Univariate Statistical Analysis of a Non-Canonical Literary Genre

Quantifying German-Language One-Act Plays (1740–1850)

Viktor J. Illmer[1,*], Dîlan Canan Çakir[1,*], Frank Fischer[1,*], Carsten Milling[2] and Lilly Welz[1]

[1]EXC 2020 Temporal Communities, Freie Univerisät Berlin, Germany
[2]CLS INFRA, University of Potsdam, Germany

Abstract

This article explores the use of metadata to analyse German-language one-act plays from 1740 to 1850, addressing the need to expand beyond canonical texts in literary studies. Utilising the Database of German-Language One-Act Plays, we examine aspects such as the number of scenes and characters as well as the role of different original languages on which the translated plays in the corpus are based. We find that one-act plays exhibit strong genre signals that set them apart from multi-act plays of the time. Our metadata-driven approach provides a comprehensive and statistically grounded understanding of the genre, demonstrating the potential of digital methods to enhance genre studies and overcome traditional limitations in literary scholarship.

Keywords

literary studies, drama, genre theory, univariate statistics, metadata

1. Introduction

It is an early promise of the digital humanities "to look beyond the canon" [19]. The specifics of how this is to be achieved often remain unclear, whether due to obstacles in obtaining approporiate material, unfamiliarity with non-canonical sources, or simply resource constraints, but the fact that this must happen has been emphasised time and again:

> "The literary scholar of the twenty-first century can no longer be content with anecdotal evidence, with random 'things' gathered from a few, even 'representative,' texts. We must strive to understand these things we find interesting in the context of everything else, including a mass of possibly 'uninteresting' texts." [15]

At the time, Matthew Jockers, from whom the quote is taken, was aiming at full-text corpora, which he examined using digital humanities methods such as stylometry and topic modelling.

✉ v.illmer@fu-berlin.de (V. J. Illmer); dilan.cakir@fu-berlin.de (D. C. Çakir); fr.fischer@fu-berlin.de (F. Fischer); milling@uni-potsdam.de (C. Milling); l.welz@fu-berlin.de (L. Welz)

🆔 0000-0002-7334-781X (V. J. Illmer); 0009-0001-0013-7205 (D. C. Çakir); 0000-0003-2419-6629 (F. Fischer); 0000-0003-0553-7512 (C. Milling); 0009-0008-0404-8147 (L. Welz)

It was clear to him and his readers that he would never have all the 19th-century English-language novels that could potentially belong to his working corpus available in full-text versions suitable for research.

Recognising this problem, the editors of the *European Literary Text Collection* (ELTeC) proposed a different kind of approach. Rather than aiming to collect as many full-texts as possible, or a representative sample, they built several corpora that would only contain 100 novels from a given language and would be balanced according to various criteria:

> "In the absence of exhaustive bibliographic records of novelistic production for most of the languages covered in ELTeC, no attempt at a randomly sampled, statistically representative corpus can reasonably be made. Instead, the corpus-composition criteria aim to ensure that the breadth and variety of novels produced during the period covered by ELTeC are well represented, while at the same time ensuring rough comparability across collections." [21]

The two corpus-based projects mentioned above have identified their own shortcomings, shortcomings that we would like to partly overcome in this article. We refrain from analysing full-text corpora and focus our attention on metadata, i.e. we use metadata to describe aspects of literary history. In doing so, we refer to two types of metadata, "the kind of descriptive, bibliographic metadata found in repositories such as library catalogues" [10] for one, but also other metadata that can be systematically collected beyond bibliographic records.

With help of the *Database of German-Language One-Act Plays 1740–1850* (*Einakter Database* hereafter) located at einakter.dracor.org, we will analyse the German-language one-act plays that were written, performed and/or printed in the mentioned time span, and we will do this on the basis of the aforementioned "exhaustive bibliographic records".

The assumption of *actually* operating on such exhaustive bibliographic records is based on our consultation of all researchable sources available to us in our field (encyclopaedias, theatre programmes, bibliographies, library catalogues) for the purpose of research – what is evidently documented is therefore included in our database. We are therefore dealing with an extensive representative sample of German-language one-act plays from 1740 to 1850, and we can describe core aspects of the genre[1] with the means of statistics. Our analysis is guided by the following five research questions:

Q1 What is the proportion of subtitle-based categories among one-act plays?

Q2 Is there a difference in the *number of scenes per act* between German one-act plays and other German plays of the same period?

Q3 Is there a difference in the *number of characters* between German one-act plays and other German plays of the same period?

Q4 What is the proportion of original languages among translated plays?

Q5 Does an estimation of the *mean number of characters* based on Pazarkaya's sample match the estimation based on the Einakter Database's sample?

---

[1]We regard one-act plays as a genre in their own right; we follow the reasoning in [3].

2. Genre theories in literary studies

Categorisation is one of the core tasks in literary studies; genre theories are almost as old as literary studies themselves [2]. Yet, analyses of characteristics are often only heuristic approximations for a genre definition. Genre characteristics are essentially summarised in three points, which describe that a genre has both necessary and alternative characteristics, that text features are recognised as a convention by a temporal community at a certain time and are marked with simple genre signals [13].

Generally, these genre signals are formulated on the basis of a small group of (canonical) texts and many paratexts and are rarely fundamentally reviewed or supplemented, usually for pragmatic reasons.

Economic scientific communication requires defined terms and genre characteristics, but classical methods of literary studies rarely offer the necessary tools or resources to describe a genre extensively, let alone exhaustively. Although the methodological alternatives are few in practice (especially in the pre-digital era), researchers are often criticised for the numerical discrepancy between the number of works known to exist and the number of works actually used for analysis [28]. This approach of considering only a few selected works as examples regarded as representative of the whole is a methodological shortcoming, without which, however, no genre theories can be developed in literary studies – at least none that can be mastered by a single person using the traditional tools of the field.

With the help of the *Einakter Database*, questions of genre can be negotiated on a new, broader and more inclusive digital and statistical basis [4]. Our aim is to analyse one-act plays from the given time frame in their entirety in order to make statements about the characteristics of the genre as a whole. Without access to full texts for the entire corpus, many questions will have to remain unexplored for now. We will therefore focus on demonstrating that essential literary-historical insights can also be derived from metadata.

One of the last comprehensive works on one-acters in the 18th and early 19th centuries was published about 50 years ago [20],[2] and there, too, an attempt was made to describe the genre with statistics and counts – however, only around 200 to 300 one-act plays were considered, whereas the *Einakter Database* has so far assembled 2,568 one-act plays for the same period, a whole order of magnitude more. As the database offers its content in a machine-readable format, we can carry out statistical analyses to describe the generic characteristics of the genre.
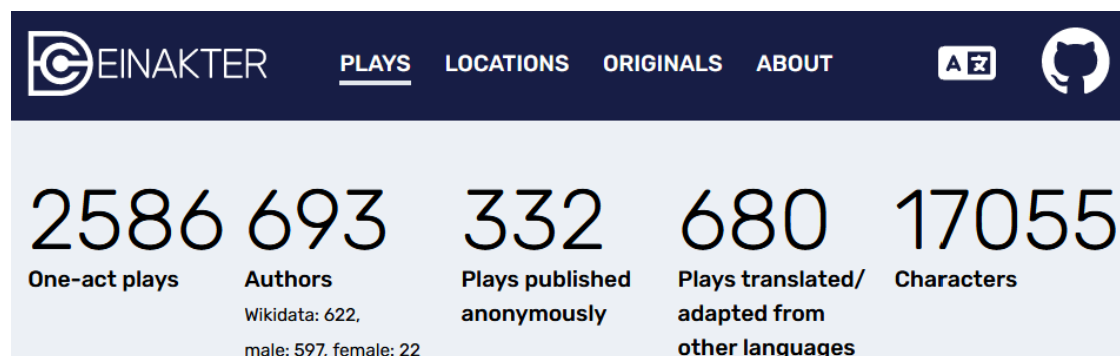
3. Einakter Database

The *Einakter Database* recognises one-act plays as a distinct dramatic genre in the 18th and early 19th centuries. Currently, it contains metadata on 2,586 one-act plays from 1740 to 1850 (Figure 1). Explicit labelling of a play as a play "in one act" (in the epitext or peritext) serves as a criterion for inclusion in the corpus. This type of marking one-act plays first emerged in the middle of the 18th century and, up to the middle of the 19th century, was associated with very specific features, which will be discussed below.

---

[2]According to its title, Pazarkaya's work only refers to the 18th century, but he does include plays from the 19th century in his corpus.

By focusing on the subtitle, the corpus for the present database is precisely defined, not merely for pragmatic reasons (for more information on inclusion criteria, see [4]). This method precludes further classification issues, as the works themselves bear the subtitle, thereby actively aligning themselves with a genre convention through their authors, editors, printers or theatre directors. This does not exclude the possibility of some structural differences between works with the same subtitles, which may lead to some being considered formally or thematically atypical. Following the concept of family resemblance, it is assumed that the works share several features, though never all simultaneously. Furthermore, this does not preclude the existence of plays that exhibit exactly the same typical features as those marked as one-act plays but do not carry such a title. Such works are not considered in this study.

As most of the plays in the corpus are non-canonical [3], only very few (114 plays, to be precise[3]) are available as full text versions. Instead, we rely exclusively on metadata to analyse the genre.

Figure 1: Screenshot of the einakter.dracor.org frontend. At the time of writing, the database contained 2,586 one-act plays.

The database contains, among other things, information on the period of origin, the first performance, the number of scenes, bibliographical details, links to digital copies, names and gender of the characters, information on the setting, links to encyclopaedias and keywords on the content of the plays (Figure 2).

4. Analysis

4.1. Subtitle categorisation

Q1 What is the proportion of subtitle-based categories among one-act plays?

Little is known about the genre of the one-act play in the period covered by this study. According to what is stated about them, the plays are mostly simple comedies [20]. This assertion

---

[3]We are referring to the plays in the German Drama Corpus at https://github.com/dracor-org/gerdracor (revision `94678457159dbfd8961dd954c24ee5d3ce8a6e35`). In contrast to the Einakter Database, however, this database does not contain all plays from the period under investigation, but a selection of mostly canonical texts [11].

Figure 2: Detailed view of a single play in the database.

can be verified with the help of the subtitles of the one-act plays in the database. The subtitles not only contain information on the number of acts ("in one act"), but often also on the subgenre ("comedy", "tragedy", etc.). They were grouped into categories using simple regular expressions to account for common spelling variations ("Komödie", "Comoedie", "Komoedie", etc.).

Table 1

Distribution of categories in the dataset with binomial confidence intervals.

Category		n	Percent	95% CI	
Lustspiel	Comedy	1433	55.41	53.47	57.34
Schauspiel		249	9.63	8.52	10.83
Posse	Farce	223	8.62	7.57	9.77
Nachspiel	Postlude	65	2.51	1.95	3.19
Drama		61	2.36	1.81	3.02
Trauerspiel	Tragedy	61	2.36	1.81	3.02
Schwank		51	1.97	1.47	2.58
Vorspiel	Prelude	46	1.78	1.31	2.37
Other	Other	397	15.35	13.98	16.80

Binomial proportion confidence intervals were calculated for these categories in an attempt to generalise the results to one-act plays of this period not in the sample. The intervals were calculated from the Clopper-Pearson interval ("exact" method) [5]. Results show *comedy* (*Komödie*) to be the most common category by a wide margin (CI = [53.5%, 57.3%]), followed by *Schauspiel* (CI = [8.5%, 10.8%]) and *farce* (*Posse*, CI = [7.6%, 9.8%]), whose order cannot be determined due to overlapping CIs (Table 1). All categories that follow (postlude, tragedy, *Drama*, *Schwank* and prelude) exhibit no meaningful rank differences among them.

In terms of literary history, this is an eminent realisation: Since this result is not merely an observation based on anecdotal knowledge from a selection of well-known (canonical) plays, but on a probability calculation on all identifiable works of a genre, we have an entirely new basis for argumentation in literary studies. Even without this statistical analysis, it has been generally assumed in 18th-century theatre research that comedies were more commonly written than tragedies [18]. This hypothesis can now be adequately supported, at least for one-act plays. It is important to emphasise that the objective was not merely to pursue surprising outcomes, but rather to uncover empirically verifiable metrics pertaining to a specific genre. This approach goes beyond the scope of conventional literary studies, which often relies on heuristic methods.

The various subgenres not only appear with varying frequency, but also at different times; the one subgenre that is relatively constantly represented is comedy (Figure 3). For the further analysis of the one-act genre, these moments of transition (such as a decrease in *Nachspiel* around 1740 or an increase in *Schwank* after 1800) would have to be considered.

4.2. Number of scenes

> Q2 Is there a difference in the *number of scenes per act* between German one-act plays and other German plays of the same period?

To answer this question, the Drama Corpora Project's *GerDraCor* corpus metadata was used for comparison. In contrast to the database of one-act plays, however, this database does not contain all plays from the period under investigation, but a selection of mostly canonical texts. This data was filtered to the time period 1740–1850 and plays with 2–5 acts, both to avoid
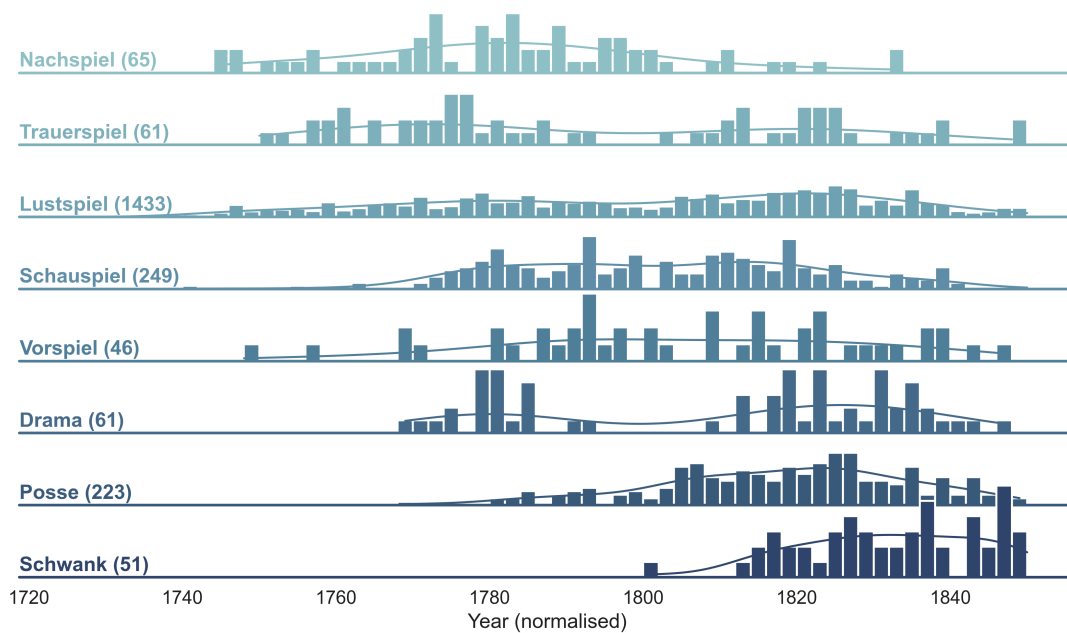
Figure 3: Histogram with kernel density estimation line of number of plays per subtitle category by normalised year (bin width: 2 years). Bar heights are normalised per category. For details on the *normalised year* variable, see [7].

duplicates between the *Einakter* and *GerDraCor* datasets and to avoid the inclusion of one-act plays that do not fall under the *Einakter* definition (cf. [3]). There are only few plays with six or more acts in *GerDraCor*, which were therefore excluded. In addition, this number of acts is rather atypical for the period of analysis and less significant for our genre questions.[4] Because *GerDraCor's* scene counts are dynamically derived from the underlying TEI encoding, the relevant variable is named `numOfSegments`. Nonetheless, segments may be interpreted as scenes for this purpose.[5]

From this data, we calculated the number of scenes per act and aggregated the results for each number of acts, which gives the *mean number of scenes per act* (Table 2). To calculate confidence intervals for the mean number of scenes per act, we first checked whether the number of scenes followed a normal distribution. A Shapiro-Wilk test for normality [22] was conducted on both the *Einakter* and *GerDraCor* datasets, which yielded p-values $p < 0.001$ for both, showing that the data is not approximately normally distributed. We therefore chose to determine the confidence intervals via bootstrapping [8]. For this purpose, we generated 10,000 resamples of size 10,000 per category, calculating the mean number of scenes per act for each. The 95% confidence intervals were calculated using the bias-corrected and accelerated (BC_a) method [9]. Additionally, we conducted the Mann-Whitney U test [16] for a pairwise

---

[4]Mentions of *GerDraCor* for the purpose of comparison will henceforth refer to this filtered subset.

[5]We also excluded Johann Nestroy's plays *Zu ebener Erde und erster Stock* and *Das Haus der Temperamente* due to idiosyncrasies in their encoding, which lead to incorrect segment counts for the purposes of this analysis.

Table 2

Mean number of scenes per act across Einakter Database and GerDraCor datasets, grouped by number of acts with bootstrapped confidence intervals.

Dataset	Number of acts	n	Mean scenes per act	95% CI	
Einakter Database	1	2306	14.20	13.95	14.46
GerDraCor	2	33	13.36	11.23	15.71
GerDraCor	3	70	10.46	9.20	11.82
GerDraCor	4	28	9.04	7.41	10.77
GerDraCor	5	140	6.97	6.35	7.68
GerDraCor	2–5	271	8.86	8.25	9.52

Table 3

Mann-Whitney U test results comparing the number of scenes per act between the Einakter Database and GerDraCor datasets. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns $p \geq 0.05$.

| Dataset	Number of acts	U	
GerDraCor	2	40,942.50	ns
GerDraCor	3	108,192.00	***
GerDraCor	4	47,753.00	***
GerDraCor	5	269,931.50	***
GerDraCor	2–5	466,819.00	***

comparison of the *Einakter Database* against *GerDraCor's* 2–5-act plays individually and in the aggregate (Table 3).

For the atypical two-act and four-act plays, sample sizes are low and confidence intervals consequently wide. Due to this, we cannot observe any difference in the mean number of scenes per act between one-act and two-act plays. The confidence interval for the mean number of scenes per act is particularly wide for two-act plays (CI = [11.23, 15.71]) and the Mann-Whitney U test is not statistically significant at $p = 0.45$. Thus, we detect no significant difference in the number of scenes per act between one-act and two-act plays.

All other comparisons suggest a highly significant ($p < 0.001$) difference between one-act and multi-act plays. Each of these groups exhibits a lower number of scenes per act compared to one-act plays, though rank differences cannot be reasonably determined. As can also be gathered from the comparisons of the distributions of the number of scenes per act for the *Einakter Database*, one-act plays exhibit a much stronger *genre signal* for this variable compared to other German-language plays of the period (Figure 4).

This observation captures the unique dramaturgy of one-act plays in the 18th and early 19th centuries. It suggests, for example, a rushed, hectic play on stage. This is because every change of scene entails a change of configuration, i.e. a change of characters on stage, which means that dialogue partners in a one-act play change more frequently than in an average act of a five-act play.
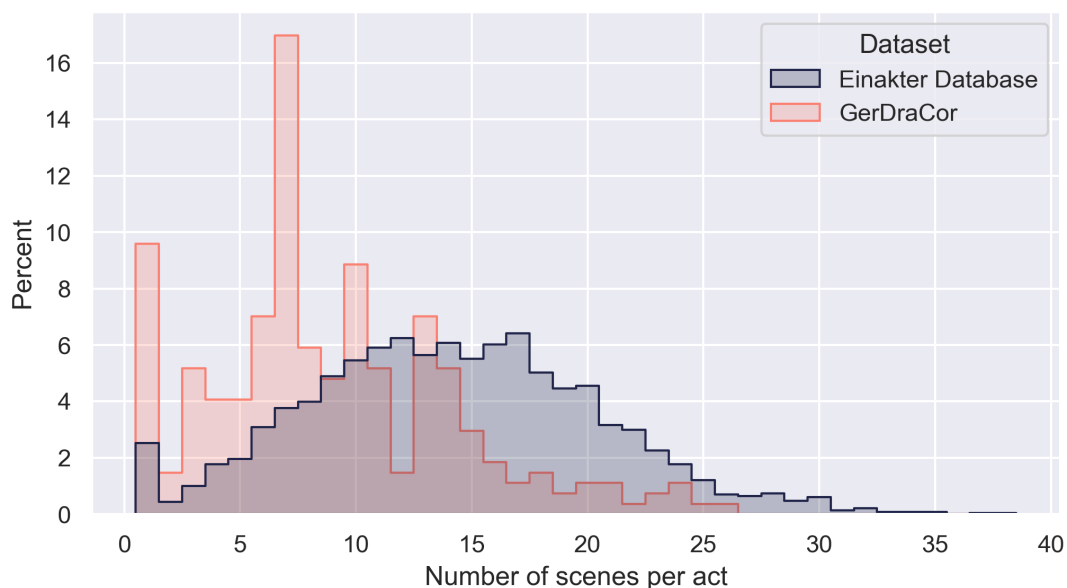
Figure 4: Distribution of number of scenes per act in the Einakter Database and GerDraCor. Bar heights sum to 100% per dataset.

4.3. Number of characters

Q3 Is there a difference in the *number of characters* between German one-act plays and other German plays of the same period?

The reduced number of characters is often referenced in attempts to define the one-act play. Some lexicons even claim that the one-act play "has only two or three characters" [6]. The significance of the number of characters for describing this genre was already observed by Pazarkaya, who began manually counting the number of characters in about 200 one-act plays in 1973. We have replicated this simple count using our corpus, which is an order of magnitude larger than Pazarkaya's, containing 2,586 one-act plays (the database contains around 150 one-act plays for which the number of characters is unknown as there is no available full-text version or cast list).

Just like for Q2, we conducted a Shapiro-Wilk test for normality on the number of characters across both *Einakter* and *GerDraCor* datasets, which again yielded results with $p < 0.001$ for both. Using sample bootstrapping with the same parameters as for Q2, we determined 95% confidence intervals for the mean number of characters (Table 4). Similarly, we calculated pairwise Mann-Whitney U values for the *Einakter Database* against each group of *n*-act plays in our filtered *GerDraCor* dataset (Table 5).

We find that there is a significant difference in the number of characters between one-act plays and each group of multi-act plays. Though confidence intervals overlap among multi-act plays, we can infer that the mean number of characters in multi-act plays is at least twice that of one-act plays. A comparison of the number of characters in one-act and five-act plays

Table 4

Mean number of characters per play across Einakter Database and GerDraCor datasets with bootstrapped confidence intervals.

Dataset	Number of acts	n	Mean number of characters	95% CI	
Einakter Database	1	2416	7.06	6.94	7.19
GerDraCor	2	34	19.47	15.35	24.59
GerDraCor	3	71	19.46	16.52	24.00
GerDraCor	4	28	21.43	17.36	33.46
GerDraCor	5	140	28.49	24.34	35.25

Table 5

Mann-Whitney U test results comparing the number of characters between the Einakter Database and GerDraCor datasets. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns $p \geq 0.05$.

Dataset	Number of acts	U	
GerDraCor	2	15,040.50	***
GerDraCor	3	18,586.50	***
GerDraCor	4	3,343.50	***
GerDraCor	5	28,585.50	***

reveals a poetic economy [1] – one-act plays have a measurably more compact structure. The reduced number of characters is often related to their shorter and less complex plot [3]. For example, the one-act plays in the period under investigation rarely contain changes of setting and, even more relevant in the context of the number of characters, there are fewer subplots than in the larger five-act play, for example.

The observation made in subsection 4.1 is also reflected in the number of characters: While the one-act play is often able to develop a plot with only two, three or four characters due to the typical comedy plot (in which, for example, a marriage with obstacles is at issue), the five-act play of the time seems to present plots in which there are never fewer than three and only rarely fewer than four characters (Figure 5).

With this type of findings we can, for example, take a canonical one-act play, such as Goethe's *Der Bürgergeneral* from 1793, and mirror some structural metadata with hard numbers "in the context of everything else", to quote Matthew Jockers. Goethe's play was both negatively received by his contemporaries and largely disregarded by scholars [27]. The play, which was based on the characters of another popular one-act play, was written to be practically implementable on stage. And structurally, in terms of the mean number of scenes and characters, the play is right in the middle: 14 scenes, 7 characters.

4.4. Translations

Q4 What is the proportion of original languages among translated plays?

We deal with the question of translations of foreign-language plays into German from the point of view of influence. A simple quantitative presentation of the data clearly shows that
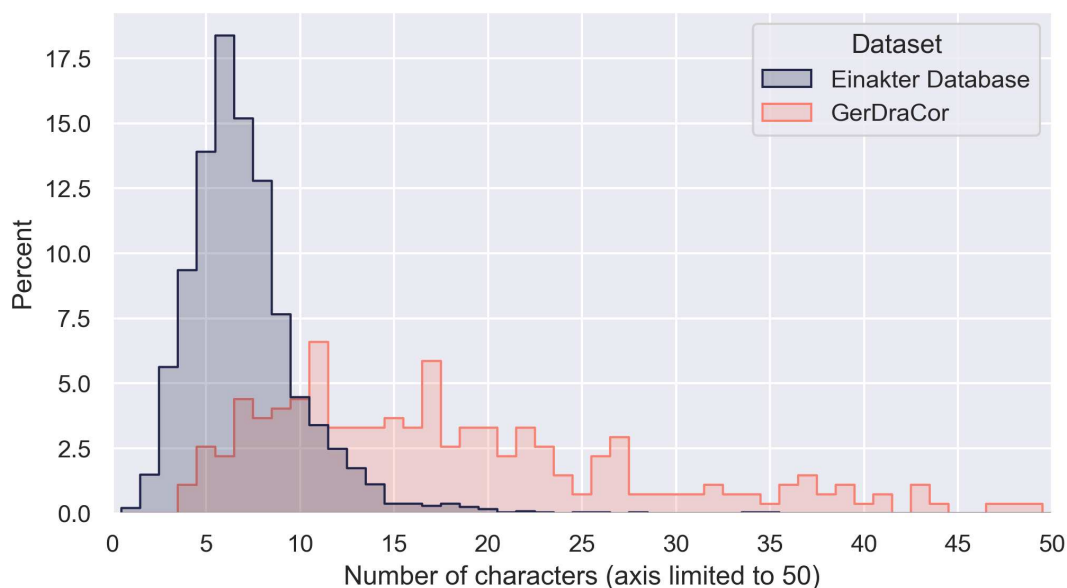
Figure 5: Distribution of number of characters in the *Einakter Database* and *GerDraCor*. Bar heights sum to 100% per dataset.

German-language one-act plays were mainly influenced by French theatre, whose one-act plays were translated en masse [3]. Numerous attempts have been made to quantify and visualise the influence of foreign-language literature on German literature. Flaischlen's *Graphische Litteratur-Tafel* (Figure 6) from 1890, a giant graph (58 × 86.5 cm) visualising German literature since its early beginnings as a stream that absorbs various other currents, is one of the many positivist attempts. It shows the influence of foreign literature – epochs, movements, authors, works – on German literature [12, 14].

The red tributaries to the main stream denote the influences of French literature. One-act plays are just one of many genres, but one name stands out, that of Eugène Scribe around the year 1830. Of all French translations contained in the *Einakter Database*, 58 were produced under Scribe's prolific (co-)authorship, who is the only author of French one-act plays mentioned in the *Litteratur-Tafel* and who stands *pars pro toto* for all the others.

As early as the 18th century, there were complaints that most authors of one-act plays merely translated from French into German [3]. But how many one-act plays are proven to be translations, and from which languages are they actually translated? Just like in subsection 4.1, we aggregated all entries with a known original by the original's language and calculated binomial proportion confidence intervals for each original language using Clopper-Pearson's method (Table 6). This shows that the overwhelming majority of originals are indeed in French (CI = [89.6%, 93.9%]). The confidence intervals for all other languages are too wide to make any meaningful statement about their order.
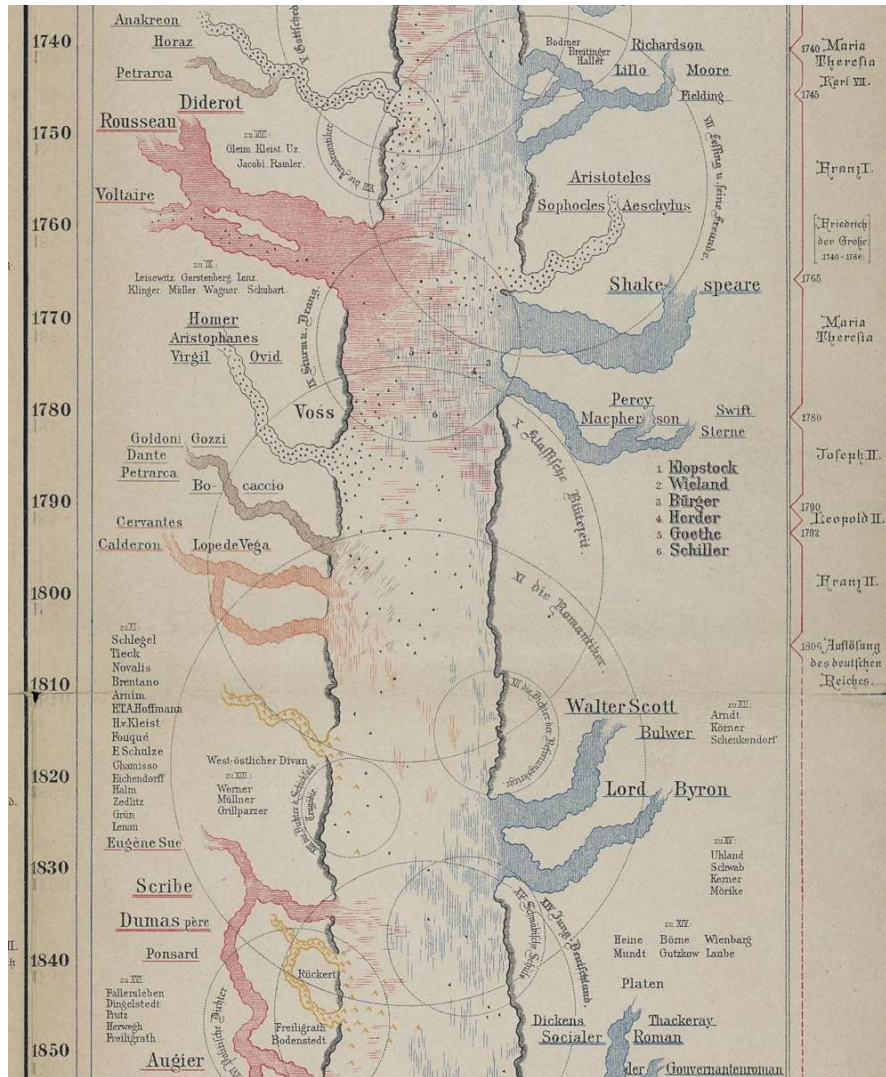
Figure 6: Detail from Cäsar Flaischlen's *Graphische Litteratur-Tafel* (1890) for the period 1740 to 1850.

4.5. Comparison with Pazarkaya's 1973 sample

Q5 Does an estimation of the *mean number of characters* based on Pazarkaya's sample match the estimation based on the *Einakter Database's* sample?

Some researchers argue that literary history may still contain undiscovered material, as only a small proportion of historical texts have been analysed to date. Others believe that, according to specific studies, certain tendencies in literature can be observed regardless of the size of the corpus and that the call to consult more non-canonical literature and larger corpora may be superfluous [23]. For a statistical analysis, however, it is sometimes necessary to use a larger corpus, as we would like to outline briefly here. We compare the analysis of the number of

Table 6
Proportion of original languages of translated plays with binomial confidence intervals.

Language	n	Percent	95% CI	
French	626	91.92	89.60	93.90
Italian	16	2.35	1.35	3.79
Danish	15	2.20	1.24	3.61
English	9	1.32	0.61	2.49
Spanish	6	0.88	0.32	1.91
Dutch	3	0.44	0.09	1.28
Latin	3	0.44	0.09	1.28
Russian	2	0.29	0.04	1.06
Czech	1	0.15	0.00	0.81

figures in Pazarkaya's corpus with those in the *Einakter Database*.

Applying the Mann-Whitney U test to compare the *Einakter Database* and Pazarkaya's sample, we received a result of $U = 313403.5$, $p < 0.001$, suggesting that there is a statistically highly significant difference between the two samples. Though we know for a fact that Pazarkaya refers to the same population of plays, the test finds strong evidence against the null hypothesis that the distributions of the two samples are the same. Looking at the mean number of characters as well as their confidence intervals and distributions (Table 7 and Figure 7), we detect no overlap in the confidence intervals of the means. This suggests that at least one of the datasets may not be a true random sample if they indeed originate from the same population.

Table 7
Mean number of characters per play in Einakter Database and Pazarkaya's samples with bootstrapped confidence intervals.

Dataset	n	Mean number of characters	95% CI	
Einakter Database	2416		7.06	6.94	7.19
Pazarkaya	221		6.12	5.85	6.41

There is some indication that Pazarkaya's dataset may be more of a *convenience sample* consisting of one-act plays available to him at the time. Pazarkaya's work, now 50 years old, was influenced by the circumstances of his time. Without digital tools and databases, his options were more limited, and his corpus was significantly smaller, consisting of about 200–300 one-act plays. These often included particularly canonical and well-preserved works that were accessible in libraries. While a similar criticism could be levelled against the *Einakter Database*, it nonetheless allows us to consider numerous first editions, including many preserved only as single copies or manuscripts, thanks to the availability of many publicly accessible digitised texts.
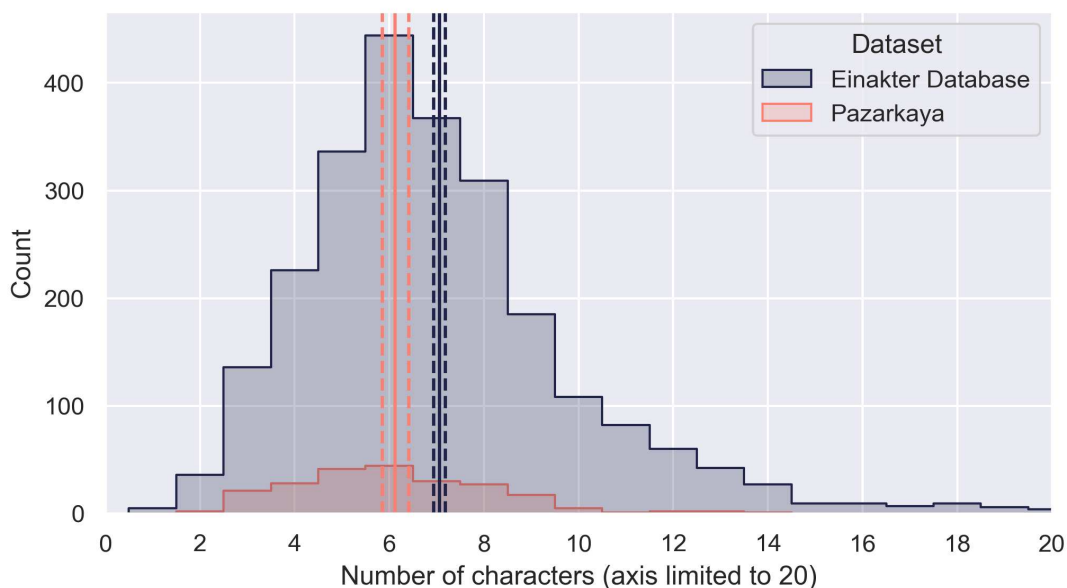
Figure 7: Distribution of number of characters in the *Einakter Database* and Pazarkaya's sample. Solid lines represent the mean, dashed lines the 95% confidence intervals of the mean.

5. Conclusion

In this paper, we demonstrate how statistical results based on the analysis of a database of German-language one-act plays can be utilised for literary historiography. The aim was not primarily to seek unexpected results, but to discover empirically determined metrics about a particular genre. While it is evident that one-act plays are "short", the question is how this shortness can be measured. The issue arises because brevity is a relative concept – something is only considered short in comparison to something else, and this relational aspect poses a significant challenge in the study of one-act plays and other short forms of literature. Addressing this issue is crucial for advancing our understanding of the genre's formal characteristics.

A subtitle analysis has drawn attention to the development of the individual subgenres of the one-act play between 1740 and 1850 (Q1). We also show that the genre signal is strongest for the mean number of characters, where one-act plays differ significantly from their German multi-act counterparts of the period. Similarly, a significant difference is found for the number of characters between one-act plays and those with three, four and five acts (Q2 and Q3). Our empirical data confirm the dominance of French plays as the preferred models for German translations by a wide margin, but also show the distribution for other original languages from which translations were produced (Q4). And finally, we were able to show that the data from an older analysis, which had a much smaller sample of one-act plays available, is not representative of the wider data, and that the accuracy of statistical findings may be increased under the conditions of the current digital landscape (Q5).

In conclusion, this paper's metadata-driven approach demonstrates the potential of digital

methods to provide a statistically grounded contribution to genre studies beyond the limitations of traditional literary scholarship.

Reproducibility

To enhance transparency, the code and data for this paper is available on GitHub at https://github.com/v-ji/einakter-chr2024-stats. We used the *Nix* package manager (nixos.org) to ensure the reproducibility of our computational environment. Nix locks the exact versions of all dependencies, including Python, the necessary libraries for executing the Jupyter notebook, and the *Einakter Database* dataset. By using Nix, we can precisely recreate the Python environment, fetch the dataset from a stable URL, and verify its integrity with a hash to ensure it matches the version used in this contribution.

The main libraries used are Python Polars 1.7.1 [24] for data manipulation, SciPy 1.14.0 [25] for statistical analysis, as well as Matplotlib 3.9.1 [17] and seaborn 0.13.2 [26] for plotting.

Acknowledgments

References

[1]   M. Bauer. "Poetic Economy: ellipsis and redundancy in literature". In: *Connotations. A Journal for Critical Debate* 21.2-3 (2011). URL: https://www.connotations.de/article/matthias-bauer-poetic-economy-ellipsis-and-redundancy-in-literature/.

[2]   M. Bies, M. Gamper, and I. Kleeberg. "Einleitung". In: *Gattungs-Wissen: Wissenspoetologie und literarische Form.* Göttingen: Wallstein, 2013, pp. 7–18.

[3]   D. C. Çakir. *Poetische Ökonomie im Drama: Einakter im 18. und frühen 19. Jahrhundert.* De Gruyter, 2024. DOI: 10.1515/9783111334059. URL: https://www.degruyter.com/document/doi/10.1515/9783111334059/html.

[4]   D. C. Çakir and F. Fischer. "Dramatische Metadaten: Die Datenbank deutschsprachiger Einakter 1740–1850". In: *DHd2022: »Kulturen des digitalen Gedächtnisses«* (2022). DOI: 10.5281/zenodo.6327977.

[5]   C. J. Clopper and E. S. Pearson. "The use of confidence or fiducial limits illustrated in the case of the binomial". In: *Biometrika* 26.4 (1934), pp. 404–413. DOI: 10.2307/2331986. URL: https://www.jstor.org/stable/2331986.

[6]   J. A. Cuddon. *A dictionary of literary terms.* Penguin reference books. London: Penguin Books, 1979.

[7]   Drama Corpora Project. *DraCor FAQ.* 2024. URL: https://dracor.org/doc/faq.

[8]   B. Efron. "Bootstrap methods: Another look at the jackknife". In: *The Annals of Statistics* 7.1 (1979), pp. 1–26. DOI: 10.1214/aos/1176344552. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-7/issue-1/Bootstrap-Methods-Another-Look-at-the-Jackknife/10.1214/aos/1176344552.full.

[9]   B. Efron. "Better bootstrap confidence intervals". In: *Journal of the American Statistical Association* 82.397 (1987), pp. 171–185. DOI: 10.1080/01621459.1987.10478410. URL: https://www.tandfonline.com/doi/abs/10.1080/01621459.1987.10478410.

[10]   M. Erlin. "From literature to metadata". In: *Goethe Yearbook* 27.1 (2020), pp. 189–196. DOI: 10.1353/gyr.2020.0005. URL: https://muse.jhu.edu/article/762245.

[11]   F. Fischer, I. Börner, M. Göbel, A. Hechtl, C. Kittel, C. Milling, and P. Trilcke. "Programmable Corpora: Introducing DraCor, an infrastructure for the research on European drama". In: *DH2019: »Complexities«* (2019). DOI: 10.5281/zenodo.4284002. URL: https://zenodo.org/record/4284002.

[12]   C. Flaischlen. *Graphische Litteratur-Tafel: Die deutsche Litteratur und der Einfluß fremder Litteraturen auf ihren Verlauf von Beginn einer schriftlichen Überlieferung an bis heute in graphischer Darstellung.* Stuttgart: Göschen, 1890. URL: http://resolver.sub.uni-goettingen.de/purl?PPN860488233.

[13]   H. Fricke. "Invarianz und Variabilität von Gattung". In: *Handbuch Gattungstheorie.* Ed. by R. Zymner. Stuttgart: J.B. Metzler, 2010, pp. 19–21. DOI: 10.1007/978-3-476-00509-0_2.

[14]   A. Hechtl, I. Börner, F. Fischer, and P. Trilcke. "Cäsar Flaischlen's ›Graphische Litteratur-Tafel‹: digitising a giant historical flowchart of foreign influences on German literature". In: *DH2017: »Access/Accès«* (2017). URL: https://dh2017.adho.org/abstracts/506/506.pdf.

[15]   M. L. Jockers. *Macroanalysis: digital methods and literary history.* Topics in the Digital Humanities. Urbana: University of Illinois Press, 2013.

[16]   H. B. Mann and D. R. Whitney. "On a test of whether one of two random variables is stochastically larger than the other". In: *The Annals of Mathematical Statistics* 18.1 (1947), pp. 50–60. URL: https://www.jstor.org/stable/2236101.

[17]   Matplotlib Development Team. *Matplotlib: visualization with Python.* 2024. DOI: 10.5281/zenodo.10916799. URL: https://zenodo.org/doi/10.5281/zenodo.10916799.

[18]   R. Meyer. *Das deutsche Trauerspiel des 18. Jahrhunderts: eine Bibliographie. mit ca. 1250 Titeln, einer Einleitung sowie Verfasser- und Stichwortregister.* München: Fink, 1977.

[19]   F. Moretti. "Conjectures on world literature". In: *New Left Review* 1 (2000), pp. 54–68. URL: https://newleftreview.org/issues/ii1/articles/franco-moretti-conjectures-on-world-literature.

[20]   Y. Pazarkaya. *Die Dramaturgie des Einakters: der Einakter als eine besondere Erscheinungsform im deutschen Drama des 18. Jahrhunderts.* Göppinger Arbeiten zur Germanistik ; Nr. 69. Göppingen: Kümmerle, 1973.

[21] C. Schöch, R. Patras, T. Erjavec, and D. Santos. "Creating the European Literary Text Collection (ELTeC): Challenges and perspectives". In: *Modern Languages Open* 1 (2021), p. 25. DOI: 10.3828/mlo.v0i0.364. URL: http://www.modernlanguagesopen.org/articles/10.3828/mlo.v0i0.364/.

[22] S. S. Shapiro and M. B. Wilk. "An analysis of variance test for normality (complete samples)". In: *Biometrika* 52.3/4 (1965), pp. 591–611. DOI: 10.2307/2333709. URL: https://www.jstor.org/stable/2333709.

[23] T. Underwood. *Distant horizons: digital evidence and literary change.* Chicago: The University of Chicago Press, 2019.

[24] R. Vink, S. d. Gooijer, A. Beedie, M. E. Gorelli, W. Guo, O. Peters, J. v. Zundert, name-exhaustion, G. Hulselmans, C. Grinstead, G. Burghoorn, Marshall, chielP, L. Mitchell, I. Turner-Trauring, M. Santamaria, D. Heres, J. Magarick, H. Harbeck, ibENPC, K. Genockey, M. Wilksch, deanm0000, J. Leitao, M. v. Gelderen, P. Barbagiannis, I. Koutsouris, J. Haag, and O. Borchert. *Python Polars 1.7.1.* 2024. DOI: 10.5281/zenodo.13754515. URL: https://zenodo.org/doi/10.5281/zenodo.13754515.

[25] P. Virtanen et al. "SciPy 1.0: fundamental algorithms for scientific computing in Python". In: *Nature Methods* 17.3 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2. URL: https://www.nature.com/articles/s41592-019-0686-2.

[26] M. Waskom. "seaborn: statistical data visualization". In: *Journal of Open Source Software* 6.60 (2021), p. 3021. DOI: 10.21105/joss.03021. URL: https://joss.theoj.org/papers/10.21105/joss.03021.

[27] W. D. Wilson. "Dramen zum Thema der Französischen Revolution". In: *Goethe Handbuch.* Ed. by T. Buck. Stuttgart: J.B. Metzler, 1996, pp. 258–287. DOI: 10.1007/978-3-476-03653-7_14. URL: http://link.springer.com/10.1007/978-3-476-03653-7%5C%5F14.

[28] T. Ziolkowski. "Review of Der moderne Einakter. Eine poetologische Untersuchung". In: *The Journal of English and Germanic Philology* 67.3 (1968), pp. 505–507. URL: https://www.jstor.org/stable/27705581.