

words to more people. While this definition of influence does not encompass what is being said by characters, it approximates the extent to which they participate in the plot of a play. In this paper, we seek to use this measure of influence to concretely see how women impact the course of Shakespeare's comedies by creating and analyzing visualizations that look at their speech patterns. The visualizations we created can be found online[1].

These visualizations demonstrate how the small, tightly knit social structures of the comedies allow women to impact the plays' characters and plots, thereby influencing their own narrative trajectories (Stiller and Hudson 2005; Algee-Hewitt 2017; Evalyn et al. 2018). The comedies' tendency to focus on an ensemble of characters and conflicts that lie 'closer to home' than the political drama of the tragedies and histories creates an environment in which women can become guiding members of their social groups (Evalyn et al. 2018; Murphy et al. 2020). Our visualizations show that many female characters address and impact more characters throughout a play as its social clusters increasingly interact. This increase in impact means that, although the comedies may frequently end with a return to patriarchal normativity, the women are able to enact power throughout the course of their plays and ultimately influence their fate within the patriarchy (Pitt 1981; Gay 1994). Indeed, the visualizations show that in all but one of the comedies, the most impactful female character speaks in the final scene, indicating that these powerful women are not silenced but maintain influence through till the very end. In addition, the visualizations allow us to see women taking part in the action of the plays during major scenes and reveal how other conventions of comedy, such as crossdressing, give female characters leeway and power with which to pursue their goals.

We have specifically structured this study around finding ways in which the women in Shakespeare's comedies enact power. This is valuable – and, in fact, a critical facet of the project's construction – for several reasons. Shakespeare's continued influence on our culture means that how we view his female characters has an impact on the modern day. In the introduction to *The Feminist Companion to Shakespeare*, Dympna Callaghan states that "language in general is crucial to any understanding of gender organization" and therefore "canonical representations of women [...] hold a hugely important place" in forming our impressions of gender relations (Callaghan 2016, p. 3). Similarly, Phyllis Rackin argues that "the way we read Shakespeare's plays matters because the cultural prestige of Shakespeare makes his plays a model for contemporary values" (Rackin 2005, p. 16). It is important to perform readings of Shakespeare in which we intentionally focus on

1 https://observablehq.com/@rhicke/let-every-word-weigh-heavy-of-her-worth.

the influence and significance of female charcters because the cultural perspective on Shakespeare's women affects our modern perspective on women's roles throughout history.

It is also important, as Rackin argues, to see the ways in which women could enact power and find agency in the past because they provide motivation for work towards a better future. Rackin states that "although we cannot afford to ignore the history of women's subjugation ...overestimating past repression can easily slip into a dangerous complacency about present progress" (Rackin 2005, p. 2). Highlighting how women in Shakespeare's plays act with autonomy and affect or even control the outcome of their narratives can help inspire a sense that change is possible. Women are not inevitably forced to be powerless and were not even fully so in a time that we look on as far more oppressive; our default assumption should not be that women were restricted to insignificant roles. It is true that the women in Shakespeare's comedies do not have as much power as the men. However, this does not mean they are *powerless* within their narratives; women may be enacting an informal influence on the plays. Whether and the extent to which female characters enact such informal influence are the questions we seek to address in this paper.

We use digital tools throughout this project in an effort to gain new insight into a topic that has already received considerable critical attention. While close reading is an important analysis tool that provides great insight into literature, it is not well-suited to processing large amounts of information. It is challenging to see high-level patterns when examining plays on a word-by-word basis. And although we may be able to develop intuitions about these large-scale trends through close reading, we cannot prove that these intuitions are correct without performing some sort of data analysis. For example, it may seem obvious that Rosalind is the most prolific speaker in *As You Like It*, but we cannot prove this without counting the number of lines or words each character in the play speaks.

Computational tools, on the other hand, allow us to handle large quantities of data and visualize them in complex and interactive ways. They let us augment traditional literary analysis "with quantitative models that can not only introduce new categories for comparative analysis but also nuance, deepen, and even fundamentally change our understanding of literary history and what a text is" (Algee-Hewitt 2017, p. 779). The models that technology allows us to produce can help reveal underlying structures in our data. By visualizing women's speech patterns throughout the comedies, we are attempting to reveal and analyze just such an underlying structure.

Visualizing data allows humans to perceive "patterns, trends or anomalies, constancy or variations" that aren't otherwise visible (Hinrichs et al. 2019, p. i81). Moreover, collecting data with the particular goal of creating a visualization can

guide researchers to interesting aspects of that data; thinking about what to present and how to present it leads to conscientious data collection. There are also great benefits to creating a specialized visualization for a specific dataset instead of using a generalized tool. Combining visualization and data collection in an iterative workflow allows both processes to inform each other and ideas to develop as intermediary versions of visualizations bring to light trends for further exploration. Using general tools can also mean one "risks reproducing unexamined assumptions and overlooking important nuances of humanistic data and inquiry" (Hinrichs et al. 2019, p. i83). Thus, we chose to center this project on data visualization and to build novel visualizations from scratch instead of attempting to use an existing tool.

We believe that computing and visualizing speech statistics will help us understand what contributes to a character's prominence as, currently, "there is little agreement on who the main character [of a play] is and which elements of a script contribute to establishing a character's leading role" (Masías et al. 2017, p. 837). Speech patterns can, to some extent, serve as a proxy for a character's impact on the action of a play because in plays most of the narrative must be conveyed through speech alone. Thus, visualizing how characters speak can indicate when and to what extent they take action. Including impact, or how many people a character speaks to, in our visualizations also allows us to examine the size of a character's network of influence. There are, of course, limits to the conclusions we can draw from examining speech in this manner; abstracting to speech patterns allows us to see overarching trends but removes the context of what exactly is being said. This methodology does not allow us to distinguish between a woman's voiced submission and an expression of independence. Yet, speech still indicates that a character has a role in the action of a scene. Moreover, an examination of the data resulting from our annotations reveals that the more independent female characters within the comedies have more prominent speech patterns. Thus, while this method does not allow us to make conclusive statements about characters' actions, it can still serve as an approximation of their narrative role.

Overall, by undertaking this project we seek to contribute a deeper understanding of how women impact the comedies and explore the ways in which we can view some of the best-known female characters in the literary canon as powerful. We also produce a dataset and interactive visualizations that will hopefully inspire future researchers to continue this work.

2 Methods

2.1 Data Collection

Although it is easy to calculate how many words each character in one of Shakespeare's comedies speaks, information on who lines are addressed to is not readily available and is difficult to calculate because of its subjectivity. Therefore, we created our own data set by hand-annotating the comedies for who each particular section of speech within each play was addressed to. This was challenging as multiple interpretations of the addressees were plausible for many sections of speech, particularly when many characters were onstage. These complications were resolved through the use of secondary materials and annotator discretion.

For the purposes of this project, we define the comedies as the fourteen plays labeled as such in the First Folio: *All's Well That Ends Well*, *As You Like It*, *The Comedy of Errors*, *Love's Labor's Lost*, *Measure for Measure*, *The Merchant of Venice*, *The Merry Wives of Windsor*, *A Midsummer Night's Dream*, *Much Ado About Nothing*, *The Taming of the Shrew*, *The Tempest*, *The Two Gentlemen of Verona*, *Twelfth Night*, and *The Winter's Tale*. We used the Folger Shakespeare Library's online text file editions as copy texts for this work and completed our initial annotations using only these versions of the texts. We then used the Arden Shakespeare Third Series editions as secondary sources for all plays except for *Twelfth Night*, for which we used the Arden second series edition as the third was not available. Using the Arden editions, we read over each of the plays again, adjusting the annotations based on new information given in the Ardens' supplementary material. The Ardens provide extensive academic footnotes on changes made to the original text and the meaning of difficult passages, which supplied extra context for the annotators to work from. When the Arden and Folger editions diverged, we used the version of the text that most closely resembled the original, except when coherence strongly suggested that emendation was necessary.

Once the annotation stage of this project was complete, we wrote a Python script to process the annotations and output the data as JSON files, which are easily compatible with the library used to create the visualizations. To allow for outside validation of our annotations and to make the dataset available for general use, we have published the annotations and processing code on Github.[2]

2 https://github.com/rmatouschekh/shakespeare_comedy_annotations.

2.2 Measuring Impact

In order to study the impact of female characters on the comedies, we define an influence metric which reflects how much of an effect a section of speech has on the underlying social network of a play. Specifically, we define influence to be the product of the number of words in a given section of text and how many characters that section of text is addressed to. Thereby, the value of a statement increases the more people it directly affects. Masías et al. found that "the 'strength' of [their] relationships affects the prominence of" a character in their social network (Masías et al. 2017, p. 837). When a section of speech is addressed to multiple characters, it can be seen as developing the relationship between the speaker and each of the addressees. Therefore, the more characters a section of speech is addressed to, the more it will raise the speaker's prominence within their social network and thus their influence. This method of calculating influence assumes that the more people a character addresses, the more people's thoughts and actions they are impacting.

If a section of speech is addressed to one or more unnamed characters, it is labeled as addressed to 'Other' and given the same weight as a section of speech addressed to one named character. This approach is necessary because it is often unclear how many unnamed characters are on stage and being addressed at once and attempting to guess a number with no basis could easily lead to faulty data. However, since unnamed characters are likely to have less impact on a play than named characters, giving them a lower influence weight makes intuitive sense. A speaker's influence over unnamed characters translates to less impact on the play's action than their influence over named characters. We also exclude any statements labeled as exclamations and those directed towards the audience from the influence measurements. Since these measurements seek to reflect a speaker's impact on the other characters in the play, sections of speech not addressed to other characters are not relevant to the metric and are instead given their own view within the visualizations.

We calculate the total influence in a play by summing the influence measurements for all the sections of text spoken throughout the play. This allows us to view each character's contribution as a percentage of the total influence in a play and thereby eliminate any skew created by varying play lengths.

2.3 Visualization Design

The final interactive visualizations produced for this study were created with the D3 library in JavaScript and have been made available on the interactive notebook

Observable. We chose Observable to host the visualizations because it simultaneously makes it easy to integrate formatted text and code, makes the visualizations broadly accessible without requiring extra software or setup, and allows viewers to explore the underlying code and data used to create the visualizations. By facilitating the exploration of our code and data we hope to increase the replicability of our work. We also want to allow those interested in expanding this research to easily extract and adapt our code for their own studies.

Each of the visualizations created for this study is a collection of area charts that show how a metric relating to characters' speech patterns changes throughout a play. A data point is included for each scene in the play. This form of visualization is meant to complement previous network analysis research. It is more difficult to examine the strength of individual relationships in these visualizations, but they allow for comparisons between different scenes in a play, a dimension which is invisible in typical networks. Although scene breaks in Shakespeare's plays are impositions by modern editors, previous research has shown that they are meaningful. For example, Stiller and Hudson (2005) find that scenes do provide "a distinct social unit within each play" (Stiller and Hudson 2005, p. 64). This finding allows us to use scene breaks in our visualizations and analysis when considering changes over the course of a play.

The visualizations allow users to examine the comedies as a unit or to study each play individually. A graph with an area chart representing the total contribution by female characters in each play is included in the general view of all the comedies. These graphs are ordered by the estimated publication date of each play. In contrast, the play-specific visualizations include a graph with an area chart representing the speech of each prominent character in that play. These graphs are ordered from greatest to least contribution by each character. The play-specific view also includes an additional line graph with the total measurement for the metric in each scene, as well as the totals for all male characters and all female characters. In addition, each play-specific visualization includes additional data on that play including the rank of each play by the percentage of speech and percentage of impact contributed by women, the first five characters ranked by speech and impact, and the weighted and regular average rank of female characters and male characters.

Both the general and play-specific visualizations have three modes: regular, weighted, and spoken to audience. The regular mode of each visualization shows the percentage of all words in the play spoken by each character of interest during each scene. A trend line in gray is included in each graph which shows the percentage of speech a character would be expected to contribute in a scene if they always spoke proportionately to the size of the scene. This trend line allows viewers to compare the relative size of each scene and note when a character's

speech pattern deviates from that of the play. This mode allows viewers to assess the basic structure of the play and compare at a fundamental level how much characters are talking in any given scene, as well as how their speech patterns change throughout a play.

The weighted mode graphs the influence measurements described above. In this view, the radius of the point representing each scene is scaled according to the square root of the number of characters the represented speaker addresses during that scene. The y-value of each point reflects the percentage of the total influence in a play the character contributes. This weighted mode is valuable because it expands what kinds of impact are visible in the visualizations. This view also has a gray trend line for each character depicting the amount of the total influence we would expect a character to contribute in a scene if their impact was always proportional to the total impact of the scene.

The final mode, spoken to audience, allows viewers to examine speech that is not directed at another character in the play. Because they aren't addressed to other characters, these asides, exclamations, and soliloquies do not have the same impact on a play's social network as speech included in the influence metric. However, soliloquies and other asides often describe to a character's emotional state. The inclusion of these speeches by female characters presents them to the audience as individuals with their own perspectives and opinions and positions them to various extents as the figures through which the audience experiences the play.

3 Discussion

Digital humanists and literary scholars alike have identified the comedies as unique amongst Shakespeare's plays. Literary, particularly feminist, critics have long pointed to the comedies as the subset of Shakespeare's plays that place the most focus on female characters and allow them the greatest freedom (Rackin 2005, p. 51; Pitt 1981, p. 76; Gay 1994, p. 3). These critics argue that in the comedies, women act as both "agents of happiness and order" (Pitt 1981, p. 76) and the "unknowable Other[s] of patriarchy" who "can refuse to obey the rules of appropriate gender behavior" and "make [their] marginal position a source of disruptive power" (Gay 1994, p. 3). They believe that female characters are central to the action of the comedies and, even more importantly, to their resolution. In this work, we search for a structural reflection of female characters' influence and power.

Digital humanists single out the comedies for their distinctive structure. Scholars have found that the comedies feature "far fewer characters" than plays

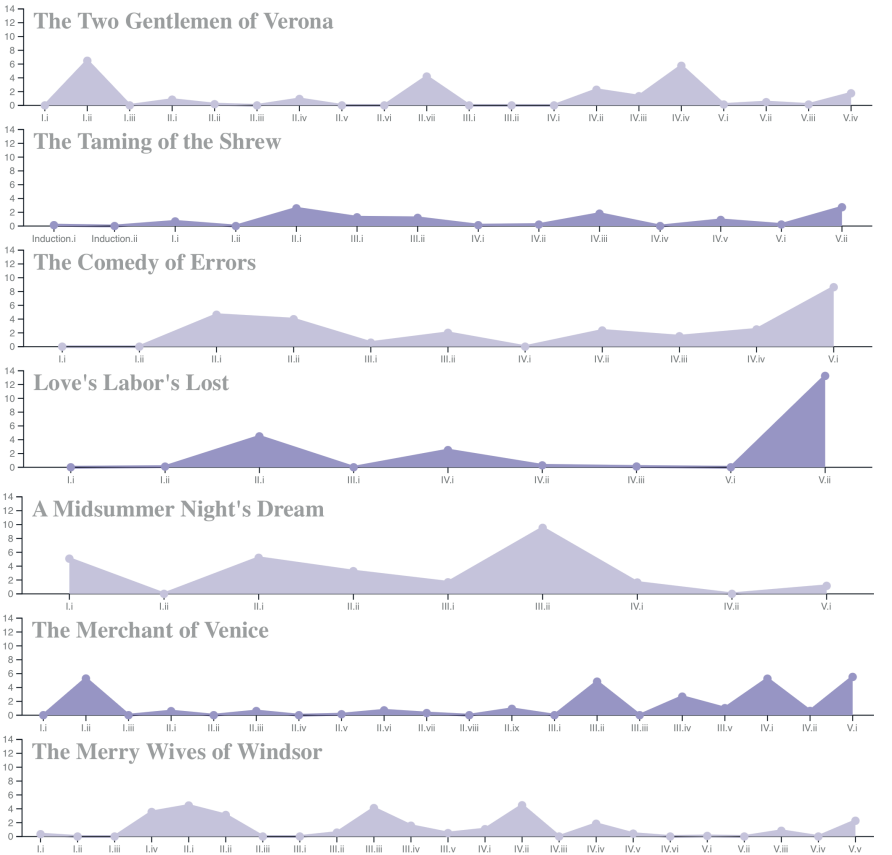


Fig. 1: Percentage of words spoken by all female characters in all comedies. The area under a dot represents the percentage of words spoken in that scene.

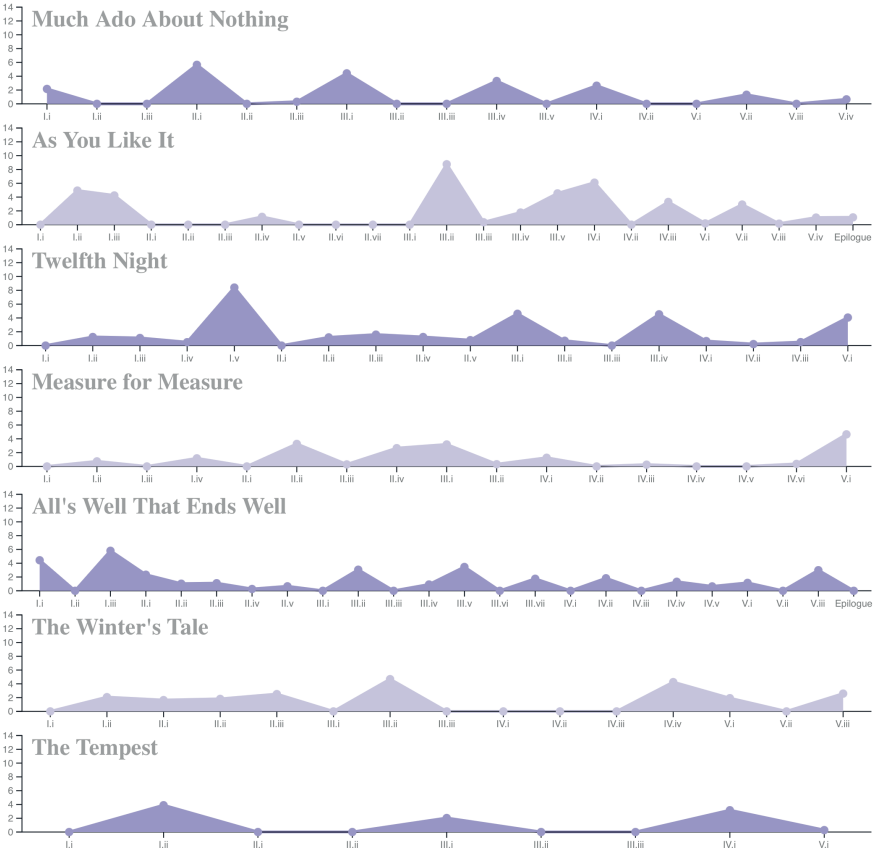


Fig. 2: Percentage of words spoken by all female characters in all comedies. The area under a dot represents the percentage of words spoken in that scene (cont.).

in the other genres and tightly knit "networks [...] in which nearly everybody speaks to nearly everybody else at some point" (Evalyn et al. 2018, p. 269). The "high degree of social clustering" (Stiller and Hudson 2005, p. 58) in the comedies means that a greater number of characters in each play can be considered important, "reflecting a focus on ensemble stories" (Evalyn et al. 2018, p. 269). Algee-Hewitt (2017) and Murphy et al. (2020) also argue that comedies tend to focus on storylines that lie 'closer to home' than histories and tragedies, "which are more oriented to sovereignty and dynasty" (Murphy et al. 2020, p. 236). We are interested in examining whether the comedies' broad attention to multiple important characters and tendency to revolve around more localized conflicts allow women, who were "excluded from the kinds of power politics" often present

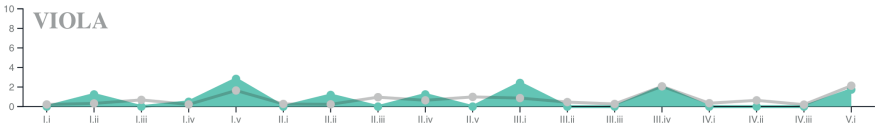


Fig. 3: Percentage of words spoken by Viola in *Twelfth Night*. The gray line represents the expected value of her speech – the percentage of the play in the scene multiplied by the total percentage of words spoken by Viola.

in the tragedies and histories, to become important and influential members of their social groups in ways they are often precluded from in the other genres (Murphy et al. 2020, p. 231). Murphy et al. also found evidence that female characters are more focused on familial and close social relationships than male characters. They examined how likely characters were to use certain words given their gender, rank, and the genre of play they are in and found that women were more likely to use words like 'husband' and 'sister,' which signify attention to local figures of authority and social, instead of political, relationships. Murphy et al. believe that "it could be argued that the female keywords and their collocates are consistent with a 'rapport' style of speech, emphasizing the building and maintaining of social relations" (Murphy et al. 2020, p. 234). These findings support the idea that the comedies' focus on local dramas is well suited to allowing women to take active and important roles in the action of a play. Although it is, of course, limiting to depict women as solely powerful within closed and homely spheres, it is impossible to ignore that this largely reflects the historical reality of when these plays were written. It is still valuable to highlight the ways in which female characters could enact power in those times even if those ways were limited.

Our visualizations demonstrate that female characters do, in fact, have a noteworthy impact on the plots of the comedies. Their speech patterns appear to reflect involvement in a social clique that usually increasingly interacts with other cliques throughout the play, allowing the women to influence a greater number of characters. Examining the visualizations shows that the speech patterns of the most impactful women in each comedy are usually 'spiky.' That is, scenes in which a female character speaks a large amount are usually followed by a scene in which they speak very little or not at all. This suggests that these women are speaking members of groups who appear in a subset of the plays' scenes. This pattern is mimicked by the plays' most impactful male characters, meaning that the female characters' absences in some scenes is not a sign of unimportance, but merely a reflection of the plays' dramatic structure.

Studying characters' speech patterns on a play-by-play basis reveals that scenes which take up a notable percentage of the play (ones in which the gray

lines spike) often correspond with a spike in speech by a female character. Moreover, just as for male characters, this spike often far exceeds the expected contribution of the female characters, as shown by the gray line (ie Scenes I.v and III.i in Figure 3). This has several implications. First, it suggests that women are members of the central social group around whom an important portion of the action in a play revolves, not only side characters in a subplot that is secondary to the main narrative. Second, it demonstrates that women speak during momentous scenes. Because speech usually signifies action in plays, this also indicates that women are likely taking part in the action during important parts of the play. The focus of the comedies on ensembles of characters combined with the tendency of Shakespeare's female characters to participate more in localized storylines may indeed help explain why we see women take action and emerge as central characters specifically in the comedies.

The visualizations also show that impactful female characters usually speak in their play's final scene and that at least one female character speaks in the final scene of every comedy (Figure 2). This is particularly interesting because many feminist critics have argued that the comedies end with a return to normativity. Penny Gay states that the comedies' heroines are "almost invariably recuperated into the patriarchal economy via marriage" by the end of each play (Gay 1994, p. 3). Additionally, although she argues that performance disrupts these theories, Gay points to the arguments that comedy is "profoundly conservative," a space in which audiences can benignly "enjoy [...] fantasies of disruption" before the play inevitably "[settles] back happily into the regulated social order of the patriarchy" (Gay 1994, p. 2). Along these same lines, Angela Pitt claims that "all women who go against prevailing conventions" in the comedies are ultimately "redeemed by the end of the play," which constitutes a "restoration of order" (Pitt 1981, pp. 75–76). However, the fact that so many female characters speak in the final scene of their play suggests that, even when women marry and thereby return to patriarchal conventions, they still take part in the play's action. The return to order may mean submission, but it does not mean submissive *silence* for female characters. Even Shakespeare's married women, held within a normative power structure, may impact the course of a comedy.

Twelfth Night provides a clear example of a woman who continues to take an active role in the play's action post marriage. As seen in Figure 4, all three of the female characters in *Twelfth Night* have notable speaking roles and both Olivia and Viola speak in the last scene of the play, addressing 5 and 10 characters respectively. It is particularly noteworthy that Olivia has such a large impact in this scene because by this point in the play she is already married to Sebastian. Instead of showing her fading into the background, this visualization reveals that Olivia's appearance post-marriage is actually her second most impactful speaking role,

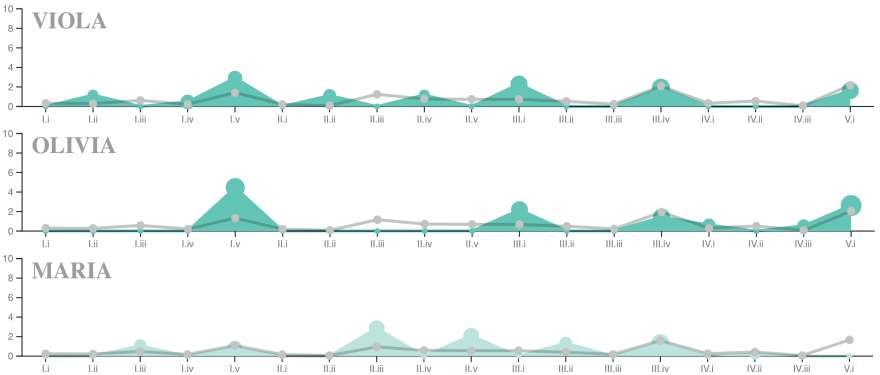


Fig. 4: Percentage of influence contributed by each female character in *Twelfth Night*.

Fig. 5: Percentage of influence contributed by the Lady Abbess in *The Comedy of Errors*.

and the scene in which she addresses the most people. Indeed, examining the text of the play we see that in this scene Olivia acts as a central figure in the resolution of both of the plays' major plotlines: the reuniting of Viola and Sebastian and the fooling of Malvolio. As one of the highest ranked characters present, Olivia is able to request and receive information from key figures in both plots. *Twelfth Night* also supplies an example of a woman who has an important role even though she does not appear in the final scene and is not as voluble as her counterparts. Maria clearly acts as what Stiller and Hudson refer to as a 'keystone character,' a figure who connects multiple social cliques. She appears in nearly every scene that Olivia and Viola do not (only two of the play's 18 scenes have no female speakers) and interacts with Olivia and Viola in multiple scenes as well. Since scenes represent social clusters, this visual pattern reflects her acting as a tie between a clique that contains Olivia and Viola, and a separate group.

Another example of a married woman who plays a very important role in the resolution of her play's drama is the Lady Abbess, or Emilia, in *The Comedy of Errors*. Figure 5 shows that the Lady Abbess only appears in one scene of the play, the last, but has a very large impact on that scene. She speaks to 12 of the 14 characters in the scene, contributes nearly 8 % of the play's total impact, and finally resolves the confusion that lies at the center of the play. She facilitates the reunion of the twins, of Adriana and her husband, and of the Lady Abbess herself and her

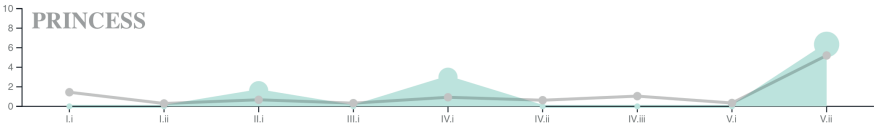


Fig. 6: Percentage of influence contributed by the Princess in *Love's Labor's Lost*.

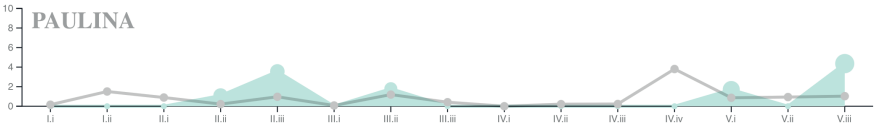


Fig. 7: Percentage of influence contributed by Paulina in *The Winter's Tale*.

husband Egeon. In the absence of her husband, Emilia has risen to a position of great influence. Her role as Lady Abbess allows her to protect Antipholus and Dromio of Ephesus and to argue their case in front of the Duke. Examining the text closely, we see that Emilia not only impacts many characters in this scene, but also that her words themselves indicate a power and control over the narrative that her husband does not have. Although Egeon has been arrested by the Duke, the highest political power in the city, upon seeing him the Lady Abbess immediately declares that "whoever bound him, [she] will loose his bonds, / and gain a husband by his liberty" (V.i. 339–340). Even after she has reunited with her husband, Emilia's speech continues to influence the movements of the other characters, moving them off-stage to a feast she organizes. The Lady Abbess serves as an example of another way in which a secondary female character may exert considerable influence over a play's narrative. In particular, her speech pattern shows that female characters must not necessarily build influence throughout a play by slowly building their network. They may enter the narrative as powerful characters. The Lady Abbess also demonstrates that married women may exert influence; although she gains her status in the absence of her husband, Emilia maintains her power even once they are reunited.

There are only two exceptions in which the woman identified as the most impactful female character in a play does not speak in its final scene. The first is *All's Well That Ends Well*, in which Helen speaks during the last scene of the play but not the Epilogue. This is merely a technical exception. The second is *A Midsummer Night's Dream*. Only Titania and Hippolyta speak in the last scene of *A Midsummer Night's Dream* and both characters only have minor speaking roles, although the scene itself constitutes a large proportion of the play.

The influence measurements for the plays' most impactful women show that many female characters not only retain status throughout their play, but, in fact,

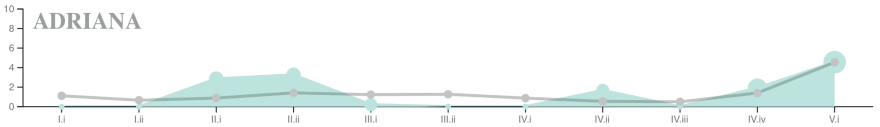


Fig. 8: Percentage of influence contributed by Adriana in *The Comedy of Errors*.

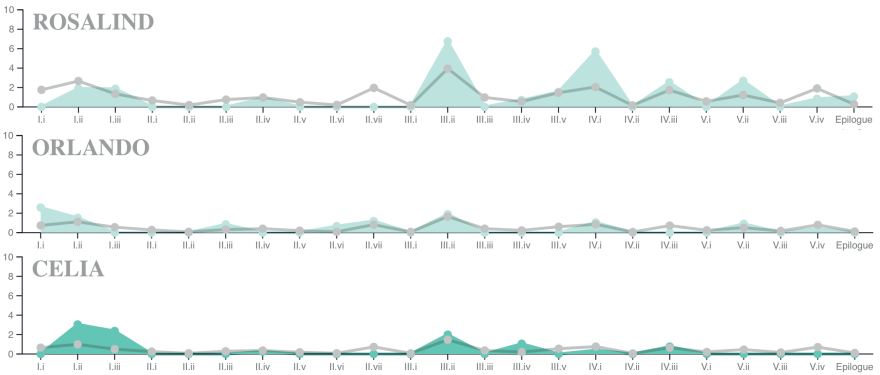


Fig. 9: Percentage of words spoken by Rosalind, Orlando, and Celia in *As You Like It*.

gain influence. These women speak with more characters in each scene they appear in. This trend is particularly obvious for the Princess in *Love's Labor's Lost* (Figure 6), Paulina in *The Winter's Tale* (Figure 7), and Adriana in *The Comedy of Errors* (Figure 8). It indicates that female characters are involved in the increasing interaction between separate cliques during the play's progression towards the "final 'resolution' [scene], which [brings] together the full cast" (Evalyn et al. 2018, p. 369). We see that women are not sidelined during this process but are instead often agents of the crossover, speaking to members of different cliques and creating bonds between the separate social groups. The particularly large number of characters addressed by many of these women in the final scene of their play, as depicted by the considerable size of the final point in some of the trend lines highlighted earlier, means that these women are likely not only involved in drawing the cliques together but also in facilitating the final resolution of the play. Specific examples from the plays seem to confirm this intuition provided by the visualizations. At the end of *As You Like It* Rosalind coordinates each of the successful marriages, Paulina in *The Winter's Tale* reunites Hermione with her family, and the Princess in *Love's Labor's Lost* arranges the ladies' departure from the play and the plan to make the King and his men prove their devotion.

The comedies also permit women to gain power by allowing for gender transgression. By crossdressing, characters like Rosalind, Portia, Nerissa, Jessica, Vi-

ola, and Julia are able to take on a man's role and thereby access forms of power women are otherwise excluded from in a patriarchy. Rosalind and Portia are the most prolific speakers in their respective plays by speech and influence and Viola and Julia are the most prolific female characters and the third most prolific characters overall in their plays, again by both speech and influence. The effect of this can be seen in the visualizations for those plays, particularly that for *As You Like It* (Figure 9). Celia speaks more than Rosalind in both scenes they appear in before they enter the forest of Arden and before Rosalind becomes Ganymede, but Rosalind outspeaks Celia by a significant amount in every subsequent scene except for one. She also speaks far more than her romantic partner, Orlando. Once she is dressed as Ganymede, Rosalind speaks at least double, and sometimes as much as triple, the percentage of words Orlando does in all but one scene. Entering into the comparative freedom of Arden and a disguise as a man allows Rosalind to interact with other characters in capacities she was not able to as a woman. In particular, acting as Ganymede allows Rosalind to escape her uncle's court safely and ultimately permits her to orchestrate her marriage to Orlando and Silvius's marriage to Phoebe by giving her the societal clout to give orders that are followed and to speak freely without needing to adhere to the rules of feminine propriety.

In fact, it appears that dressing as men allows each of these female characters to pursue their happy endings: Portia and Nerissa win the legal case that protects their husbands' friend, Jessica runs away with Lorenzo, Viola survives as an orphaned woman alone and enters Orsino's service, and Julia follows Proteus and acts as his page, allowing them to eventually reunite as lovers after Proteus's reformation. Additionally, each play that features a cross-dressed woman appears in the top-half of the comedies when they are ranked by female characters' contribution to impact (*Twelfth Night*: #2, *As You Like It*: #3, *The Two Gentlemen of Verona*: #5, *The Merchant of Venice*: #7). Interestingly, examining the individual visualizations for these plays reveals that many of the female characters who dress as men are relatively verbose speakers even before they disguise themselves. A possible interpretation of this trend is that even already powerful female characters must access influence that is withheld from them as women in order to fully achieve their goals. This pattern of prominent women who cross-dress demonstrates another way in which the conventions of the comedies allow women to exert influence.

Additionally, some interesting trends appear when we broadly compare female and male speaking characters in the comedies. Each of the plays features more, usually considerably more, male than female speaking characters. However, in half of the plays the women on average speak more and have a larger impact. Many of the extra male characters seem to be minor characters with only very small speaking roles. The worlds of the comedies appear to be populated

primarily with men, but the female characters who do feature usually contribute significantly to the plays.

4 Conclusion

In this study, we produce visualizations of characters' speech in Shakespeare's comedies. These visualizations demonstrate that the women of the comedies are prominent characters with influence over their own narrative trajectories, perhaps influenced by the comedies' social structures and thematic foci. However, it is important to note that there are limitations to the definition of influence we use throughout this paper. Because our metric only takes into account the volume of speech, and not its content, it is impossible to see *how* female characters are interacting with the plays' narratives, whether passivity or activity is suggested by their language, and whether the ways in which they interact with other characters change after marriage. Yet, our metric still provides us with a sense of when and how much characters are participating in a play's dialogue and action, which allows us to examine some subtle and informal ways in which women may impact a play.

Overall, the visualizations suggest that the comedies' tendency to feature more localized drama allows female characters, whose influence is often built through social rapport with closer relations like friends, family, servants, and employers instead of militaristic or political power, to become influential members of the ensemble a comedy focuses on. This influential involvement in social cliques can be seen in the 'spikes' in graphs of women's speech, as these spikes often align with major scenes in a play. This indicates that female characters are involved in the plays' important moments. The visualizations also show that women usually continue to speak and take part in a play's action through the last scene, often gaining influence and helping to orchestrate a play's final resolution. The influence we see female characters exerting provides evidence against the argument that the comedies usually end with women subdued and reincorporated powerlessly into appropriately patriarchal structures. Even if women's trajectories end with a marriage, their visible influence on the plot up to that moment demonstrates that they have helped to shape the path that led them to this marriage. That is, even if marriage represents a return to normativity, the women have been able to impact how this return occurs. It is especially interesting to note that, because these plays are comedies, the endings women negotiate for themselves are frequently happy. While it is in many ways unsatisfying to discover that female characters most heavily impact localized narratives, this study does

reveal that the female characters in these plays frequently hold major, impactful speaking roles and have access to avenues of power which are otherwise often witheld from them.

5 Future Work

This work could be continued and expanded on in many fruitful ways. Without any further annotation or coding, these visualizations could be used to identify and analyze commonalities in speech patterns. Even during a cursory examination of the visualizations interesting patterns appear. For example, in many of the plays there are two women who nearly always appear as a pair. Sometimes these women are related (Rosalind and Celia in *As You Like It*), sometimes they are friends (Mistress Page and Mistress Ford in *The Merry Wives of Windsor*), and sometimes they are a lady and her maid (Portia and Nerissa in *The Merchant of Venice*). Studying how these pairs of women interact could lead to new insights, and with further analysis more interesting trends are likely to appear. Other interesting avenues for future study include examining who characters speak to in the comedies and comparing the comedies with the tragedies and histories.

Acknowledgment: We would like to thank Professor George Shuffelton at Carleton College for acting as primary advisor in the initial stages of this project and Professors Pierre Hecker and Eric Alexander at Carleton College for their support and advice throughout its creation.

Data and Code: Annotations and processing code are available at https://github.com/rmatouschekh/shakespeare_comedy_annotations, the visualizations used can be found at https://observablehq.com/@rhicke/let-every-word-weigh-heavy-of-her-worth.

References

Algee-Hewitt, Mark (2017). "Distributed Character: Quantitative Models of the English Stage, 1550–1900." In: *New Literary History* 48 (4), pp. 751–782. URL: https://doi.org/10.1353/nlh.2017.0038.

Callaghan, Dympna (2016). "Introduction." In: *A Feminist Companion to Shakespeare*. Ed. by Dympna Callaghan. 2nd ed. Wiley Blackwell, pp. 1–18.

Dash, Irene G. (1981). "Introduction." In: *Wooing, Wedding, and Power: Women in Shakespeare's Plays*. Ed. by Irene G. Dash. Columbia University Press, pp. 1–6.

Evalyn, Lawrence, Susan Gauch, and Manisha Shukla (2018). "Analyzing Social Networks of XML Plays: Exploring Shakespeare's Genres." In: *Digital Humanities 2018. Puentes–Bridges. Book of Abstracts*. Pp. 368–371.

Gay, Penny (1994). "Introduction." In: *As She Likes It: Shakespeare's Unruly Women*. Ed. by Penny Gay. Taylor & Francis Group, pp. 1–16.

Hinrichs, Uta, Stefania Forlini, and Bridget Moynihan (2019). "In Defense of Sandcastles: Research Thinking through Visualization in Digital Humanities." In: *Digital Scholarship in the Humanities* 34, pp. i80–i99. URL: https://doi.org/10.1093/llc/fqy051.

Masías, Victor Hugo, Paula Baldwin, Sigifredo Laengle, Augusto Vargas, and Fernando A. Crespo (2017). "Exploring the Prominence of Romeo and Juliet's Characters Using Weighted Centrality Measures." In: *Digital Scholarship in the Humanities* 32 (4), pp. 837–858. URL: https://doi.org/10.1093/llc/fqw029.

Murphy, Sean, Dawn Archer, and Jane Demmen (2020). "Mapping the Links between Gender, Status and Genre in Shakespeare's Plays." In: *Language and Literature* 29 (3), pp. 223–245. URL: https://doi.org/10.1177/0963947020949438.

Pitt, Angela (1981). *Shakespeare's Women*. David & Charles.

Rackin, Phyllis (2005). *Shakespeare and Women*. Oxford University Press.

Stiller, James and Matthew Hudson (2005). "Weak Links and Scene Cliques Within the Small World of Shakespeare." In: *Journal of Cultural and Evolutionary Psychology* 3, pp. 57–73. URL: https://doi.org/10.1556/JCEP.3.2005.1.4.

David Mareček, Marie Nováková, Klára Vosecká, Josef Doležal, Tomáš Musil, and Rudolf Rosa

Annotation and Automated Classification of Dramatic Situations

Abstract: In this paper, we describe the process of annotation of dramatic situations on 52 selected play scripts. We updated the list of the original Polti's 36 situations and use 58 situations in total. We show that the task of selection and annotation of dramatic situations is not easy, as annotators often disagree. In the second part of the paper, we propose and evaluate a method of automatic detection of 5 selected dramatic situations in the play scripts.

1 Introduction

Automatic analysis and generation of literary works is a big challenge in the field of Natural Language Processing. The main problem is understanding the plot and motivations of the characters.

In the THEaiTRE project (Section 2), we aim to automatically interactively generate theater play scripts in the Czech language using an artificial intelligence system based on GPT-2 neural network. As a part of this project, we focused on identifying dramatic situations in the play scripts, as the dramatic situation is considered the building block of drama (Section 3). Our intention was to develop a generation process that operates on the level of dramatic situations instead of individual script lines.

As we have not been able to find any computationally usable data resource containing annotations of dramatic situations in scripts, we decided to create such a corpus, which could serve for the training of tools for the recognition and generation of specific dramatic situations. The theatrical experts from the Academy of Performing Arts in Prague (DAMU), managed to mark dramatic situations in 52 play scripts (Section 4). Devising a list of dramatic situations to use was also an inherent part of this process. The team of annotators consisted of Marie Nováková as the lead, Klára Vosecká and Josef Doležal; each member selected plays to annotate on their own, but the repertory of situations to use was discussed and agreed upon by the whole team.

David Mareček, Tomáš Musil, Rudolf Rosa, Charles University, Prague
Marie Nováková, "Tygr v tísni" theatre, Prague
Klára Vosecká, Josef Doležal, The Academy of Performing Arts in Prague, Prague

https://doi.org/10.1515/9783111071824-006

We decided to look at the concept of a *well-made play* which derives from nineteenth century tendencies towards realism and is a traditional model of play construction in the West. The well-made play is built on the model of Aristotle's ideal Greek tragedy outlined in his Poetics, involving a tight plot and a climax that takes place close to the end of the play. Another important aspect of these works is that they are written in prose and not in verse. This eliminated a large number of works that pre-date authors such as Chekhov or Ibsen.

Subsequently, we used the resulting corpus for training models that would automatically identify dramatic situations in scripts (Section 6). Since the results are not very satisfactory, we think many more annotated plays would be necessary for successful training.

2 The THEaiTRE Project

THEaiTRE (www.theaitre.com) is an interdisciplinary project directly combining theater and computer science. The project team is led by Rudolf Rosa, an expert on computational linguistics and natural language processing; however, the team is composed of both computational linguists as well as theater experts, under the leadership of Daniel Hrbek, the director of the Švanda Theater in Prague. The goal of the project is to explore the potential of current artificial intelligence techniques to be incorporated into theater practice and to directly confront the general public with the outcome while explaining the process behind the creation of the play and thus educating the audience about the current state and capabilities of the techniques used. We found that this immersive experience can spark a lot of interest both among artificial intelligence enthusiasts as well as among people generally ignorant of the current developments in the field. Our goal is to demystify AI, to explain the rather simple basic principles behind machine learning based on textual training data, and hopefully to bring the AI closer to people so that they neither glorify it nor fear it irrationally, but rather understand its basic principles and thus set reasonable expectations and precautions, making it clear that AI should be approached with caution, but can be very useful for some tasks in practice. The project bears some similarities to the theater play *Lifestyle of the Richard and Family* (Helper 2018), the musical *Beyond the Fence* (Colton et al. 2016), the short movie *Sunspring* (Benjamin 2016) or the performances of the Improbotics theater group (Mathewson and Mirowski 2018).

3 Dramatic Situations

A dramatic situation leads characters to necessary and immediate action. For our needs, we are mostly interested in classifications and abstractions over theater play scripts or their parts. In the field of theater studies, there is a vast amount of research on the structure and interpretation of theater plays. Unfortunately, the results of such research are not made available in forms and formats that would easily allow us to use these as data and annotations in machine learning approaches.

The *Thirty-Six Dramatic Situations* by Polti ([1895] 1921),[1] originally proposed in 1895, is a classic work which is highly respected in theater studies. The author presented a supposedly ultimate list of all categories of possible dramatic situations that can occur in a theater play (e.g., "adultery" or "conflict with a god"), further subclassified into 323 situational possibilities. While being a good starting point, this list seems somewhat outdated for today's plays, as Polti based his list primarily on an analysis of classical Greek texts, as well as some classical French works.

Although not directly related to theater plays, the work of Propp ([1928] 1968) is also essential. Propp analyzed Russian folk tales and identified 31 *functions*, similar to Polti's situations but somewhat more down-to-earth (e.g., "villainy" or "wedding"), as well as 7 abstract character types (e.g., "villain" or "hero") and other abstractions.

Polti's and Propp's categorizations are sometimes used in analyzing and generating narratives, although typically not in drama. The works closest to our focus are probably that of Gervás et al. (2016) or Lombardo et al. (2018), who devised an ontology of abstractions for annotating scripts, based on both of the mentioned works, as well as on more recent plot categorization studies (Booker 2004; Tobias 2011).

There are also works producing drama analyses in the form of networks, capturing various relations between the characters in the play (Moretti 2014; Fischer et al. 2019; Horstmann 2019).

4 Annotation Process

The annotation process consisted of continual joint annotation of play scripts with dramatic situations and refinement of the repertoire of situations to use for the an-

1 https://en.wikipedia.org/wiki/The_Thirty-Six_Dramatic_Situations.

Sean: Would you like some music – shall I put a cassette on? I've got a few cassettes? What would you like?
Sean starts looking through some cassettes in a box.
Sean: Madness. Everybody likes Madness. Do you like Madness?
Lisa: I don't know. I've never heard them before.
Sean: Never heard Madness. Everybody's heard Madness. Baggy Trousers – da da da da.
Lisa: I don't know it. Put it on if you want.
Sean: Maybe not. It's maybe a bit lively. Gary Numan.
Lisa: I really don't mind.
Lisa gets up, goes over to him.
Sean: I don't know what people like these days.
[BEGIN Interrogation]
Lisa: The woman in the photograph – on the windowsill? Who is she?
Sean: She's – my ex-wife. She's – quite a long time ago.
Lisa: Is that your boy in the picture?
Sean: Yeah. That's why I have the picture out. Because it's a picture of him.
Lisa: Do you still see them?
Sean: No.
Lisa: That's a shame.
Sean: Yeah.
Lisa: What's his name?
Sean: Conor.
[END Interrogation]

Fig. 1: An example of dramatic situation "Interrogation" in a play script.

notation. The plays were annotated by three experts on drama from the Academy of Performing Arts in Prague.

As for the selection of plays to annotate, we left it up to the annotators. They themselves chose play scripts that they knew well, or which they even authored. We justify this decision with the fact that a play script itself does not directly capture the feelings or intentions of the individual characters. Therefore, it may have many different readings, often contradictory. In addition, the plays that the authors know can be annotated much faster, and therefore much more annotations may be produced in the same time. The disadvantage is that such annotations can be heavily biased compared to annotations made by people who do not know the play well. However, such cases do not occur in our data. All the plays annotated are well-known by their annotators.

The annotators primarily chose contemporary theater or plays that provided a relatively realistic depiction of human communication. They excluded highly stylized text, verse, monologue, or absurd drama, because these work with metaphor, figurative language, puns, and complex references, involving a breakdown of words where artificial intelligence would have difficulties recognizing individual situations.

For annotation purposes, we define a dramatic situation as a continuous section of the play script. The annotators named each situation (based on a list of

situations we had previously compiled together), marking the beginning and end of each one. See the example in Figure 1.

Naturally, there may be sections of the play scripts not annotated by any dramatic situations. There are several instances in a play that are "nondramatic." This includes *didaskalia* and specific information or description of a setting or a character. A scene might be important for our understanding of a character or situation but not in and of itself dramatic. These "non-situations" are present throughout and are often necessary to ensure the rhythm of the text. If we look back at the Greeks, their plays are purely dramatic in the sense that each block of text can be identified as a dramatic situation. Polti's list of dramatic situations alone – influenced by the Greeks – isn't an adequate tool for analyzing contemporary plays.

The annotation is line-oriented, so the situations cannot begin or end in the middle of a line. In inevitable cases in which the situation changes in the middle of the line, the annotator is asked to split the line into two. Also, the situations cannot intersect each other, however, it is possible that the same section in the script may capture multiple different situations. Typically, the situation may differ from the point of view of individual characters. In these cases, annotators are allowed to annotate the block with multiple situations. However, the situation in the first position should be the most suitable one.

Dramatic situations typically do not cross scene boundaries. However, it is not forbidden and it is possible that one situation spans two scenes.[2]

The annotators didn't use any specific software. They simply got a preprocessed text file and added new lines into them marking the beginnings and the ends of situations. These added lines were formatted so that they can be automatically differentiated from the standard lines of the play scripts.

4.1 First Phase – Forming the Set of Situations

We took Polti's 36 situations as the starting point. While applying them to contemporary and realistic (well-made) theater plays, our suspicion that the list needs updating was confirmed. Not only did the list contain situations that are highly unlikely to appear in a modern dramatic text (such as *Slaying of kin unrecognized*), but it was missing a number of situations which contemporary authors use frequently (e.g., *Interrogation*).

2 The way how plays are divided into scenes varies greatly. Some plays are not structured into scenes at all, another plays are structured into a lot of very short scenes. Therefore, we decided not to tie scenes and annotated situations in any way.

Tab. 1: List of the original Polti's 36 dramatic situations and our modifications.

Supplication	~~Slaying of kin unrecognized~~
Deliverance	Self-sacrifice for an ideal
Crime pursued by vengeance	Self-sacrifice for kin
Vengeance ~~taken for kin upon kin~~	~~All sacrificed for passion~~
Pursuit	Necessity of sacrificing loved ones
Disaster	~~Rivalry of superior vs. inferior~~
Falling prey to cruelty/misfortune	Adultery
Revolt	Crimes of love
Daring enterprise	Discovery of the dishonour of a loved one
Abduction	Obstacles to love
The enigma	An enemy loved
Obtaining	Ambition
Enmity ~~of kin~~	Conflict with a god
Rivalry ~~of kin~~	~~Mistaken jealousy~~
~~Murderous adultery~~	Erroneous judgment
Madness	Remorse
~~Fatal imprudence~~	Recovery of a lost one
Involuntary crimes of love	Loss of loved ones

Tab. 2: List of added situations.

Seduction	Accusation	Breaking the taboo	Murder
Intruder	Fight	Jealousy	Curse
Interrogation	Humiliation	Reconciliation	Extortion
Revelation	Bad news	Succumb	Betrayal
Parting	Fear	Passing	Rape
Admission	Ruse-trap-fraud	Break up	Confession
Intimidation	Unfulfilled desire	Bad premonition	Capture

In the first phase of annotation, in which we annotated 19 play scripts, we removed six obsolete situations, update the list as shown in Table 1 and introduced another 28 situations (Table 2) that might not be as archetypal as Polti's, but better reflect the kind of texts we were working with.

Eventually, we arrived at a list of 58 situations. We are still perfecting the list, trying to merge some of the situations so that each situation is sufficiently supported by the annotated data. Our repertory of situations is thus based on a combination of theatrological knowledge and statistical data processing.

4.2 Evaluation and Inter-Annotator Agreement

To evaluate the quality of the annotations and agreement between the annotators, we selected two plays from the first phase to be annotated by all three annotators. The evaluation of the agreement is not straightforward, because even if the annotators agree on the type of situation for a particular part of the script, they usually differ in marking the exact place where the situation begins or ends. To mitigate these disagreements, we compute the agreement score in a line-by-line fashion: each line of the script either belongs to one specific situation or does not belong to any situation. The inter-annotator agreement is then computed as the number of lines in the script that were annotated with the same situation divided by the total number of lines.

If we measure agreement on the full set of 58 situations, the average agreement between pairs of annotators is 60 %. If we count only lines where all three annotators agreed, the score is 47 %. We also computed Fleiss' kappa for assessing the reliability of the agreement between three annotators. It reached 54 %, which may be considered a moderate agreement. Therefore, we can say that determining dramatic situations is difficult even for humans, as the annotators often could not agree with each other.

The confusion matrix of the annotated situations for the play *Věc Makropulos* is shown in Figure 2.

This analysis allowed us to see the differences between annotators' individual interpretations. We noticed that they fundamentally depend on the perspective from which the annotators had read the text or the character through whom they had viewed the situation. They often annotated a single situation under two different names: for example, "Interrogation" vs. "Admission" or "Daring Enterprise" vs. "Madness." This was not because they were unable to decide, but precisely because of the subjectiveness of interpretation through a specific character. If we read a situation in which character X asks questions and character Y answers under pressure, we can perceive it from the perspective of character X, where we describe the situation as Interrogation, or from the perspective of character Y, where we can describe it as "Admission." Other competing situations are "Interrogation" vs. "Revelation," or "Daring Enterprise" vs. "Fight" vs. "Ambition."

All of these disagreements show that developing a set of situation labels that could be clearly assigned to the situations in the play scripts is really challenging.

A possible way of dealing with this problem could be to organize the types of situations hierarchically. We leave that for future work.

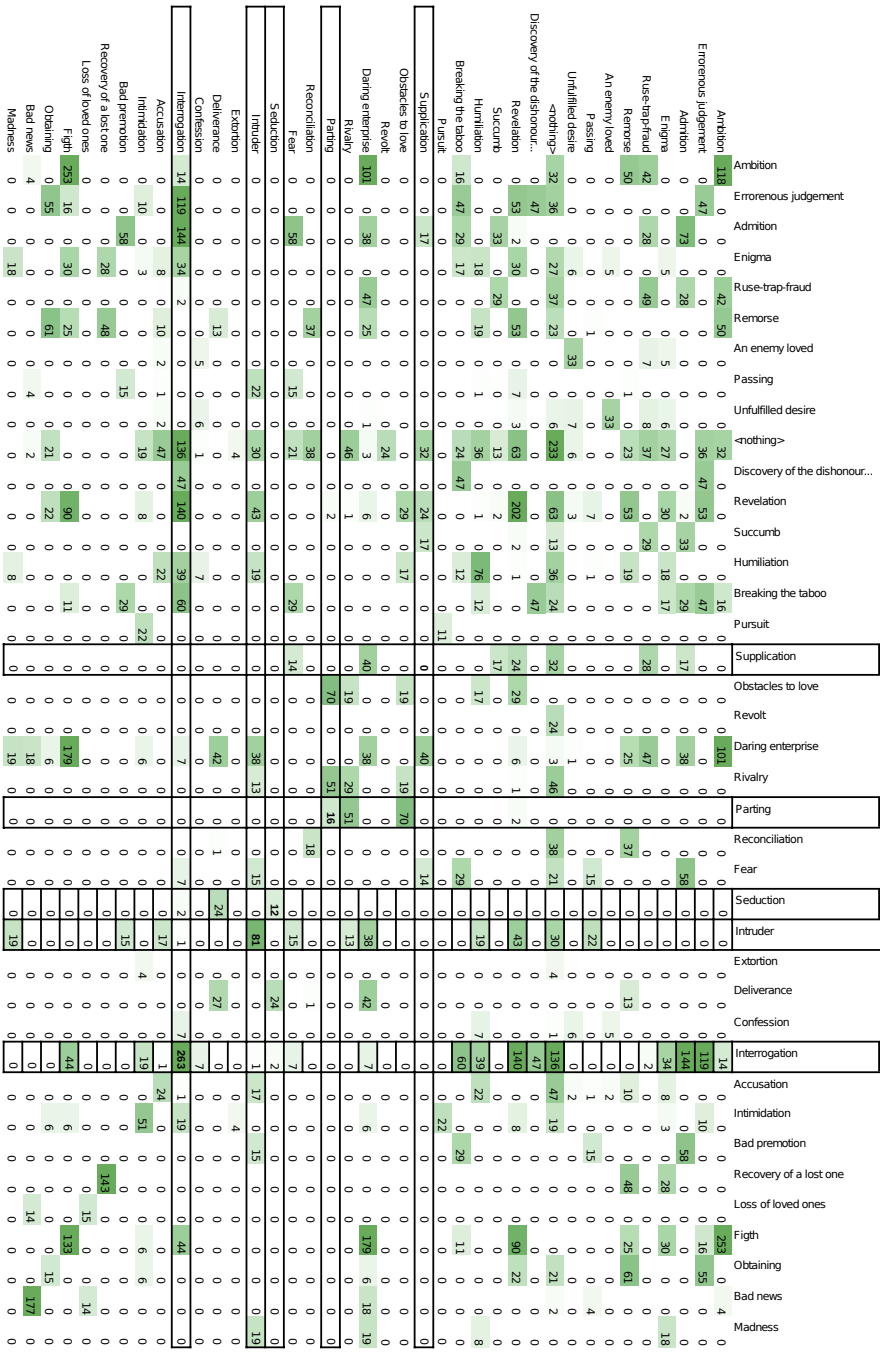


Fig. 2: Confusion matrix for the inter-annotator agreement on the play *Věc Makropulos*. The number of lines where the pair of annotators agreed/disagreed.

Tab. 3: Average pair-wise annotator agreement on the five selected situations. The percentage shows the average agreement, i.e., for example number of lines annotated by "Interrogation" by both the annotators divided by the number of lines annotated by "Interrogation" by one annotator. 100 % agreement means that all the lines annotated by one annotator were just the lines annotated by the other one.

Situation	Agreement
Interrogation	98 %
Intruder	72 %
Parting	63 %
Supplication	63 %
Seduction	59 %

4.3 Second Phase – Five Selected Situations

The play annotation phase was enormously lengthy. During the process, we understood that it is impossible to collect as much data as needed, because the analysis of one play takes several hours on average.

To make the annotations both more efficient and more useful, in the second phase, we decided to focus on a few frequent situations which we found to be the easiest to decipher in the scripts as the annotator agreement was high on them: *Supplication*, *Intruder*, *Seduction*, *Parting* and *Interrogation*. These situations appear frequently and regularly in plays and have stylistic characteristics and linguistic commonalities.

If we only measure the inter-annotator agreement on the five selected situations, the average agreement between a pair of annotators is 69.7 %; Table 3 details the annotator agreement for each of the five selected situations.

This significantly sped up the annotation work, because the annotators could only annotate plays in which they were sure this type of situation would be abundantly represented.[3] Over time we developed a manual to isolate these five situations, using typical words, scene notes, and typical dialogue structure.

We annotated another 33 play scripts with only these 5 situations; approximately 20 % of each script can be categorized into these five situations, leaving 80 % of each script unmarked on average.

The reason for a higher annotator agreement with these particular situations was that they could often be codified by either a particular word, symbol, or stage direction. An example of this could be that a situation is most likely to be an *in-*

3 For example, in Molière's *Don Juan* we would find many situations of Seduction, in Chekhov's *Cherry Orchard* there would be situations of Parting, etc.

terrogation if one character is asking questions (resulting in a number of question marks) while the other is answering. *Parting* could be suggested by a character leaving, saying "farewell" or waving goodbye. This method obviously cannot be applied across the board, as dramatic situations are far more nuanced, however, it offered us a pattern that could be studied.

5 The Annotated Corpus

We eventually annotated 52 play scripts (consisting of 625 739 words in total) with a set of 58 dramatic situations. One play is written in English, all the other plays are in Czech.

The annotations were then sanitized and converted to a JSON data format so that they are easy to process and can be utilized for further experiments. Each file in the dataset includes one play script annotated by one annotator and contains the following items:

- the title of the play,
- the author of the play,
- an array of acts. Each act is an array of scenes. Each scene is an array of contents. Content may be:
  - a line described by a character (speaker) and a text,
  - a stage direction,
  - a mark showing the beginning or the end of a dramatic situation. It is structured as follows:
    - id of the situation,
    - the situation label,
    - an array of other alternative labels of this situation.
    - For some situations, we also annotated a short synopsis.

Due to copyright, it is not possible to publish the entire dataset at this time, so we are releasing only part of the annotated data. The newest release consists of nine older play scripts that can be already freely distributed and ten plays for which we succeeded to obtain a license with the right to make them publicly available for non-commercial use in artificial intelligence research only. One play released is annotated by all three annotators. You can find the statistics of the situations annotated in this portion in the last column of Table 4. The data can be downloaded from the LINDAT repository.[4]

4 http://hdl.handle.net/11234/1-4930.

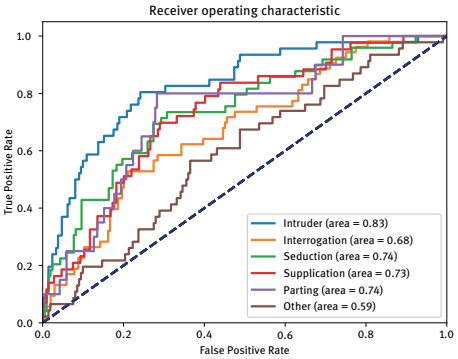


Fig. 3: ROC curves for the classifiers for particular type of situation.

6 Experiments in Predicting the Situations

The purpose of creating these annotations was to detect them automatically in a much larger set of play scripts. Such bigger datasets could be further analyzed and used for the generation of play scripts with given dramatic situations.

In preliminary experiments, we first simplified the task of detecting a dramatic situation by entering a section of the script that we know corresponds to a situation with the goal to assign it. The task is therefore simplified by skipping the search for the beginning and the end of the situation (i.e., the boundaries of the situation).

On the created data, the trivial solution (always assigning the most common type of the situation, which is *Seduction*) has an accuracy of 26 %. To solve the task, we tried to use various basic and advanced methods of data processing and machine learning (tokenization, bag of words, stopwords, anonymization, leave-one-out filtering, frequency-based filtering, TF.IDF, word embeddings, Naive Bayes, Support Vector Machine, Multilayer Perceptron, and others). We tried many variants of the procedure, but the highest accuracy we were able to achieve in this way was 42.97 % with the following pipeline: for each situation, we created a vector of word-counts (ignoring character names), the counts were weighted with TF-IDF and then classified with a linear Support Vector Machine optimized with stochastic gradient descent (weighted average precision was 47 %, recall 44 %, F1 score 0.44). The training and testing was performed as cross-validation, where for each play, the leave-out set was the situations coming from that play.

The *Intruder* situation was the easiest to classify, the area under the curve (AUC) reached 0.83. It is followed by *Seduction* and *Parting* (0.74 AUC), *Supplication* (0.73 AUC), and *Interrogation* (0.68 AUC). See Figure 3 for the ROC curves.

Such accuracy is too low, the tool thus returns the wrong answer in most cases. Moreover, it is the accuracy of an artificially simplified task with knowledge of the boundaries of the situations, which is impossible to assume for the actual data. Therefore, part of the task must also include searching for the boundaries of the situations, which would reduce the accuracy further, probably very significantly.

One of our hypotheses for the bad result was based on the observation that the annotators often characterize a relatively long section of the script as a dramatic situation. However, in such cases, a part of the marked section is rather a kind of lead-in for the given situation, from which the actual situation is not yet well recognizable, and only at the end of the situation does it gradate to such extent that it is obvious which kind of situation it is (typically, the dynamics are rising in this way, with the exception of the *Intruder* situation which, on the contrary, is usually the strongest at its beginning, when the intruder appears). That is why, together with the annotators, we carried out a further revision of the annotations, where the annotators added emphasis on the core of the situation in which it is strongest. Subsequent experiments using such enriched annotations, however, only led to a slight improvement in the results.

7 Conclusion

We believe that the task of identifying dramatic situations is considerably more difficult than we originally estimated. We relied on theatrical theories and insights which say that a play takes place in dramatic situations, that the author, the director, and the actor think in such situations, and thus it is a concept, which is well established and usually universally understood and shared. Until now, however, it has been more of a tradition based on theory and introspection, which has never been empirically verified on a larger number of scripts and methodically evaluated.

Experiments carried out by us have shown that even if limited to 5 selected situations, the paired inter-annotator agreement is only 70 %, which suggests that the concept of a dramatic situation is not nearly as straightforward and obvious as the theory claims.

Nevertheless, the experiments in the automatic identification of the situations fall far below this theoretically achievable limit. We conclude that this is mainly due to the too small amount of available training data, and we assume that for a task that proved to be so difficult, a significantly larger amount of data (certainly at least ten times, probably more than 100 times the current data) would be required, which is not feasible within this project.

We still consider the concept of dramatic situations to be substantial and potentially useful, but we believe that for any meaningful automatic work with them it would be necessary to create significantly larger annotated data.

Subsequently, we have run preliminary experiments on automatically classifying the dramatic situations in the scripts. We have trained a multi-layer perceptron classifier on the annotated data, obtaining slightly promising results.

Funding: The project TL03000348 *THEaiTRE: Umělá inteligence autorem divadelní hry* is co-financed with the state support of Technological Agency of the Czech Republic within the ETA 3 Programme. The work described herein has been using data, tools and services provided by the LINDAT/CLARIAH-CZ Research Infrastructure (https://lindat.cz), supported by the Ministry of Education, Youth and Sports of the Czech Republic (Project No. LM2018101).

References

Benjamin (2016). *Sunspring*. Short film.

Booker, Christopher (2004). *The Seven Basic Plots: Why We Tell Stories*. A&C Black.

Colton, Simon, Maria Teresa Llano, Rose Hepworth, John Charnley, Catherine V. Gale, Archie Baron, François Pachet, Pierre Roy, Pablo Gervás, Nick Collins, Bob Sturm, Tillman Weyde, Daniel Wolff, and James Robert Lloyd (2016). "The Beyond the Fence musical and Computer Says Show documentary." In: *Proceedings of the Seventh International Conference on Computational Creativity*. URL: https://www.computationalcreativity.net/iccc2016/proceedings-2016/.

Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling, and Peer Trilcke (2019). "Programmable Corpora. Die digitale Literaturwissenschaft zwischen Forschung und Infrastruktur am Beispiel von DraCor." In: *DHd 2019 Digital Humanities: multimedial & multimodal. Konferenzabstracts*, pp. 194–197. DOI: 10.5281/zenodo.2600812.

Gervás, Pablo, Raquel Hervás, Carlos León, and Catherine V Gale (2016). "Annotating musical theatre plots on narrative structure and emotional content." In: *7th Workshop on Computational Models of Narrative (CMN 2016)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Helper, Roslyn (2018). *Lifestyle of the Richard and Family*. Theater Play.

Horstmann, Jan (2019). "DraCor: Drama Corpora Project." In: *forTEXT. Literatur digital erforschen*. URL: https://fortext.net/ressourcen/textsammlungen/dracor-drama-corpora-project.

Lombardo, Vincenzo, Rossana Damiano, and Antonio Pizzo (2018). "Drammar: A comprehensive ontological resource on drama." In: *International Semantic Web Conference*. Springer, pp. 103–118.

Mathewson, Kory W. and Piotr Mirowski (2018). "Improbotics: Exploring the Imitation Game using Machine Intelligence in Improvised Theatre." In: *Fourteenth Artificial Intelligence and Interactive Digital Entertainment Conference*. DOI: 10.1609/aiide.v14i1.13030.

Moretti, Franco (2014). "'Operationalizing': or, the Function of Measurement in Modern Literary Theory." In: *Pamphlets of the Stanford Literary Lab* 6. URL: https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf.
Polti, Georges [1895] (1921). *The Thirty-six Dramatic Situations*. JK Reeve.
Propp, Vladimir [1928] (1968). *Morphology of the Folktale*. University of Texas Press.
Tobias, Ronald B (2011). *20 MASTER Plots: And How to Build Them*. Penguin.

Appendix

Tab. 4: List of situations and their frequencies.

Situation label	Total	Phase 1	Phase 2	Release
seduction	146	37	109	62
intruder	136	50	86	51
interrogation	135	41	94	42
supplication	96	27	69	26
revelation	57	56	1	38
parting	57	18	39	23
admission	49	49	0	15
intimidation	47	44	3	14
accusation	46	37	9	26
fight	45	45	0	22
humiliation	42	42	0	22
bad news	36	36	0	18
fear	34	34	0	15
ruse-trap-fraud	31	31	0	15
enigma	30	29	1	27
unfulfilled desire	27	27	0	20
breaking the taboo	27	27	0	7
revolt	25	25	0	12
daring enterprise	25	25	0	15
remorse	23	23	0	13
jealousy	23	23	0	17
rivalry of kin	22	22	0	11
reconciliation	21	21	0	12
obtaining	20	20	0	11
madness	19	19	0	13
succumb	18	18	0	9
loss of loved ones	17	17	0	13
enmity	17	17	0	10
disaster	16	16	0	4
passing	15	15	0	5
erroneous judgment	15	15	0	8
break up	15	14	1	6
ambition	15	15	0	10
obstacles to love	13	13	0	11
deliverance	13	13	0	6
vengeance	12	12	0	3
falling prey to cruelty/misfortune	12	12	0	5
bad premonition	12	12	0	12
recovery of a lost one	11	11	0	68
pursuit	9	9	0	4
murder	9	9	0	2
conflict with a god	7	7	0	6
the necessary sacrifice of loved ones	6	6	0	4
discovery of the dishonor of a loved one	6	6	0	3
curse	6	6	0	4
extortion	5	4	1	1
betrayal	5	5	0	0
self-sacrifice for kin	4	4	0	3
an enemy loved	4	4	0	3
rape	3	3	0	1
crime pursued by vengeance	3	3	0	1
adultery	3	3	0	2
self-sacrifice for an ideal	2	2	0	1
crime of love	2	2	0	2
confession	2	2	0	2
capture	2	2	0	1
involuntary crimes of love	1	1	0	1
abduction	1	1	0	0

Benjamin Krautter and Janis Pagel

The Schemer in German Drama

Identification and Quantitative Characterization

Abstract: Dramatic characters frequently fill out different role types and act according to traits conventionally attributed to their role. One of these role types is the "schemer," characterized by intervening in a play's main plot and driving forward the plot's main conflicts. In our study, we utilized secondary literature to identify 50 characters as schemers and extracted a wide range of features which are likely to distinguish "schemers" from "non-schemers." Using machine learning, we trained a model to automatically classify characters according to these two classes and performed a number of analyses in order to identify the most contributing features. Our model is able to reliably detect schemers, utilizing features that cover information about stage presence and content of character speech, but exhibits a rather low precision. We show that this can partially be attributed to the heterogeneous nature that characterizes the group of schemers.

1 Introduction

When marquess Marinelli casually mentions that Count Appiani will marry Emilia Galotti on this very day, Prince Hettore Gonzago feels completely "lost, and [...] no longer" able to "live," as he ardently "love[s]" and "adore[s]" Emilia (Lessing 1800, p. 11). The dialogue between the Prince and Marinelli in the first act of Gotthold Ephraim Lessing's tragedy *Emilia Galotti* (1772) discloses one page of the play's dramatic conflict and, at the same time, sparks the further action. To prevent the marriage Marinelli proposes to "devise some scheme" and asks the Prince for his permission: "[W]ill you let me act as I please? Will you approve all I do?" (Lessing 1800, p. 13).

In German literary history, Marinelli is considered a typical representative of the evil schemer (cf. Alt 2004; Martens 1995; Asmuth 2016, p. 126). He is described as a slick, unscrupulous, compliant, cynical, and cunning advisor – a character type that was well established in Lessing's days (cf. Martens 1995, pp. 69–70, 78). Lessing's Marinelli, however, does not primarily act as an instigator that incites the Prince's evil desires by pointing him to the apparent limitlessness of his power. According to Alt (2004, p. 8), he instead embodies a dynamic principle with the

Benjamin Krautter, Janis Pagel, University of Cologne

https://doi.org/10.1515/9783111071824-007

function of organising a silent crime that leaves the originator, i.e., the Prince, in the dark. In the end, his 'scheme' causes a fundamental destabilization of relationships, as it not only separates bride and groom, but also father-in-law and groom, father and mother, mother and daughter, and, lastly, even father and daughter.

Over the course of the early modern period, the role of the schemer has been established as a distinct role type (Rollenfach)[1] in German theater practice (cf. Asmuth 2016, p. 124). The intrigue itself has always been an important element of drama, e.g., regarding Aristotelian concepts such as reversal *(peripeteia)* and recognition *(anagnorisis)*. With the schemer being one of the more central character types to the plot's action, we are, therefore, focusing on automatically identifying and quantitatively analyzing different schemers in German plays from the eighteenth and nineteenth centuries. For our article's undertaking, we can build upon principles that have already been discussed in the field of Computational Literary Studies (CLS) in recent years. Automatically detecting protagonists or main characters in literary texts, for example, has become a more and more reliable task (cf. Moretti 2011; Jannidis et al. 2016; Reiter et al. 2018; Fischer et al. 2019; Krautter and Pagel 2019).

In the 1970s Manfred Pfister (1988, p. 165) proposed a first operationalization to detect the main characters of a play in terms of "[q]uantitative relations of dominance." Pfister intended to distinguish main from secondary characters by using two criteria: namely, "the length of time that a dramatic [character] spends on stage and the extent of its participation in the primary text" (Pfister 1988, p. 165). He stresses, however, that neither do the two criteria he describes have to coincide nor are they perfectly reliable. Moreover, a more precise categorization of the dramatis personae, e.g., according to a character's function, would lack "sophisticated preliminary studies of the 'grammar' of plot," as Pfister (1988, p. 166) has argued.

In our contribution, we will extend existing computational approaches and thus aim at a more subtle differentiation of dramatic characters than Pfister has called for. Doing so, our goal is to automatically detect schemers on the basis of a number of well-established quantitative methods of Natural Language Processing.[2] As schemers are often characterized by their changing roles of acting and observing, it becomes clear that identifying dramatic characters as schemers is a complex and multi-dimensional task. Once a model for classifying individual characters is established, our corpus analytical perspective enables us to analyze

1 The role types are considered to be a central principle for the structure of European theater practice (cf. Detken and Schonlau 2014, p. 7). For an overview on role types, cf. Harris 1992.

2 To do this, we draw on related groundwork published in Krautter et al. (2020, 2022).

overarching developments: how does the dramatis personae change with regard to schemers? Can we observe differences in genre or in period?

To this end, we will cursorily outline the characteristics of schemers from a literary studies perspective in a first step (Section 2). In a second step, we set forth our methodological premises in establishing a data set of 50 schemers in total. This data set serves as training data for our machine learning model (3). Following this, we explain our computational operationalization, which over several steps transmits properties of dramatic characters into measurable indicators through methods of quantitative text analysis (4). These indicators serve as features for our machine learning models. In a series of experiments, we attempt to take a closer look at schemers from a quantitative point of view (5.1). We go on to evaluate and compare individual features to provide insights into the conception of schemers in dramatic texts (5.2) and finally discuss our results (6).

2 The Schemer in Literary Studies

The schemer is probably one of the more heterogeneous role types in the theater practice of the eighteenth and nineteenth centuries. It is also one of the longest lasting role types (cf. Schonlau 2017, p. 180). Within the role type system, the schemer is part of the so-called character role – a role that is specifically tailored to more complex, dynamic, rounded and individual characters.[3] These character roles serve as an addition to the otherwise rather clearly socially defined characters of the role type system, such as fathers, mothers or lovers. Thus, the system includes characters that partially undermine the basic idea of the system to have roles that are easily transferable from one play to another.[4] Therefore, it seems particularly challenging to identify schemers automatically. The schemer's heterogeneity has been discussed as an inherent element of the intrigue itself. The intrigue is understood as a genuine dramatic principle that makes use of social transformations and social rearrangements to bring about change by stimulating action (cf. von Matt 2006, p. 34). Consequently, the schemer can be described as a character of intervention (cf. Alt 2004, pp. 1, 8). For audience and characters alike, the intrigue presupposes different degrees of awareness (cf. Pfister 1988, pp. 49–

3 The corresponding entry in the *Allgemeines Theater-Lexikon* from 1839 gives a historical perspective on character roles in German theater plays (cf. Storch 1839, p. 122).

4 As soon as theater groups started to settle into cities, they had to expand their repertoire of plays to maintain visitor interest. Consequently, the role type system provided certain conventions regarding the different types of roles an actor could play (cf. Harris 1992, pp. 222–225).

57). Typically, the audience in the outer communication system is informed about the plans of the schemer. His victims in the inner communication system, however, do not know about the scheme he is planning, as the schemer's actions are determined by a moment of deception. Accordingly, he can be described as unpredictable, opaque, or ambiguous in his acting (cf. Alt 2004, p. 1).

The intrigue itself is open to a wide variety of manifestations and is also not specific to gender – although the majority of schemers in plays is believed to be male (cf. Schonlau 2017, p. 179). Moreover, in the context of literary history the intrigue is morally indifferent. It can serve good intentions and purposes, but it can just as well trigger the play's catastrophe. Thus, a distinction between destructive and constructive intrigues can be made, which at the same time can delineate different characterizations and strategies of schemers (cf. Matt 2002, 34). While intrigues in satirical comedies oftentimes pretend to respond to a character's mistake, most of the time they are actually directed against the very same character (cf. Steinmetz 1978, p. 42). The tragic intrigue, according to Aristotle's reasoning, demands both surprise and credibility. In this context, the intrigue is essentially not a means of self-empowerment but serves as a tool for destabilization and selection (cf. Alt 2004, p. 3).

One of the most prominent tools of scheming or intriguing machinations is the letter: Peter Pütz describes the letter as a reliable vehicle of intrigue. Firstly, characters can disguise themselves in letters without giving away facial expressions and the sound of their voice. Secondly, a letter can be faked in multiple ways: in relation to the originator, the addressee, and the content. And thirdly, a letter can – unintentionally or intentionally – fall into other hands and, subsequently, be used as means of blackmail (cf. Pütz 1970, p. 82). Looking at the characters, servants can frequently be found as the originators of an intrigue (cf. Eibl 1971, p. 151). Chamberlain Marinelli, for instance, is a prime example of a scheming courtier servant. In addition to generally vicious behavior of schemers, individual passions can serve as a central motivation for intrigue, whether that is envy, hatred, or jealousy (cf. Schonlau 2017, pp. 181–182).

3 Methodology and Corpus

In our article, we follow an approach that can be labeled as a top-down-procedure. More precisely, we make use of secondary literature to establish a data set of schemers. We then go on to use this data set as training data to train a machine learning model. As Pichler and Reiter (2022, p. 9) have put it, such "a proceeding is consistent with an established and widely used research practice in data

sciences. It consists of using pre-existing gold standard data, based on some concept(s), as the training data of a machine learning algorithm."[5] In our case, however, the data set, which contains the 'gold standard data,' must be put together first. To do this, we lean on Franco Moretti's original idea of distant reading. In his essay *Conjectures on World Literature* 2000, Moretti envisages distant reading as a second-hand criticism. In Moretti's controversial view, literary history should be dealt with as a "patchwork of other people's research, *without a single direct textual reading*" (Moretti 2000a, p. 57). For his ambitious focus on world literature, he proposes to skip reading primary literature entirely and instead concentrate on secondary literature or the expertise provided by research networks (Moretti 2005, p. 5). This rather polemical suggestion is an attempt to find an appropriate method for analyzing literary history without depending "on its canonical fraction, which is not even one per cent of published literature" (Moretti 2000a, p. 55; cf. Moretti 2000b, p. 226).

Although Moretti's idea is not highly specified from a methodological point of view, it seems to be useful for our approach of identifying and analyzing literary characters as schemers. As the automatic detection of schemers requires a sound data set to train our computational model, we can take advantage of established research to find the respective characters. In order to constitute our data set of schemers in German plays of the eighteenth and nineteenth centuries, we followed Moretti's suggestion: we did second-hand criticism. We used attributions made by other researchers to label a character as schemer. We searched through handbooks, articles, monographs, literary histories, and encyclopedias to find characters that have been referenced as schemers. Table 5 (see appendix) gives an overview of our data set, which consists of 50 characters that were labeled as schemers.[6] The table lists all the characters, the plays they are part of, and a reference to secondary literature. The 50 schemers in our data set come from a total of 38 different plays, which comprise 848 characters. Our data set, however, only contains positive examples of schemers. As plays can incorporate more than one schemer, we cannot rule out that all characters that are not listed in our data are, in fact, not schemers. For the analysis of the 38 plays, we resort to the *German Drama Corpus* (Fischer et al. 2019).

5 A prominent example of such an approach can be found in Andrew Piper's (2016) article on *Fictionality*.

6 Since our approach does not call for a consensual categorization of the characters, the established data do not reflect ground truth but instead serve as heuristic.

4 Operationalization

To automatically identify schemers in a set of dramatic characters, we first need to operationalize the concept computationally. We understand operationalization as the development of algorithmic routines for the unambiguous detection or measurement of a concept's instantiations in data (cf. Pichler and Reiter 2022, p. 2, 2021, p. 4). This can be done both manually and computationally. As we derive our training data from literary studies' secondary literature, we can do without manually annotating schemers in dramatic texts. Instead, we can focus on our computational operationalization. In literary studies, schemers are oftentimes described according to their function for the plot's action: the schemer, then, is regarded as the character that instigates an intrigue. To automatically identify a schemer as the originator of an intrigue presupposes that we can either computationally find scenes, which include an intrigue, or that all relevant intrigues have been annotated manually. To bypass these requirements, we operationalize schemers indirectly by using several sub-concepts. Most of these sub-concepts, in turn, are not directly measurable, either. Thus, we have to operationalize them first. This incremental process leads to a hierarchical order with directly measurable indicators at the bottom of it.

Figure 1 gives an overview of our operationalization hierarchy. To detect schemers, we employ six sub-concepts to distinguish between different dramatic characters: 'sentiment,' 'aboutness,' 'interaction,' 'stage presence,' 'action' and 'character speech stylistics.'[7] These sub-concepts are further divided into different methods that measure or identify instances of further constricted sub-concepts. Looking at the 'interaction' of characters, for instance, we employ network analysis to create co-presence networks. For every character in our corpus of plays we calculate network metrics such as *degree* or *betweenness centrality*. Indirect operationalizations must always be undertaken with regard to the operationalizability of the sub-concepts. For this, both pragmatic and theoretical aspects are relevant. Consequently, our operationalization of schemers is a compromise between ambition and feasibility. A concept like interaction is obviously not limited to the co-presence of characters. In our case, the edges that are connecting the nodes represent that "two characters are listed as speakers within a given segment of a text (usually a 'scene')" (Trilcke et al. 2015, p. 1). Operationalized in this way, the co-presence of two characters is only one aspect of 'interaction.' It is an approx-

7 Some of the concepts mentioned are influenced by insights from critical discourse analysis, e.g., regarding so-called turn-length and turn-taking or the importance of topics (cf. Bennison 1998, pp. 70–75).

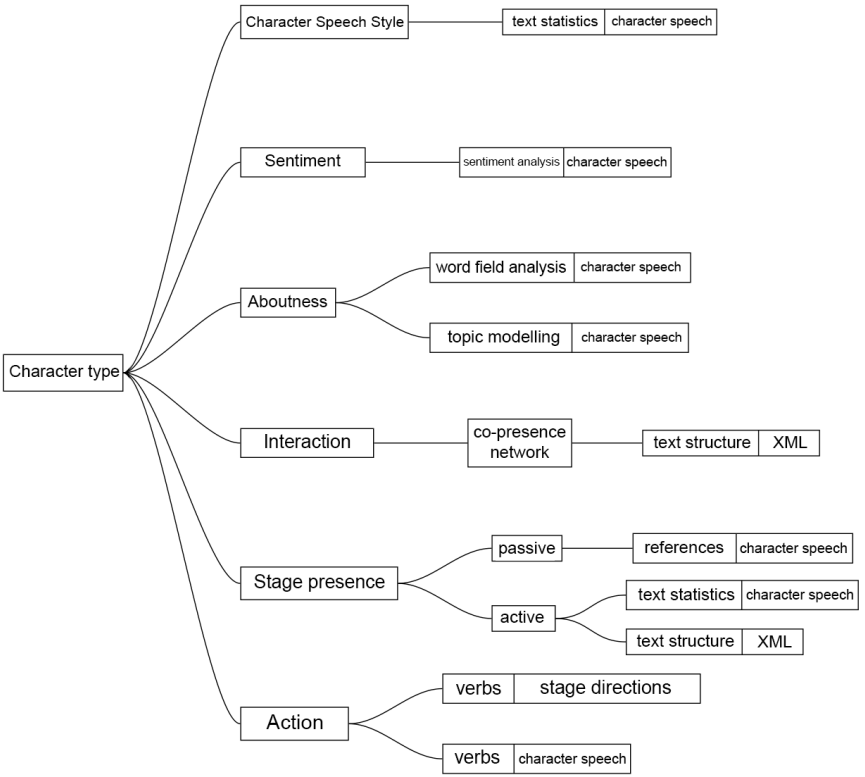


Fig. 1: Computational operationalization of dramatic character types.

imation of the concept of 'interaction.' At the same time, the sub-concepts we employ cannot always be separated sharply. To give an example: the stage presence of dramatic characters obviously has some overlap with the co-presence of characters. There are further questions at the lower end of the operationalization hierarchy to which this article can only give cursory answers. To give an example: The detection of references in (literary) texts, which computational linguistics has been dealing with for many years under the notion of coreference resolution, is far from a trivial problem (cf. Pagel and Reiter 2021; Schröder et al. 2021).

In the following, we explain the indicators, which we subsequently employ as 'features' in our machine learning model, from a technical perspective.

4.1 Features

The features can be divided into several groups based on the methods that are used: 'text statistics,' 'network metrics,' 'stage presence,' 'word fields,' 'topics,' 'action verbs,' 'sentiment,' and 'priors.' In the following, we describe these features and their implementation in our model in greater detail.

4.1.1 Text Statistics

We use several features that describe certain statistics about the character speech:
- *tokens*: the number of tokens a character utters. This feature is normalized by the total number of tokens in a play.
- *utterances*: the number of utterances of a character. An utterance is a span of text that a character speaks before and/or after another character is speaking.
- *utteranceLengthMean*: the mean length of an utterance measured in tokens.
- *utteranceLengthSd*: the standard deviation of a character's utterances.
- *TTR*: the type-token-ratio of a character, which is the number of types divided by the number of tokens in a certain span of text. We opted for a segment length of 200 tokens and averaged the result.

4.1.2 Network Metrics

We create networks by treating characters as nodes and the co-presence of characters in a scene as edges between the nodes (cf. Marcus 1973; Trilcke 2013). We calculate several metrics on these co-presence networks, namely:[8]
- *degree* centrality
- *weighted degree* centrality
- *betweenness* centrality
- *closeness* centrality
- *eigenvector* centrality

4.1.3 Stage Presence

We measure the presence of characters on the stage in four different ways:

8 For a detailed discussion and overview of these metrics, see Wasserman and Faust (1994).

- *active* presence: the number of scenes in which a character is present on stage, divided by the total number of the play's scenes.
- *passive* presence: the number of scenes in which a character is mentioned by name, but not present on stage, divided by the total number of the play's scenes.
- *firstBegin*: the first time a character utters a word, measured in character offsets.
- *lastEnd*: the last time a character utters a word, measured in character offsets.

4.1.4 Word Fields

To infer the semantics of the character speech, we use seven different word fields (WF) that cover the areas of *family* (Familie), *war* (Krieg), *love* (Liebe), *politics* (Politik), *reason* (Ratio), *religion* (Religion) and *economy* (Wirtschaft).[9] A word field contains lemmatized entries for terms of a certain domain that we consider relevant for the time span we analyze (roughly 1730–1850). For each character, we count the number of times the lemma of a spoken token is found in one of the word lists. It is then divided by the total number of tokens spoken by a character. The word lists were created manually in an iterative process (cf. Willand and Reiter 2017). We started by examining a few plays, where we identified important domains and associated words. We then gradually enriched the word lists by looking at further plays and dictionaries, e.g., Dornseiff, a dictionary for German vocabulary that is based on thematic groups.[10]

4.1.5 Topics

In addition to the manually created word fields, the character speech may contain further topics, which we attempt to capture using topic modeling (cf. Blei et al. 2003). For this purpose, we make use of Latent Dirichlet Allocation (LDA) with Gibbs sampling. We use the *German Drama Corpus* (Fischer et al. 2019) as input and divide the individual plays into segments of 1000 tokens each. We only consider those tokens that have been identified as 'noun,' 'verb,' 'adjective,' or 'ad-

9 The word lists are publicly available: https://github.com/quadrama/metadata/tree/master/fields.

10 Reiter and Willand (2022) conducted an evaluation study in which they compare the performance of word fields and topic modeling for a series of experiments that focus on character speech in dramatic texts.

Tab. 1: List of the top 12 lemmas in *T19*, both the original German words and an English translation.

	Topic 19
Lemma	**English Translation**
zar	tsar (male)
kirche	church
heilig	holy
zarin	tsar (female)
meister	master
Sir	sir
plötzlich	suddenly
rein	pure
kreuz	cross
nackt	naked
ewig	eternal
berühmt	famous

verb.' Proper nouns are also removed. The number of topics is set to 20, resulting in the topics *T1–T20*. The pre-processing and the parameters have been refined over a couple of studies (cf. Reiter et al. 2018; Krautter and Pagel 2019; Krautter et al. 2020) and we have trained various models with different numbers of topics. After manual inspection, we chose a model with 20 topics as it provided a good compromise between covering a broad range of different topics and – at the same time – providing clearly distinguishable and interpretable topics. Table 1 shows the top 12 lemmas for *T19* as an example. As can be seen *T19* mostly consists of lemmas related to religion and authority. As a feature value, each character is given the posterior probability of a topic, i.e., the probability of a character to utter tokens of a given topic. The posterior probability of a character is determined on the basis of all of his utterances.

4.1.6 Verbs of Action

Both word fields and topics should, to some extent, reflect the content or the aboutness of the characters' speech and thus indirectly allow us to draw conclusions about a character's function for the plot of the play. Another feature that attempts to quantify the function of characters within the plot is the distribution of verbs in both the primary and the secondary text of the plays. To utilize this feature, we extract the ten most frequent verbs in the character speech and the stage directions for each character in our corpus of plays while not considering modal and

Tab. 2: Most frequent verbs in character speech and stage directions.

Action verbs in utterances (utt.)		Action verbs in stage directions (sd.)	
Verb	**Frequency**	**Verb**	**Frequency**
sagen (to say)	2088	gehen (to go)	762
lassen (to let)	1892	treten (to tread)	490
sehen (to see)	1634	stehen (to stand)	335
machen (to make)	1623	sehen (to see)	291
kommen (to come)	1614	kommen (to come)	263
geben (to give)	1280	nehmen (to take)	196
gehen (to go)	1178	fallen (to fall)	183
tun (to do)	1037	setzen (to sit)	176
wissen (to know)	824	geben (to give)	142
hören (to hear)	751	werfen (to throw)	132

auxiliary verbs. The frequency of uttered verbs is determined for each character and divided by the total number of tokens a character speaks. Table 2 shows a list of the most frequent verbs in the 33 plays we analyze. The list distinguishes between character speech (utt.) and staged directions (sd.) in descending order of frequency.

4.1.7 Sentiment

To calculate the sentiment of a character's utterances, we make use of a static list of words with positive and negative connotations from the German SentiWS corpus (Remus et al. 2010). The list contains lemmatized entries sorted by part-of-speech as well as weights and possible inflectional forms. Similarly to the procedure for the word fields, we match the automatically lemmatized tokens spoken by a character with both lists and sum all weights when a match occurs. The two resulting final values (positive and negative) are then divided by the number of tokens spoken by a character to ensure comparability. This results in the two features, a positive *(posRatio)* and a negative *(negRatio)* ratio.

4.1.8 Priors

A prior is in our case an information that is linked to the play and its creation rather than to individual characters of the play. These priors, nevertheless, can

serve as criteria to group plays and thus contextualize properties of characters. We use two priors:

- *decade*: for every character, we store the decade in which the play was written or performed.
- *prose*/lines: a boolean value that stores whether the play is (mainly) written in prose or in lines. This is done automatically by containing the number of specific TEI-encoded tags in the source files of the *German Drama Corpus*: a majority of <l> and <lg> tags (line and line group) determines the play to be in lines, a majority of <p> tags (paragraph) determines the play to be written in prose.

4.2 Feature Overview

In order to look at the impact of different types of information, we group these features according to our operationalization presented earlier. Table 3 gives an overview of all features and the group they belong to.

5 Experiments

We use the features from section 4.1 as input to train a machine learning model. For the model, we opt for the Random Forest (RF) algorithm (Ho 1995). This model is trained by using the extracted features to predict whether a character is likely to be classified as 'schemer' or not.

Before training the model, we perform two preprocessing steps: Since highly correlated features contribute redundant information, all features are first checked for correlation in a pairwise manner. For pairs with a Pearson correlation coefficient larger than 0.7, the feature that correlates most strongly with all others is removed. Furthermore, since our data is unevenly distributed – there are many more negative than positive instances of schemers –, we use SMOTE (Chawla et al. 2002) to dynamically generate new training samples. Starting from nearest neighbors in the feature values, SMOTE generates artificial data points. This way, SMOTE helps to minimize imbalances in classes.

We use 10-fold-cross validation in all the following experiments to split the data into ten different folds. In each fold, we create a different division of the data into 80 % training and 20 % test data, so that in the end all data points are used once as test data. We pass the resulting training data to the RF algorithm and obtain ten different models based on each set of training data. We apply each model

Tab. 3: Overview of all features and their groups.

Group	Feature
Character speech stylistics	TTR utteranceLengthMean utteranceLengthSd
Stage presence	actives passives firstBegin lastEnd tokens utterances
Interaction	degree weightedDegree betweenness closeness eigenvector
Aboutness	Word fields (love, family, war, reason, religion, economy, politics) Topic model (T1-20)
Sentiment	posRatio negRatio
Action	utt.geben, utt.gehen, utt.hören, utt.kommen, utt.lassen, utt.machen, utt.sagen, utt.sehen, utt.tun, utt.wissen sd.fallen, sd.geben, sd.gehen, sd.kommen, sd.nehmen, sd.sehen, sd.setzen, sd.stehen, sd.treten, sd.werfen
Priors	decade prose/lines

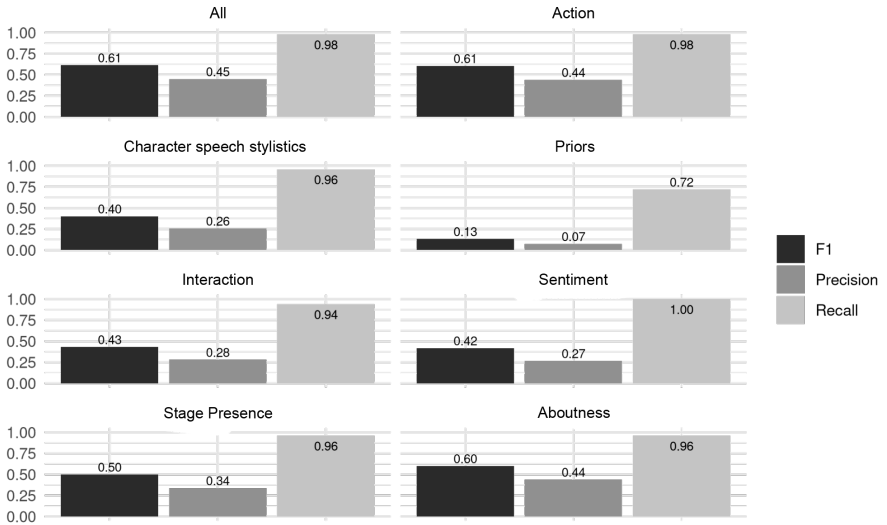


Fig. 2: Classification results for classifying "schemer" vs. "non-schemer" for different models.

to its respective test set and obtain a final test result by taking the average of all ten models. To evaluate the model, we use Precision, Recall, and F1-Score as metrics.

5.1 Classification Results

Figure 2 shows the results of the classification. The full model 'All,' which contains all features in combination, achieves the highest F1 score (0.61). It is followed by the models that focus on individual sub-concepts: 'Action' (0.61), 'Topics' (0.60), and 'StagePresence' (0.50). While the recall is consistently high for all models, the values for precision are rather low. This suggests that the models are indeed capable of correctly identifying characters as schemers but also detect many characters as schemers that are not identified as such in our gold data.[11]

11 We also performed preliminary experiments using transformer models, namely HuggingFace's German BERT (https://huggingface.co/bert-base-german-cased) and ELECTRA (https://huggingface.co/german-nlp-group/electra-base-german-uncased) models, by extracting all utterances and associated stage directions for each character and fine-tuning the two models on the resulting texts. However, we found the results not satisfactory, as the models achieved an overall F1-score of 0.48 for BERT and 0.49 for ELECTRA.

5.2 Error Analysis and Model Interpretability

In order to better understand the classification results, we look at several aspects of the classification process. Firstly, we look at the feature (or variable) importance, which quantifies how 'important' a feature was for the model in order to make a classification decision. Figure 3 shows all features of the model 'All' and their respective feature importance. The most important features are 'tokens,' 'T19,' 'utterances,' 'T15,' 'T18,' 'T14,' 'T17,' 'T5,' 'T20,' 'T3,' 'religion,' and so on. The feature importance seems to correspond to the classification results, as 'Aboutness' was the highest-performing sub-concept. Interestingly – and also in correspondence with the classification results –, the calculated network metrics, which we used to operationalize the sub-concept 'interaction,' have only limited importance for the classification decision. Looking at the schemer from a theoretical point of view, this is rather surprising, as the interaction with other characters would seem to be an important criterion to establish the intrigue. As Peter-André Alt has argued, the schemer represents various forms of intervention that fundamentally alter contexts of communication and interaction (cf. Alt 2004, p. 1).

We can also look at these features in terms of their value distribution with regard to the classification results. This can be seen in Table 4. TP stands for 'true positive,' meaning that the character is recognized as a schemer in literary studies' research and was also correctly classified as schemer by our model; TN stands for 'true negative,' a character was correctly classified as 'non-schemer'; and FP stands for 'false positive,' a character is incorrectly classified as schemer. A false negative (FN) only occurs once and can therefore not be used for a value distribution analysis.

For all three of the represented features, the results of the value distribution analysis confirm that the schemer not only tends to be a heterogeneous role type, but that this heterogeneity is also perceivable when looking at the schemers' character speech. The standard deviation – especially regarding the features 'tokens' and 'utterances' – is rather high, which complicates the automatic detection. Table 4 shows that the mean values for TP and FP are quite similar. This leads to two central questions: are the sub-concepts and, consequently, the features we have chosen to operationalize character types suited to identify schemers? Or are the schemers in our gold standard, which we derived from literary studies research, too heterogeneous in terms of their quantitative properties to categorize them as a group of similar characters?

To further investigate these questions, we perform a Principal Component Analysis (PCA). A PCA compresses the multi-dimensional features into two dimensions while preserving the most important properties of all features. This allows

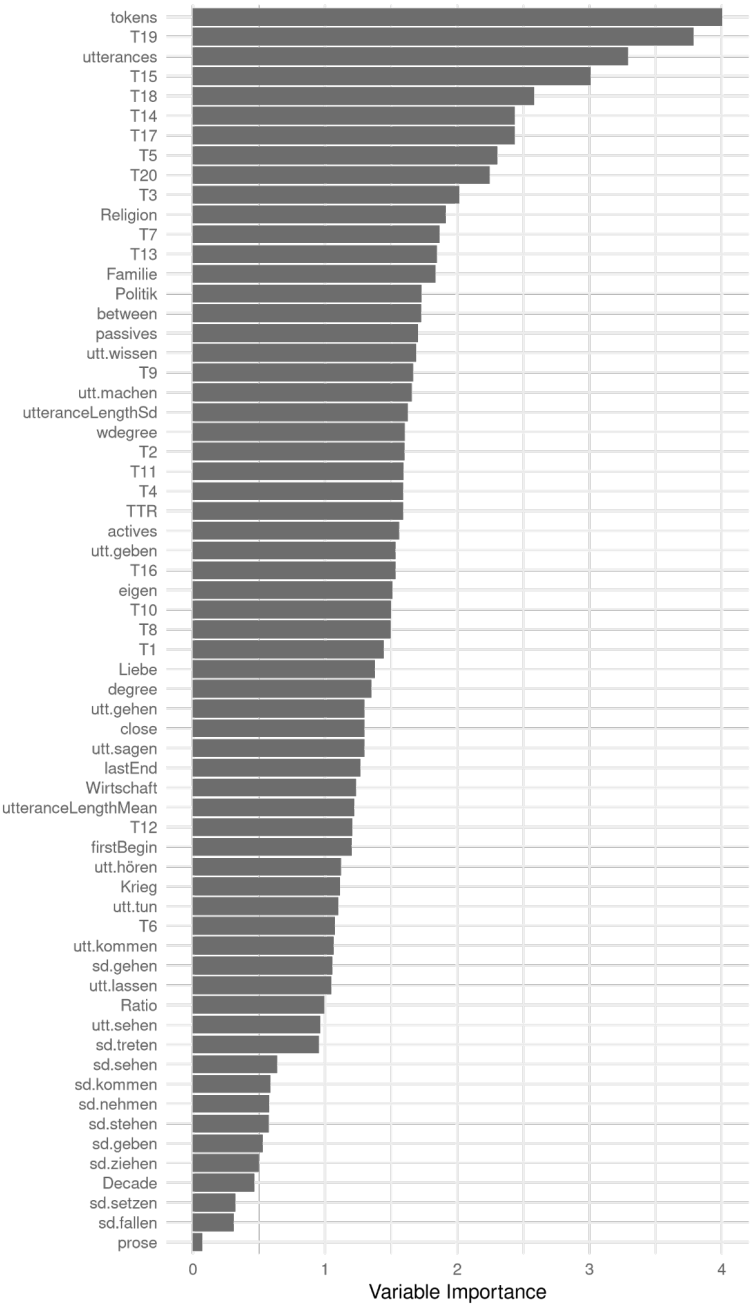


Fig. 3: Feature importance of model "All."

Tab. 4: Mean, max, and min value and standard deviation (Sd) for three top-performing features. The values are calculated based on the evaluation outcome: true positive (TP), true negative (TN) and false positive (FP).

Feature	Outcome	Mean	Min	Max	Sd
tokens	TP	0.133	0.015	0.411	0.09
	TN	0.025	0.000 066	0.396	0.05
	FP	0.126	0.0255	0.366	0.081
utterances	TP	128.69	11.00	364.00	83.53
	TN	27.66	1.00	509.00	49.55
	FP	121.87	38.00	341.00	72.02
religion	TP	0.000 018	0.00	0.000 065	0.000 014
	TN	0.000 028	0.00	0.000 94	0.000 087
	FP	0.000 022	0.000 002 5	0.000 071	0.000 015

to plot classes into a two-dimensional space while still showing a representation that resembles the totality of all features.

The PCA in Figure 4 demonstrates that schemers seem to have a rather close relation with regard to PC1 but are spread apart pretty far in relation to PC2. However, there seems to be a closer core group of schemers that is located in between the values 3.5 to -3.5 (PC2) and 1.25 to 3.75 (PC1). To give these observations more context, we also look at a PCA in which the false positives (FP) from our classification results are highlighted by color. This allows us to inspect whether characters that prompted the model to falsely classify them as schemers are closely located to schemers from our gold standard data. As we can see from the data in Figure 5, this seems to be the case. From this we can derive that the predictions of our model seem to be reasonable in view of the different character properties.

6 Discussion

Our classification results and our subsequent analyses have shown that automatically identifying schemers in German plays is a demanding task that confronts us with several difficulties at once. Firstly, creating or annotating a gold standard data set of schemers is not straightforward. Instead of manually reading and annotating the different plays and characters, we opted for an approach that resorts to second-hand criticism. While this brings the advantage of bypassing the time-consuming development of annotation guidelines and the effort of applying these guidelines to dramatic texts, it comes at the cost of – at least potentially – less con-

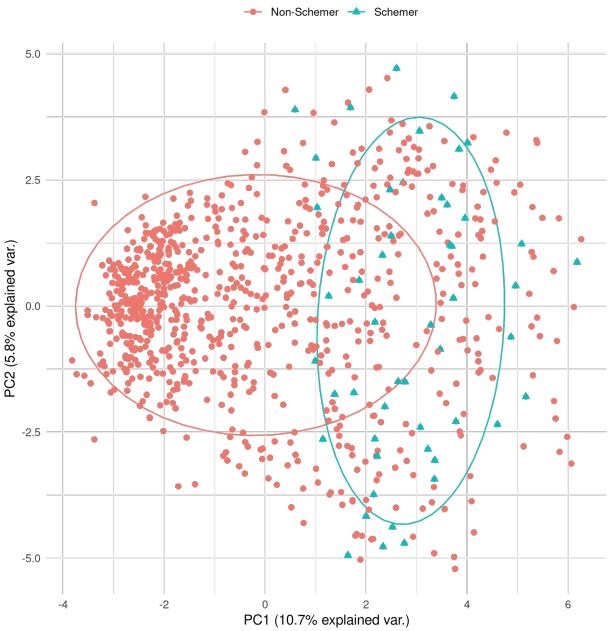


Fig. 4: PCA comparing schemers and non-schemers.

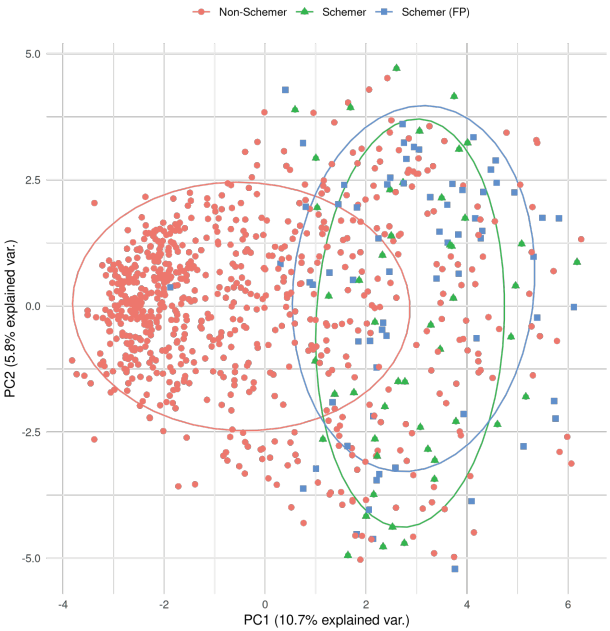


Fig. 5: PCA comparing schemers and non-schemers and false positives (FP).

sistency. It is safe to say that various literary scholars might weigh or apply criteria in different ways to categorize characters as schemers. This might add further inconsistencies to a class of already heterogeneous characters. Secondly, characters that are perceived as schemers can also be part of other groups of characters. They might as well be a father, a lover, or a servant. Consequently, some of their properties would then overlap with other characters that are not perceived as schemers. Thirdly, our indirect operationalization of character types focuses on several sub-concepts that are not specific to schemers. Instead of using an individually tailored feature set to delimit schemers from other characters, we aimed at creating a more general set of sub-concepts that can also be used to identify other character types (cf. Krautter et al 2020). We expected that – from a bird's-eye view – a synthesis of these sub-concepts could grasp abstract patterns in the schemers' character speech, in their form of interaction with other characters, or in their presence on stage that would distinguish them quantitatively from other characters.

Our analyses have shown that our model is able to accurately identify the characters taken from our gold standard data as schemers. However, the model has its difficulties in delimiting the class of schemers from other characters. As Figures 4 and 5 and Table 4 have shown the source of these difficulties is probably down to a combination of reasons. On the one hand, the class of schemers, as it is designed by our approach of using secondary literature to create a gold standard, might be too heterogeneous to delimit it accurately from all other characters. On the other hand, the features we use to operationalize character types might be too general and not distinct enough for schemers. We are therefore planning to adjust our experiments in the future according to these findings in at least two ways. We will try a different approach to creating the gold standard dataset that starts from prototypical schemers such as Lessing's Marinelli to then gradually expand the data set. Furthermore, we would like to supplement our feature set, e.g., by creating a word field that is specifically tailored to schemers.

7 Conclusion

In our study, we presented annotations for 38 plays, classifying characters into 'schemers' and 'non-schemers.' We based our annotation decisions on secondary literature, utilizing the results of research that has done in-depth analyses on the plays in question. Using the *German Drama Corpus* as a basis, we extracted several features from the characters' speech and from the stage directions. The features, in turn, were used to train a machine learning model that learned to classify characters into being a schemer or not being a schemer. The results of this classi-

fication show that while the model is already able to correctly identify characters as schemers (high recall), it requires better features or an improved dataset to reliably delimit schemers (true positives) from other characters (false positives) in the dataset (low precision). Features that capture information about stage presence and the content of character speech proved to be most useful to the model. In its current state – with a rather low precision – our model can be used as a heuristic to find promising candidates for schemers. These instances, however, would then have to be inspected manually to use them for further analysis.

Funding: This work was funded by the Volkswagen foundation and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 424244162.

Data and Code: We published the data and source code for all experiments described in this article under https://doi.org/10.5281/zenodo.7469269.

References

Alt, Peter-André (2004). "Dramaturgie des Störfalls: Zur Typologie des Intriganten im Trauerspiel des 18. Jahrhunderts." In: *Internationales Archiv für Sozialgeschichte der deutschen Literatur* 29, pp. 1–28.

Asmuth, Bernhard (2016). *Einführung in die Dramenanalyse*. 8th edition. J.B. Metzler.

Bennison, Neil (1998). "Accessing Character Through Conversation: Tom Stoppard's 'Professional Foul'." In: *Exploring the Language of Drama. From Text to Context*. Ed. by Jonathan Culpeper, Mick Short, and Peter Verdonk. Routledge, pp. 67–82.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003). "Latent Dirichlet Allocation." In: *Journal of Machine Learning Research* 3, pp. 993–1022. URL: https://jmlr.csail.mit.edu/papers/v3/blei03a.html.

Brandt-Schwarze, Ulrike (1996). "Intriganten, Giftmischer und Meuchelmörder: Die Handlanger des Bösen im Trauerspiel des frühen 19. Jahrhunderts." In: *Klassik, modern: Für Norbert Oellers zum 60. Geburtstag (= Sonderheft der Zeitschrift für deutsche Philologie 115)*. Ed. by Georg Guntermann, Jutta Osinski, and Hartmut Steinecke. Erich Schmidt Verlag, pp. 78–93.

Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer (2002). "SMOTE: Synthetic Minority Over-sampling Technique." In: *Journal of Artificial Intelligence Research* 16, pp. 321–357. DOI: 10.1613/jair.953.

Detken, Anke and Anja Schonlau (2014). "Das Rollenfach: Definition, Theorie, Geschichte." In: *Rollenfach und Drama: Europäische Theaterkonvention im Text*. Ed. by Anke Detken and Anja Schonlau. Narr, pp. 7–30.

Drews, Jörg, Richard Mellein, Meinhard Prill, Michael Schmidt, and Ruth Florack (2020). "Grabbe, Christian Dietrich: Das dramatische Werk." In: *Kindlers Literatur Lexikon (KLL): Living Edition*. Ed. by Heinz Ludwig Arnold. J.B. Metzler, pp. 1–6.

Drux, Rudolf (2005). "Kunigundes künstlicher Körper. Zur rhetorischen Gestaltung und Interdiskursivität eines 'mosaischen' Motivs aus Heinrich von Kleists Schauspiel 'Das Käthchen von Heilbronn'." In: *Kleist-Jahrbuch* 2005, pp. 92–110.

Eibl, Karl (1971). *Gotthold Ephraim Lessing: Miss Sara Sampson. Ein bürgerliches Trauerspiel*. Athenäum Verlag.

Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling, and Peer Trilcke (2019). "Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama." In: *Digital Humanities 2019: Book of Abstracts*, pp. 1–6. DOI: 10.5281/zenodo.4284002.

Gess, Nicola (2016). "Choreographie der Intrige. Zum dramatischen Rhythmus in Hofmannsthals 'Der Turm'." In: *Hofmannsthal Jahrbuch zur Europäischen Moderne* 24, pp. 197–228.

Grimm, Gunter E. (2020). "Lessing, Gotthold Ephraim: Jugendlustspiele." In: *Kindlers Literatur Lexikon (KLL): Living Edition*. Ed. by Heinz Ludwig Arnold. J.B. Metzler, pp. 1–5.

Grzesiuk, Ewa (2004). "'Ich reime, dächt' ich, doch noch ziemlich zusammen, was zusammengehört': Intriganten und Intrigen in Lessings Emilia Galotti." In: *Lügen und ihre Widersacher: Literarische Ästhetik der Lüge seit dem 18. Jahrhundert*. Ed. by Hartmut Eggert and Janusz Golec. Königshausen & Neumann, pp. 72–83.

Harris, Edward P. (1992). "Lessing und das Rollenfachsystem: Überlegungen zur praktischen Charakterologie im 18. Jahrhundert." In: *Schauspielkunst im 18. Jahrhundert: Grundlagen, Praxis, Autoren*. Ed. by Wolfgang F. Bender. Franz Steiner Verlag, pp. 221–235.

Hering, Christoph (1966). *Friedrich Maximilian Klinger. Der Weltmann als Dichter*. De Gruyter.

Ho, Tin Kam (1995). "Random decision forests." In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. 3rd International Conference on Document Analysis and Recognition. Vol. 1, pp. 278–282. DOI: 10.1109/ICDAR.1995.598994.

Horn, Eva (2011). "Herrmanns 'Lektionen'. Strategische Führung in Kleists 'Herrmannsschlacht'." In: *Kleist-Jahrbuch* 2011, pp. 66–90. DOI: 10.1007/978-3-476-00712-4_8.

Jannidis, Fotis, Isabella Reger, Markus Krug, Lukas Weimer, Luisa Macharowsky, and Frank Puppe (2016). "Comparison of Methods for the Identification of Main Characters in German Novels." In: *Digital Humanities 2016: Book of Abstracts*, pp. 578–582. URL: https://dh2016.adho.org/abstracts/static/dh2016_abstracts.pdf.

Kiss, Orsolya (2010). "Reinventing the Plot: J.C. Gottsched's 'Sterbender Cato'." In: *Deutsche Vierteljahrsschrift für Literaturwissenschaft und Geistesgeschichte* 84.4, pp. 507–525.

KLL (2020). "Iffland, August Wilhelm: Der Spieler." In: *Kindlers Literatur Lexikon (KLL): Living Edition*. Ed. by Heinz Ludwig Arnold. J.B. Metzler, pp. 1–2.

Knorr, Heinz (1951). "Wesen und Funktionen des Intriganten im deutschen Drama von Gryphius bis zum Sturm und Drang." PhD thesis. Universität Erlangen.

Kopp, Detlev (2016). "Grabbe. Ein Dramatikerleben: I. Von seiner Geburt bis zum Erscheinen der Dramatischen Dichtungen (1827)." In: *Grabbe-Jahrbuch* 35, pp. 7–37.

Krautter, Benjamin and Janis Pagel (2019). "Klassifikation von Titelfiguren in deutschsprachigen Dramen und Evaluation am Beispiel von Lessings 'Emilia Galotti'." In: *Konferenzabstracts DHd 2019 Digital Humanities: multimedial & multimodal*, pp. 160–164. DOI: 10.5281/zenodo.2596095.

Krautter, Benjamin, Janis Pagel, Nils Reiter, and Marcus Willand (2020). "'[E]in Vater, dächte ich, ist doch immer ein Vater': Figurentypen im Drama und ihre Operationalisierung." In: *Zeitschrift für digitale Geisteswissenschaften* 5. DOI: 10.17175/2020_007.

Krautter, Benjamin, Janis Pagel, Nils Reiter, and Marcus Willand (2022). "Properties of Dramatic Characters: Automatically Detecting Gender, Age and Social Status." In: *Computational*

Stylistics in Poetry, Prose and Drama. Ed. by Anne-Sophie Bories, Petr Plecháč, and Pablo Ruiz Fabo. De Gruyter, pp. 179–202.

Lessing, Gotthold Ephraim (1800). *Emilia Galotti. A Tragedy, in Five Acts*. Trans. by Benjamin Thompson. Vernor & Hood.

Linhardt, Marion (2008). "Bauernfeld und Nestroy, oder: Übertretung der Ordnung. Konzepte für ein nicht-ernstes Wort- und Körpertheater im Wien der 1830er Jahre." In: *Nestroyana* 28.1, pp. 8–27.

Löffelmann, Daniel (2015). "Geist und Sinnlichkeit. Zur dialektischen Transformation eines anthropologischen Dualismus in Grabbes 'Don Juan' und 'Faust'." In: *Grabbe-Jahrbuch* 34, pp. 75–98.

Marcus, Solomon (1973). *Mathematische Poetik*. Trans. by Edith Mândroiu. Athenäum Verlag.

Martens, Wolfgang (1995). "Zum Marinelli-Typus vor Lessing." In: *Ethik und Ästhetik: Werke und Werte in der Literatur vom 18. bis zum 20. Jahrhundert. Festschrift für Wolfgang Wittkowski zum 70. Geburtstag*. Ed. by Richard Fisher. Peter Lang, pp. 69–79.

Meier, Albert (2018). "'Sieh da der Neger.' Christian Dietrich Grabbes Herzog Theodor von Gothland als romantisierendes Schiller-Pastiche." In: *Grabbe-Jahrbuch* 37, pp. 31–40.

Memmolo, Pasquale (1995). *Strategen der Subjektivität: Intriganten in Dramen der Neuzeit*. Königshausen & Neumann.

Moretti, Franco (2000a). "Conjectures on World Literature." In: *New Left Review* 1, pp. 54–68.

Moretti, Franco (2000b). "The Slaughterhouse of Literature." In: *Modern Language Quarterly* 61.1, pp. 207–227.

Moretti, Franco (2005). *Graphs, Maps, Trees. Abstract Models for Literary History*. Verso Books.

Moretti, Franco (2011). "Network Theory, Plot Analysis." In: *Pamphlets of the Stanford Literary Lab* 2, pp. 1–12. URL: https://litlab.stanford.edu/LiteraryLabPamphlet2.pdf.

Müller-Seidel, Walter (1990). "Verschwörungen und Rebellionen in Schillers Dramen." In: *Schiller und die höfische Welt*. Ed. by Achim Aurnhammer, Klaus Manger, and Friedrich Strack. Max Niemeyer Verlag, pp. 422–446.

Pagel, Janis and Nils Reiter (2021). "DramaCoref: A Hybrid Coreference Resolution System for German Theater Plays." In: *Proceedings of the Fourth Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2021)*, pp. 36–46. URL: https://aclanthology.org/2021.crac-1.4/.

Pape, Walter (2011). "'Das Otello-Kleeblatt wird ein grimmiges Spectakel machen'. Dramatische Spiegelungstechnik." In: *Nestroyana* 31.3, pp. 133–149.

Pfister, Manfred (1988). *The Theory and Analysis of Drama*. Cambridge University Press.

Pichler, Axel and Nils Reiter (2021). "Zur Operationalisierung literaturwissenschaftlicher Begriffe in der algorithmischen Textanalyse. Eine Annäherung über Norbert Altenhofers hermeneutische Modellinterpretation von Kleists 'Das Erdbeben in Chili'." In: *Journal of Literary Theory* 15.1, pp. 1–29.

Pichler, Axel and Nils Reiter (2022). "From Concepts to Texts and Back: Operationalization as a Core Activity of Digital Humanities." In: *Journal of Cultural Analytics* 7.4, pp. 1–18. DOI: https://doi.org/10.22148/001c.57195.

Pütz, Peter (1970). *Die Zeit im Drama: Zur Technik dramatischer Spannung*. Vandenhoeck & Ruprecht.

Reiter, Nils, Benjamin Krautter, Janis Pagel, and Marcus Willand (2018). "Detecting Protagonists in German Plays around 1800 as a Classification Task." In: *Abstracts of EADH: Data in the Digital Humanities*. DOI: 10.18419/opus-10162.

Reiter, Nils and Marcus Willand (2022). "What are they talking about? A Systematic Exploration of Theme Identification Methods for Character Speech in Dramatic Texts." In: *Digitale Literaturwissenschaft*. Ed. by Fotis Jannidis. J.B. Metzler. DOI: 10.1007/978-3-476-05886-7_20.

Remus, Robert, Uwe Quasthoff, and Gerhard Heyer (2010). "SentiWS – a Publicly Available German-language Resource for Sentiment Analysis." In: *Proceedings of the 7th International Language Resources and Evaluation Conference*. LREC, pp. 1168–1171. URL: http://www.lrec-conf.org/proceedings/lrec2010/pdf/490_Paper.pdf.

Roselli, Antonio (2013). "'Nichts steht auf Erden fest'. 'Ende der Welt' und 'Ende einer Welt' in Grabbes 'Herzog Theodor von Gothland', 'Napoleon oder die hundert Tage' und 'Hannibal'." In: *Grabbe-Jahrbuch* 32, pp. 27–60.

Saße, Günter (1994). "Aufrichtigkeit: Von der empfindsamen Programmatik, ihrem Kommunikationsideal, ihrer apologetischen Abgrenzung und ihrer Aporie, dargestellt an Gellerts 'Zärtlichen Schwestern'." In: *Texttyp, Sprechergruppe, Kommunikationsbereich: Studien zur deutschen Sprache in Geschichte und Gegenwart. Festschrift für Hugo Steger zum 65. Geburtstag*. Ed. by Heinrich Löffler, Karlheinz Jakob, and Bernhard Kelle. De Gruyter, pp. 105–120.

Sautermeister, Gert (2020a). "Schiller, Friedrich von: 'Kabale und Liebe'." In: *Kindlers Literatur Lexikon (KLL): Living Edition*. Ed. by Heinz Ludwig Arnold. J.B. Metzler, pp. 1–3. DOI: 10.1007/978-3-476-05728-0_19148-1.

Sautermeister, Gert (2020b). "Schiller, Friedrich von: 'Wallenstein': Ein dramatisches Gedicht." In: *Kindlers Literatur Lexikon (KLL): Living Edition*. Ed. by Heinz Ludwig Arnold. J.B. Metzler, pp. 1–4.

Schneider, Helmut J. (2009). "Der zerbrochne Krug." In: *Kleist Handbuch: Leben – Werk – Wirkung*. Ed. by Ingo Breuer. J.B. Metzler, pp. 33–41.

Schonlau, Anja (2017). *Emotionen im Dramentext: Eine methodische Grundlegung mit exemplarischer Analyse zu Neid und Intrige 1750–1800*. De Gruyter.

Schröder, Fynn, Hans Ole Hatzel, and Chris Biemann (2021). "Neural End-to-end Coreference Resolution for German in Different Domains." In: *Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021)*, pp. 170–181. URL: https://aclanthology.org/2021.konvens-1.pdf.

Steglich, Dana (2017). "'So lang die Jungfer schweigt'. Verleumdete Frauenfiguren in 'Much Ado About Nothing' und 'Der zerbrochne Krug'." In: *Kleist-Jahrbuch* 2017, pp. 166–176.

Steinmetz, Horst (1978). *Die Komödie der Aufklärung*. 3rd edition. J.B. Metzler.

Stephan, Inge (2015). "'Er hatte einen entschiedenen Hang zur Intrige': Überlegungen zu J. M. R. Lenz, seiner Rezeption und seinen Werken." In: *Goethe Yearbook* 22, pp. 247–260.

Storch, Ludwig (1839). "Charakter-Rollen." In: *Allgemeines Theater-Lexikon oder Encyklopädie alles Wissenswerthen für Bühnenkünstler, Dilettanten und Theaterfreunde*. Ed. by Robert Blum, Karl Herleßsohn, and Hermann Marggraff. H. A. Pierer, p. 122.

Trilcke, Peer (2013). "Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft." In: *Empirie in der Literaturwissenschaft*. Ed. by Philip Ajouri, Katja Mellmann, and Christoph Rauen. Mentis, pp. 201–247.

Trilcke, Peer, Frank Fischer, and Dario Kampkaspar (2015). "Digital Network Analysis of Dramatic Texts." In: *Digital Humanities 2015: Book of Abstracts*. DOI: 10.5281/zenodo.3627711.

Vonhoff, Gert (2011). "Maria Stuart. Trauerspiel in fünf Aufzügen (1801)." In: *Schiller Handbuch: Leben – Werk – Wirkung*. Ed. by Matthias Luserke-Jaqui. J.B. Metzler, pp. 153–168.

Von Matt, Peter (2006). *Die Intrige: Theorie und Praxis der Hinterlist*. Carl Hanser Verlag.

Wasserman, Stanley and Katherine Faust (1994). *Social Network Analysis: Methods and Application*. Cambridge University Press.

Willand, Marcus and Nils Reiter (2017). "Geschlecht und Gattung. Digitale Analysen von Kleists 'Familie Schroffenstein'." In: *Kleist-Jahrbuch* 2017, pp. 177–195. DOI: 10.1007/978-3-476-04516-4_16.

Wittkowski, Wolfgang (1990). "Höfische Intrige für die gute Sache: Marquis Posa und Octavio Piccolomini." In: *Schiller und die höfische Welt*. Ed. by Achim Aurnhammer, Klaus Manger, and Friedrich Strack. Max Niemeyer Verlag, pp. 378–397.

Appendix

Tab. 5: List of schemers in our data set.

Schemer	Play	Author	Reference
Pharnaces	Der sterbende Cato (1731)	J. Chr. Gottsched	Memmolo 1995, pp. 128–129, Kiss 2010, p. 518.
Ulfo	Canut (1746)	J. E. Schlegel	Memmolo 1995, p. 132
Siegmund	Die zärtlichen Schwestern (1747)	Chr. F. Gellert	Saße 1994, p. 107
Henley	Der Freigeist (1758)	J. W. v. Brawe	Memmolo 1995, p. 143, Schonlau 2017, p. 233
Northumberland	Lady Johanna Gray (1758)	Chr. M. Wieland	Knorr 1951, p. 75.
Lisette	Der junge Gelehrte (1748)	G. E. Lessing	Memmolo 1995, p. 204
Hilaria	Der Misogyn (1755)	G. E. Lessing	Grimm 2020
Marwood	Miß Sara Sampson (1755)	G. E. Lessing	Schonlau 2017, p. 226
Franziska	Minna von Barnhelm	G. E. Lessing	Asmuth 2016, pp. 125–126
Orsina	Emilia Galotti (1772)	G. E. Lessing	Grzesiuk 2004, pp. 72–73
Marinelli	Emilia Galotti (1772)	G. E. Lessing	Alt 2004, pp. 7–8; Grzesiuk 2004, pp. 72–73
Hasenpoth	Die Kindermörderin (1776)	H. L. Wagner	Alt 2004, p. 9; Pütz 1970, p. 82
Guido	Julius von Tarent (1776)	J. A. Leisewitz	Schonlau 2017, pp. 286–287
Mary	Die Soldaten (1776)	J. M. R. Lenz	Stephan 2015, pp. 251–252
Haudy	Die Soldaten (1776)	J. M. R. Lenz	Stephan 2015, pp. 251–252
Weislingen	Götz von Berlichingen (1773)	J. W. Goethe	Schonlau 2017, p. 258
Adelheid	Götz von Berlichingen (1773)	J. W. Goethe	Alt 2004, p. 15; Memmolo 1995, pp. 213–214, Schonlau 2017, p. 179
Carlos	Clavigo (1774)	J. W. Goethe	Alt 2004, pp. 17–19
Alba	Egmont (1788)	J. W. Goethe	Memmolo 1995, p. 226
Amtmann	Die Jäger (1785)	A. W. Iffland	Detken and Schonlau 2014, p. 22
Geheimerath	Der Spieler (1796)	A. W. Iffland	KLL 2020

continued on next page

continued from previous page

Schemer	Play	Author	Reference(s)
Guelfo	Die Zwillinge (1776)	F. M. Klinger	Schonlau 2017, p. 314
Drullo	Die neue Arria (1776)	F. M. Klinger	Hering 1966, p. 86
Franz von Moor	Die Räuber (1781)	F. Schiller	Alt 2004, p. 22; Memmolo 1995, pp. 282–290
Fiesko	Die Verschwörung des Fiesko zu Genua (1782)	F. Schiller	Memmolo 1995, pp. 295–299, von Matt 2006, p. 327
Muley Hassan (Mohr)	Die Verschwörung des Fiesko zu Genua (1782)	F. Schiller	von Matt 2006, pp. 325–327
Wurm	Kabale und Liebe (1784)	F. Schiller	Alt 2004, p. 23; Sautermeister 2020a; Memmolo 1995, p. 276
Marquis von Posa	Don Karlos (1787)	F. Schiller	von Matt 2006, p. 202; Gess 2016, p. 213
Prinzessin Eboli	Don Karlos (1787)	F. Schiller	Schonlau 2017, p. 331, Müller-Seidel 1990, p. 435
Domingo	Don Karlos (1787)	F. Schiller	Schonlau 2017, p. 331
Herzog Alba	Don Karlos (1787)	F. Schiller	Schonlau 2017, p. 331, Müller-Seidel 1990, p. 435
Octavio	Die Piccolomini (1799)	F. Schiller	Sautermeister 2020b; Wittkowski 1990, p. 380
Gräfin Terzky	Die Piccolomini (1799)	F. Schiller	Alt 2004, p. 15
Octavio	Wallensteins Tod (1799)	F. Schiller	Sautermeister 2020b; Wittkowski 1990, p. 380
Gräfin Terzky	Wallensteins Tod (1799)	F. Schiller	Alt 2004, p. 15
Leicester	Maria Stuart (1800)	F. Schiller	Vonhoff 2011, p. 161; von Matt 2006, pp. 328, 380
Mortimer	Maria Stuart (1800)	F. Schiller	von Matt 2006, p. 328
Gertrude	Die Familie Schroffenstein (1803)	H. Kleist	Alt 2004, p. 16
Hermann	Die Hermannsschlacht (1808)	H. Kleist	Drews et al. 2020; Horn 2011, p. 80
Adam	Der zerbrochne Krug (1808)	H. Kleist	Schneider 2009, pp. 35–37; Steglich 2017, p. 166
Kunigunde	Das Käthchen von Heilbronn (1810)	H. Kleist	Alt 2004, p. 16; Drux 2005, pp. 109–110
Ratcliff	William Ratcliff (1823)	H. Heine	Brandt-Schwarze 1996, pp. 86–87
Der Teufel	Scherz, Satire, Ironie und tiefere Bedeutung (1827)	Chr. D. Grabbe	Kopp 2016, p. 21
Berdoa	Herzog Theodor von Gothland (1827)	Chr. D. Grabbe	Roselli 2013, 39 and 43; Kopp 2016, p. 15
Arboga	Herzog Theodor von Gothland (1827)	Chr. D. Grabbe	Meier 2018, p. 33
Don Juan	Don Juan und Faust (1829)	Chr. D. Grabbe	Löffelmann 2015, p. 83
Unruh	Bürgerlich und Romantisch (1835)	E. v. Bauernfeld	Linhardt 2008, p. 20

continued on next page

continued from previous page

Schemer	Play	Author	Reference(s)
Hutzibutz	Das Haus der Temperamente (1837)	J. Nestroy	Pape 2011, p. 144
Schlankel	Das Haus der Temperamente (1837)	J. Nestroy	Pape 2011, p. 144
Santinelli	Monaldeschi (1841)	H. Laube	Brandt-Schwarze 1996, p. 91

Artjoms Šeļa, Ben Nagy, Joanna Byszuk,
Laura Hernández-Lorenzo, Botond Szemes, and Maciej Eder

From Stage to Page: Stylistic Variation in Fictional Speech

Abstract: Stylometry is mostly applied to *authorial* style. More recently, researchers have begun investigating the style of *characters*, finding that although there is detectable stylistic variation, the variation remains within authorial bounds. In this article, we address the stylistic distinctiveness of characters in drama. Our primary contribution is methodological; we introduce and evaluate two non-parametric methods to produce a summary statistic for character distinctiveness that can be usefully applied and compared across languages and times. This is a significant advance – previous approaches have either been based on pairwise similarities (which cannot be easily compared) or indirect methods that attempt to infer distinctiveness using classification accuracy. Our first method is based on bootstrap distances between 3-gram probability distributions, the second (reminiscent of 'unmasking' techniques) on word keyness curves. Both methods are validated and explored by applying them to a reasonably large corpus (a subset of DraCor): we analyze 3301 characters drawn from 2324 works, covering five centuries and four languages (French, German, Russian, and the works of Shakespeare). Both methods appear useful; the 3-gram method is statistically more powerful, but the word keyness method offers rich interpretability. Both methods are able to capture phonological differences such as accent or dialect, as well as broad differences in topic and lexical richness. Based on exploratory analysis, we find that smaller characters tend to be more distinctive and that women are cross-linguistically more distinctive than men, with this latter finding carefully interrogated using multiple regression. This greater distinctiveness stems from a historical tendency for female characters to be restricted to an 'internal narrative domain' covering mainly direct discourse and family/romantic themes. It is hoped that direct, comparable statistical measures will form a basis for more sophisticated future studies, and advances in theory.

Artjoms Šeļa, Ben Nagy, Joanna Byszuk, Maciej Eder, Polish Academy of Sciences
Laura Hernández-Lorenzo, University of Seville
Botond Szemes, Institute for Literary Studies Budapest
Artjoms Šeļa, University of Tartu

https://doi.org/10.1515/9783111071824-008

1 Introduction

Since Vladimir Propp's work, structural narratology has approached fictional characters mainly through their role or function – by what they do or what is done to them (Eder et al. 2010). This character typology relied on recurring functions in the narrative (lover, villain, victim, detective, etc.) and the same perspective was often adopted in computational research, where characters in novels were modeled on the basis of narrative passages rather than dialogue (Bamman et al. 2014; Bonch-Osmolovskaya and Skorinkin 2017; Underwood et al. 2018; Stammbach et al. 2022).

In dramatic texts, however, the dominant device for characterization is an utterance. While the script usually contains some stage directions, the specifics of characterization and style of performance are not determined by the text itself, but developed by a specific theater, director or a troupe. Over the course of history, many plays were written for specific theater stages, and it was common practice to write characters for specific actors (Fischer-Lichte 2002). Of course, this kind of 'outsourced characterization' was supported by dramatic conventions and formulas. Viewers' expectations could be shaped without a single word being uttered on stage, just by a character wearing a costume, operating a puppet, or changing a dell'arte stock mask. At the same time, the things characters say and how they say them are the main textual source of information about them. It is reasonable to assume that dramatists make significant efforts to create linguistic distinctions between princes and paupers, lovers and schemers, aristocrats and merchants. Tragic monologue is written differently from a comedic exchange between servants. Some previous computational works treat linguistic distinctiveness of characters from the perspective of this stylistic continuum (Vishnubhotla et al. 2019), noting that it can be influenced by genre, character gender, or their social and professional dispositions.

A parallel narratological tradition, tied to Bakhtin's ideas of heteroglossia, focuses not on abstract character roles, but on the words characters say (Sternberg 1982; Culpeper 2001; Bronwen 2012). The modern novelistic space of dialogic exchange, 'educated conversation' (Moretti 2013, p. 20) and the clash of styles in reported discourse become central here. Available stylometric research on fictional speech and micro-stylistic variation suggests that characters within a text are often distinguishable by their local linguistic patterns without obscuring the global authorial trace (Burrows 1987; Hoover 2017). As put by Burrows and Craig: "Characters speak in measurably different ways, but the authorial contrasts transcend this differentiation. The diversity of styles within an author always remains within bounds" (Burrows and Craig 2012, pp. 307–308).

Conceptually and methodologically, the majority of previous works examined not the *distinctiveness* of characters, but their (pairwise) *similarity*. Similarity measures are meaningful in pairwise contexts but cannot be analyzed and compared as individual summary statistics. Since Burrows' seminal study of speech patterns in Jane Austen's characters (Burrows 1987), these approaches focused on calculating similarity within a collection of characters: how different is character X from character Y, and each of them from character Z. Burrows measured the correlation between characters' usage of 30 most frequent words (technically, he fit a linear regression for two sets of log-frequencies); later, similarity was most often inferred through clustering based on pairwise distance calculations (Reeve 2015; Craig and Greatley-Hirsch 2017; Hoover 2017). Sometimes linguistic similarity served as a basis for arguing functional similarity as well. A recent study that linked Bakhtin's dialogism and the stylistic diversity of characters' speech (Vishnubhotla et al. 2019) proposed the analysis of distinctiveness rather than similarity using supervised classification. Instead of using a network of pairwise relationships, the authors asked how well a classifier can recognize character X as being written by author A. Classification accuracy in this scenario becomes an explicit summary statistic for distinctiveness that can be assigned to a character (or, in an aggregated manner, to a play or an author). However, the supervised approach, proposed by Vishnubhotla et al., is data hungry: it suffers from extreme class imbalance, an abundance of short samples (most characters speak only a little), and is dependent on language-specific feature construction procedures.

By contrast, this paper will present a simple, non-parametric measure of character distinctiveness that is based on bootstrapped probability distributions representing a character and all others present in a given play: an approach largely informed by authorship verification techniques. This measure is language-independent and relies only on the context of a single work, which, in turn, minimizes problems of language variation, authorial signal and chronological change in a comparative setting. Individual distinctiveness scores can then be tested against other measures and metadata categories in a hypothesis-driven manner, not only across languages, but also across genres (e.g., novel vs. drama). Do comedies tend to employ more distinct characters? Does distinctiveness increase (authors get better), or decrease (social and linguistic homogenization occurs) over time? Is there a difference between the distinctiveness of fictional women and men? If so is it the direct result of perceived gender differences, or is it constructed by imagined differences in social and professional status?

Lacking good descriptive metadata on the dramatic characters, this paper will not answer the above-mentioned questions in any satisfying way. Instead, we focus on presenting and justifying the measure of distinctiveness and exploring sev-

Tab. 1: A summary of the corpus. All word and 3-gram counts are for the filtered corpus (characters that speak at least 2000 words) only.

Corpus	Total Characters	Characters Analyzed	Unique 3-grams	Unique Words	Total 3-grams	Total Words
French	15462	1744	9896	79994	29.79 m	5.47 m
German	14010	1182	14341	150956	24.80 m	4.31 m
Russian	3707	248	12542	71217	4.05 m	0.72 m
Shakespeare	1431	127	5921	19595	2.16 m	0.43 m

eral factors that might shape the final scores (like the year of composition, character gender and characters' sample size).

2 Materials

As the beginning of our exploration of cross-linguistic variation, we examined four dramatic corpora from DraCor (Fischer et al. 2019): Shakespeare, French, German, and Russian. DraCor is a project that gathers dramatic corpora in various languages, primarily European, encoded in TEI-XML. With 15 corpora available so far, including the Shakespeare corpus available both in English and German, DraCor facilitates large scale analysis of dramatic conventions across language traditions and offers a wide variety of useful metadata at the level of both plays and characters. While the analysis of all DraCor corpora would be possible with the methods we developed, for the purpose of this preliminary study we focused on the languages and dramatic traditions well-known to the members of our team, eventually selecting the full corpora for Shakespeare, French, German, and Russian: a total of 2324 texts, the majority of which come from French and German. The corpus is summarized in Table 1.

3 Methods

3.1 General Approach and Definitions

Our understanding of character distinctiveness is largely informed by 'authorship verification' approaches, which center around verifying that a text is written by a target author. This problem is more general than 'authorship attribution,' which tries to identify the nearest stylistic neighbor for a text (Halvani et al. 2019). In-

stead, authorship verification asks about the relative *magnitude* of similarity: is a target text more similar to same-author samples or different-author samples? With this in mind, we define a character's 'distinctiveness' as the degree to which the style of their speech differs from that of other characters. We understand 'style' here instrumentally, as a deviation from an unobserved average language (Hermann et al. 2015), and do not introduce aggressive feature filtering, allowing both 'grammatical' and 'thematic' signals to contribute to the final measures. We anchor our distinctiveness measure in the context of the specific text in which a character appears. In theory, the frame of reference could be all plays from one author, or all plays from the same period, or even some external corpus – however, all of these would greatly complicate any comparative study.

3.2 Bootstrap 3-Gram Distinctiveness

Based on our definition of distinctiveness above, we considered a character's style to be an idiolect sampled from a frequency distribution of character 3-grams. As a natural language distribution, this was expected to be generally Zipfian, a family of heavy-tailed distributions, so non-parametric methods were seen to be important. 3-grams were preferred to words for a number of reasons: first, they capture sub-word information, which means they will reflect general sonic preferences (so they can capture things like accent) and, particularly in inflected languages, also reflect some grammatical style; second, as a practical matter, they effectively expand the sample data, since a string of text produces approximately one 3-gram per character. This increased sample size should reduce the variance of the statistics. Finally, the number of unique 3-grams in a language is considerably smaller than the number of words, so the frequency data is less sparse, which again is expected to increase robustness. To now operationalize the distinctiveness, as defined, we used standard bootstrap methods to measure the median energy distance (Székely and Rizzo 2013) with bootstrap confidence intervals between the two distributions (character 3-gram frequencies vs. 'other' 3-gram frequencies). The energy distance is one of a family of related metrics that are commonly used to measure the difference between probability distributions.

Some limitations and choices were required. As mentioned, we measured distinctiveness only within the context of a single work (even for authors with multiple works). To expand beyond single works would produce very mismatched sample sizes, since some authors were prolific and some produced just one play; even with non-parametric methods, hugely mismatched sample sizes are problematic. Furthermore, the plays span four languages and roughly five centuries, making the 'distant' context seem ridiculous. As well as the selected distinctiveness statis-

tic (median energy distance) we also recorded a 'baseline' distinctiveness, this being each character's distance from themselves. The theoretical baseline is, of course, zero, but the sample baselines will not be, meaning that this gives us an idea of the inherent variance of the samples. Finally, when selecting characters to examine, we chose a minimum size of 2000 words. Sample sizes are somewhat arbitrary, and are matters of debate (Eder 2015, 2017), but this seemed to be a reasonable, or perhaps even slightly aggressive, lower bound.

3.3 Area under Keywords

Our second, supplementary approach was informed by 'unmasking' techniques often employed in stylometric research (Koppel and Schler 2004; Kestemont et al. 2016; Plecháč and Šeļa 2021). Unmasking refers to a range of methods that share one goal: to measure and compare the *depth* of the differences between two sets of texts. For example, an author might write both high fantasy fiction and historical novels: a classifier would have little difficulty distinguishing one genre from another by simply using superficial features (e.g., 'dragons,' 'magic,' 'elves'). However, by assumption, if these most distinctive features are removed, the classifier will have more trouble determining which text came from which pool, because the texts share one deep similarity – a common authorial style. Conversely, if we compare books by two different fiction writers, these texts will also have superficial differences. However, while removing more and more distinctive features, the classifier should remain confident in distinguishing the authors from each other, because the texts do not share an authorial style that is deeply rooted in common linguistic elements and distributed over many features. By comparing the speed with which the rates of accuracy decay, we can approach authorship verification problems, i.e., how plausible is it that this text belongs to author A?

We applied the same thinking to fictional characters, as opposed to authors: the distinctiveness of a character may rely on a small number of catch-phrases ('Gadzooks!' or 'Cowabunga!') or it may be driven by non-stylistic, referential factors (Mary, speaking to John, is not likely to use the word 'Mary,' but is likely to use the word 'John,' and vice-versa). On the other hand, there are characters whose speech systematically differs from the neutral language: such as when the author imitates dialects, slang, regionalism, speech and phonetic idiosyncrasies. In the former case, an imaginary classifier should quickly lose accuracy (since John and Mary speak quite similarly), but in the latter case the removal of a small number of features would not be enough to disrupt classification.

In our case, it was impractical to use 'standard,' supervised (i.e., classifier-based) unmasking because individual characters, as samples, were simply too

small. Instead we used word keyness – a character's relative preference for a word in the context of a given drama – to calculate an alternative distinctiveness score together with a bag of easily interpretable features per character. First, we use weighted log-odds (Monroe et al. 2008) to calculate keywords for a character relative to the speech pool of the rest of the cast; second, we represented each character by their 100 words with the highest keyness, arranged by rank; finally, we measured the area under this curve, which we interpret as distinctiveness – characters with just a few key words will exhibit less area under the keyness curve. By comparing these final areas, we can compare the relative difference between each character and rest of the speech in the play. In a similar manner to the bootstrapped approach, we upsample each character's word pool to match the size of the rest of the words in the play to minimize the effect of the sample size as much as possible.

4 Results

Overall, the distinctiveness energy statistic appears useful. The baseline (character vs. self) is quite stable cross-linguistically, although it is slightly higher for characters with a very large share of dialogue (Figure 1). Note also that the distinctiveness statistic appears roughly Gaussian (see Appendix B for more discussion) and its range is relatively consistent between languages (peaking at roughly 0.20), although this consistency does not apply at the level of authors. The obvious issue is that there is a strong negative correlation between character size and distinctiveness, but this is not only a limitation of the method – lead characters naturally set the dominant style of a text (and possibly inherit more of the 'true' authorial voice). Importantly, distinctiveness does not increase with the number of speakers in a play. The method works best when there are reasonable sample sizes for both the examined character and the 'other' class. This is illustrated by the 'U' curve visible in the French corpus in Figure 1 as the examined characters' dialogue share passes 50%. As hoped, the energy-distance method does appear to capture characters who are written with distinctive idiolects, representing things like foreign accents or social class. For a discussion of this see Section 5.

As seen in Figure 2, there is no clear correlation between the date of composition and character distinctiveness, which suggests that language change does not disturb the measure. The finding that seems clear is that women are written differently from men. Female characters are generally more distinctive in all corpora (Figure 3b), although this is not visible using the keyness AUC measure – leading us to conclude that the keyness measure has lower power. This difference in the

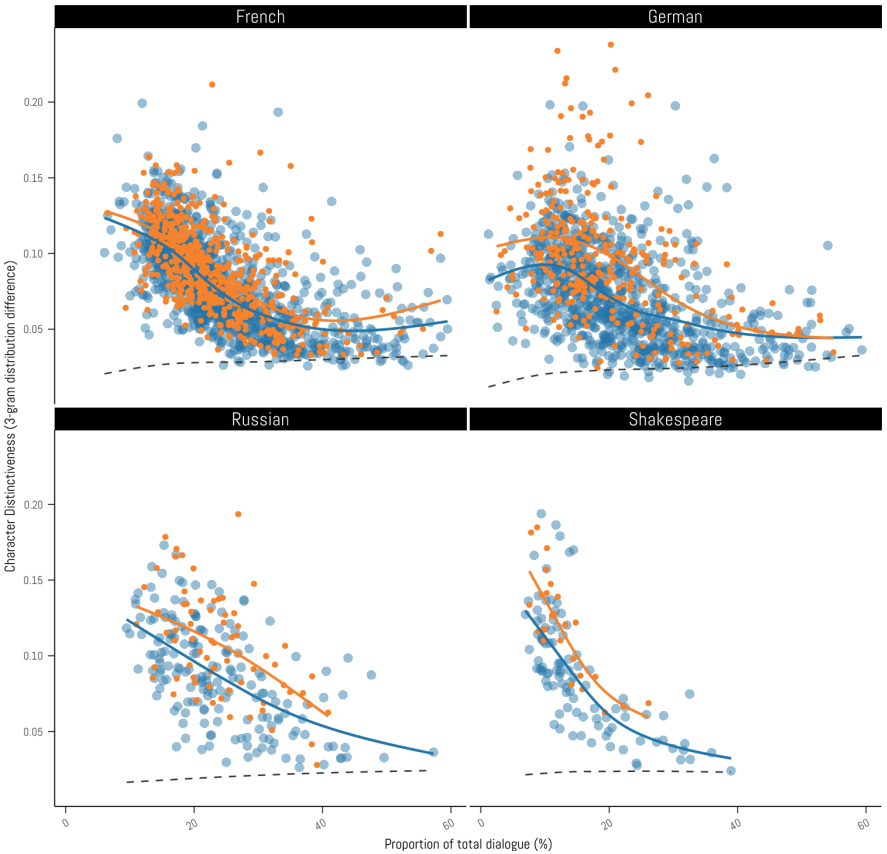


Fig. 1: Character distinctiveness, per corpus, versus % Dialogue. Women are shown smaller and in orange, men (and undefined) larger and in blue. GAM (Generalized Additive Model) trendlines are superimposed in the same colors. Baseline data (GAM trend for distinctiveness of character vs. self) is shown as a dashed line.

distinctiveness of female characters can partly be explained by the fact that they tend to have smaller parts (Fig 3a), and smaller characters in general are more distinctive (Figure 1), but that is not the whole story. Female parts have more restricted 3-gram vocabularies (Figure 3c), suggesting that they are also restricted in their semantic fields. This becomes clearer when the relative frequencies of their (word) vocabularies are examined. As well as the stereotypical tendencies (women say 'love,' men say 'sword'), the female characters, cross-linguistically, seem to be less likely to reference the 'external world' of the drama. As seen in Appendix A, relatively more frequent words for women are dominated by personal

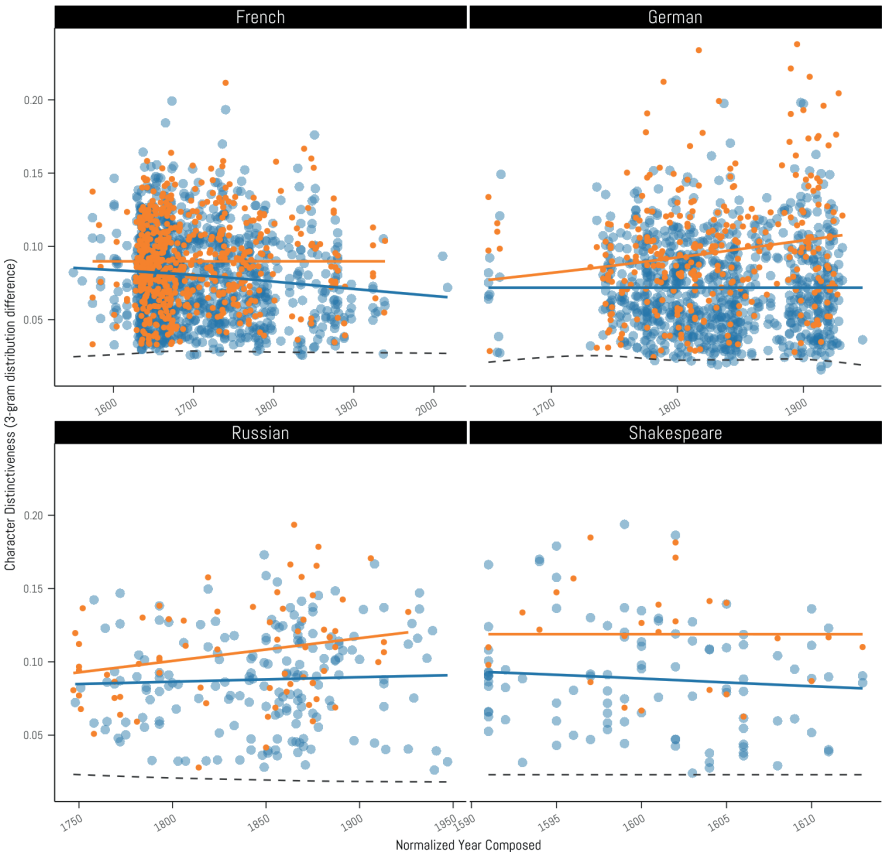
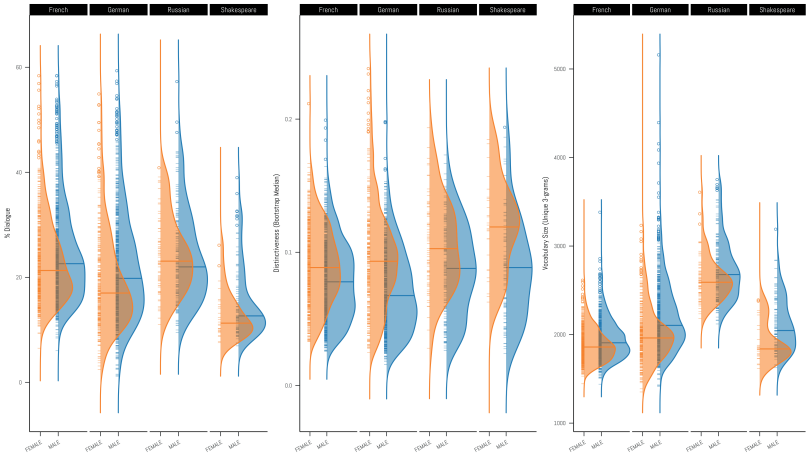


Fig. 2: Character distinctiveness, per corpus, versus year composed (DraCor data). Women are shown smaller and in orange, men (and undefined) larger and in blue. GAM (Generalized Additive Model) trendlines are superimposed in the same colors. Baseline data (GAM trend for distinctiveness of character vs self) is shown as a dashed line.

pronouns representing 'I,' 'me,' 'you' etc. or words relating to family. The male lists are dominated by indicative articles and political terms ('law,' 'noble,' 'king' etc.).

The higher distinctiveness of female characters is further supported by a formal linear model: we fit a Bayesian multiple regression where distinctiveness was conditioned on both gender and size (characters' percentage of total dialogue). A direct gender effect is present in all corpora, as expected from Figure 3a, but when we account for variation among authors, the effect may be less pronounced than it appears (for analysis and more detailed discussed, the posterior estimates are described in Appendix B). Our finding interlocks with the observation by Under-



(a) % Dialogue **(b)** Distinctiveness **(c)** Vocab size (3-grams)

Fig. 3: An analysis, per corpus, of the distribution of various features by gender. Distributions are estimated, with the median shown as a solid line. Actual points are shown as rug plots with outliers 'o' plotted for points outside 3Q + 2×IQR.

wood et al. (2018) that female characters found in English eighteenth- to twentieth-century fiction displayed high distinctiveness due to the particular way they were narrated, suggesting a pervasive authorial mentality.

5 Discussion

The measures of stylistic character distinctiveness that were proposed in this paper appear to be effective in capturing a *degree* to which characters stand out from others. The most distinctive characters, by both of our metrics, often have systematically different speech, in the form of dialects, regionalisms, or class markers. For example, Shakespeare's Captain Fluellen (*Henry V*) is Welsh and his accent is written for comedic effect. The systematic replacements b→p and d→t make him the most distinctive Shakespearean character according to both the 3-gram and word measures:

> FLUELLEN
> Your grandfather of famous memory, an't please your majesty, and your great-uncle Edward the Plack Prince of Wales, as I have read in the chronicles, fought a most prave pattle here in France.

KING HENRY V
They did, Fluellen.
FLUELLEN
Your majesty says very true: if your majesties is remembered of it, the Welshmen did good service in a garden where leeks did grow, wearing leeks in their Monmouth caps; which, your majesty know, to this hour is an honourable badge of the service; and I do believe your majesty takes no scorn to wear the leek upon Saint Tavy's day.

Regional differences also contribute to high distinctiveness in the German corpus. For example, Emerike in *das Manuskript*, written by Johanna von Weißenthurn, uses -ey instead of -ei (zwey, bey, Freylich) which is a form indicative of pre-standardized Southern German spelling. John, in Hauptmann's *Die Ratten*, speaks Plattdeutsch, a variant heavily influenced by Dutch, e.g., 'Det hat er jesacht, det ick noch ma hin müßte und janz jenau anjeben.'

In the French corpus, the most distinctive character by keyness is Gareau, from *Le Pédant Joué* (Cyrano de Bergerac), who speaks a 'patois' or rural dialect. In his critical edition, Frédéric Lachèvre comments on this distinct idiolect when Gareau is first introduced (Cyrano de Bergerac 1921, p. 25):

> Cyrano a fabriqué de toutes pièces le patois de Gareau. Le manuscrit de la BN donne un langage tout différent que celui imprimé en 1654, la pronociation des mots n'est pas tout à fait la même. Nous avons naturellement maintenu pour Gareau le texte de 1654 publié par Cyrano lui-même.
>
> Cyrano created the patois of Gareau from scratch. The manuscript of the [Bibliothèque Nationale] offers quite a different language to the one printed in 1654, the pronunciation of the words is not quite the same. We have naturally maintained for Gareau the text of 1654 published by Cyrano himself.

The most distinctive Russian characters come from Ostrovskii, who gave the main stage to Muscovite merchants and their families with their vernacular, non-aristocratic language. Tolstoy's Nikita (high on both the 3-gram and keyness lists) from *The Power of Darkness* has heavily stylized speech suggestive of Western or Southern Russian dialects, e.g., featuring a word-initial [w].

It must be borne in mind, however, that dialects or accents do not automatically cause high distinctiveness – what is being detected is the *difference* in speech patterns. In a text where everyone speaks Welsh, an English character would score highly on distinctiveness, and vice versa. Cross-linguistic inference must also account for systematic language differences: the lexical and morphological features of the various languages lead naturally to different probability distributions for

both words and N-grams (although the exact nature of those differences is too complex to grapple with here). Word-based distinctiveness measures permit easier interpretation but appear less (statistically) powerful. In addition, word-based measures operate in much higher dimensions, with all the usual problems that entails (sparsity, the 'curse of dimensionality,' etc. See, for example Moisl 2011). Finally, word-based measures naturally invite lemmatization for highly inflected languages (like Russian and German), which might cause problems for future work dealing with languages that are non-standard, historical, or otherwise less well-resourced.

We have noted that our distinctiveness measure has a strong negative correlation to the size of the character. This relationship should not be understood as a simple artifact that renders our measurement useless. Distinctive speech is always a construct, a subset of linguistic and stylistic reality. If a minor character has just a few lines about gallows and graves – like Shakespeare's gravedigger – we will never know more about their language. However, *Hamlet* is not *only* about gallows and graves; if we imagine bootstrapping the gravedigger's speech, it would be endlessly populated by these few words: we don't know how the gravedigger would speak when ruling a country or murdering their uncle. From this perspective, a protagonist is more likely to represent the lexical and stylistic norm, while minor characters will sample the Other in their ethnic, dialectal, or professional distinctiveness.

Despite the few limitations, we hope that these measures of character distinctiveness will support improved theories about style, characterization, and history. The most important question to be asked concerns the source(s) of this representational distinctiveness that authors instill in their characters. To even begin to address this issue, we need much richer annotation for characters: their social class, profession, region of origins, and age. Determining the drivers of distinctiveness will not be easy. Even to carefully verify the effect of character gender was quite complicated. We know that part of the effect comes from size: women are more likely to be minor characters. However, it is reasonable to assume that gender difference can also be confounded by genre (e.g., in comedies, there are more women playing larger roles) and social class (rural people speak more in comedies). There is also the effect of time: changing the relative dynamics of character sizes (Algee-Hewitt 2017), improving the representation of women as dramatists and altering the depiction of social class, all of which complicates the analysis even further. However, having a clear summary measure for a character's stylistic distinctiveness may help us to refine our theories on the speech of fictional characters, leading in turn to better causal models.

Funding: AŠ, JB, LHL and ME were funded by the "Large-Scale Text Analysis and Methodological Foundations of Computational Stylistics" project (SONATA-BIS 2017/26/E/HS2/01019). BN is also grateful to QuaDramA, funded by Volkswagen Foundation and the DFG-priority programme Computational Literary Studies, for financing the presentation of the paper at the workshop.

Data and Code: The details of our approach, including data acquisition and preprocessing, were published in a Zenodo repository, allowing for full replication of all reported results: https://doi.org/10.5281/zenodo.7383687.

References

Algee-Hewitt, Mark (2017). "Distributed Character: Quantitative Models of the English Stage, 1550–1900." In: *New Literary History* 48.4, pp. 751–782. DOI: 10.1353/nlh.2017.0038.

Bamman, David, Ted Underwood, and Noah A. Smith (2014). "A Bayesian Mixed Effects Model of Literary Character." In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 370–379. DOI: 10.3115/v1/P14-1035.

Bonch-Osmolovskaya, Anastasia and Daniil Skorinkin (2017). "Text Mining war and Peace: Automatic Extraction of Character Traits from Literary Pieces." In: *Digital Scholarship in the Humanities* 32.suppl_1, pp. i17–i24. DOI: 10.1093/llc/fqw052.

Bronwen, Thomas (2012). *Fictional Dialogue: Speech and Conversation in the Modern and Postmodern Novel*. University of Nebraska Press.

Burrows, John (1987). *Computation Into Criticism: a Study of Jane Austen's Novels and an Experiment in Method*. Clarendon Press.

Burrows, John and Hugh Craig (2012). "Authors and Characters." In: *English Studies* 93.3, pp. 292–309. DOI: 10.1080/0013838X.2012.668786.

Craig, Hugh and Brett Greatley-Hirsch (2017). *Style, Computers, and Early Modern Drama: Beyond Authorship*. Cambridge University Press. DOI: 10.1017/9781108120456.

Culpeper, Jonathan (2001). *Language and Characterisation: People in Plays and Other Texts*. Longman.

Cyrano de Bergerac, Savinien (1921). *Les oeuvres libertines*. Ed. by Frédéric Lachèvre. Edouard Champion.

Eder, Jens, Fotis Jannidis, and Ralf Schneider, eds. (2010). *Characters in Fictional Worlds: Understanding Imaginary Beings in Literature, Film, and Other Media*. De Gruyter. DOI: 10.1515/9783110232424.

Eder, Maciej (2015). "Does Size Matter? Authorship Attribution, Small Samples, Big Problem." In: *Lit. Linguist. Computing* 30.2, pp. 167–182. DOI: 10.1093/llc/fqt066.

Eder, Maciej (2017). "Short Samples in Authorship Attribution: A New Approach." In: *Digital Humanities 2017. Conference Abstracts*. URL: https://dh2017.adho.org/abstracts/341/341.pdf.

Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling, and Peer Trilcke (2019). "Programmable Corpora: Introducing DraCor, an Infrastructure for

the Research on European Drama." In: *Proceedings of DH2019: "Complexities", Utrecht, July 9–12, 2019*. DOI: 10.5281/zenodo.4284002.

Fischer-Lichte, Erika (2002). *History of European Drama and Theatre*. Routledge. URL: https://www.taylorfrancis.com/chapters/edit/10.4324/9781315872193-37/erika-fischer-lichte-history-european-drama-theatre-hans-van-maanen.

Halvani, Oren, Christian Winter, and Lukas Graner (2019). "Assessing the Applicability of Authorship Verification Methods." In: *arXiv:1906.10551 [cs, stat]*. DOI: 10.1145/3339252.3340508.

Herrmann, J. Berenike, Karina van Dalen-Oskam, and Christof Schöch (2015). "Revisiting Style, a Key Concept in Literary Studies." In: *Journal of Literary Theory* 9.1, pp. 25–52. DOI: 10.1515/jlt-2015-0003.

Hoover, David L (2017). "The Microanalysis of Style Variation." In: *Digital Scholarship in the Humanities* 32.suppl_2, pp. ii17–ii30. DOI: 10.1093/llc/fqx022.

Kestemont, Mike, Justin Stover, Moshe Koppel, Folgert Karsdorp, and Walter Daelemans (2016). "Authenticating the writings of Julius Caesar." In: *Expert Systems with Applications* 63, pp. 86–96. DOI: https://doi.org/10.1016/j.eswa.2016.06.029.

Koppel, Moshe and Jonathan Schler (2004). "Authorship Verification as a One-class Classification Problem." In: *Proceedings of the Twenty-first International Conference on Machine Learning*, p. 62. DOI: 10.1145/1015330.1015448.

Moisl, Hermann (2011). "Finding the Minimum Document Length for Reliable Clustering of Multi-Document Natural Language Corpora." In: *Journal of Quantitative Linguistics* 18.1, pp. 23–52. DOI: 10.1080/09296174.2011.533588.

Monroe, Burt L., Michael P. Colaresi, and Kevin M. Quinn (2008). "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." In: *Political Analysis* 16.4, pp. 372–403. DOI: 10.1093/pan/mpn018.

Moretti, Franco (2013). *Distant Reading*. 1st edition. Verso.

Plecháč, Petr and Artjoms Šeļa (2021). "Applications." In: *Versification and Authorship Attribution*. Karolinum Press, pp. 69–91. DOI: 10.14712/9788024648903.5.

Reeve, Jonathan (2015). *Imperial Voices: Gender and Social Class among Shakespeare's Characters, a Stylometric Approach*. URL: https://jonreeve.com/2015/03/imperial-voices/.

Stammbach, Dominik, Maria Antoniak, and Elliott Ash (2022). "Heroes, Villains, and Victims, and GPT-3." In: *WNU 2022*, p. 47.

Sternberg, Meir (1982). "Proteus in Quotation-Land: Mimesis and the Forms of Reported Discourse." In: *Poetics Today* 3.2, pp. 107–156. DOI: 10.2307/1772069.

Székely, Gábor J. and Maria L. Rizzo (2013). "Energy Statistics: a Class of Statistics Based on Distances." In: *Journal of Statistical Planning and Inference* 143.8, pp. 1249–1272. DOI: 10.1016/j.jspi.2013.03.018.

Underwood, Ted, David Bamman, and Sabrina Lee (2018). "The Transformation of Gender in English-Language Fiction." In: *Journal of Cultural Analytics* 3.2, p. 11035. DOI: 10.22148/16.019.

Vishnubhotla, Krishnapriya, Adam Hammond, and Graeme Hirst (2019). "Are Fictional Voices Distinguishable? Classifying Character Voices in Modern Drama." In: *Proceedings of the 3rd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pp. 29–34. DOI: 10.18653/v1/W19-2504.

Appendix

A Relatively-More-Frequent Words

French		German		Shakespeare	
Female	**Male**	**Female**	**Male**	**Female**	**Male**
vous	diable	ach	der	husband	the
époux	la	o	die	you	of
mère	ami	du	teufel	alas	this
amant	les	vater	und	love	sir
mari	parbleu	mutter	ein	husbands	and
tante	maître	er	des	me	we
hélas	morbleu	mich	in	romeo	king
coeur	des	liebe	den	lysander	our
rivale	amis	mama	kerl	willow	their
ne	morgué	papa	kaiser	pisanio	duke
malheureuse	serviteur	nein	ihr	sister	three
quil	belle	dat	euch	nerissa	her
mon	vin	mein	auf	yours	to
me	un	herz	dem	o	whom
maman	heureux	gemahl	wir	pray	lordship
frère	leur	gott	könig	mother	in
fâchée	rome	geliebter	sache	nurse	stand
père	peuple	kind	also	i	noble
obligée	boire	ihn	hm	woman	ha
sûre	soldats	lieber	majestät	malvolio	dog
dorante	peste	nicht	oder	prithee	certain
il	prêt	nich	volk	my	kate
soeur	rival	sie	euer	orlando	master
amour	dé	mann	das	boyet	sword
lui	sénat	weh	unter	do	follow
heureuse	ça	dir	im	false	soldiers
que	messieurs	dich	zum	ring	his
pleurs	coquin	mellefont	freund	emilia	caesar
cruel	gens	ja	krieg	refuse	us
lamour	du	ihm	durch	troilus	law
chevalier	allons	angst	hölle	pilgrim	friends
seule	beauté	freundin	zu	windsor	york
aimée	au	wat	gnaden	would	money
lingrat	lhomme	doch	wein	rosalind	pompey
valère	par	mamachen	heer	such	england
aime	obligé	so	mit	weep	present
aimer	lé	mir	bürger	faith	warwick
maime	bon	fritz	jeder	suit	great
hans	cents	arme	herren	am	heads
chère	bâton	gurli	rom	diana	ready
ingrat	quatre	lieb	land	never	business

B Bayesian Regression Models: Effect of Gender on Distinctiveness

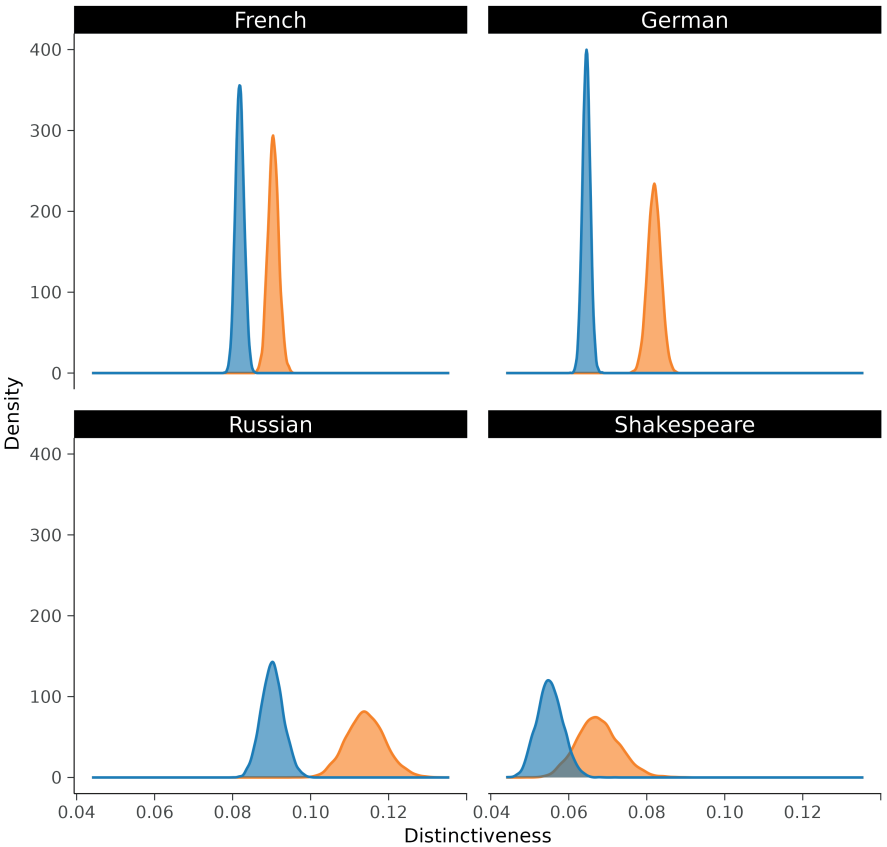


Fig. 4: Character distinctiveness, predicted from posterior, estimate of grand mean (no group-level effects), 6000 draws. Predictions are made for a counterfactual "median" character role, who has 20.9 % of dialogue share. Predictions are presented at natural scale.

Is the perceived gender effect 'real'? In technical terms, what is the direct influence of character gender (G) on distinctiveness scores (D) across traditions (T), conditioned on the share of dialogue they have (S)? To answer this, we fit a Bayesian multilevel multiple regression with group-level estimates for individual plays (P). We chose to model at the level of plays both because our D statistic is tied to the context of a single play, and because character features coming from the same

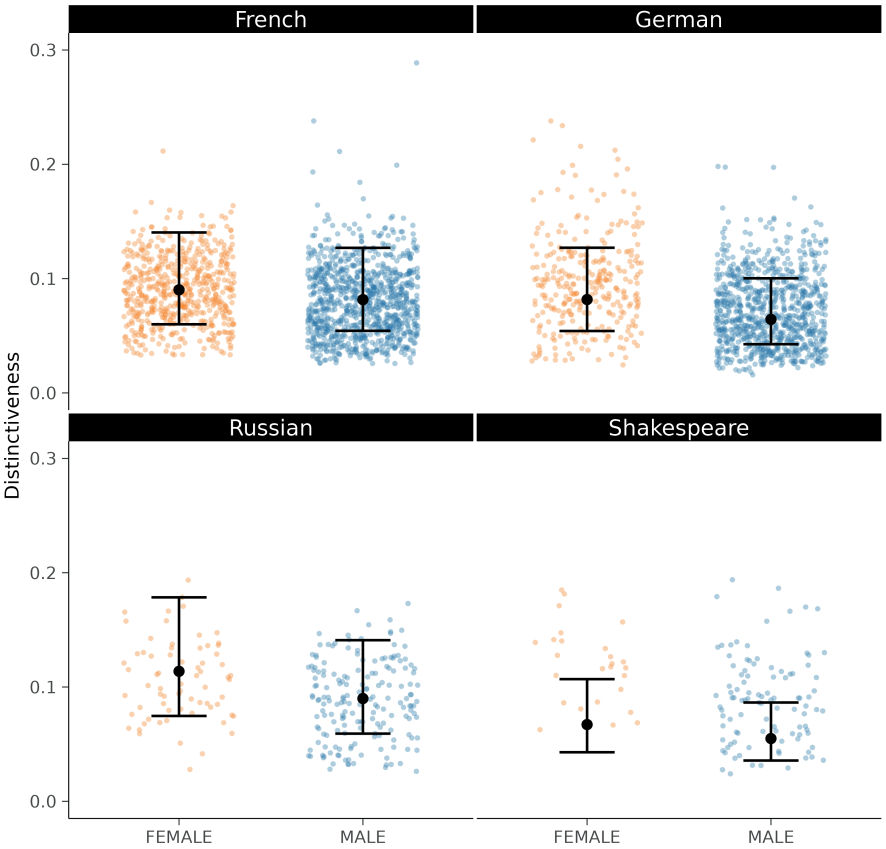


Fig. 5: Posterior predictions for gender, marginal of individual plays. Errorbars show .95 CI. Empirical data is plotted in colour, 5 extreme cases (>0.3) are filtered out. Predictions are presented at natural scale.

play are not independent (e.g. there cannot be two characters with 60 % of the dialogue). Modelling this way also significantly improved predictions. Gender is allowed to interact by corpus, yielding a single, cross-linguistic model that makes compatible predictions for different traditions. In brms formula syntax:

$$\log(D) \sim G * T + T * (S + I(S\hat{}2)) + (1|P) \tag{1}$$

Based on sample observation, we used a Gaussian prior for log-transformed D scores. We could have also fitted the original values, but D scores have extreme outliers that extend the tail: the model has much easier time with sampling and chain convergence on a log-transformed domain. We chose a quadratic term for S,

because the relationship between D and S is U-shaped. Importantly, 'unknown' gender entities are filtered, because often (but not always) this is not data that is missing, but entries that are incompatible with a binary classification:[1] primarily collective or compound entities (people, choirs, soldiers). It would have been possible to use standard strategies, like imputation, to 'repair' the data, but that approach would be incorrect.

Posterior estimates for distinctiveness by gender are shown in Figure 4. Based on the figure, we can be most confident about the difference in German and least confident in Shakespeare (few characters and, specifically, few women with large dialogue shares). The differences in means, however, appear consistent. As calculated from the posterior: in French, female characters are more distinctive by only .009 ($\pm$ .003); in German, by .017 ($\pm$ .003); in Russian by .023 ($\pm$ .009, the widest CI); and in Shakespeare by .012 ($\pm$ .008).

To understand the full extent of variation across different plays, it is useful to look at the marginal posterior means of the plays (Fig. 5). Here, the difference in distinctiveness between genders remains visible, but there is a better estimation of the global uncertainty and variation across different texts. Note that the confidence intervals in Fig. 5 are asymmetric (wider on the upper arm), having been transformed from symmetric intervals on a log domain.

1 In modern terms, it is vexing to be forced to reduce characters to a gender binary, but since gender non-conforming characters are virtually unrepresented in this predominantly historical corpus, the point is moot.

Botond Szemes and Bence Vida

Tragic and Comical Networks

Clustering Dramatic Genres According to Structural Properties

Abstract: There is a growing tradition in the joint field of network studies and drama history that produces interpretations from the character networks of the plays. The potential of such an interpretation is that the diagrams provide a different representation of the relationships between characters as compared to reading the text or watching the performance. Our aim is to create a method that is able to cluster texts with similar structures on the basis of the play's well-interpretable and simple properties, independent from the number of characters in the drama, or in other words, the size of the network. Finding these features is the most important part of our research, as well as establishing the appropriate statistical procedure to calculate the similarities between the texts. Our data was downloaded from the DraCor database and analyzed in R (we use the GerDraCor and the ShakeDraCor sub-collection). We want to propose a robust method based on the distribution of words among characters, distribution of characters in scenes, average length of speech acts, or character-specific and macro-level network properties such as clusterization coefficient and network density. Based on these metrics, a supervised classification procedure is applied to the sub-collections to classify comedies and tragedies using the Support Vector Machine (SVM) method. Our research shows that this approach can also produce reliable results on a small sample size.

1 Introduction

Do the various dramatic traditions and genres have any statistically verifiable structural features in common, i.e., are there any features of dramatic structure that guide or even presuppose the process of creating plays irrespective of the authors' individual style? Considering how interesting these questions are, the amount of debate it generates in the field of quantitative drama analysis should come as no surprise. Indeed, an exploration of this problem has the potential to offer not only a novel approach to drama and theater studies but also access to the broader mechanisms of culture. To that end, the question might be reformulated

Botond Szemes, Research Centre for the Humanities, Institute for Literary Studies Budapest
Bence Vida, Eötvös Loránd University, Budapest

https://doi.org/10.1515/9783111071824-009

in the following way: how are the relationships between people or other actors arranged according to the type of story or, vice versa, what types of stories does each arrangement make possible? An adequate methodology needs to be designed to answer these questions and identify the characteristics that make a comedy a comedy or a tragedy a tragedy. This should provide a new perspective to this comparison, especially as earlier research distinguished comedies from tragedies mainly based on their thematic differences, such as the extent to which historicity determines the setting and the period the events take place in, the emphasis on moral issues, the social status and speech patterns of the characters, or the conclusion of the plot. (Asmuth 2016; for a theme-based classification of dramatic genres with topic modelling, see Schöch 2017).

The present paper suggests that the primary key to answering these questions is the character networks, which allows the plot (Moretti 2011), interpersonal relations and societal world (Algee-Hewitt 2017), and/or dramatic form (Trilcke and Fischer 2018, and also see their study in the present volume) of a given play to be viewed simultaneously on the plane of a surface instead of in the process of reading a play or watching a performance, which inevitably unfold over time. In our study, we use the concept of *structure*, by which we mean the system of relations of characters in the broadest sense (and not only network relations, as will be explained later). This concept is closest to *form*, but it is also related to (and thus brings together) the other approaches mentioned – e.g., to plot: the analysis of structure also deals with the relationships between characters as a result of the progression of the plot (for an analysis of one aspect of this, see Section 4.1) Moreover, following the considerations and methods of SNA (Social Network Analysis), character networks can also be seen as structural representations of the contemporary cultural-historical context of a given drama; and thus plays can be understood as the author's – not necessarily intentional – attempt to model a society (Stiller et al. 2003; Stiller and Hudson 2005).

Furthermore, these networks, beyond being a cognitive tool to get a comprehensive overview of a text, can also be compared and classified statistically. The main implication of this is that the approach and terminology of graph theory can be applied to character networks drawn from the text; in our case, the characters represent the nodes and the edges are created by characters appearing in the same scene (in the case of weighted graphs, the more often the characters appear together, the thicker the edge between them). This allows the use of several numerical measures to specify the properties of the network. While specific analyses conducted on a small population sample and at the level of individuals are better suited to the SNA approach (Freeman 2004; Borgatti et al. 2009), we endeavored to use or create numerical measures that describe the plays in their entirety, so that they can be compared to each other. Our hypothesis is that the plays' plot es-

tablishes a network based on interactions (dialogue and action), which the author subjects to the criteria of the genre they chooses.

Earlier similar research also drew mainly on character networks when trying to explore the differences and similarities between dramatic genres (comedies and tragedies). These projects, however, suffer from two crucial problems that call into question the usefulness of their results. First, most papers rely heavily on the size of the networks to classify genres, which, instead of capturing the structural design of networks, simply compares the number of characters in the plays.[1] In what follows, an attempt will be made to present a method that does not take the size of the networks into account, i.e., it will be based exclusively on the system of relations between characters and, through this, on the plays' dramatic structure. The other pitfall to be avoided is the application of overly complicated mathematical procedures to investigate similarities between networks (e.g., Tsitsulin et al. 2018; Raymond 2020). The problem with this approach is that the results obtained this way cannot be attributed to real properties of the plays; such results would only show whether two works have a similar structure. By contrast, our goal is not simply to confirm that there is such a similarity but also to explain what this similarity consists of. Accordingly, the analyses below will be founded on data that truly describes how these works are structured (i.e., how they elaborate relationships and what kind of hierarchy they create between the elements): first, size-independent numerical measures of character networks, and second, values that we generated pertaining mainly to the distribution of speech and stage time among characters, which thus can describe relations in a different way, in terms of *ratios*.

2 Corpora

To explore the issue, we needed a large number of plays available in a structured form, which was provided by the database of the DraCor Project (Fischer et al. 2019). This project, a product of an international collaboration, acts as the corpora of the dramatic literature of various nations. More than a mere collection of texts, it also supplies important tools for analyzing these texts, most notably the automatic mapping of character networks and access to the data behind them. This makes it possible to use network theory measures for describing the relationships in the works, which in turn lays the foundation for a quantitative

1 See Trilcke (2013), Grandjean (2015), and Shukla (2018). In Algee-Hewitt's paper (2017), which is a milestone in quantitative drama analysis, it is unclear how independent the numerical measures are from the number of characters.

comparison of the structure of plays.[2] For the present research, the 37 plays of the Shakespeare collection and the "GerDracor" database of German plays were used. Beyond the fact that the Shakespeare collection attracts special attention in quantitative drama analysis, like in so many other fields, it also provides a well-defined sub-corpus on which to carry out tests, and its small size allows for the interpretation of the results on specific texts. GerDracor, by contrast, supplies a large number of texts, which is ideal for statistical analysis. At the moment, this database stores 591 plays, but only those with a clear genre designation (comedy or tragedy)[3], more than five characters, and more than two scenes will be analyzed below; in all other cases, character networks cannot be said to be indicative of dramatic structure. Applying these reductions, the collection finally contains 253 dramas. History plays from the Shakespeare corpus were included in the tragedy genre, since Shakespeare criticism suggests that a strong similarity can be detected between these two genres (Kott 1974; Bate 1998). Therefore, we used a binary classification (comedy vs non-comedy) in this case as well. The two corpora have different characteristics because of the divergences in their size and diversity. For example, as we will see, the number of characters in Shakespeare's plays correlates more strongly with other structural features of the genres, whereas in GerDracor the metrics are more size-independent – although it is important to bear in mind that GerDracor is much more spread out in terms of authors and time and thus represents a less coherent unit. In what follows, we take the results for the GerDracor as a baseline, as it shows statistically more reliable results on a larger scale, but the variation in the corpora also indicates that it is always worthwhile to design the method according to the corpora under study.

2 DraCor provides a standardized framework that can be used across all languages, with its core being the dramatic texts encoded in TEI XML format. In the process of coding, the annotators identify the plays' structural elements – the acts and scenes –, label the characters separately, and also indicate which characters appear in a given act or scene. This helped create the networks that were extracted for the present study (using the simple network visualization tool of the interface) to complement the lists of characters available in .csv format and downloadable one by one for each play. Therefore, the raw data contains the structure of each play, along with unique identifiers of characters appearing in the same scene. In the present study, connections were established between those characters that appear at the same time in the same scene according to the annotated list of characters.

3 The metadata in GerDracor's corpus also specify the so-called "normalized genre tags," which standardize the different designations that refer to the same genre tradition (e.g., *Lustspiel* and *Komödie*).

3 Numerical Measures

To determine the size-independent but distinctive properties of plays, the following numerical measures were considered and in some cases developed for this study. The first four measures are considered standard in SNA studies, and they are featured in most papers dealing with quantitative drama analysis.

Average Clustering Coefficient: This measure indicates the proportion of nodes connecting to distinct sub-units that have a high number of edges in a given network. The small-world principle states that most social networks are highly clustered (Watts and Strogatz 1998).[4]

Density: A ratio of the number of actual edges in a network to the number of possible edges. The denser the network, the more connections can be observed in it, which provides information on the network's complexity and type. For example, a democratic network (informal group of friends) has a higher density than a hierarchical one (chain of command).

Average Path Length: In the network, a path can be drawn between any two nodes through the edges. This measure shows the average length of the shortest path that can be drawn between any two nodes (the path that touches the least number of nodes).

Diameter: Finding the shortest path between nodes in a network is crucial to defining its basic properties. The diameter represents the longest of these paths, which is also the distance between the furthest nodes.

Maximum Betweenness: The betweenness centrality for node A indicates the number of possible shortest paths that connect any two nodes of the network and touch A, relative to all possible shortest paths. In other words, betweenness is the centrality measure that assesses the mediating role of a given node and thus answers the question of how many other nodes are involved in information transfer or path formation between nodes. The Maximum Betweenness indicates the maximum value of these centralities. The measure is particularly useful for identifying nodes that may act as bridges between clusters in a network.[5]

Average Degree – Max Degree Ratio: Degree connotes the number of nodes that any given node is connected to via edges. From this number, the average and maximum degree for the network as a whole can be calculated. This is done so that

4 Stiller et al. (2003) demonstrate that a large cast of characters increases the clustering of the play's character network, which can thus be described as a small world. Clustering can also be used to identify weak ties, an important subject in SNA that measures the resilience of social networks and the effectiveness of information diffusion (cf. Granovetter 1983; Albert et al. 2000).

5 This measure was also used by Algee-Hewitt (2017) in his research.

statements can be made that pertain to the network itself. This measure therefore gives the proportion of the average degree number relative to the maximum degree number and, in a sense, resembles density. The extent to which this measure can be considered an independent characteristic that describes networks from a unique point of view will be discussed later.

Maximum Degree – Number of Characters Ratio: In this case, to get the total number of characters correct, a value of one must be subtracted from the actual number of characters, since a node does not interact with itself. The ratio shows what percentage of the entire cast the node with the highest degree interacts with.

High Speech, Medium Speech, Low Speech: These three measures refer to the proportion of characters who speak 'much,' 'little,' or 'a moderate amount' during the play. This was calculated according to the procedure outlined in the Stanford Literary Lab's *Pamphlet 14* (Kanatova et al. 2017), which used the k-means clustering procedure to classify movies into 3 groups (k=3) based on the amount of analepses and prolepses they contained, thus distinguishing works that scored 'high,' 'medium' and 'low' from each other (i.e., the group with the highest mean value constituted the 'high' category and so on.) In the present paper, k-means clustering was employed (k=3, likewise) to distinguish between characters in plays who are not just members of a group in the stage (for which DraCor uses the annotation "isGroup = True", e.g., 'street vendors'). The three groups were then ranked according to their mean value and the number of characters in each category adjusted to the total number of non-group characters in the play. This ratio makes it possible to compare plays with different numbers of characters, since it shows not only the number of characters who speak a lot/medium amount/less often in a text, but also the proportion of characters who do so, which essentially gives us an idea of the distribution of linguistic complexity among characters.

High Weighted Degree, Medium Weighted Degree, Low Weighted Degree: The same method was employed to calculate the weighted degree of characters as nodes. Here, the proportion of characters with many, some or few connections in the network was examined. Since the result is always adjusted to the total number of characters, this measure is also size-independent, which allows us to compare works with different numbers of characters. Our previous measures showed that the metrics of weighted degree (which takes into account not only the number of edges leading into a node but also their 'weight,' in this case the frequency of appearing together in the same scene) better captures the differences between genres and the specific characteristics of plays than using only the degree for the calculations.

Average Character Speech: This shows the average number of words spoken by a character in a play. Only characters who speak more than 10 times were considered.

Average Character Per Scene: First, the average number of characters in one scene is calculated. To make texts of different lengths comparable, this number is divided by the total number of scenes.

Number of Connected Components: The number of groups in the network within which any node is connected to any other node by edges (Minimum = 1).

Before presenting the results, a discussion on how these measures relate to each other – whether a correlation can be detected between them – is in order. Are these measures independent of each other and of the size of the networks? The upper part in the correlation matrix of Figure 1 created from the values of 253 plays of GerDraCor shows these connections, while the lower part shows the results based on the Shakespeare collection. In the upper triangle (which serves as baseline to our study) the 'number of connected components' is strongly correlated with the number of speakers (i.e., the size of the network), which is hardly surprising, since it is almost exclusively found in the case of plays with a high number of characters, which means the network can separate into multiple unconnected units. Thus, this measure was not used in subsequent analyses. In addition the correlation between density, average path length and average degree – maximum degree ratio is very strong (>0.9). For this reason, the latter two measures were also excluded from the study, since density is one of the most important and informative properties of character networks, which is why it was deemed a more characteristic, more descriptive measure than the other two. Furthermore, the exclusions make sure that the strong correlation between diameter and average path length (0.88) do not influence the subsequent results. Although there are still some pairs of values that are correlated (e.g., density – average character in one scene), the relationship between them is not exclusive, i.e., they cannot be said to describe the same phenomenon or phenomena that are in a deterministic relationship (see Figure 2). Regardless, it should be noted that they reinforce each other and have a joint effect in the clustering process (cf. Shukla 2018).

The lower triangle of Figure 1 shows that a smaller corpus of one author is organized in a different way. Here we can observe a stronger correlation between density and the size of the networks. Perhaps this, together with the fact that almost exclusively Shakespeare's texts were studied previously in similar research, made it important to also involve the number of characters into the classification process. However, it seems that for measurements on a larger scale, the features are less dependent on the number of characters, which would also play a less important role in the classification (see later).

The differences between the 13 variables (which remained after taking into account the correlations based on GerDracor) across genres are shown in Figure 3 and can be considered significant based on the Wilcoxon Rank Sum Test, except for the category of the proportion of the medium weighted degree (this test was

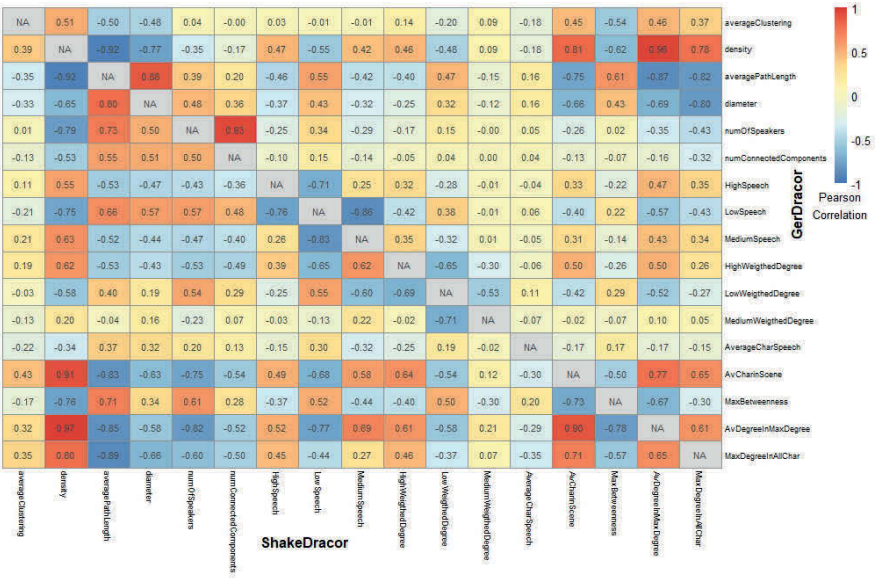


Fig. 1: Correlation matrix of the values used in the research. Upper triangle based on the GerDracor, lower triangle based on the ShakeDarcor database.

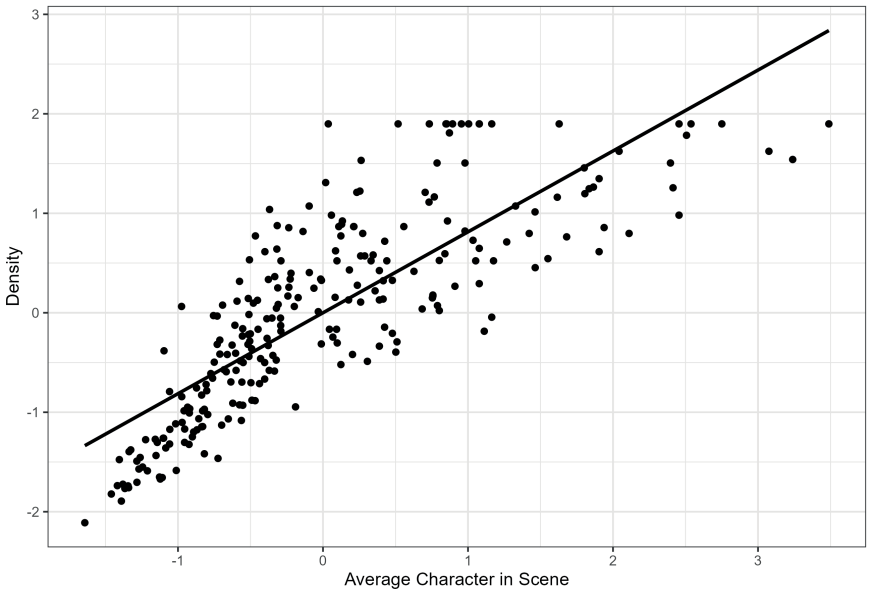


Fig. 2: Correlation between density and average character per scene based on the GerDracor database. Normalized values presented (z-scores).

chosen instead of the t-test because the variables do not follow a normal distribution). The calculations were then performed using the normalized values (z-score) of the 13 variables for the corpus in question, so that differences of different scales would be given equal weight when determining genre groups.

4 Results

As a first experiment, a supervised classification procedure was applied to the GerDracor sub-collection using the leave-one-out (LOO) method. In other words, one of the 253 dramas was always withheld at the beginning of the experiment, then the remaining 252 works were divided into comedies and tragedies using the Support Vector Machine (SVM) classification procedure. This procedure can be viewed as if each play were a data point in a 13-dimensional space, where the SVM 'drew' the hyperplane that best separates the categories (since we used a linear kernel, this plane is created without curvature). We then examined which side of the space (comedy/tragedy) the withheld work would lie on based on the values assigned to the variables, and assessed the correctness of the binary classification on this basis. This withholding was repeated with every single play (separately for tragedies and comedies). The accuracy of the classification was finally gauged by the average of these numbers, which is represented by the mean accuracy score – which in this binary case is the same as recall.

Since the corpus contained an uneven number of tragedies and comedies, the probability for random classification was not 0.5 but 0.54 (136/253) for comedies and 0.46 (117/253) for tragedies. The experiment produced an accuracy rate of 0.82 (F1 score = 0.77, precision = 0.73, recall = 0.82) for comedies and 0.65 (F1 score = 0.7, precision = 0.76, recall = 0.65) for tragedies, which suggests that the model is more accurate than random classification and its performance is quite good, but far from flawless. If we include the number of characters as a feature to the model, the performance is roughly the same: 0.83 mean accuracy rate for comedies and 0.64 for tragedies; furthermore, this feature proved to be the second least important in the classification after testing the role the number of characters plays in the model (methods described later). This means that in this case, the size of the network is far less relevant in determining the genre of a play than previous studies suggested. Moreover, the promising results of the original model imply that the properties examined here can indeed confirm the existence of a 'genre fingerprint' that shapes the dramatic structure of tragedies and comedies. However, it is worth pointing out that this fingerprint does not determine structure categorically but rather provides a scheme from which individual authors often deviate.

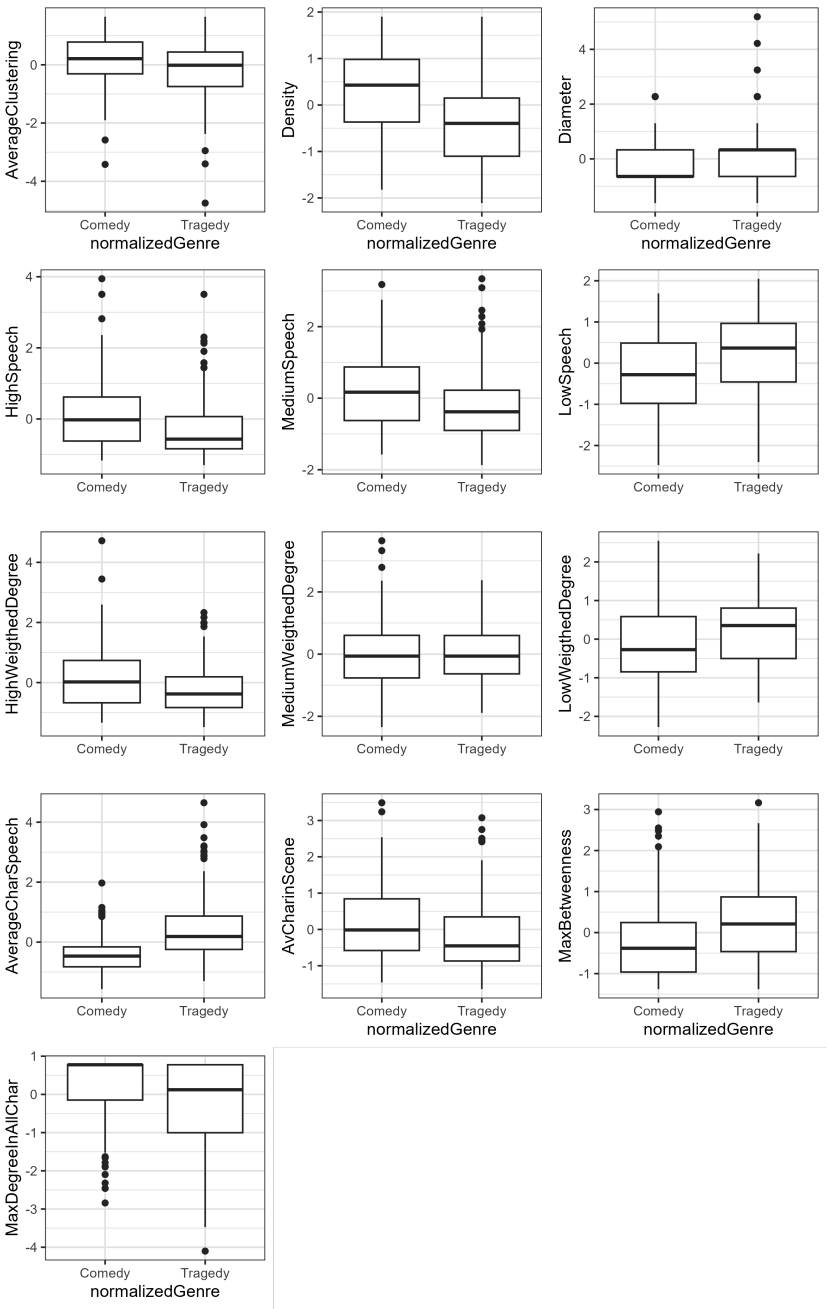


Fig. 3: Differences in the features by genre based on the GerDracor database. Normalized values presented (z-scores).

Figure 4, generated by principal component analysis (PCA), attests to the same phenomenon: there is no striking difference between comedies and tragedies in their structural properties, but one can detect a tendency of the comedies to gravitate towards the left side of the figure and the tragedies to the right side. The outliers are also positioned along the same lines. Furthermore, the figure shows which side each variable can influence on the plane generated by the PCA's reduction of dimensions, i.e., how the variables affect the location of data points. The possibility of such a visualization and the satisfactory results of the SVM require us to keep all the variables in the model – although the RFE (Recursive Feature Elimination) procedure for feature selection[6] suggests leaving the features of High, Medium and Low Medium Weighted Degree out of the research, as they play a less prominent role in distinguishing genres. But the difference between the accuracy scores of the remaining 10 features (0.72) and of the whole set (0.7) is not as relevant as the importance of taking into account the most probable aspects in defining the specificities of the genres – and thus obtaining a ground for more complex interpretations. According to the PCA graph, for example, most of the comedies are marked by dense character networks and scenes with multiple characters, they have more characters who appear with many characters in the same scene, more characters who talk a lot throughout the play, and finally the character connected with the highest number of characters can be linked with a high percentage of the entire cast. By contrast, in prototypical tragedies there are more characters who speak very little and appear with few others in the same scene, but the length of utterances per person is longer (i.e., there are more monologues), the diameter of the network is larger (i.e., a character might mediate between two other characters who do not interact with each other directly), and one or two characters with a special 'connecting' function are more often responsible for the stability of the network, i.e., they are typically the ones through whom information is passed between the different subgroups (maximum betweenness).

An examination of the Shakespeare corpus with the same setup yields similar results, which can more easily be put into practice at the level of individual dramas. Here, when included, the number of characters becomes the most important feature of the classification, which has an accuracy of 0.71 for comedies (random at 0.37) and 0.83 for tragedies (random at 0.62). Without the number of characters, tragedies stay at 0.83; however, comedies decline to 0.6, which also shows the

6 The RFE works in two steps: first it calculates the importance of the features (based on the role they play in the classification), then it removes the least important ones iteratively and calculates the accuracy results for each scenario.

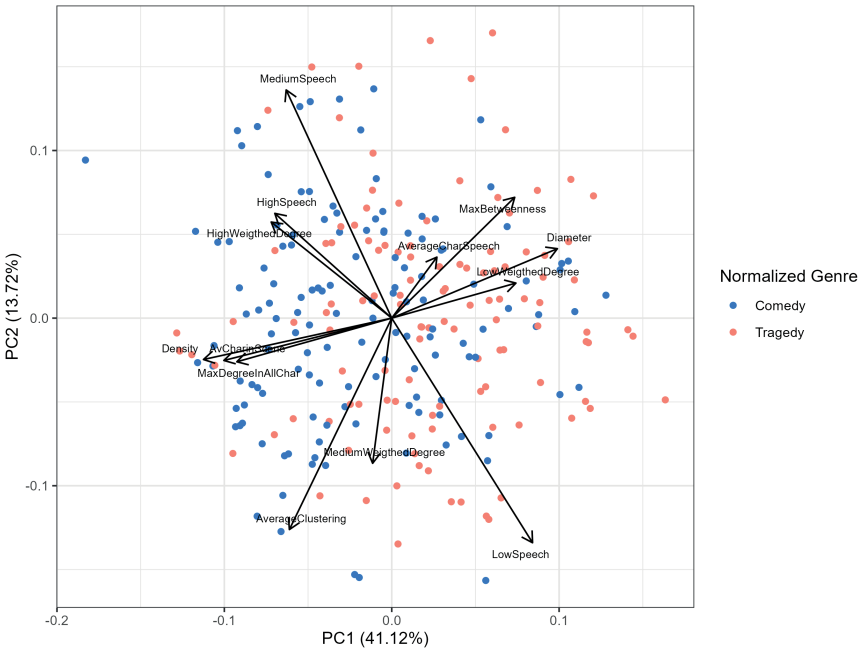


Fig. 4: Principal Component Analysis based on the measured features – 253 dramas from GerDraCor.

importance of the network size in this case. But we ought to bear in mind as well that the small size of the Shakespeare corpus provides less reliable results than the GerDracor corpora. More interesting is the PCA diagram (Figure 5), which shows a clearer distinction between genres than the German sub-corpus, even though the variables exert their influence on very similar areas. In the figure, the prototypical comedies can be easily identified (*The Merry Wives of Windsor, Love's Labour's Lost, A Midsummer Night's Dream, Twelfth Night, or What You Will, The Comedy of Errors* and, somewhat surprisingly, *The Tempest*, which most scholars would argue is a blend of genres) as well as the non-comedies (incidentally, their location on the figure confirms our hypothesis that history plays and tragedies do not differ, at least not in terms of their structure). Works that depart from the schema are also conspicuous in the figure. The reason why *The Taming of the Shrew* falls close to tragedies most probably lies in its use of frame scenes, since that increases the size of networks, decreases its density, and reduces the number of characters who are connected to each other and speak frequently or moderately often – all of which evoke the structure of tragedies.

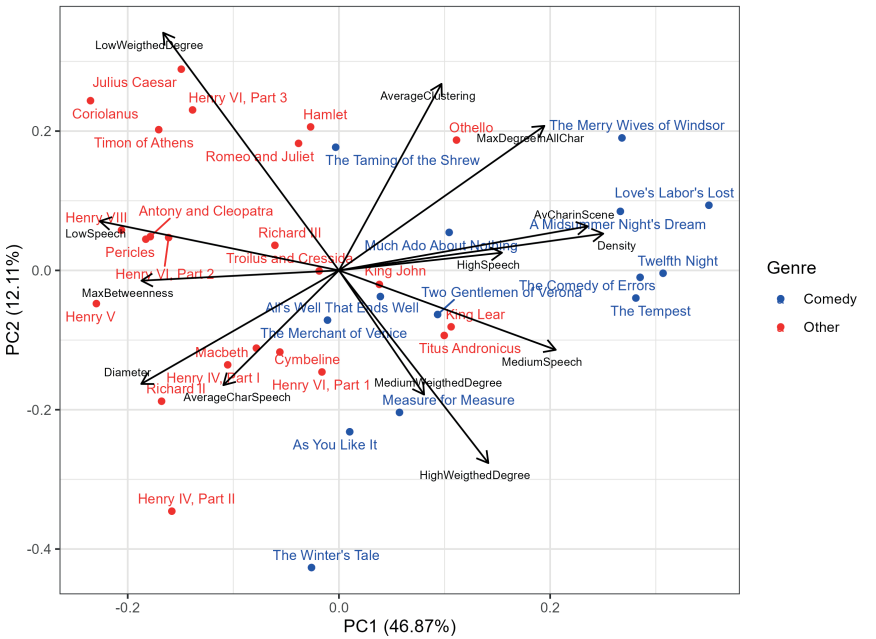


Fig. 5: Principal Component Analysis based on the measured features – 37 dramas of the Shakespeare corpus.

The Winter's Tale also unmistakably stands apart from the rest of the plays, which can be related to its elusion of distinct genres; accordingly, even its classification as a comedy is debatable. More interesting are the tragedies that appear among the comedies, such as *Titus Andronicus*, *King Lear*, and *Othello*. *Titus Andronicus*, one of Shakespeare's early plays, is a long revenge tragedy in barely structured acts and scenes featuring a high number of minor characters.[7] On several occasions, the ill-structured scenes include situations in which sub-groups of characters exit or enter the stage without interacting with each other. Unfortunately, since our model relies on the annotations from DraCor, it does not detect this movement, as it assumes a connection between all the characters in a scene. In the future, a solution to deal with such situations could be to introduce a method that focuses on such subgroups and the two-way nature of interpersonal dialogues; such a method would indicate a transition even within a scene if the set of speakers changes (see Seung-Bo et al. 2012; Labatut and Bost 2019). However, the case is different with the other two tragedies. There, the tragic plot un-

7 Shukla (2018) also struggles with classifying this work; they treat it as a comedy.

folds in a framework that would also be suitable for comedies. Ultimately, the tragic nature of these plays stems precisely from the fact that the not particularly tragic arrangement of social relations (i.e., the exchange of information between characters appears not to be overseen by one or two characters) still leads to fatal mistakes and punishment. The plot of *Othello* is set almost entirely within one subgroup, the household of the main characters is shattered only as a culmination of the dramatic twists in the last scenes. Shakespeare's trick is to pick a domestic environment as a setting for what is essentially a tragedy played out in the public sphere and motivated by a lust for political power. The conflict upsetting the social order thus remains hidden throughout, lurking in the depths of long dialogues, only to suddenly tear apart the tense structure of the surrounding network. Part of the reason Iago can carry out his plan without a problem is that he divides the characters' trust-based, prototypically non-tragic system of relations into subgroups that can be played against each other.[8] Susan Snyder drew attention to the comedy-like nature of the drama as early as 1979 (Snyder 1979, pp. 70–74), which is confirmed not only by the structural features but also by the linguistic characteristic of the text (Hope and Witmore 2010, p. 374) – which also highlights the relationship between these levels. A similar dramatic device is at work in *King Lear*. A number of elements evoke the world of comedies on a thematic level: the populous family where the king is not at the center (at least according to the centrality measure, which is also low for the other characters), the figure of the Fool who probes issues associated with language philosophy and ontology in playful prattles, and the other characters, who come to resemble him more and more (especially Kent, who pretends to be a fool), the misunderstandings caused by the disguises, and the forest as a setting, which in comedies is the space where one can indulge in desires repressed in the social sphere. The measuring we conducted confirm the affinity with comedies. The structure of King Lear promises a comedy, albeit not a prototypical one: the fact that it does not follow that path is the source of its tragedy. Lear is unable to 'live happily ever after' even though his prospects seem good, as the play opens with what comedies usually end with: a marriage. It ends, however, in the traditional setting of tragedies – among graves (István 1999). Thus, the play ambitiously traces an arc from comedy to tragedy, raising questions about the conventions of both.

8 Stiller and Hudson (2005) describe this phenomenon from a different perspective by focusing on edges in their analysis of *Othello*: Iago consistently tries to manipulate those partners of his who, like himself, are in a "connectional" role (Cassio, Othello), in such a way as to slowly take control of the flow of information, thus eroding the trust of the others and the stability of the network. His work succeeds at the climax of the play, as the network falls apart from within in the escalating conflict.

This leads us to a question of the connection between plot development and the metrics used in this study. This is the question captured by an often-cited observation ascribed to Bernard Shaw, according to whom the two genres are distinguished in Shakespeare according to whether they end with a wedding (where all the characters are present, so the network of characters becomes denser) or the death of the characters (causing a less dense network) (Trilcke et al. 2015). From this angle, *King Lear* could be understood as transitioning from one genre to the other even on a structural level, because both plot points are emphatically present in it. But how can we model the connection computationally?

4.1 Results Without the Last Act

To put this question to the test, we repeated our earlier measurements first for the Shakespeare corpus (because this is the domain where the question most often arises), but this time the character network of the plays was based only on the relational structure of four acts – i.e., one act was always disregarded. We examined how this would change the measures describing the network as a whole (in this case density) for tragedies and comedies and whether the difference between the two remains significant. First and foremost, we were interested in the role of the last act, which is of particular importance to the plot (data for the other acts is available in the project's GitHub repository).

The importance of the last act lies in the fact that, according to some researchers, tragedies typically end with the isolation of the protagonist, who gradually becomes a lonely figure and suffers losses, whereas the conclusion of comedies is a large crowd scene on stage (Smith 2019). Lee and Lee (2017), by contrast, seek to demonstrate that the final scene is always a crowd scene, since the conclusion of tragedies also requires multiple characters and the real difference hinges on the type of the relationship between them. In their view, the tragic conclusion is caused precisely by the close and dense arrangement of the characters, which allows the conflicts to unfold. If true, this stresses the importance of the nodes that create weak links, since it is their role to disrupt the stability of the network: it is because of them that the initial network of the play turns into that of the denouement (for a similar interpretation about *Hamlet* and Horatio's role see: Moretti 2011).[9] This theory is not confirmed by the results of the measurements, which rather support the distinction made in Shakespeare

9 This argument can be linked to an idea that has been put forward from time to time in Shakespeare scholarship, which argues that the overarching motif of the Shakespeare corpus is the destruction of the established social order. Network analysis shows that this destruction is not

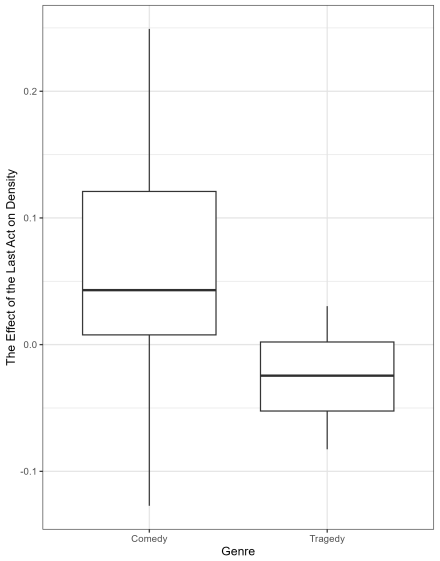


Fig. 6: The effect of the last act on density in the character networks of comedies and tragedies in Shakespeare. First, we calculated scores for each play that show the difference between the values based on the whole drama, leaving out the last act. If the number is positive, the score for the whole play is greater than without the last act (positive effect). If the number is negative, the value of the shortened drama is greater (the ending reduces the density). We grouped the results by genre and visualized them as a boxplot.

criticism. Indeed, removing the last act on average decreased the density of the networks in comedies and increased it in tragedies (see Figure 6, which shows the effect of the last act; by simply subtracting the score without the final act from the total density, a positive number indicates that the measure is higher when the conclusion of the plot is included). This means that in the Shakespeare corpus, the structure of the networks is indeed influenced by the genre-specific closure of the plot. This is not apparent in the case of GerDracor, but for both corpora, on average only the last act has a positive effect on density in comedies (Figure 7) – which can be explained also by the fact that new characters are rarely introduced in the last acts of comedies. But at the same time, a nuance in the picture is that the difference in density also remains significant without the last act based on the Wilcoxon-test (p-value for the whole drama: 0.0003702; p-value without the last acts: 0.00002745). This supports the idea that the structural properties of the plays (at least in the Shakespeare corpus) are influenced by the development of the plot, although not exclusively. The plot is only one component, though an essential one, in the formation of the relationships between the characters.

self-serving or anarchic but transformative, with the key agents of the new structure always doing their work before the audience, bringing the outcome closer, gossip by gossip, intrigue by intrigue.

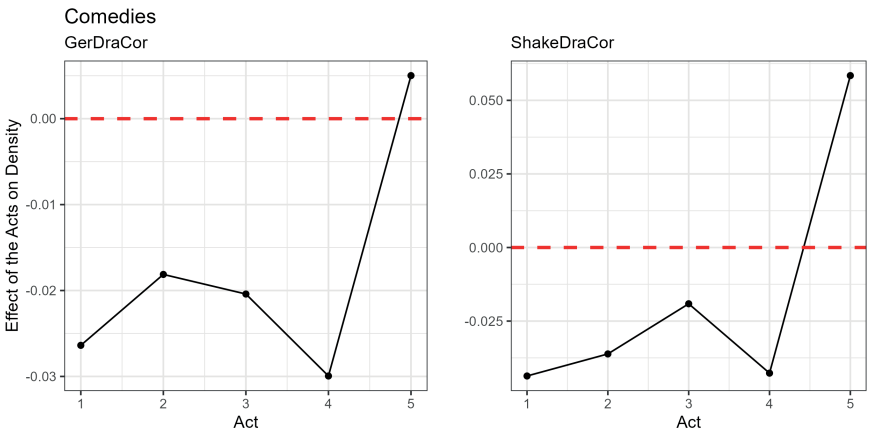


Fig. 7: The effect of the acts on the density of the character networks in comedies. From GerDracor we only used the five-act comedies/tragedies (128 plays).

5 Conclusion

It can be concluded that the measures of network theory and the distribution of speech and stage time between characters can be helpful in identifying a genre fingerprint that can be used to distinguish between comedies and tragedies. We have reached this conclusion by using a larger corpus than those of earlier relevant studies, by drawing on a number of measures, and by testing their combined use, always keeping in mind that our results should be as independent as possible of the size of the networks.

This suggests that prototypical comedies are characterized by denser character networks and fewer prominent characters who control the exchange of information, talk notably more than their peers, or have significantly more connections than others. In such a network, multiple pieces of information circulate simultaneously, not only leading to misunderstandings and frequent situational comedy but also preventing the will of a single character and a single worldview from prevailing in the play. The structure thus avoided is the logic of tragedies, in which networks are broken down into subgroups, one or two key figures act as the link between them, and the majority of the characters play only a peripheral role in the development of the plot (both in terms of number of links and amount of speech). While these arrangements allow for more effective communication by allowing the spread of a single or small number of truths and values (thus limiting misunderstandings and duplicities but increasing vulnerability and the possibility

of deliberate deception), they are also more fragile, since the failure of this 'one truth' can lead to the permanent disintegration of the whole network.

However, these genre features are characteristic of ideal types – not all comedies and tragedies follow these rules so neatly. This can be witnessed in Figure 4 and Figure 5, which show overlapping categories rather than a clear divide. Although the aim of all research in relevant fields of digital humanities is to develop clustering procedures that are as efficient as possible, the discipline should not abandon the investigation of categories that are based on family resemblance and prototype theory. Indeed, beside the classic approach of the philosophy of science, a principle of categorization suggested by cognitive psychology should also be considered (Rosch 1975; Rosch and Warren 1977). According to the former, "categories are defined by necessary and sufficient conditions; properties are binary (either present in the instance or not); belonging to a category is a matter of yes-no decision; categories have precise boundaries, all members of the categories are equal, the subordinate concept has all properties of the superordinate category" (Tolcsvai Nagy 2005, pp. 6–7). By contrast, cognitive psychology argues that in some cases "specimens are categorized on the basis of recognized characteristics; the main principle of classification is family resemblance [...]; categorization is a matter of degree, with some specimens being central, 'good,' and others less good, corresponding to fewer characteristics; the outlines of categories are not defined, they may overlap" (Tolcsvai Nagy 2005, pp. 6–7). An examination of 253 plays from the GerDracor corpus clearly suggests that prototype theory is better suited to describe the difference between comedies and non-comedies. In other words, the categories of genre classification are not absolute but part of a spectrum: even plays that are clearly considered tragedies contain features associated with comedies – such as *King Lear* and *Othello* in the Shakespeare corpus. According to the x axis in the PCA plot of Figure 5, we can mention *Comedy of Errors* and *Henry V.* as examples of central members of the comedy and tragedy prototypes, respectively. The former builds up a denser network of multiple protagonists and moves more characters into the scene at once; the latter has a larger network diameter and is characterized by more supporting characters and/or characters with high betweenness centrality. This can be seen in Figures 8 (a) and (b): while in *Henry V*, the sub-groups of characters are linked in a chain through some central, mediating character, in *The Comedy of Errors*, there is a single tightly linked group at the center of the network, loosely connected to the more peripheral characters. Of course, the procedure presented here could be improved and replaced by a more effective model, yet the point of studies in this field – unlike, for instance, authorship attribution studies – may not be the making of strict distinctions but the detection of prototypical features and the identification of instances that conform to or deviate from the scheme.

Finally, the interdependence between the plays' structure and the plot should be noted. Measurements made without the final acts have shown that the evolution of character networks in the Shakespeare corpus is influenced by the outcome of the plot (densely populated scenes of marriages vs. the gradual accumulation of character deaths). In the German corpus, such a strong correlation could not be observed between the parts of the plays and the change in the different features, but the relationships of the characters (whether we mean by this term the networks or the distribution of stage time and speech) in the plays cannot be considered independent of the unfolding of the story. We rather prefer to think of this as a reciprocal connection. It cannot be argued that it is the structural features of comedies and tragedies that create comic and tragic plots; neither can it be argued, that the characters' system of relations is determined exclusively by the plot development. Rather, there is an interaction between the two sides, certain stories can unfold within a certain framework, but in doing so they also influence the structure of that framework. There have also been examples of a plot of one genre combining with the structure of another genre, in which case it is the examination of the resulting tension that can lead to important insights.

In the future, it is worth adding another aspect of temporality to these results: are genre differences considered to be constant over time, or do these differences evolve historically? With such a study it would then become possible to examine how dramatic forms change and are preserved within a framework of cultural evolution.[10]

Contributions: Botond Szemes: concept, statistics, visualization, writing; Bence Vida: TEI XML.

Data and Code: The data extracted from the DraCor database was analyzed in the R programming environment; the data for testing the effect of the acts on the network structure (4.1.) was produced in python – the codes and the results were made publicly available in the following GitHub repository: https://github.com/SzemesBotond/drama_cluster_genre.

Acknowledgment: Szemes Botond was supported by the ÚNKP-22-4 New National Excellence Program of the Ministry for Culture and Innovation (Hungary) from the source of the National Research, Development and Innovation Fund.

10 This discipline studies the change, persistence, and diffusion of information acquired through social learning – that is, it considers cultural processes to be captured along the dynamics of innovation and preservation. To model this, it draws on the insights and terminology of evolutionary theory (see e.g. Mesoudi 2017).

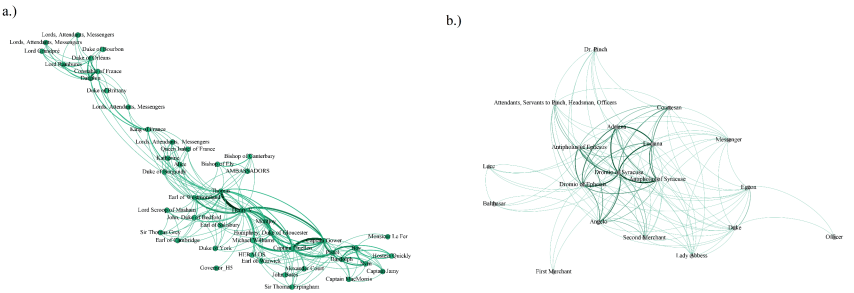


Fig. 8: Weighted character networks based on co-occurrence in a scene. Network data was downloaded from the Dracor website and processed in the Gephi software using the ForceAtlas2 algorithm – a) *Henry V.*, b) *Comedy of Errors*.

The authors are grateful for the help of Mihály Nagy (ELTE University) in the computations of the network metrics without given acts; and Jessie Labov (Research Centre for the Humanities, Institute for Literary Studies Budapest) in proofreading the text.

References

Albert, Réka, Hawoong Jeong, and Albert-László Barabási (2000). "Error and attack tolerance of complex networks." In: *Nature* 406, pp. 378–382.

Algee-Hewitt, Mark (2017). "Distributed Character: Quantitative Models of the English Stage, 1550–1900." In: *New Literary History* 48.4, pp. 751–782. DOI: 10.1353/nlh.2017.0038.

Asmuth, Bernhard (2016). *Einführung in die Dramenanalyse*. J.B. Metzler. DOI: 10.1007/978-3-476-05472-2.

Bate, Jonathan (1998). *The Genius of Shakespeare*. Oxford University Press.

Borgatti, Stephen, Ajay Mehra, Daniel Brass, and Giuseppe Labianca (2009). "Network Analysis in the Social Sciences." In: *Science* 323, pp. 892–895. DOI: 10.1126/science.1165821.

Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling, and Peer Trilcke (2019). "Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama." In: *Proceedings of DH2019: „Complexities*. Utrecht University, pp. 1–6. URL: https://dracor.org/.

Freeman, Linton (2004). *The Development of Social Network Analysis*. Booksurge Publishing.

Grandjean, Martin (2015). *Network Visualization: Mapping Shakespeare's Tragedies*. URL: http://www.martingrandjean.ch/network-visualization-shakespeare/.

Granovetter, Mark (1983). "The Strength of Weak Ties: A Network Theory Revisited." In: *Sociological Theory* 1, pp. 201–233. DOI: 10.2307/202051.

Hope, Jonathan and Michael Witmore (2010). "The Hundredth Psalm to the Tune of 'Green Sleeves': Digital Approaches to Shakespeare's Language of Genre." In: *Shakespeare Quarterly* 61.3, pp. 357–390. DOI: 10.1353/shq.2010.0002.

István, Géher (1999). *Shakespeare-olvasókönyv*. Corvina. URL: https://www.libri.hu/konyv/geher_istvan.shakespeare-olvasokonyv-1.html.

Kanatova, Maria, Alexandra Milyakina, Tatyana Pilipovec, Artjoms Šeļa, Oleg Sobchuk, and Peeter Tinits (2017). "Broken Time, Continued Evolution: Anachronies in Contemporary Films." In: *Pamphlets of the Stanford Literary Lab* 14. URL: https://litlab.stanford.edu/LiteraryLabPamphlet14.pdf.

Kott, Jan (1974). *Shakespeare Our Contemporary*. Norton.

Labatut, Vincent and Xavier Bost (2019). "Extraction and Analysis of Fictional Character Networks: A Survey." In: *ACM Computing Surveys, Association for Computing Machinery* 52.5, p. 89.

Lee, James Jaehoon and Jason Lee (2017). "Shakespeare's Social Network; or Why All the World's a Stage." In: *Digital Humanities Quarterly* 11.2. URL: https://www.digitalhumanities.org/dhq/vol/11/2/000289/000289.html.

Mesoudi, Alex (2017). "Pursuing Darwin's Curious Parallel: Prospects for a Science of Cultural Evolution." In: *Proceedings of the National Academy of Sciences* 114.30, pp. 7853–7860.

Moretti, Franco (2011). "Network theory, plot analysis." In: *New Left Review*, pp. 80–102.

Raymond, Cameron (2020). *Classifying Shakespeare with Networks*. URL: https://cameronraymond.me/blog/shakespeare-netlsd/.

Rosch, Eleanor (1975). "Cognitive Representations of Semantic Categories." In: *Journal of Experimental Psychology: General* 104.3, pp. 192–233. DOI: 10.1037/0096-3445.104.3.192.

Rosch, Eleanor and Neil Warren (1977). *Human Categorization*. Academic Press.

Schöch, Christoph (2017). "Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama." In: *DHQ* 11.2. DOI: https://doi.org/10.1073/pnas.1620741114.

Seung-Bo, Park, Oh Kyeong-Jin, and Jo Geun Sik (2012). "Social network analysis in a movie using character-net." In: *Multimedia Tools and Applications* 59.2, pp. 601–627.

Shukla, Manisha (2018). "Theatrical Genre Prediction Using Social Network Metrics." In: *Graduate Theses and Dissertations*. URL: https://scholarworks.uark.edu/etd/2920.

Smith, Emma (2019). *Not Shakespeare: Elizabethan and Jacobean Popular Theatre*. Podcast. URL: https://podcasts.ox.ac.uk/series/not-shakespeare-elizabethan-and-jacobean-popular-theatre.

Snyder, Susan (1979). *The Comic Matrix of Shakespeare's Tragedies: Romeo and Juliet, Hamlet, Othello, and King Lear*. Princeton University Press.

Stiller, James and Matthew Hudson (2005). "Weak Links and Scene Cliques Within the Small World of Shakespeare." In: *Journal of Cultural and Evolutionary Psychology* 3, pp. 57–73. DOI: 10.1556/JCEP.3.2005.1.4.

Stiller, James, Daniel Nettle, and Robin Dunbar (2003). "The small world of Shakespeare's plays." In: *Human Nature* 14, pp. 397–408. DOI: 10.1007/s12110-003-1013-1.

Tolcsvai Nagy, Gábor (2005). "A nyelvi kategorizáció kognitív nyelvészeti keretbe." In: *Az Eszterházy Károly Főiskola tudományos közleményei*. Új sorozat 32, pp. 5–20. URL: http://publikacio.uni-eszterhazy.hu/id/eprint/4501.

Trilcke, Peer (2013). "Social Network Analysis (SNA) als Methode einer textempirischen Literaturwissenschaft." In: *Empirie in der Literaturwissenschaft*. Ed. by Katja Mellmann, Philip Ajouri, and Christoph Rauen. Mentis, pp. 201–247.

Trilcke, Peer and Frank Fischer (2018). "Literaturwissenschaft als Hackathon. Zur Praxeologie der Digital Literary Studies und ihren epistemischen Dingen." In: *Wie Digitalität die Geisteswissenschaften verändert: Neue Forschungsgegenstände und Methoden. Sonderband der Zeitschrift für digitale Geisteswissenschaften* 3. Ed. by Sybille Krämer and Martin Huber. DOI: 10.17175/sb003_003.

Trilcke, Peer, Frank Fischer, Mathias Göbel, and Dario Kampkaspar (2015). *Comedy vs. Tragedy: Network Values by Genre*. URL: https://dlina.github.io/Network-Values-by-Genre/.

Tsitsulin, Anton, Davide Mottin, Panagiotis Karras, Alex Bronstein, and Emmanuel Müller (2018). *NetLSD: Hearing the Shape of a Graph*. DOI: 10.1145/3219819.3219991.

Watts, Duncan J. and Steven H. Strogatz (1998). "Collective Dynamics of 'Small-world' Networks." In: *Nature* 393.6684, pp. 440–442. DOI: 10.1038/30918.

Katrin Dennerlein, Thomas Schmidt, and Christian Wolff

Emotions in Stage Directions in German Drama of the Early Modern Period

Explorations via Computational Emotion Classification

Abstract: We present results about emotions in stage directions in German drama of the early modern period, an epoch, in which drama became the most important literary genre and the major cultural outlet to present emotion concepts in action. To investigate the representation of emotions in stage directions, we apply computational emotion classification of character emotions for a corpus of 245 dramatic texts from the period from 1600 to 1815. We use context-sensitive prediction results of a trained *gbert* model which achieves accuracies of 73 % for the single-label emotion classification of 14 classes (13 emotions types and *no emotion*). We apply the classification on all 439 678 sentences of our corpus and obtain 190 241 sentences classified with an emotion, which we evaluate separately by stage directions (12 268 sentences) and spoken text (177 973 sentences).
We show that the proportion of emotion representation in stage directions increases notably from 1740 onwards and thus clarify earlier qualitative studies that attribute such growth only to the period from 1770 onwards. Furthermore, we analyze which emotions in German drama from 1600 to 1815 are expressed in spoken texts and which are portrayed through acting. Finally, we explore the three most frequent emotions in stage directions ('suffering,' 'joy,' 'being moved') semantically.

1 Research Questions and Related Work

> Pickelhäring schläget Bullabutän an den hals, Bullabutän schläget ihm hergegen die wand um den kopff, sie kriegen einander bey den haaren und zerren sich hurtig auff dem schauplatz herumb, worüber die wand schier gantz in stücken gehet.[1]

1 Andreas Gryphius, Absurda Comica. Oder Herr Peter Squenz, 1657, III ("Pickelhäring hits Bullabutän on the neck, Bullabutän hits him in return with the wall on the head, they catch one another by the hair and quickly drag each other around the scene, over which the wall almost falls to pieces," own translation).

Thomas Schmidt, Christian Wolff, University of Regensburg
Katrin Dennerlein, University of Würzburg

https://doi.org/10.1515/9783111071824-010

Stage directions (SD) provide important additional information to the spoken text (ST) of dramatic texts (meaning the character speeches). They inform about the entrance and exit of characters, place and time as well as the mode of speech, bodily expressions, and characters' emotions as is the case in the example above. In this snippet from Andreas Gryphius' play *Absurda Comica oder Herr Peter Squenz* the characters Pickelhäring and Bullabutän have already insulted each other in the previous speeches. By means of the fight described in the SD, the anger of the two characters is not expressed on another level.

SD are a specific part of the 'side text' or 'secondary text' of a drama (Pfister 1988, p. 15; Ingarden [1931] 1973, pp. 377–396). While the primary text consists of the spoken text, the secondary text comprises, among others, the title of the play, announcements of act and scene, and SD as well as indications of the speakers of speeches. The secondary text fulfills explanatory as well as descriptive and narrative functions for the primary text (Pfister 1988, p. 15). As is the case for the entire secondary text, SD shall be attributed to the implicit author in the course of interpretation.[2] As Suchy rightly points out in her study, SD are thus to be read "as part of a play's fiction" (Suchy 1991, p. 71).[3] They can, for example, make the dialogue text more precise or clarify the disguises of the characters.[4] In addition, SD may be descriptions, reflections, and narrations containing information that is added to the dialogue text.[5]

SD have been investigated both systematically (Issacaroff 1998; Aston and Savona 2013; Aebischer 2003; Pfister 1988, pp. 13–17; Tonger-Erk 2018) and historically. Historical studies have been carried out, for example, on Shakespearian theater (Dustagheer and Woods 2018), French classicist drama (Gallèpe 1998), and Russian drama (Sperantov 1998). A few studies work quantitatively (Gallèpe 1998; Sperantov 1998; Trilcke et al. 2020).

2 According to Booth, the term 'implicit author' refers to the idea that the reader creates of the author as the intentional instance of a fictional text, based on the reading of the text (Booth 1991, 74). It is a text-based construct, which is ontologically different from the real author.

3 In this paper, Suchy provides some systematic and very illuminating thoughts on SD as utterances.

4 Pfister sees SD as a source of information on how to stage a text. He treats them in the context of other pieces of information that give clues about the way in which a text is performed. He suggests distinguishing between SD that refer to the actor and those that refer to the visual or acoustic context of the actor (Pfister 1988, p. 15). The distinction is followed by further explanations about the possibility of assigning staging information in the dialogue text as well (cf. Pfister 1988, p. 15).

5 This is the case even if the dramatic texts are not intended to be performed, cf. the so-called 'reading dramas' (Weber 2018). For different functions of the secondary text cf. Issacaroff (1998) and Tonger-Erk (2018).

For the German drama from 1600 to 1815, which is the subject of this paper, there are predominantly qualitative studies on single periods or genres. Constanze Baum has recently refuted the assumption that SD hardly occurred in Baroque drama (Baum 2018). By drawing on dramatic texts by Jakob Ayrer, Andreas Gryphius, Johann Christian Hallmann, and Christian Weise, she shows that secondary texts can already be found in all genres of drama in the early modern period. She argues that these texts cannot only be read in terms of a reconstruction of historical stage practice, but also fulfill a meaningful function for the communication of events. In his study from the 1960s, Walter Lehr takes a look at the SD of a genre which is highly under-researched in the history of German drama, the so-called 'Alt-Wiener Volkskomödie' (Lehr 1960). For the period from 1710–1760, he states that the indications referring to acting are becoming increasingly rare in SD during the period. From the end of the 1740s onwards, the information on acting is limited to the bare essentials. Lehr sees this as a sign that the acting is improving, so that ever less SD are necessary (Lehr 1960, p. 539). In the context of her study on emotions in drama, Anja Schonlau refers to the importance of SD, the evaluation of which, especially in relation to ST, she also recommends in her model for emotion-related text analysis (cf. Schonlau 2017, pp. 76 f., 131, 146). She argues that the increase in emotion representation in SD in the second half of the eighteenth century shows the rise of character concepts which are more complex in psychological terms (cf. Schonlau 2017). The largest study on SD in German drama examines SD in selected dramatic texts from 1750 to 1800 (from Gotthold Ephraim Lessing, Jakob Michael Reinhold Lenz, Johann Wolfang von Goethe, and Friedrich Schiller as well as from Friederike Caroline Neuber, Christiane Karoline Schlegel and August von Kotzebue) (Detken 2009). Detken, the author of the study, understands SD as written body language. Her thesis holds that acting and SD are initially determined almost exclusively by rhetoric and that they become more oriented towards body language and a natural art of acting around 1770 (cf. Detken 2009, p. 31).

To our knowledge, there is only one quantitative study on SD for German dramas. Trilcke et al. (2020) investigate SD in a corpus of 384 German plays between 1730 and 1930. They examine different aspects like the distribution of the token counts and mean sentence lengths as well as the epification of SD. They also state a steady increase of emotions in SD throughout their period of study. For this purpose, they calculated the percentage of words (adjectives, nouns, verbs) from the semantic field 'GEFUEHL' (emotions) from the lexical-semantic network GermaNet, a semantic word net for German similar to the English-based WordNet.[6]

6 https://www.clarin-d.net/de/germanet-de; https://wordnet.princeton.edu/.

In this paper, we focus on emotions in SD, more precisely characters' emotions in a corpus of 245 texts from the period of 1600–1815. We chose this period because during this time, dramatic texts became a major cultural outlet to present emotion concepts in action. These concepts are developed in several disciplines but are brought to a broader public only by dramatic texts (see Grimm 1980). There are a few publications in traditional literary studies on single emotions like 'fear,' 'envy,' or 'suffering' (Schings 1971; Wiegmann 1987; Schulz 1988; Zeller 2005; Schonlau 2017) and studies on the relationship between the emergence of subgenres and single emotions like 'admiration,' 'pity,' and 'moving' (Schings 1980; Schlott 1996). These studies are based on the evaluation of few dramatic texts. However, we built upon this work by using state-of-the art computational methods to automatically classify emotions in a much larger amount of texts.[7] We accomplished this by training large deep learning based language models with manual emotion annotations.

In previous computational research on historical dramatic texts, however, emotion classification was predominantly performed with lexicon-based approaches (working with predefined lists of annotated words). For instance, character relationships and their developments in the plays of Shakespeare with regard to emotions and sentiment have been investigated (Mohammad 2011; Nalisnick and Baird 2013; Yavuz 2021). Schmidt and Burghardt (2018) evaluated sentiment analysis approaches on plays by Gotthold Ephraim Lessing. In the context of this research, small annotated corpora for sentiment analysis were created (Schmidt et al. 2018, 2019). Examining the state of this field shows that most of the research is focused on the analysis of valence or polarity and mostly on individual authors (Mohammad 2011; Nalisnick and Baird 2013; Schmidt and Burghardt 2018). In their survey paper, Kim and Klinger (2019) point out that (1) the majority of the applied methods are indeed lexicon-based, which is regarded as outdated in the Natural Language Processing (NLP) community, and that (2) there is a lack of annotated corpora to perform more advanced ML-based approaches and evaluate the applied methods. The method of emotion classification we describe in the following sections is meant to remedy both desiderata by annotating a large corpus to train and evaluate transformer-based models for emotion classification and applying this to a larger amount of text. With this new method, we can for the first time not only detect explicit, but also implicit emotion representation in this research area.

After explaining our method of emotion data collection, we will address three questions:

7 Cf. for example our paper on the relationship of emotions and genres (Dennerlein et al. 2023).

1. When does the amount of emotion representation in SD in German drama from 1600 to 1815 increase notably?
2. Which emotions in German drama from 1600 to 1815 are expressed in SD, which in ST?
3. Which semantic changes can be observed in the most frequent emotions in SD during the period 1600–1815?

2 Emotion Definition

There are several different terms for affective states in the period of study. To have a term on the meta level to cover all the different concepts like 'Affekt' (affect) 'Gefühl' (feeling), 'Passion' (passion), 'Leidenschaft' (passion) etc., we choose the term 'emotion.' Following Schwarz-Friesel we understand emotions as

> [...] multidimensional, internally represented and subjectively experienced syndrome categories, which can be registered by the individual ego-related and introspectively-mentally as well as physically, whose experiential values are linked to a positive or negative evaluation and which are (can be) realized for others in perceptible expression variants. (Schwarz-Friesel 2007, p. 55, own translation)[8]

As syndromes, emotions are phenomena composed of different symptoms. In the case of emotions, these are both mental and physical and the individual may be aware of them to varying degrees. In addition, emotions are 'multidimensional' because they can be determined by several content-related characteristics. We assume that emotions manifest themselves physically, linguistically, as well as in behavior. Our object of investigation are the intended emotions of the characters in a drama. We do not analyze the emotions that authors or recipients feel[9] but those emotions which are experienced and/or expressed by characters.[10]

8 „[...] mehrdimensionale, intern repräsentierte und subjektiv erfahrbare Syndromkategorien, die sich vom Individuum ichbezogen und introspektiv-geistig sowie körperlich registrieren lassen, deren Erfahrungswerte an eine positive oder negative Bewertung gekoppelt sind und die für andere in wahrnehmbaren Ausdrucksvarianten realisiert werden (können)."

9 See the distinction between production-, reception-, text-, and context-related approaches to emotions in literary studies (Winko 2003).

10 Although fictional characters are not real persons, they are designed as anthropomorphic constructs to which model readers are expected to attribute a consciousness and feelings (cf. Jannidis 2004).

3 Set of Emotions

In NLP, emotion analysis is usually based on the classification systems of psychology (Plutchik 1980; Wood et al. 2018a, 2018b). For our purposes, we focused on the historical setting, and in doing so we selected a set of 13 emotions that only partially coincide with those of today's psychology. In the period under study, there are many different proposals for the systematization of emotions, most of which show a mixture of virtues and affects (short, strong surges of emotion) (Grimm 1980). Especially in philosophy, the proper number and nature of affects has been discussed with great intensity (Zeller 2005, p. 692). In order to be able to capture the change in the weighting of emotions in the period under investigation, it is important to abstract from the historical category systems. The main criterion for the selection of individual emotions was therefore their usefulness for mapping change in literary history, especially in genre differences. For this reason, emotions such as 'compassion' that have become important from a certain point in literary history onward were also included (cf. Schings 1980). The schema consists of the main categories of 'affection,' 'pleasure,' 'anxiety,' 'rejection,' 'suffering/empathy,' 12 subtypes, and one single emotion ('being moved'):[11]

- Emotions of affection / Zuneigung
  - desire / Lust (+)
  - love / Liebe (+)
  - friendship / Freundschaft (+)
  - admiration, reverence / Verehrung, Bewunderung (+)
- Emotions of pleasure / Freude und Glück
  - joy / Freude (+)
  - schadenfreude (+)
- Emotions of anxiety / Angst und Sorge
  - fear / Angst (–)
  - despair / Verzweiflung (–)
- Emotions of rejection / Ablehnung
  - anger / Ärger (–)
  - abhorrence / Abscheu, Wut, Hass (–)
- Emotions of suffering and empathy / Leid
  - suffering / Leid (–)

11 More information on this can be found in our annotation guidelines (Dennerlein et al. 2022c). This annotation scheme was improved and changed several times after testing it during annotation as is common and recommended for the humanities (Reiter 2020).

  - compassion / Mitleid (–)
- being moved / emotionale Bewegtheit (undetermined)

The plus and minus signs refer to the evaluation of the emotions by characters. Plus stands for a positive, minus for a negative evaluation by a character. In the following, we refer to this concept (including being moved) as polarity, the main categories as main emotion classes, and the 12 subtypes as well as the single emotion 'being moved' as sub-emotions.

4 Corpus

The following explorations refer to our corpus of German-language dramatic texts from the period of 1600–1815. It comprises 245 works stemming from the *GerDraCor* corpus (Fischer et al. 2019), from the platform *TextGrid*[12], and *Kasperl* plays from the *Leopoldstädter Theater* in Vienna. These are operettas that had enormous commercial success and were widely distributed throughout the German-speaking area.

In order to be able to detect changes, we consider three different periods for some questions. According to our knowledge of literary history of that period, we chose periods of different length:

- Period I (1600–1730) covers the period roughly from the 30 Years' War up to the beginning of the Enlightenment period.
- Period II (1731–1770) covers the period of the Enlightenment, i.e., Enlightenment and sentimental drama.
- Period III (1771–1815) begins with the revolutionary period of Storm and Stress, covers classical and romantic drama but also chivalric plays and family dramas, and lasts up to 1815. In this last period, the variety of genres is particularly great.

Table 1 gives an overview of the genre distribution in general and for the three time periods. In total, the corpus consists of 120 comedies, 66 tragedies and 54 *Schauspiel* plays. Classification of these genres was done manually based on knowledge of literary history, which was particularly important in the case of non-specific genre designations in the subtitles of the dramas. Table 2 illustrates general corpus statistics of the overall corpus. The corpus consists of 439 678 sentences that amount to almost five million tokens. Indeed, the vast majority of sentences of the

12 https://textgridrep.org.

Tab. 1: Genre distribution for 245 plays of the corpus.

Timespan/Genre	Comedies	Tragedies	Schauspiel	Overall
1600–1730	18	17	4	39
1731–1770	30	12	7	49
1771–1815	72	37	48	157
Overall	120	66	59	245

Tab. 2: General corpus statistics concerning SD and ST of the overall corpus. Avg # tokens refers to the average number of tokens per sentence.

Text type	# Sentences	%	# Tokens	Avg. # tokens
Stage directions (SD)	61 052	14	365 191	5.98
Spoken text (ST)	378 626	86	4 519 292	11.94
Overall	439 678	100	4 884 483	11.11

plays are unsurprisingly ST (86 %). SD are on average much shorter (5.98 tokens) than character speeches (11.11 tokens). Table 3 shows general corpus statistics for the different time periods.

5 Annotation

In the following chapter, we report on the annotation process and the results of the annotation of 17 plays. The annotated corpus was published via GitHub.[13]

Tab. 3: General corpus statistics for the three time periods.

Timespan	# Plays	# Tokens	# Sentences	# SD	# ST
1600–1730	39	793 617	55 966	5337	50 629
1731–1770	49	942 399	83 835	8106	75 729
1771–1815	157	3 148 467	299 877	47 609	252 268
Overall	245	4 884 483	439 678	61 052	378 626

13 https://github.com/lauchblatt/Emotions_in_Drama.

5.1 Annotation Process

The annotation was performed with the CATMA tool (Gius et al. 2020).[14] The annotators, students of German Literary Studies, had an annotation guideline (Dennerlein et al. 2022c) and went through numerous training sessions in pilot annotations (please refer to Schmidt et al. 2021b; Dennerlein et al. 2022a, 2022b for more information about the annotations process). Each of the 17 dramatic texts was annotated independently by two annotators. The task was to annotate, both in ST and in SD, the emotions experienced by a character and/or attributed to them. Multiple and overlapping annotations were explicitly allowed. Both the immediately preceding speeches and the entirety of the work were to be taken into account as the context of the interpretation. The length of the annotated text could be chosen freely, from single words to long text passages (at maximum the entire ST or SD). This procedure of variable annotation settings, which is very rare in NLP (Wood et al. 2018a, 2018b), was chosen in order to do justice to the variability and ambiguity of literary texts.

5.2 Annotated Dramatic Texts

We annotated the following 17 plays from different authors, periods, and genres to give a representative sample of the overall corpus.

- *Catharina von Georgien* by Andreas Gryphius (1657, tragedy)
- *Der Welt Erschröckende Attila* anonymous (after 1682, Schauspiel)
- *Massaniello* by Christian Weise (1683, tragedy)
- *Ein wunderliches Schau-Spiel vom niederländischen Bauer* by Christian Weise (1669, comedy)
- *Die getreue Sclavin Doris* anonymous (1720, Schauspiel)
- *Das Testament* by Luise Adelgunde Victorie Gottsched (1745, comedy)
- *Canut* by Johann Elias Schlegel (1746, tragedy)
- *Die zärtlichen Schwestern* by Christian Fürchtegott Gellert (1747, comedy)
- *Lucie Woodvil* by Johann Gottlieb Benjamin Pfeil (1757, tragedy)
- *Der Freigeist* by Joachim Wilhelm von Brawe (1758, tragedy)
- *Minna von Barnhelm* by Gotthold Ephraim Lessing (1767, comedy)
- *Der Postzug* by Cornelius von Ayrenhoff (1769, comedy)
- *Kabale und Liebe* by Friedrich Schiller (1784, tragedy)
- *Kasperl' der Mandolettikrämer* by Ferninand Eberl (1789, comedy)
- *Menschenhass und Reue* by August von Kotzebue (1790, comedy)

14 https://catma.de.

Tab. 4: Average inter-rater agreement across all plays. Average κ refers to *Cohen's* κ and average % is the pairwise agreement among annotators.

Categorical system	Average κ	Average %
Polarity	0.5	68%
Main class	0.4	62%
Sub-emotion	0.4	58%

- *Wallensteins Lager* by Friedrich von Schiller (1800, tragedy)
- *Faust* by Johann Wolfgang von Goethe (1807, tragedy)

Most of the plays stem from the *GerDraCor* corpus (Fischer et al. 2019), *Catharina von Georgien* from the *TextGrid* repository. The play *Kasperl' der Mandolettikrämer* was acquired from an open web repository[15], *Die getreue Sclavin Doris*, *Der Welt Erschröckende Attila*, and *Ein wunderliches Schau-Spiel vom niederländischen Bauer* from separate editions (Weise 1986; Noe 2007). These plays had to be further prepared for the annotation process.

5.3 Annotation Results

We collected 20 297 emotion annotations overall and calculated the inter-rater agreement on SD and ST level. To cope with varied annotation lengths and overlapping annotations and to calculate agreement metrics, we attribute to each SD and ST (per annotator) the emotion category that is annotated the most (measured in number of tokens). The agreement among the two annotators per play is presented in Table 4 as measured by the average of Cohen's κ values (values for Krippendorff's α are similar). The agreement ranges from values of 0.5 (*polarity*) to 0.4 (*sub-emotions*), which is a moderate level of agreement according to Landis and Koch (1977). These results are in line with similar research on historical and narrative text genres (Alm and Sproat 2005; Sprugnoli et al. 2015; Schmidt et al. 2018, 2019). In our understanding this is due to the ambiguity of literature.

Due to low to mediocre agreement metrics, we removed every annotation upon which the annotators disagreed. Table 5 shows the distribution of annotated text units against non-annotated units. This distribution includes all separated annotations independent of the length, which can range from one word to a maximum length of an entire SD or ST unit. Non-annotated text is treated as an

15 http://lithes.uni-graz.at/maezene/eberl_mandolettikraemer.html.

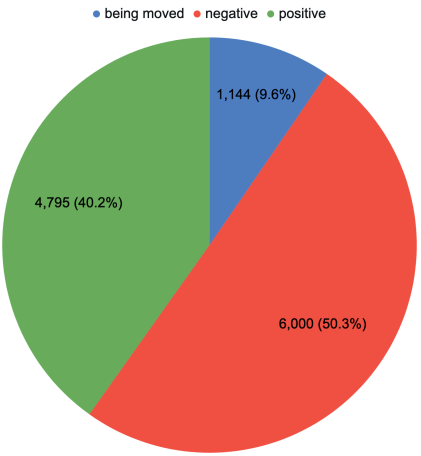


Fig. 1: Distribution of polarity classes among emotion annotations. Percentages are rounded.

Tab. 5: Distribution of annotation vs. no emotion annotation. # refers to the absolute number of separated full annotations (or non-annotations for no emotion), which can range from one word to the entire character speech or stage direction.

Class	#	%
Emotion annotations	11 939	51
No emotion	11 349	49
Overall	23 288	100

annotation with no emotion (a SD or ST that is not annotated is therefore counted as one no emotion annotation). Figure 1 illustrates the annotation distribution for the polarity classes and Table 6 for the main classes and sub-emotions; only the material with emotion annotations was used, excluding the non-annotated material.

We have collected 11 939 annotations after removing disagreements (8358 annotations were removed) and 11 349 no emotion annotations. Table 6 illustrates the emotion distributions among the emotion annotations. The most frequently annotated main class are the *emotions of rejection* (24 %). As for the sub-emotions, the most frequently annotated sub-emotions are 'suffering' (15 %), 'anger' (14 %), 'joy' (14 %), and 'love' (13 %). 'Desire,' 'friendship,' and 'schadenfreude' are the least common ones. The annotators make significant use of the varied annotation spans, ranging from one-word annotations to multiple sentences (mean = 25.15 to-

Tab. 6: Distribution of main classes and sub-emotion categories in annotated plays among emotion annotations (excluding no emotion annotations). The sums for the *main classes* (MC) are listed, followed by the *sub-emotions*. Percentages are rounded.

Emotion category	#	%
MC: emotions of affection	2804	23
desire	71	1
love	1569	13
friendship	240	2
admiration, reverence	924	8
MC: emotions of pleasure	1991	17
joy	1689	14
schadenfreude	302	3
MC: emotions of anxiety	989	8
fear	739	6
despair	250	2
MC: emotions of rejection	2856	24
anger	1621	14
abhorrence	1235	10
MC: emotions of suffering and empathy	2155	18
suffering	1760	15
compassion	395	3
Being moved	1144	10
Overall	11 939	100

kens; max = 578 tokens; Std = 29.41).[16] Most common is an annotation span across 2–3 sentences.[17]

6 Computational Emotion Classification

We define the emotion classification task as a single-label emotion classification on text sequences of varied lengths, in our case the sentences of the plays, since this is the linguistic unit that represents the annotation behavior the most. Our

16 We have analyzed token distributions of the annotations via the *NLTK Punkt* Tokenizer. https://www.nltk.org/_modules/nltk/tokenize/punkt.html.

17 Please note that more information about the annotation process and results can be found in previous publications (Schmidt et al. 2021b; Dennerlein et al. 2022c).

emotion scheme consists of all sub-emotions: 13 emotions and one no emotion class, leading to a 14-class setting. The implementation and selection of an emotion classification method was based on previous research (Schmidt et al. 2021a, 2021c) analyzing and comparing various established methods like lexicon-based methods (Schmidt and Burghardt 2018), traditional machine learning approaches, static word embeddings (Bojanowski et al. 2017), and predominantly transformer-based language models like BERT (Devlin et al. 2019), since they are currently regarded as state-of-the-art in the area of emotion classification (Shmueli and Ku 2019; Cao et al. 2020). We experimented with the most established and well-known German language transformer-based models from the platform Hugging Face[18] (Wolf et al. 2020). Since these models are trained primarily on contemporary language, we also included available models that were trained from scratch (e.g., the *German EuropeanaBERT*[19] by Schweter 2020) or further pretrained on historical or narrative language (e.g., a German BERT model by Brunner et al. 2020). We also developed models by further training base models with the texts of our corpus since this sort of domain-adaptive pretraining has been shown to be beneficial in certain special domains (Gururangan et al. 2020).

As training data for the fine tuning, we use all annotations (including non-annotated material as no emotion class) and filter them by removing disagreeing annotations of annotators, which results in the corpus described in Table 6. The best-performing model for the classification tasks is the large German BERT model *gbert-large* by *deepset* (Chan et al. 2020). This model achieves an accuracy of 73 % and weighted f1-score of 72 % for the classification task of 14 classes and up to 86 % when the class system is reduced to polarity. Due to class imbalances in our training data, we apply methods of over- and undersampling to achieve these accuracies. Nevertheless, f1-scores for low-frequency classes are on average lower. Please refer to previous research for more details regarding the technical implementations and experiments (Schmidt et al. 2021a, 2021c, 2022). The model was trained and evaluated on the filtered corpus in a 5x5 stratified setting and fine-tuned to the specific classification task for 4 epochs, a batch size of 32, a learning rate of 4e-5 and the *Adam* optimizer as optimization algorithm with a *Tesla P100* GPU.

We perform the final classification on sentences of the plays. For this, we segmented the SD and ST into sentences, using the *NLTK Punkt* sentence segmentation.[20] The SD and ST were either already annotated for plays of the *GerDraCor*

18 https://huggingface.co.

19 https://huggingface.co/dbmdz/bert-base-german-europeana-cased.

20 https://www.nltk.org/_modules/nltk/tokenize/punkt.html.

or *TextGrid* corpus or added in an additional processing step and manually post-corrected. We chose the sentence as classification text span since it resembles the annotated text spans the most, although both units are not necessarily the same.

7 Emotion Classification Results for the Overall Corpus

Applying the emotion classification on all 439 678 sentences of the corpus results in a total amount of 190 241 sentences classified with an emotion, of which 12 268 are SD and 177 973 are ST.[21] 57 % of all sentences are labeled with no emotion and 43 % with an emotion (see overall in Figure 2). The difference is bigger for SD, for which 80 % of sentences are classified as representing no emotion.

Table 7 shows the emotion classification for sub-emotions and main classes among all sentences that were classified as emotions (that means excluding no emotion classifications). The most frequent main classes are the emotions of 'rejection' (22.5 %) and the emotions of 'suffering and empathy' (20.6 %), while the most frequent sub-emotions are 'suffering' (18.4 %), 'joy' (13.6 %), and the special category 'being moved' (13.3 %). Certain emotions like 'desire' (0.1 %) and 'friendship' (1.4 %) are rarely classified.

Table 8 illustrates the emotion distribution for all emotion classifications on the overall corpus, differentiated between SD and ST. The most striking difference concerns the category 'being moved,' which is more frequently classified in regard to percentage points (10) in SD than in ST. This results in the fact that most emotions are more frequent for ST than SD, with a few exceptions like 'fear.'

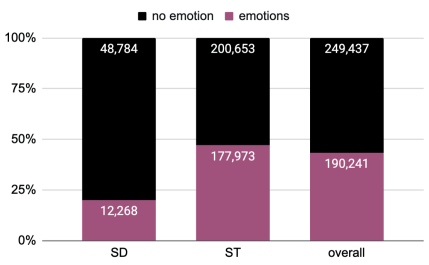


Fig. 2: Distribution of classification results concerning emotion and no emotion classifications among sentences for SD and ST.

21 More data and detailed results for the following results chapters can be found here: https://github.com/lauchblatt/Emotions_in_Drama.

Tab. 7: Distribution of classification results concerning main classes and sub-emotions in the sentences of the overall corpus that are classified as emotions.

Emotion category	# Sentences	%
MC: emotions of affection	37 893	19.92
desire	167	0.09
love	20 575	10.82
friendship	2703	1.42
admiration, reverence	14 448	7.59
MC: emotions of pleasure	28 282	14.87
joy	25 843	13.58
schadenfreude	2439	1.28
MC: emotions of anxiety	16 055	8.44
fear	11 800	6.2
despair	4255	2.24
MC: emotions of rejection	43 446	22.84
anger	24 029	12.63
abhorrence	19 417	10.21
MC: emotions of suffering and empathy	39 206	20.61
suffering	35 020	18.41
compassion	4186	2.2
Being moved	25 359	13.33
Overall	190 241	100

Tab. 8: Distribution of emotion classifications among sentences in SD and ST.

Emotion category	SD (%)	ST (%)
MC: emotions of affection	12.23	20.45
desire	0.17	0.08
love	6.81	11.09
friendship	0.36	1.49
admiration, reverance	4.89	7.78
MC: emotions of pleasure	17.63	14.68
joy	15.89	13.43
schadenfreude	1.74	1.25
MC: emotions of anxiety	10.60	8.29
fear	9.24	5.99
despair	1.35	2.30
MC: emotions of rejection	18.83	23.11
anger	14.38	12.51
abhorrence	4.45	10.60
MC: emotions of suffering and empathy	17.69	20.81
suffering	17.27	18.49
compassion	0.42	2.32
Being moved	23.03	12.66
Overall	100	100

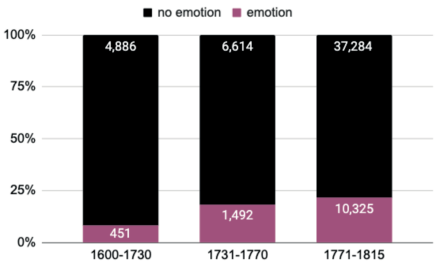


Fig. 3: Distribution of classification results for sentences concerning emotions and no emotions among SD sentences throughout three time periods.

8 Results: Emotions in Stage Directions

In the following, we explore the three major research questions concerning emotions in SD as introduced in the opening chapter.

8.1 When Does the Amount of Emotion Representation in SD in German Drama from 1600 to 1815 Increase Notably?

It can be seen that in the entire corpus (1600–1815), there are much less emotion classifications in SD than in ST: Figure 2 shows that 177 973 of the sentence classifications in ST are of emotions, which results in a share of 46.8 %. In SD the proportion is only 19.9 % (12 268 sentences). The value just reported is in regard to the entire corpus, but the change in this proportion was also measured in the three time periods. The results show an increase of emotions in SD over time: As can be seen in Figure 3, 8.5 % (451) of the SD in period I, 18.4 % (1492) of the SD in period II, and 21.7 % (10 325) of the SD in period III contain emotions.

This means that in our data, we observe a deviation from Detken's thesis (Detken 2009, p. 31). While Detken is of the opinion that emotions in SD increase considerably from 1770 onwards, our results show that the strongest increase can already be observed between period I and period II. Here, the proportion of period I to period II has more than doubled (from 8.5 % to 18.4 %).

Furthermore, the difference of our approach in contrast to a lexicon-based method becomes quite clear now. While Trilcke et al. (2020) only observe an increase in emotion words of 1.5 % for the period of 1740–1815 with a lexicon-based method,[22] the methodology we use yields an increase of 13.2 % for those SD that contain an explicit or implicit emotion representation (Figure 3).

22 We interpreted their plot of Figure 8 based on the assumption that 1,00 equals 100 %.

Tab. 9: Distribution of classification in sentences classified with an emotion for SD and ST throughout three time periods in percentage.

Timespan	1600–1730		1731–1770		1771–1815	
Class	SD (%)	ST (%)	SD (%)	ST (%)	SD (%)	ST (%)
MC: emotions of affection	11.09	21.9	11.13	20.88	12.44	19.96
desire	0.22	0.15	0	0.03	0.19	0.08
love	4.66	10.18	5.9	11.33	7.03	11.24
friendship	0.67	0.83	0.27	2.04	0.36	1.49
admiration, reverence	5.54	10.74	4.96	7.48	4.85	7.16
MC: emotions of pleasure	22.39	12.45	25.07	13.47	16.35	15.58
joy	19.29	11.13	24.26	12.38	14.53	14.3
Schadenfreude	3.1	1.32	0.8	1.09	1.82	1.28
MC: emotions of anxiety	9.09	7.93	9.18	7.73	10.87	8.55
fear	7.98	6.65	8.78	5.45	9.37	6
despair	1.11	1.28	0.4	2.28	1.5	2.55
MC: emotions of rejection	31.93	23.61	22.92	23.6	17.67	22.85
anger	20.4	10.41	18.1	13.18	13.58	12.81
abhorrence	11.53	13.2	4.83	10.41	4.09	10.03
MC: emo. of suffering & empathy	20.84	27.49	15.35	21.19	17.89	19.08
suffering	20.4	26.1	14.88	18.45	17.48	16.66
compassion	0.44	1.4	0.47	2.74	0.41	2.42
Being moved	4.66	6.63	16.35	13.13	24.79	13.98
Overall	100	100	100	100	100	100

8.2 Which Emotions in German Drama from 1600 to 1815 are Expressed in SD, Which in ST?

We investigated the emotion distribution in SD in comparison with the emotion distribution in ST. The numbers for the entire period can be found in Table 8. There are some emotions with similar proportions in SD and ST, for example 'suffering.' However, there are clear differences for some emotions. 'Abhorrence,' for example, is classified twice as often in ST as in SD, 'friendship' four times, and 'compassion' six times. These emotions seem to be more likely to be expressed by a character's speech. The state of 'being moved,' on the other hand, is classified almost twice as often in SD as in ST. Obviously, authors see the need to state the condition of 'being moved' by additional comments so that recipients and actors can grasp it.

These results will now be differentiated according to the time periods (cf. Table 9). So far, there are no hypotheses in literary studies on the distribution of emotions in the three different time periods. The results of our study can provide

important new insights for literary historical research of this period. However, the results of our investigations can be interpreted well on the basis of historical knowledge about the history of drama. In the following, we will discuss these results by presenting the percentage-wise values of ST and SD as seen in Table 9 in chronological lists.

The values for ST remain at about the same level, while the value of 'schadenfreude' for SD decreases in the middle period and increases again in the last period: SD 3.1 %, 0.8 %, 1.82 % (ST 1.3 %, 1.1 %, 1.3 %). The high value for the first period could be due to the strong presence of comic characters in subplots, with which the authors aim at a comedy of laughs.[23] The last increase could be related to the fact that we have the *Kasperl* plays in the corpus (24 of the 157 plays in period III). In these plays, 'schadenfreude' is expressed through the facial expressions, gestures, and behavior of typical comic characters. Here is an example with a mean landlord: "Wirt: (geht mit verätherischen Grimassen ab)."[24]

For two sub-emotions, we see an almost steady decrease over time in SD. For 'abhorrence,' the values are 11.5 %, 4.8 %, 4.1 % (SD) and 13.2 %, 10.4 %, 10 % (ST). 'Anger' is classified with 20.4 %, 18 %, 13.6 % (SD) and 10.4 %, 13.2 %, 12.8 % (ST). These findings can be explained by the well-known development in drama away from deterrence and admiration towards empathy and moderate ridicule and by the poetics of compassion and friendship in sentimental comedy, which is a very important genre in period II (cf. Pikulik 1966, Glaser 1969).

In the following cases, we observe an increase of the frequency only in ST. On the one hand, 'compassion' is classified with 0.4 %, 0.5 %, 0.4 % (SD) and 1.4 %, 2.7 %, 2.42 % (ST) and 'friendship' with 0.7 %, 0.3 %, 0.4 % (SD) and 0.8 %, 2 %, 1.5 % (ST). The increases for the numbers for these two emotions in ST are probably also due to the poetics of 'compassion' and 'friendship' in sentimental comedy (period II), according to which characters should express these feelings verbally. A result we cannot explain is the decrease of 'despair' in SD in period II: 'despair': SD 1.1 %, 0.4 %, 1.5 % (ST 1.3 %, 2.3 %, 2.6 %). Why is 'despair' represented much more often in ST than in SD in period II? To be able to better explain these differences, one would probably have to distinguish between comic and non-comic occurrences of 'despair.'

23 For the relevance of this character in the history of German comedy, see: Dennerlein (2021).
24 Ferdinand Huber: Der eifersüchtige Schuster, 1791, I,4, "Landlord: (walks off with tell-tale grimaces)," own translation.

Tab. 10: General corpus statistics for SD in the three time periods.

Timespan	# Sentences	# Tokens	Avg. # tokens
1600–1730	5337	34 555	6.47
1731–1770	8106	42 714	5.26
1771–1815	47 609	287 922	6.05

8.3 Semantic Explorations of 'Suffering,' 'Joy,' and 'Being Moved' in SD over Time

In this last section we analyze the semantics of emotions in SD. For this, we only regard the SD proportion of our overall corpus (see Table 10 for corpus statistics). We take the three most frequent emotions in SD, 'suffering,' 'joy,' and 'being moved,' and explore the lexical differences in their representation. For this, we analyzed the most frequent words in the emotion-classified sentences.[25] This method required some preliminary steps: (1) lemmatization, (2) removal of stop words (e.g., articles, pronouns), and (3) removal of character names. For the lemmatization, we use the *Hanover Tagger* (Version 1.1.0)[26] by Wartena (2019). Different inflections of a word are reduced to a base form, which is beneficial for the semantic analysis. For the removal of stop words, we use the German stop words list by the NLTK[27] and extend it with further high frequency words bearing no semantic meaning: we remove character names as identified via the speaker tag in our XML-based corpus from the texts. We also split a speaker name if it contains white space to filter first and last names of characters or similar name combinations. Based on this method, we see the following changes in the language of emotions of the three most frequent sub-emotions in SD.

8.3.1 Suffering (Figure 4)

The most frequent word in emotion representations of 'suffering' in period I is the word 'gehen' ('to walk'), while in periods II and III it is 'weinen' ('to cry'). This shows that from 1730 on, 'suffering' is no longer expressed only by physical restlessness or exits of the characters, but specifically by crying. Looking at the top

25 Additional data like word clouds and frequency lists ca be found in our repository: https://github.com/lauchblatt/Emotions_in_Drama.

26 https://github.com/wartaal/HanTa.

27 https://www.nltk.org.

(a) SD (1600–1730) **(b)** SD (1730–1770) **(c)** SD (1770–1815)

Fig. 4: Most frequent words for 'suffering.'

five, we can see a semantic change from descriptive emotion representation to emotion representation through acting which makes the suffering much more concrete, but also greatly increases the demands on the actors:

- While SD indicate in period I that a character is speaking or behaving 'traurig' ('sadly'), SD in period II and III indicate that characters cry and sigh ('weinen,' 'weinend,' 'seufzen'). In Andreas Gryphius' play *Papinianus*, for example, the singing of the spirits leaves the emperor mournful: "Die Geister verschwinden zugleich / der Käyser erwachet und gehet traurig ab."[28]
- Periods II and III refer even more explicitly to body language by using 'Auge' ('eye') and 'Hand' ('hand'): 'Hand' ('hand') appears in period II on position 3, where it remains for period III. 'Auge' ('eye') occurs initially in period III under the top five on position 4. The following SD contains an occurrence of 'eye': "steht in einer trostlosen Stellung, er schweigt, er seufzt, und heftet seine Augen unbeweglich bald auf den Himmel, bald auf den Boden."[29]

8.3.2 Joy (Figure 5)

Concerning the emotion representation of 'joy' in SD, the top five of the most frequently used words in period I show one very specific action, which is singing. In the Christian tragedy *Liberata*, for example, "A singing angel" comes to Liberata, joyful in the face of the divine salvation he promises her in the song that follows.[30] Moreover, the emotionally underspecified movement verb 'gehen' ('to walk') and the descriptive lemmas 'frölich' and 'freudig' can be found in the top five (both

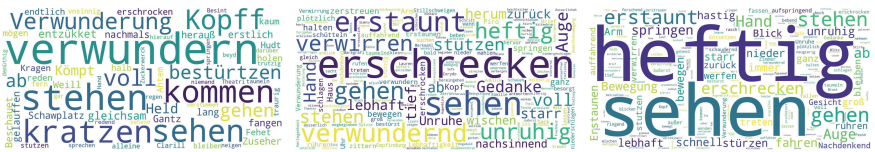
28 Andreas Gryphius, Papinianus, 1659, IV ("The spirits disappear all at once / the emperor awakes and sadly exits," own translation).

29 Christoph Martin Wieland, Klementina von Poretta 1759, III,4 ("stands in a desolate position, he is silent, he sighs, and fixes his eyes without moving sometimes on the sky, sometimes on the ground," own translation).

30 Johann Christian Hallmann, Die unüberwindliche Keuschheit oder Die groszmüthige Prinzeszin Liberata, 1700, III,9.

(a) SD (1600–1730) **(b)** SD (1730–1770) **(c)** SD (1770–1815)

Fig. 5: Most frequent words for 'joy.'



(a) SD (1600–1730) **(b)** SD (1730–1770) **(c)** SD (1770–1815)

Fig. 6: Most frequent words for 'being moved.'

'happy/happily'). In period II and III, laughing and smiling comes to the fore, expressed through the German adverbs 'lachend' and 'lächelnd.' According to the rise of the sentimental drama in the second half of the eighteenth century, 'umarmen' ('to hug') becomes the most frequently used word in the emotion representation of 'joy' in period III. The verb 'springen' ('to jump') appears in the top five for the first time in period III. In these cases, it is not another word for 'dancing,' but denotes a spontaneous movement associated with a sudden strong feeling of joy. In the tragedy *Eid und Pflicht* by Johann Jakob, the rapid change from despair to joy is indicated in the SD by a sudden jump:

> Eduard: [...] aber er röchelt! er stirbt! – – Alle meine Sinne, mein Blut – *Neben ihm niederstürzend.* O um Gotteswillen! Ich werde denn doch sein Mörder! Selbst dadurch, dass ich mich ihm aufopfre, sein Mörder! – – *Wieder aufspringend und froh.* Er regt sich wieder. Es ist noch Leben in ihm. [...][31]

8.3.3 Being Moved (Figure 6)

We use the term 'being moved' in the sense of a strong emotional arousal, which is, however, unspecified in terms of content. The need to include this category

31 Johann Jakob Engel, Eid und Pflicht, 1776, IV, 5 ("but he is gasping! he is dying! – All my senses, my blood – *Falling down beside him.* O for God's sake! I will become his murderer after all! Even by offering myself to him, his murderer! – *Jumping up again and joyfully*. He is moving again. There is still life in him," own translation).

arose in the course of the close reading for the annotation of 17 dramatic texts in all of which this phenomenon frequently appears. In period I, 'being moved' is expressed by the characters stopping in astonishment and scratching their heads ('verwundern,' 'stehen,' 'kratzen,' 'Kopf'), as in the subsequent example: "Peter Squentz: [...]. Hierauf verstummt er und kratzt sich im kopff."[32]

In period II, 'being moved' is expressed more vehemently by using the verb 'erschrecken' and the adjective 'erstaunt' ('to startle,' 'astonished'). In addition, the characters in this period are described as 'unruhig' and 'verwirrt' ('restless,' 'confused'). In period III, an adjective used from the second period onward is now most often used to denote emotional movement: 'heftig' ('violently'). The characters now either stop when they are stunned to see something surprising or walk around when something stirs them up ('erstaunt,' 'sehen,' 'gehen' – 'astonished,' 'to see,' 'to walk'). In the following example the character Eduard's inner turmoil when he is confirmed that his mortally ill father will soon be taken away to prison is illustrated by the following SD: "Eduard: *schnell und heftig.* Ha! – So errieth ich's? So soll er fort?"[33]

9 Conclusion

Our findings may be summarized as follows:

- SD have been depicting significantly more emotion since 1730 already and not only since 1770.
- In some cases, SD follow what we would expect from the drama poetics: decrease of 'abhorrence' and 'anger' in SD (correlated with an increase of 'compassion' and 'friendship' in ST of tragedies).
- In part, they show interesting shifts (the shift of despair into the character speeches), which makes further emotion-related analyses necessary.
- Semantic shifts in the language of emotions in SD point to the development of a much more vivid body language. As we investigated this only for three sub-emotions it would be of great interest to include more sub-emotions. To learn more about the semantics of emotions in SD, it would be instructive to investigate distinctive words of sub-emotions, to examine the combined oc-

32 Absurda Comica oder Herr Peter Squenz, 1757, III ("He then falls silent and scratches his head," own translation).

33 Johann Jakob Engel, Eid und Pflicht, 1776, I, 5 ("Eduard: *quickly and violently.* Ha! – That's what I guessed? Is he supposed to go away like that?," own translation).

currence of lemmata up to multi-word expressions as well as to include the results of part-of-speech tagging in the evaluation.

Funding: We developed these methods in the project "Emotions in Drama." This project was funded by the DFG (German Research Association) as part of the priority program Computational Literary Studies (SPP 2207/1) with two grants (project number 424207618; grants DE 2188/3-1 und WO 835/4-1). https://dfg-spp-cls.github.io/projects_en/2020/01/24/TP-Emotions_in_Drama/.

References

Aebischer, Pascale (2003). "Didascalia and Speech in the Dramatic Text." In: *Journal of Dramatic Theory and Criticism* 17.2, pp. 25–44. URL: https://journals.ku.edu/jdtc/article/view/3471/3347.

Alm, Cecilia Ovesdotter and Richard Sproat (2005). "Emotional Sequencing and Development in Fairy Tales." In: *Proceedings Affective Computing and Intelligent Interaction (ACII 2005)*, pp. 668–674. DOI: 10.1007/11573548_86.

Aston, Elaine and George Savona (2013). *Theatre as Sign System: A Semiotics of Text and Performance*. Routledge.

Baum, Constanze (2018). "Nebentexte als Bedeutungsträger im Drama der Frühen Neuzeit." In: *Zeitschrift für Literaturwissenschaft und Linguistik* 48, pp. 445–461.

Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov (2017). "Enriching Word Vectors with Subword Information." In: *Transactions of the Association for Computational Linguistics* 5, pp. 135–146. DOI: 10.1162/tacl_a_00051.

Booth, Wayne C. (1991). *The Rhetoric of Fiction*. 2. ed., repr. Penguin.

Brunner, Annelen, Ngoc Duyen Tanja Tu, Lukas Weimer, and Fotis Jannidis (2020). "To BERT or not to BERT — Comparing Contextual Embeddings in a Deep Learning Architecture for the Automatic Recognition of four Types of Speech, Thought and Writing Representation." In: *Proceedings of the 5th Swiss Text Analytics Conference (SwissText) & 16th Conference on Natural Language Processing (KONVENS)*. URL: https://ceur-ws.org/Vol-2624/paper5.pdf.

Cao, Lihong, Sancheng Peng, Pengfei Yin, Yongmei Zhou, Aimin Yang, and Xinguang Li (2020). "A Survey of Emotion Analysis in Text Based on Deep Learning." In: *2020 IEEE 8th International Conference on Smart City and Informatization (iSCI)*, pp. 81–88. DOI: 10.1109/iSCI50694.2020.00020.

Chan, Branden, Stefan Schweter, and Timo Möller (2020). "German's Next Language Model." In: *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 6788–6796. DOI: 10.18653/v1/2020.coling-main.598.

Dennerlein, Katrin (2021). *Materialien und Medien der Komödiengeschichte. Zur Praxeologie der Werkzirkulation zwischen Hamburg und Wien von 1678–1806*. De Gruyter.

Dennerlein, Katrin, Thomas Schmidt, and Christian Wolff (2022a). "Emotion Courses in German Historical Comedies and Tragedies." In: *Book of Abstracts, Digital Humanities 2022 (DH2022). Responding to Asian Diversity*, pp. 193–197. URL: https://dh2022.dhii.asia/dh2022bookofabsts.pdf.

Dennerlein, Katrin, Thomas Schmidt, and Christian Wolff (2022b). "Emotionen im kulturellen Gedächtnis bewahren." In: *DHd 2022 Kulturen des digitalen Gedächtnisses. 8. Tagung des Verbands "Digital Humanities im deutschsprachigen Raum" (DHd 2022)*, pp. 93–98. DOI: 10.5281/zenodo.6327957.

Dennerlein, Katrin, Thomas Schmidt, and Christian Wolff (2022c). *Figurenemotionen in deutschsprachigen Dramen annotieren*. Annotation guidelines. DOI: 10.5281/zenodo.6228151.

Dennerlein, Katrin, Thomas Schmidt, and Christian Wolff (2023). "Computational emotion classification for genre corpora of German tragedies and comedies from 17th to early 19th century." In: *Digital Scholarship in the Humanities*, pp. 1–16. DOI: 10.1093/llc/fqad046.

Detken, Anke (2009). *Im Nebenraum des Textes: Regiebemerkungen in Dramen des 18. Jahrhunderts*. De Gruyter. DOI: 10.1515/9783110230031.

Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186. DOI: 10.18653/v1/N19-1423.

Dustagheer, Sarah and Gillian Woods (2018). *Stage Directions and Shakespearean Theatre*. Bloomsbury.

Fischer, Frank, Ingo Börner, Mathias Göbel, Angelika Hechtl, Christopher Kittel, Carsten Milling, and Peer Trilcke (2019). "Programmable Corpora: Introducing DraCor, an Infrastructure for the Research on European Drama." In: *Book of Abstracts. Digital Humanities 2019: "Complexities"*. DOI: 10.5281/zenodo.4284002.

Gallèpe, Thierry (1998). *Didascalies: les mots de la mise en scène*. L'Harmattan.

Gius, Evelyn, Jan Christoph Meister, Marco Petris, Malte Meister, Christian Bruck, Janina Jacke, Mareike Schuhmacher, Marie Flüh, and Jan Horstmann (2020). *CATMA*. DOI: 10.5281/zenodo.4353618.

Glaser, Horst Albert (1969). *Das bürgerliche Rührstück. Analekten zum Zusammenhang von Sentimentalität mit Autorität in der trivialen Dramatik Schröders, Ifflands, Kotzebues und anderer Autoren am Ende des achtzehnten Jahrhunderts*. J.B. Metzler.

Grimm, Hartmut (1980). "Affekt." In: *Ästhetische Grundbegriffe*. Ed. by Karlheinz Barck, Martin Fontius, Dieter Schlenstedt, Steinwachs Burkhart, and Friedrich Wolfzettel. Vol. 1, pp. 16–49.

Gururangan, Suchin, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith (2020). "Don't Stop Pretraining: Adapt Language Models to Domains and Tasks." In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 8342–8360. DOI: 10.18653/v1/2020.acl-main.740.

Ingarden, Roman [1931] (1973). *The Literary Work of Art: An Investigation of the Borderlines of Ontology, Logic and Theory of Literature. With an Appendix on the Functions of Language in the Theatre.* Northwestern University Press.

Issacaroff, Michael (1998). "Stage Codes." In: *Performing Texts*. Ed. by Michael Issacaroff and Robin F. Jones. University of Pennsylvania Press, pp. 59–74.

Jannidis, Fotis (2004). *Figur und Person. Beitrag zu einer historischen Narratologie.* De Gruyter.

Kim, Evgeny and Roman Klinger (2019). "A Survey on Sentiment and Emotion Analysis for Computational Literary Studies." In: *Zeitschrift für digitale Geisteswissenschaften*. DOI: 10.17175/2019_008.

Landis, J. Richard and Gary G. Koch (1977). "The Measurement of Observer Agreement for Categorical Data." In: *Biometrics* 33.1, pp. 159–174.

Lehr, Walter (1960). "Die szenischen Bemerkungen in den Dramen des Altwiener Volkstheaters bis 1752." PhD thesis. University of Vienna.

Mohammad, Saif (2011). "From Once Upon a Time to Happily Ever After: Tracking Emotions in Novels and Fairy Tales." In: *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pp. 105–114.

Nalisnick, Eric T. and Henry S. Baird (2013). "Character-to-Character Sentiment Analysis in Shakespeare's Plays." In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 479–483. URL: https://www.aclweb.org/anthology/P13-2085.

Noe, Alfred (2007). "Italienische Spieltexte: Aus unveröffentlichten Handschriften." In: *Spieltexte der Wanderbühne*. Ed. by Manfred Brauneck and Alfred Noe. Vol. V, 1+2. De Gruyter.

Pfister, Manfred (1988). *The Theory and Analysis of Drama*. Cambridge University Press.

Pikulik, Lothar (1966). *Bürgerliches Trauerspiel und Empfindsamkeit*. Böhlau.

Plutchik, Robert (1980). *Emotion, a Psychoevolutionary Synthesis*. Harper & Row.

Reiter, Nils (2020). "Anleitung zur Erstellung von Annotationsrichtlinien." In: *Reflektierte algorithmische Textanalyse: Interdisziplinäre(s) Arbeiten in der CRETA-Werkstatt*. Ed. by Nils Reiter, Axel Pichler, and Jonas Kuhn. De Gruyter, pp. 193–202. DOI: 10.1515/9783110693973-009.

Schings, Hans-Jürgen (1971). "Consolatio Tragoediae. Zur Theorie des barocken Trauerspiels." In: *Deutsche Dramentheorien* 1, pp. 1–44.

Schings, Hans-Jürgen (1980). *Der mitleidigste Mensch ist der beste Mensch: Poetik des Mitleids von Lessing bis Büchner*. C. H. Beck.

Schlott, Michael (1996). "Albert Meier, Dramaturgie der Bewunderung. Untersuchungen zur politisch-klassizistischen Tragödie des 18. Jahrhunderts." In: *Arbitrium* 14.3, pp. 334–339. DOI: 10.1515/arbi.1996.14.3.334.

Schmidt, Thomas and Manuel Burghardt (2018). "An Evaluation of Lexicon-based Sentiment Analysis Techniques for the Plays of Gotthold Ephraim Lessing." In: *Proceedings of the Second Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pp. 139–149. URL: https://www.aclweb.org/anthology/W18-4516.

Schmidt, Thomas, Manuel Burghardt, and Katrin Dennerlein (2018). "Sentiment Annotation of Historic German Plays: An Empirical Study on Annotation Behavior." In: *Proceedings of the Workshop on Annotation in Digital Humanities 2018 (annDH 2018)*, pp. 47–52. URL: https://ceur-ws.org/Vol-2155/schmidt.pdf.

Schmidt, Thomas, Manuel Burghardt, Katrin Dennerlein, and Christian Wolff (2019). "Sentiment Annotation for Lessing's Plays: Towards a Language Resource for Sentiment Analysis on German Literary Texts." In: *2nd Conference on Language, Data and Knowledge (LDK 2019)*, pp. 45–50.

Schmidt, Thomas, Katrin Dennerlein, and Christian Wolff (2021a). "Emotion Classification in German Plays with Transformer-based Language Models Pretrained on Historical and Contemporary Language." In: *Proceedings of the 5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*, pp. 67–79. DOI: 10.18653/v1/2021.latechclfl-1.8.

Schmidt, Thomas, Katrin Dennerlein, and Christian Wolff (2021b). "Towards a Corpus of Historical German Plays with Emotion Annotations." In: *3rd Conference on Language, Data and Knowledge (LDK 2021)*. Vol. 93, 9:1–9:11. DOI: 10.4230/OASIcs.LDK.2021.9.

Schmidt, Thomas, Katrin Dennerlein, and Christian Wolff (2021c). "Using Deep Learning for Emotion Analysis of 18th and 19th Century German Plays." In: *Fabrikation von Erkenntnis: Experimente in den Digital Humanities*. Ed. by Manuel Burghardt, Lisa Dieckmann, Timo Steyer, Peer Trilcke, Niels-Oliver Walkowski, Joëlle Weis, and Ulrike Wuttke. Melusina Press, p. 34. DOI: 10.26298/melusina.8f8w-y749-udlf.

Schmidt, Thomas, Katrin Dennerlein, and Christian Wolff (2022). "Evaluation computergestützter Verfahren der Emotionsklassifikation für deutschsprachige Dramen um 1800." In: *DHd 2022 Kulturen des digitalen Gedächtnisses. 8. Tagung des Verbands "Digital Humanities im deutschsprachigen Raum" (DHd 2022)*. DOI: 10.5281/zenodo.6328169.

Schonlau, Anja (2017). *Emotionen im Dramentext: eine methodische Grundlegung mit exemplarischer Analyse zu Neid und Intrige 1750-1800*. De Gruyter.

Schulz, Georg-Michael (1988). *Tugend, Gewalt und Tod: das Trauerspiel der Aufklärung und die Dramaturgie des Pathetischen und des Erhabenen*. De Gruyter. DOI: 10.1515/9783111384795.

Schwarz-Friesel, Monika (2007). *Sprache und Emotion*. Francke.

Schweter, Stefan (2020). *Europeana BERT and ELECTRA models*. DOI: 10.5281/zenodo.4275044.

Shmueli, Boaz and Lun-Wei Ku (2019). *SocialNLP EmotionX 2019 Challenge Overview: Predicting Emotions in Spoken Dialogues and Chats*. Preprint. DOI: 10.48550/arXiv.1909.07734.

Sperantov, V. V. (1998). "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)]." In: *Philologica* 5.11/13, pp. 9–48. URL: https://rvb.ru/philologica/05/05sperantov.htm.

Sprugnoli, Rachele, Sara Tonelli, Alessandro Marchetti, and Giovanni Moretti (2015). "Towards sentiment analysis for historical texts." In: *Digital Scholarship in the Humanities* 31, pp. 762–772. DOI: 10.1093/llc/fqv027.

Suchy Patricia, A. (1991). "When Worlds Collide: The Stage Direction as Utterance." In: *Journal of Dramatic Theory and Criticism* 6.1, pp. 69–82.

Tonger-Erk, Lily (2018). "Das Drama als intermedialer Text. Eine systematische Skizze zur Funktion des Nebentextes." In: *Zeitschrift für Literaturwissenschaft und Linguistik* 48, pp. 421–444.

Trilcke, Peer, Christopher Kittel, Nils Reiter, Daria Maximova, and Frank Fischer (2020). "Opening the Stage. A Quantitative Look at Stage Directions in German Drama." In: *Digital Humanities 2020*. URL: https://dh2020.adho.org/wp-content/uploads/2020/07/337_OpeningtheStageAQuantitativeLookatStageDirectionsinGermanDrama.html.

Wartena, Christian (2019). "A Probabilistic Morphology Model for German Lemmatization." In: *Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019)*, pp. 40–49. DOI: 10.25968/opus-1527.

Weber, Alexander, ed. (2018). *Zur Druckgeschichte und Intermedialität frühneuzeitlicher Dramen*. Lit.

Weise, Christian (1986). *Sämtliche Werke*. Ed. by Hans-Gert Roloff and Gerd-Hermann Susen. De Gruyter, pp. 231–390.

Wiegmann, Hermann, ed. (1987). *Die ästhetische Leidenschaft: Texte zur Affektenlehre im 17. und 18. Jahrhundert*. Olms.

Winko, Simone (2003). *Kodierte Gefühle: Zu einer Poetik der Emotionen in lyrischen und poetologischen Texten um 1900*. Erich Schmidt Verlag.

Wolf, Thomas, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush (2020). *HuggingFace's Transformers: State-of-the-art Natural Language Processing*. Preprint. DOI: 10.48550/arXiv.1910.03771.

Wood, Ian, John McCrae, Vladimir Andryushechkin, and Paul Buitelaar (2018a). "A Comparison of Emotion Annotation Approaches for Text." In: *Information* 9.5, p. 117. DOI: 10.3390/info9050117.

Wood, Ian, John P. McCrae, Vladimir Andryushechkin, and Paul Buitelaar (2018b). "A Comparison Of Emotion Annotation Schemes And A New Annotated Data Set." In: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. URL: https://www.aclweb.org/anthology/L18-1192.

Yavuz, Mehmet Can (2021). "Analyses of Character Emotions in Dramatic Works by Using EmoLex Unigrams." In: *Proceedings of the Seventh Italian Conference on Computational Linguistics, CLiC-it'20*.

Zeller, Rosemarie (2005). "Tragödientheorie, Tragödienpraxis und Leidenschaften." In: *Passion, Affekt und Leidenschaft in der Frühen Neuzeit*. Ed. by Johann Anselm Steiger. Vol. 2. Harrassowitz, pp. 691–704.

List of Contributors

Melanie Andresen
University of Stuttgart, **melanie.andresen@ims.uni-stuttgart.de**

Melanie Andresen is a computational linguist working as a postdoctoral researcher and lecturer at the Institute for Natural Language Processing at the University of Stuttgart. She received her PhD in corpus linguistics from the University of Hamburg and has since focused on digital humanities and computational literary studies.

Delphine Bernhard
Université de Strasbourg, **dbernhard@unistra.fr**

Delphine Bernhard has been a lecturer in computer science at the University of Strasbourg's Faculty of Modern Languages since September 2011. Her research interests lie in the fields of natural language processing, text mining and language resources. In particular, she focuses on less-resourced languages and the challenges involved in developing appropriate methods and resources for these languages.

Andrew Briand
University of Washington, **abriand@uw.edu**

Andrew Briand received his MS in Computational Linguistics from the University of Washington in 2022 and is now a developer technology engineer at NVIDIA. His research interests lie in natural language processing and its application to less-resourced languages.

Ingo Börner
University of Potsdam, **ingo.boerner@uni-potsdam.de**

Ingo Börner studied Russian and German Philology at the University of Vienna. He is involved in the development of the DraCor platform focusing mainly on data modeling and schema development. In July 2021 he joined the University of Potsdam as a research associate in the EU Horizon2020 funded project "CLS INFRA" to continue the work on DraCor and explore the potential of "Programmable Corpora" for literary studies.

Joanna Byszuk
Institute of Polish Language (Polish Academy of Sciences), **joanna.byszuk@ijp.pan.pl**

A PhD candidate and a full-time researcher at the Institute of Polish Language (Polish Academy of Sciences). Her interests include computational stylistics and discourse analysis, in particular their methodology, limitations, and application to audiovisual and historical works. She fre-

quently works on examining authorship in texts written in collaboration and on cross-lingual problems, and has her ongoing project applying stylometry to the study of authorship patterns in audiovisual works.

Katrin Dennerlein
University of Würzburg, **katrin.dennerlein@uni-wuerzburg.de**

Katrin Dennerlein is lecturer at the University of Würzburg. She is currently working as project leader of the computational literary studies project "Emotions in Drama" where she explores emotions with historical hermeneutic as well as computational methods. She did her PhD on a narratology of space and wrote her second book (Habilitation) on German comedy in the seventeenth and eighteenth century. Among her research interests are German drama and opera from 1500 to 1850, novels around 1800, mixed methods in digital humanities, and narrated space and mobility in novellas and novels from the nineteenth century to our days.

Josef Doležal
Academy of Performing Arts in Prague, **jodoleza@gmail.com**

Josef is a student of directing and dramaturgy at Academy of Performing Arts in Prague. He has studied acting at the conservatory. He also likes to write theatre plays.

Maciej Eder
Institute of Polish Language (Polish Academy of Sciences), Institute of Polish Studies at the Pedagogical University of Kraków, Poland, **maciej.eder@ijp.pan.pl**

He is interested in European literature of the Renaissance and the Baroque, classical heritage in early modern literature, and scholarly editing (his most recent book is a critical edition of sixteenth-century Polish translations of the *Dialogue of Salomon and Marcolf*). For at least a couple of years, he's been interested in computer-based stylometry and non-traditional authorship attribution. His work is now focused on a thorough re-examination of current attribution methods and applying them to non-English languages, e.g. Latin and Ancient Greek.

Pablo Ruiz Fabo
Université de Strasbourg, **ruizfabo@unistra.fr**

Pablo Ruiz Fabo has been a lecturer in computational linguistics at the University of Strasbourg's Faculty of Modern Languages since September 2018. His research interests lie in language technology applications in digital humanities and computational literary studies.

Frank Fischer
Free University of Berlin, **fr.fischer@fu-berlin.de**

Frank Fischer is Professor of Digital Humanities at Freie Universität Berlin. He holds Master's Degrees in Computer Science and German Studies from Leipzig University and received his PhD from the University of Jena with a study on revenge drama in the Enlightenment. From 2017 to 2021 he was director of DARIAH-EU, the pan-European digital research infrastructure for the Arts and Humanities. He is founder and editor-in-chief of DraCor (https://dracor.org/), a multilingual platform dedicated to digital research on European drama.

Beatrice Maria Vittoria Gavazza
University of Freiburg, **beatrice.gavazza@altphil.uni-freiburg.de**

Beatrice Maria Vittoria Gavazza is Assistentin at the chair of Greek Studies at the University of Freiburg. She is author of a commentary on the surviving work of the tragedian Agathon (Narr Francke Attempto 2021). Her actual research focuses on fragmentary plays by the comic poet Menander as part of the project "Kommentierung der Fragmente der griechischen Komödie" (Heidelberger Akademie der Wissenschaften). Since 2019, she has been working on the project "Lexicon of objects of the Greek comedy."

Leon A. Glaser
University of Freiburg, **leon.a.glaser@gmail.com**

Leon Glaser studied classics, philosophy and history in Freiburg, Bordeaux and Thessaloniki and is currently a Master's student in Vienna. Since 2019, he has been working at the database "Lexicon of objects of the Greek comedy."

Laura Hernández-Lorenzo
Spanish National Distance University (UNED), **laurahlr@gmail.com**

She is currently a Juan de la Cierva postdoctoral researcher. Previously she was a postdoctoral researcher at POSTDATA ERC project (UNED, Spain), at the Institute of Polish Language (Krakow) and at the University of Seville. She holds a PhD in Spanish Literature, which has been awarded as best PhD in Digital Humanities by the BBVA Foundation and the Spanish Association of Digital Humanities. Her research focuses on the application of Digital Humanities, mainly quantitative, Computational Stylistics and Stylometry methodologies to Spanish Literature, and especially to Spanish Poetry.

Rebecca M. M. Hicke
Cornell University, **rmh327@cornell.edu**

Rebecca Hicke is a PhD student in Computer Science at Cornell University. She is advised by David Mimno. Her work lies at the intersection of Machine Learning and the Digital Humanities and focuses on what we can learn by applying ML techniques to the study of humanities datasets. She graduated from Carleton College with a B.A. in English and Computer Science in 2022.

Benjamin Krautter
University of Cologne, **benjamin.krautter@uni-koeln.de**

Benjamin Krautter is a PhD student at the Department of German Studies at the University of Heidelberg and a member of the Q:TRACK project at the University of Cologne. His research interests lie in the operationalization of literary concepts for quantitative drama analysis and the combination of quantitative and qualitative methods for the analysis and interpretation of literary texts.

David Mareček
Charles University, **marecek@ufal.mff.cuni.cz**

David is an assistant professor at Charles University, Faculty of Mathematics and Physics. He is interested in interpretation and biases in large neural language models, text generation and unsupervised machine learning.

Virginia Mastellari
University of Freiburg, **virginia.mastellari@altphil.uni-freiburg.de**

Dr. Virginia Mastellari is Assistentin of Ancient Greek Literature at the University of Freiburg since 2020 and researcher of the Heidelberger Akademie der Wissenschaften in the project "Kommentierung der Fragmente der griechischen Komödie" since 2017. She has extensively researched Greek fragmentary comedy and its reception in the Latin theater. Since 2019, she has been working at the database "Lexicon of objects of the Greek comedy," regarding objects in Greek comedy, and has been recently interested in the functions of objects on the Greek stage.

Carsten Milling
University of Potsdam, **carsten.milling@uni-potsdam.de**

Carsten Milling is a software and web developer. He is technical lead of "DraCor" (https://dracor.org/) and currently part of work package 7 "Building the Ecosystem of and for Programmable Corpora" in the research project "Computational Literary Studies Infrastructure," funded by the European Commission.

David Mimno
Cornell University, **mimno@cornell.edu**

David Mimno is an associate professor in the department of Information Science at Cornell University. He holds a PhD from UMass Amherst and was previously the head programmer at the Perseus Project at Tufts and a researcher at Princeton University. His work has been supported by the Sloan foundation, the NEH, and the NSF.

Tomáš Musil
Charles University, **musil@ufal.mff.cuni.cz**

Tomáš is a research assistant and PhD student at Charles University, Faculty of Mathematics and Physics in computational linguistics, with background in artificial intelligence and philosophy of language.

Benjamin Nagy
Institute of Polish Language (Polish Academy of Sciences), **benjamin.nagy@ijp.pan.pl**

Ben Nagy is a researcher in computational stylometry, computational poetics, and a doctoral candidate at the IJP PAN. He specialises in authorship attribution questions in classical Latin verse.

Marie Nováková
"Tygr v tísni" theatre, Prague, **masanovak@email.cz**

Marie is a dramaturge, an author and a director. She is also a Co-founder and dramaturge of the Prague theater group called Tygr v tísni and the theater scene VILA Štvanice. In her dramaturgical and authorial work, she deals with adaptations of non-dramatic materials. She cooperates with Český rozhlas and many Czech theaters. She is a lecturer in theater various workshops.

Anna Novokhatko
University of Thessaloniki, **anna.novokhatko@altphil.uni-freiburg.de**

Anna A. Novokhatko is Associate Professor of Classical Philology at the Aristotle University Thessaloniki. She has published numerous articles on comedy and is the author of a monograph on the embodiment of scholarly discourse on comic stage (De Gruyter 2023). She is a co-editor of, among others, Classics@20: Digital Text Analysis of Greek and Latin sources; Methods, Tools, Perspectives, 2022, https://classics-at.chs.harvard.edu/volume/classics20-digital-text-analysis-of-greek-and-latin-sources/ and Digitale Altertumswissenschaften: Thesen und Debatten zu Methoden und Anwendungen (Propylaeum Universitätsbibliothek Heidelberg 2020). In her recent research she concentrates on cognitive studies and their interaction with classics, as well as methodologies and development of Digital Classics.

Janis Pagel
University of Cologne, **janis.pagel@uni-koeln.de**

Janis Pagel is a final year PhD student at the Institute for Natural Language Processing at the University of Stuttgart and research associate at the Department for Digital Humanities at the University of Cologne. He studied German studies and linguistics in Bochum, and computational linguistics in Stuttgart and Amsterdam. His research focuses on the application of computational linguistic methods to literary studies and coreference resolution on literary texts.

Nils Reiter
University of Cologne, **nils.reiter@uni-koeln.de**

Nils Reiter is professor for Digital Humanities and Computational Linguistics at the University of Cologne. His research interests are the use of computational methods in the humanities, and the operationalization of humanities research questions and concepts. In the past years, he has specifically worked in the area of computational literary studies.

Rudolf Rosa
Charles University, **rosa@ufal.mff.cuni.cz**

Rudolf is a researcher at Charles University, Faculty of Mathematics and Physics, and is an expert in multilingual language processing & robopsychologist. Rudolf has participated in many national and international research projects and supervises several theses in language generation.

Artjoms Šeļa
Institute of Polish Language (Polish Academy of Sciences), **artjoms.sela@ijp.pan.pl**

Assistant Professor at the Department of Methodology at IJP PAN and Computational Stylistics Group. He works in computational literary studies, with a primary focus on semantics and morphology of European verse.

Thomas Schmidt
University of Regensburg, **thomas.schmidt@informatik.uni-regensburg.de**

Thomas Schmidt holds a Master of Science in media informatics from the University of Regensburg. He works as research assistant and lecturer at the chair for media informatics at the University of Regensburg since 2018. He published research in a wide range of topics including information behavior, human-computer interaction, natural language processing and digital humanities. In the context of digital humanities his main research interests are computational literary studies, social media analysis and quantitative movie analysis. He is currently working on his PhD about computational emotion classification in German plays from the seventeenth to the nineteenth

century as part of his work as research assistant in the DFG project "Emotions in Drama." More information: https://go.ur.de/thomas-schmidt.

Botond Szemes
HUN-REN Research Centre for the Humanities, **szemes.botond@abtk.hu**

Member of the Department of Literary Theory in HUN-REN Research Centre for the Humanities and The Research Group of Stylistics at Eötvös Loránd University. In addition to stylometry and quantitative drama analysis, he researches the praxeology of literary studies and the history of data visualisation.

Peer Trilcke
University of Potsdam, **trilcke@uni-potsdam.de**

Peer Trilcke is professor of modern German literature at the University of Potsdam and director of the Theodor Fontane Archive. He is founder and spokesperson of the Potsdam Network for Digital Humanities and currently lead of work package 7 "Building the Ecosystem of and for Programmable Corpora" in the research project "Computational Literary Studies Infrastructure," funded by the European Commission. Together with Evelyn Gius and Christof Schöch, he founded the "Journal of Computational Literary Studies" (https://jcls.io) in 2021.

Evgeniya Ustinova
Charles University Prague, **eugustinova@gmail.com**

Evgeniya Ustinova is a Master's student of the Erasmus Mundus European Masters Program in Language and Communication Technologies (LCT) at the University of Malta and Charles University in Prague. She holds a Bachelor's Degree in Applied Mathematics from the National Research University "Higher School of Economics," where she joined the DraCor project and started doing research in the area of Digital Humanities. In her Bachelor's thesis, she used statistical analysis to study quantitative dominance relations in Russian Drama.

Bence Vida
Eötvös Loránd University, **vida.bence@btk.elte.hu**

A member of The National Laboratory for Digital Heritage, Hungary. He is working on the reconstruction of early modern communication networks in Eastern-Europe.

Klára Vosecká
Academy of Performing Arts in Prague, **klara.vosecka@gmail.com**

Klára is an undergraduate student of directing and dramaturgy at Theatre Academy of Performing Arts in Prague.

Carole Werner
Université de Strasbourg, **wernerc@unistra.fr**

Carole Werner received her PhD in German Studies at the University of Strasbourg's Faculty of Modern Languages in December 2022. She is now a postdoctoral researcher at the same faculty. Her research interests lie in historical sociolinguistics, sociolinguistics and dialectology. In particular, she focuses on language contact in the written production of Alsatian less experienced writers between 1681 and 1914.

Christian Wolff
University of Regensburg, **christian.wolff@informatik.uni-regensburg.de**

Christian Wolff has been professor for media informatics at the University of Regensburg since 2003. He studied Information Science, History, General Linguistics, and English and American Studies at the Universities of Regensburg and Bielefeld. In 1994 he earned a doctorate in information science in Regensburg and in 2000 a habilitation degree in computer science from the University of Leipzig. As an information and computer scientist, he has a long record of research in information behaviour, usability engineering / human-computer interaction, legal tech, and digital humanities. Since the 1990ies, he has worked on projects in corpus linguistics, text technology, and natural language processing. Besides the B.A. and M.Sc. degree programs in media informatics, he has helped establish the M.A. Program in Digital Humanities and the LL. M. Legal Tech executive degree at the University of Regensburg. More information: https://go.ur.de/christian-wolff.