

Philology: scientific researches

Правильная ссылка на статью:

Zenkov A.V. The numbers reveal the author: a stylometric comparison of German-language modernist texts // Филология: научные исследования. 2024. № 11. DOI: 10.7256/2454-0749.2024.11.72167 EDN: PDWIOX URL: https://nbpublish.com/library_read_article.php?id=72167

The numbers reveal the author: a stylometric comparison of German-language modernist texts / Числа указывают на автора: стилометрическое сопоставление немецкоязычных модернистских текстов

Зенков Андрей Вячеславович

ORCID: 0000-0002-1233-9082

кандидат физико-математических наук

доцент, кафедра "Моделирование управляемых систем"; Уральский федеральный университет
старший научный сотрудник; Уральский федеральный университет

620002, Россия, Свердловская область, г. Екатеринбург, ул. Мира, 19, оф. 434

✉ zenkow@mail.ru



[Статья из рубрики "Автоматическая обработка языка"](#)

DOI:

10.7256/2454-0749.2024.11.72167

EDN:

PDWIOX

Дата направления статьи в редакцию:

01-11-2024

Аннотация: Настоящее исследование относится к стилометрии (и, шире, к количественной лингвистике). Количественный метод изучения авторского стиля литературных текстов, основанный на анализе статистики встречающихся в них числительных, применен к литературным текстам на немецком языке. Разработана компьютерная программа для поиска в немецкоязычном тексте количественных и порядковых числительных, выраженных как цифрами, так и словесно в разных словоформах. Программа автоматически убирает из текста фразеологизмы и устойчивые сочетания, случайно (без авторского замысла) содержащие числительные. Предварительно текст вручную очищается от служебных числительных, таких, как пагинация, номера глав, перечисления и т. п. Показано, что числительные, используемые автором в (художественном) тексте, для каждого автора индивидуальны;

их совокупность является характерным признаком (авторским инвариантом), различающим тексты разных авторов. Анализ авторского использования числительных в литературном тексте почти не зависит от перевода текста на другой язык. Структура языка может оказывать лишь небольшое влияние на статистику числительных. Это позволяет при недоступности оригинального текста на данном языке воспользоваться его доступным переводом, а также количественно сопоставлять тексты авторов, творивших на нескольких языках. Выполнен сопоставительный стилометрический анализ литературных текстов Т. Манна, Г. Броха, Р. Музиля, Э. Канетти – представителей немецкоязычного литературного модернизма XX века. Обнаружены существенные различия в использовании авторами числительных. Результаты анализа подвергнуты иерархической кластеризации. Используются манхэттенская метрика, метод дальнего соседа (Complete linkage method) и метод межгрупповых связей (Group average method). Кластерный анализ правильно распределил тексты в соответствии с авторством. Таким образом, новый метод стилометрии способен успешно атрибутировать литературные тексты.

Ключевые слова:

стилометрия, стилеметрия, количественная лингвистика, атрибуция текстов, авторство текстов, числительные в тексте, Т. Манн, Г. Брех, Р. Музил, Э. Канетти

Исследование выполнено за счет средств гранта Российского научного фонда № 23-28-00750, <https://rscf.ru/project/23-28-00750/>, проект «Разработка нового метода стилометрии на основе статистики использования числительных в авторских текстах».

The research was supported by the grant No. 23-28-00750 from the Russian Science Foundation; see <https://rscf.ru/en/project/23-28-00750/>.

Introduction

The present study has two main objectives: firstly, to provide new examples to support our approach to the problems of stylometry [1–8], and secondly, based on this approach, to conduct a quantitative analysis of the works by T. Mann, H. Broch, R. Musil, and E. Canetti – classics of 20th-century German-language modernist literature.

Stylometry (and, more broadly, quantitative linguistics) still does not have a completely satisfactory universal working method [9, 10]: some studies consider the frequencies of occurrence of content and function words (prepositions, conjunctions), average word and sentence lengths; in a pair of analyzed texts, one compares the most frequently used words common for both texts (the well-known “Burrows delta” [11]) and even letter combinations (oddly enough, the latter approach often demonstrates good results). Unfortunately, different methods often lead to controversial conclusions, so it is more reliable to use several methods together.

Promising results have been obtained using neural networks, and it seems that soon, artificial intelligence will be able to successfully solve problems in quantitative linguistics [12]. Nevertheless, meaningful interpretation of results within this approach remains problematic, since the method itself is a *black box*.

The study of apocrypha (starting with the biblical [13] and Shakespearean [14]), cases of dubious authorship (M. Ageyev [2, 15], B. Traven [16]) and fictitious authorship (Émile Ajar

[17]), forged memoirs (Misha Defonseca [18]) are examples of tasks in which stylometric methods can be useful.

We have developed an original stylometric method for analyzing authorial texts based on the authors' use of numerals in their texts [1–8]. Among the content words, numerals by their nature are the easiest to quantify. With regard to literary texts, the content of which is not rigidly tied to real-life events, but generated by free imagination, it is natural to assume that the use of numerals is associated with the author's psychological features which imperceptibly for him influence the result of his creative work. Consequently, the manner of numerals use is an author-specific feature (or fingerprint), which allows, under certain circumstances, to solve the problem of text authorship attribution.

Note that, unlike all the methods listed above, it is the analysis of the use of numerals that is almost independent of the translation of the text into another language (the structure of the language may have a *slight* effect on the statistics of numerals: in the English phrase *tenth anniversary* a numeral will be found, while in its German equivalent *zehnjähriges Jubiläum* it will not). This makes it possible, when the original text in a given language is unavailable, to use its accessible translation, as well as to quantitatively compare the texts of authors who wrote in several languages (A. Strindberg, S. Beckett, V.V. Nabokov, ...).

The study of the works of several dozens of authors in Russian, Czech, and English revealed tangible authorial features in the use of numerals in texts, the influence of genre, style, and artistic direction on them [1–8]. Thus, the results of the analysis allow for a meaningful philological interpretation.

By now we have developed a computer program that identifies numerals in German-language texts, and in this work the objects of study will be German-language literary texts for the first time. We will analyze some works by Thomas Mann (1875–1955), Hermann Broch (1886–1951), Robert Musil (1880–1942), and Elias Canetti (1905–1994) from the point of view of the use of numerals.

T. Mann is recognized as one of the most prominent representatives of German literary modernism (with all the vagueness of this concept) [19–24]. In Austria, such a description could be given to Musil [24–33] (less well-known to the general public and less prolific as a writer, but comparable to Mann in the artistic merits of his works) and Broch [24, 33–37] (author of prose, poetry, philosophical and political essays). Their younger contemporary Canetti, who is considered more of a postmodernist, was distinguished by the versatility of his work: from a novel, plays, and artistic autobiography to an extensive compilation treatise "Crowds and Power" claiming to be scientific [38–43].

Hardly anything can be added to the existing literary and critical examination of the works of these writers. In this paper, we will apply a formal quantitative approach to their texts, which, to our knowledge, has not been done before.

Method and objects of research

Our computer program scans the German-language text for cardinal and ordinal numerals, expressed both in numbers and in words in various forms. The program automatically eliminates idiomatic expressions (*im siebenten Himmel*) and fixed phrases (*die fünfte Kolonne*), which *accidentally* contain numerals.

The numerals not related to the authors' creative ideas were manually deleted from the text beforehand – such as page and chapter numbering, itemizations 1), 2), 3), etc.

The following texts were analyzed:

T. Mann:

- *Königliche Hoheit* («Royal Highness»), 1909 – novel;
- *Bekenntnisse des Hochstaplers Felix Krull* («Confessions of Felix Krull»), 1922–54 – novel;
- *Der Zauberberg* («The Magic Mountain»), 1924 – novel;
- *Lotte in Weimar*, 1939 – novel;
- *Doktor Faustus* («Doctor Faustus»), 1947 – novel;
- *Erzählungen* – a collection of short stories [\[44\]](#) including *Herr und Hund*, *Der Knabe Henoch* (Fragment), *Die vertauschten Köpfe*, *Die Betrogene*, *Fiorenza*, *Gesang vom Kindchen*.

H. Broch:

- *Die Schlafwandler* («The Sleepwalkers»), 1932 – novel;
- *Die Entsöhnung* («The atonement»), 1933 – a play;
- *Die Verzauberung* («The Spell»), published 1976 – novel;
- *Gedichte* – The complete collection of poems [\[45\]](#).

R. Musil:

- *Die Verwirrungen des Zöglings Törless* («The Confusions of Young Törless»), 1906 – novel;
- *Der Mann ohne Eigenschaften* («The Man without Qualities»), 1932 – novel.

E. Canetti:

- *Masse und Macht* («Crowds and Power»), 1962 – non-fiction prose;
- *Die gerettete Zunge* («The Tongue Set Free»), 1977 – a fictionalized autobiography.

Some numerical characteristics of the texts are presented in Table 1.

The choice of texts for analysis was influenced by their availability for free download on the Internet. Unfortunately, some important works were not available, even though the copyright protection period for most of them had long expired.

Results

For a primary assessment of the similarity/differences in the authors' use of numerals, we calculated the inverse density of numerals for each text (the result of dividing the volume of the text by the number of numerals contained in it). The lower the inverse density, the more often numerals appear in the text.

Noteworthy is the significantly lower value of the inverse density for Musil's texts (which turned out to be the same to within tenths of a fraction in both analyzed works!) compared to the texts of other authors: Musil *more often* resorts to numerals.

As for Canetti's texts, with their very different densities of numerals, this result provides a preliminary answer to the question discussed in literary criticism: whether *Crowds and*

Power should be attributed to fiction or to texts modeled on scientific ones. It is indeed a text that claims to be scientific. We will return to this issue later.

The texts by Mann and Broch differ slightly in the inverse density of numerals. Poems (by Broch), quite expectedly, have the highest inverse density of numerals: in poetry, numerals are less common than in prose.

After examining the general use of numerals in the texts, we proceeded to a separate account of each numeral. The differences in the authors' use of numerals become apparent when we apply hierarchical cluster analysis [46, 47], which groups objects (here: texts) into clusters based on their similarity – in our case, the similarity of the absolute frequencies of occurrence of the numerals 1, 2, 3, 4, 5 in the texts (these numerals are present without exception in all analyzed texts, subsequent numerals are found with gaps). Since the texts differ significantly in volume (see Table 1), correction factors had to be introduced to make the frequencies comparable. Mann's *Der Zauberberg* served as the reference text for comparison. Therefore, for example, for *Königliche Hoheit* the frequencies were multiplied by $2,075,077 / 751,961 = 2.76$, and for *Der Mann ohne Eigenschaften* by Musil – by $2,075,077 / 4,437,225 = 0.47$.

The measure of similarity in cluster analysis is the metric ρ ("distance"): the smaller the "distance" between objects, the more similar they are. We applied the Manhattan metric

$$\rho(\mathbf{x}, \mathbf{y}) = \sum_i^n |x_i - y_i|, \quad (1)$$

where $\mathbf{x}$ and $\mathbf{y}$ are n -dimensional vectors, the components of which are the corrected absolute frequencies of the first n natural numbers occurring in the two analyzed texts (here $n = 5$).

In the clustering process, we used the far neighbor method, also known as the Complete linkage method [48], which leads to the formation of compact isolated clusters.

In the initial phase, we grouped together only literary works by Mann, Broch, and Musil. They were reasonably distributed into clusters according to authorship (Figure 1).

Conclusions:

- 1) The uniqueness of Musil's writings is confirmed. But now it becomes clear which numeral is responsible for the high frequency of numerals. It is *ein* ("one") in different word forms; unfortunately, in German it is formally and semantically impossible to distinguish it from the indefinite article. Our program has taken into account all instances of *ein* appearing in the text.
- 2) In general, the texts by Mann and Broch do not differ greatly in the use of specific numbers.
- 3) Our approach to the problems of stylometry is based on the assumption that each writer has an individual manner of using numerals; this would seem to be contradicted by the alternation of Mann and Broch microclusters. However, firstly, there is no universal stylometric method that perfectly distributes texts according to authorship; secondly, these microclusters merge into an intermediate cluster at a high altitude (10 – in Figure 1) which is still 2.5 times less than the height of the formation of the final supercluster (with Musil's texts participating).

How stable is the dendrogram structure with respect to the addition of new texts from other

authors? We now add the texts of the fourth author, E. Canetti, and re-run the clustering (Fig. 2).

Based on Fig. 2, we can draw several observations:

- 1) The general appearance of the dendrogram has not changed much (the program has only reordered the low-level clusters).
- 2) The two texts by Canetti were clustered not just separately, but in branches of the dendrogram that merge at the maximum height. This indicates a fundamental distinction between the texts: while *Die gerettete Zunge* is a work that adheres to the conventions of fiction, *Masse und Macht* is not. The abundance of factual numbers places this text in a shared cluster with Musil's texts, albeit at a high height (an addition regarding Canetti's works will be provided below).
- 3) Adding Canetti's texts literally loosens the dendrogram: the heights of merging increase (note that the maximum height is always normalized to 25).

Additional information about the author's use of numerals can be derived from Figure 3, which shows a fragment of the frequency distribution of numerals from the range [1; 30] in some works of the authors under consideration:

- 1) The frequency of numerals decreases rapidly as the numerals increase.
- 2) Local maxima are observed at the round numbers 10, 20, 30, and so on. They can be explained by a well-known psychological phenomenon of preference for "round" numbers.
- 3) The differences between the texts by Mann and Broch become noticeable: Mann has a greater variety and frequency of numerals in texts (with the exception of the numeral *ein* (one), which, however, can also be an article, as noted above).

In our opinion, the most significant indicator of the author's desire for subjective "accuracy" of the narrative is the inclusion of specific dates in the text. According to this indicator, Canetti's works are the leaders among all the analyzed texts. Although they are little similar in the use of numerals in general, they are very close in the abundance of dates found in them.

The choice of metric and clustering method cannot be definitively justified, yet they can have a substantial impact on the outcomes of clustering. We performed clustering of texts by the same authors as in Figure 1, but using not the far neighbor method, as in the previous attempt. Instead, we utilized the group average method (Between-groups linkage) [47], still using the Manhattan metric (Figure 4). In our case, the results were quite consistent, and all the conclusions remained valid. Even when we used different combinations of metrics and clustering techniques, the dendrogram only changed slightly.

Table 1

Occurrence of numerals in the studied texts

No.	Author, text	Size (bytes, UTF encoding)	Number of numerals	Inverse density of numerals
1	Mann, <i>Königliche Hoheit</i>	751 961	2865	262
2	Mann, <i>Bekenntnisse des</i>	866 978	3271	265

	<i>Hochstaplers Felix Krull</i>	2 075 077	7697	270
3	Mann, <i>Der Zauberberg</i>	2 075 077	7697	270
4	Mann, <i>Lotte in Weimar</i>	842 414	3073	274
5	Mann, <i>Doktor Faustus</i>	1 410 387	5554	254
6	Mann, <i>Erzählungen</i>	849 390	2893	294
7	Broch, <i>Die Schlafwandler</i>	1 622 049	6156	263
8	Broch, <i>Die Entsöhnung</i>	212 697	759	280
9	Broch, <i>Die Verzauberung</i>	771 187	2843	271
10	Broch, <i>Gedichte</i>	97520	316	309
11	Musil, <i>Die Verwirrungen des Zöglings Törless</i>	337 625	1617	209
12	Musil, <i>Der Mann ohne Eigenschaften</i>	4 437 225	21 182	209
13	Canetti, <i>Masse und Macht</i>	1 230 512	5532	222
14	Canetti, <i>Die gerettete Zunge</i>	760 747	2927	260

Conclusions

The approach to stylometry problems we are developing, based on the analysis of numerals statistics in texts, despite its simplicity, demonstrates high efficiency and sensitivity. Texts by T. Mann, H. Broch, R. Musil, and E. Canetti, which have so far been analyzed only through the traditional descriptive philological methods, were subjected to a formal stylometric analysis for the first time. The analysis correctly distributed the texts by authors and revealed some features of the literary style. Appreciable authorial differences in the manner of using numerals were discovered. The use of various clustering methods for text analysis enhances the significance of the results obtained and confirms their non-random nature. The method is suitable for text attribution.

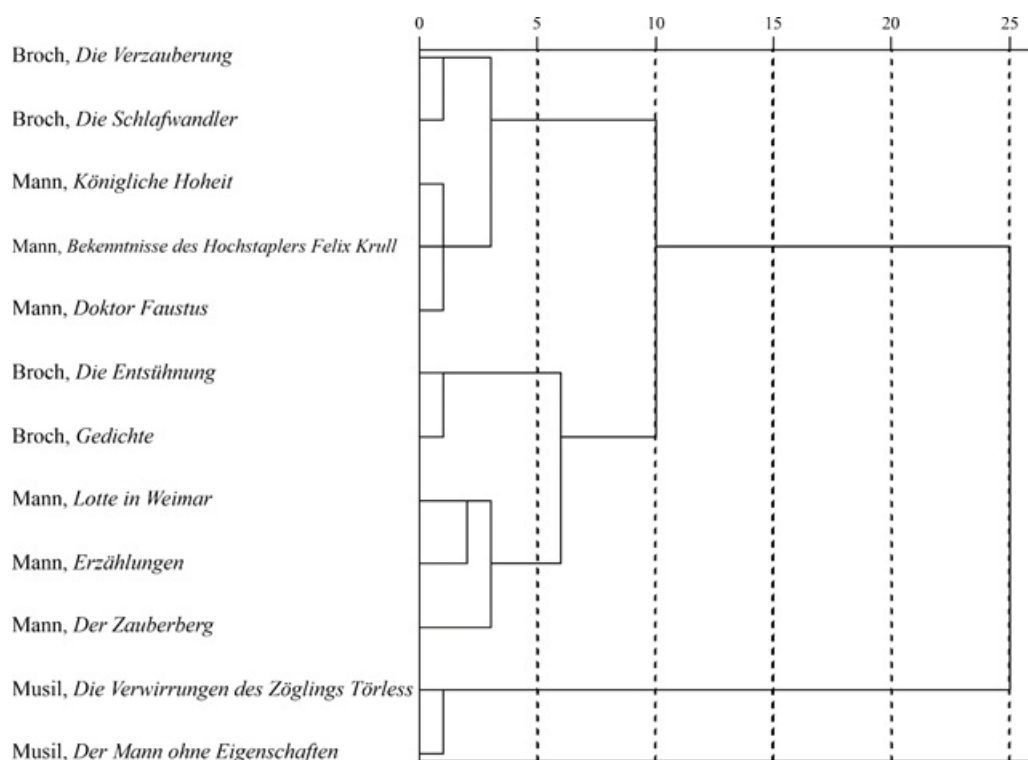


Figure 1 – The result of applying hierarchical cluster analysis to the texts by T. Mann, H. Broch, and R. Musil (clustering uses the far neighbor method and the Manhattan metric).

The horizontal axis indicates the "distance" in arbitrary units

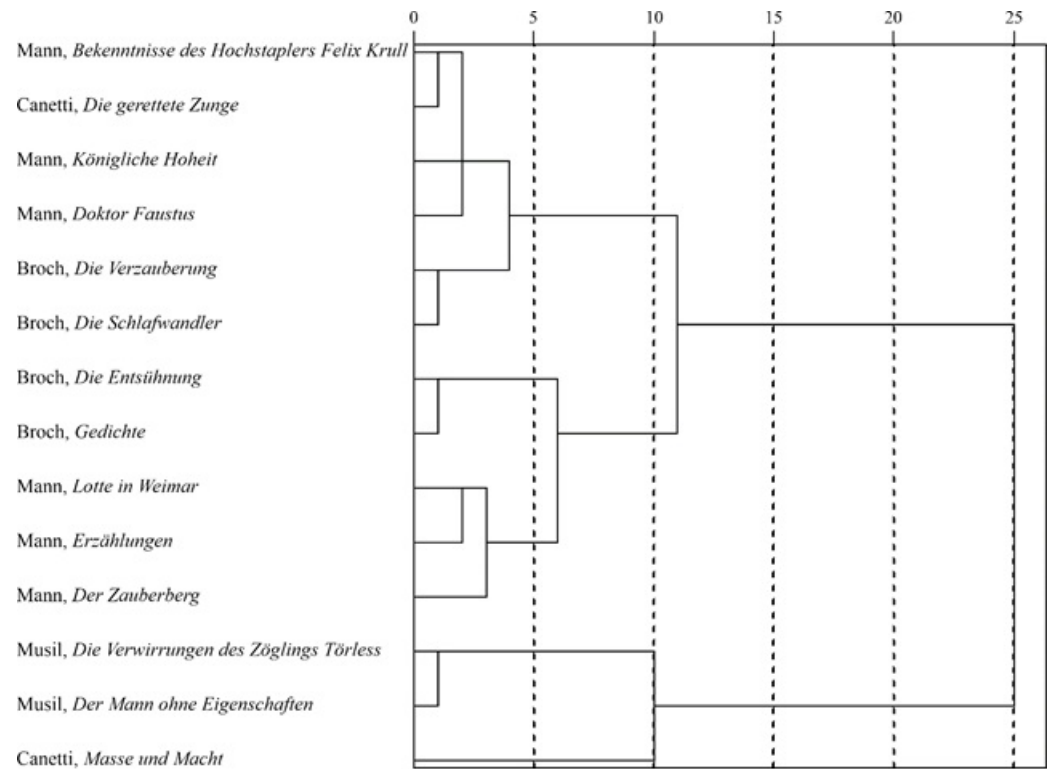


Figure 2 – The result of applying hierarchical cluster analysis to the texts by T. Mann, H. Broch, R. Musil, and E. Canetti (clustering uses the far neighbor method and the Manhattan metric – the same as in Figure 1). The horizontal axis indicates the "distance" in arbitrary units

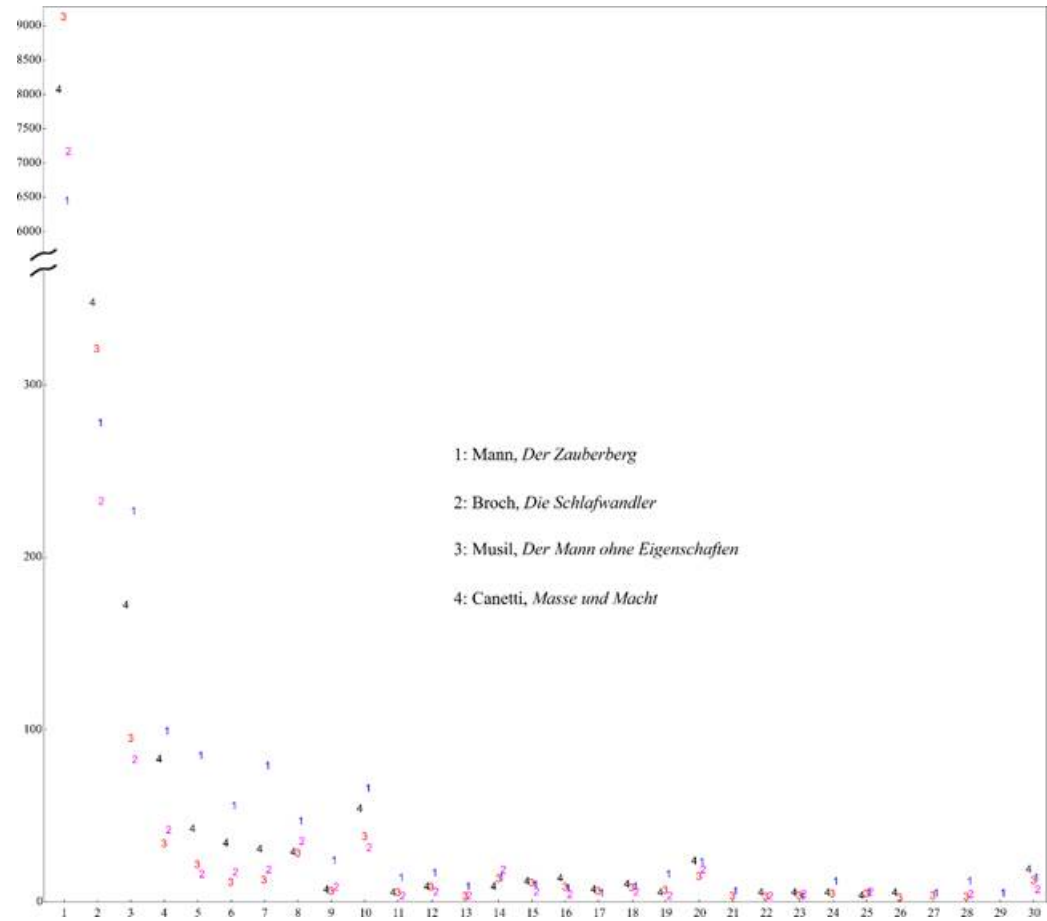


Figure 3 – A fragment of the frequency distribution of numerals from the range [1; 30] in some works by T. Mann, H. Broch, R. Musil, and E. Canetti. The vertical axis shows the frequency of numerals after introducing correction factors to account for the different sizes of texts. The axis is broken to save space

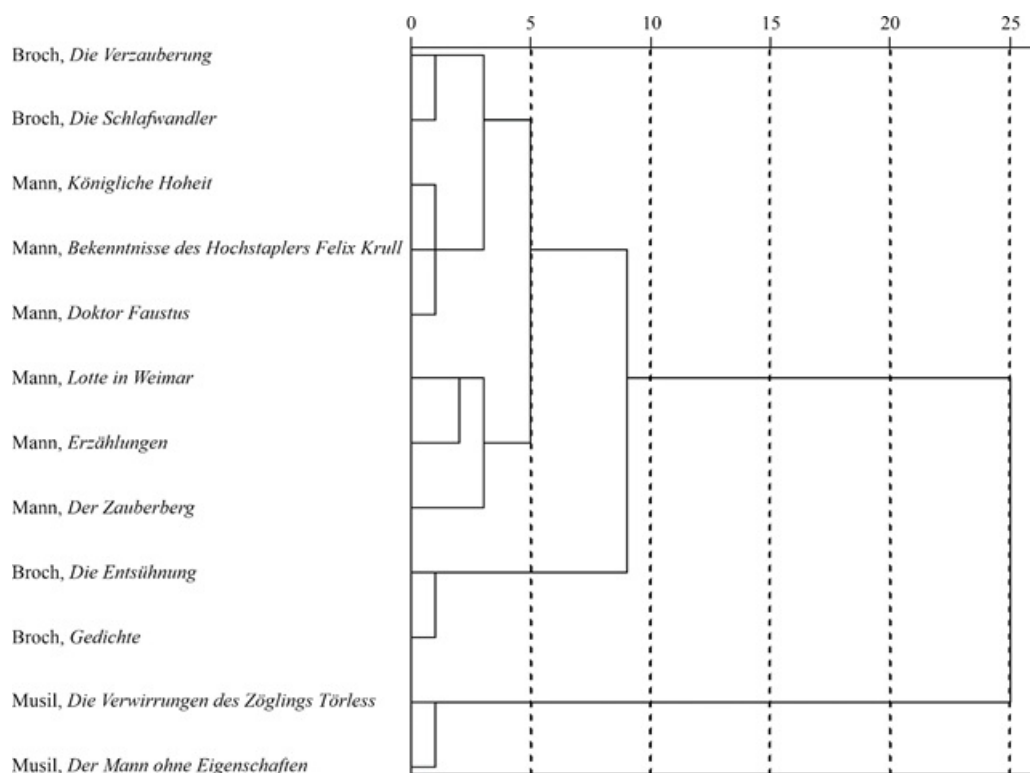


Figure 4 – The result of applying hierarchical cluster analysis to the texts by T. Mann, H. Broch, and R. Musil (unlike Figure 1, the clustering uses the group average method, but still the Manhattan metric). The horizontal axis indicates the "distance" in arbitrary units

Библиография

1. Зенков А. В. Новый метод стилеметрии на основе статистики числительных, Компьютерные исследования и моделирование, 2017, Т. 9, № 5. С. 837–850.
2. Zenkov A.V. A Method of Text Attribution Based on the Statistics of Numerals // J. of Quantitative Linguistics. 2018. No. 25(3). Pp. 256–270.
3. Zenkov A.V., Místecký M. The Romantic Clash: Influence of Karel Sabina over Mácha's Cikáni from the Perspective of the Numerals Usage Statistics // Glottometrics. 2019, No. 46, Pp. 12–28.
4. Zenkov A.V. Stylometry and Numerals Usage: Benford's Law and Beyond // Stats 2021. No. 4. Pp. 1051–1068.
5. Zenkov A., Místecký M. Young Vladimír Vašek? – A Numerals Analysis Contribution to the Bezruč–Hrzánský Identity Issue // Naše řeč, 2022. No. 105(3). Pp. 151–161.
6. Зенков А.В. Литературные мистификации и авторское использование числительных // Филологические науки. Вопросы теории и практики. 2023. № 16(11). С. 3696–3709. URL: <https://doi.org/10.30853/phil20230568>.
7. Zenkov A.V. Under a False Flag: Literary Hoaxes and the Use of Numerals // Litera. 2023. № 10. С. 86–109. DOI: 10.25136/2409-8698.2023.10.68743 EDN: TYDRFD URL: https://e-notabene.ru/fil/article_68743.html
8. Зенков А.В., Ермаков Н.Е. Числительные в текстах как характерная особенность авторского стиля // Russian Linguistic Bulletin. 2023. № 45(9). URL: <https://doi.org/10.18454/RULB.2023.45.28>.
9. Stamatatos E. A survey of modern authorship attribution methods // J. Amer. Soc. for

Information Science and Technology. 2009. No. 60(3). Pp. 538–556.

10. Tempestt N., Kalaivani S., Aneez F., Yiming Y., Yingfei X., and Damon W. Surveying Stylometry Techniques and Applications // ACM Comput. Surv. 2017, No. 50(6), Article 86, 36 pages.

11. Burrows J. Delta: a Measure of Stylistic Difference and a Guide to Likely Authorship / J. Burrows // Literary and Linguistic Computing. – 2002. – 17(3). – P. 267–287.

12. La Inteligencia Artificial ayuda a descubrir una obra desconocida de Lope de Vega en los fondos de la BNE, Biblioteca Nacional de España, <https://www.bne.es/es/noticias/inteligencia-artificial-ayuda-descubrir-obra-desconocida-lope-vega-fondos-bne> (Accessed: October 25, 2024).

13. Schröter, J. (2020). Die apokryphen Evangelien: Jesusüberlieferungen außerhalb der Bibel. Munich: C. H. Beck.

14. Vickers, B. (2002). 'Counterfeiting' Shakespeare: Evidence, Authorship and John Ford's Funeral Elegy. Cambridge: Cambridge University Press.

15. Сорокина М. Ю., Суперфин Г. Г. «Был такой писатель Агеев...»: версия судьбы или о пользе наивного биографизма // Минувшее: Исторический альманах. Вып. 16. М., СПб: Феникс-Атенеум, 1994. С. 265–289.

16. Dammann, G. (ed.) (2012). B. Traven, Autor – Werk – Werkgeschichte. Würzburg: Königshausen & Neumann.

17. Bellos, D. (2010). Romain Gary: A Tall Story. London: Harvill Secker.

18. Hupertz, H. (2021). Wie eine Frau sich als Holocaust-Überlebende ausgab. Frankfurter Allgemeine, 23 November. Available at <https://www.faz.net/aktuell/feuilleton/medien/fraugab-sich-als-holocaust-ueberlebende-aus-dokumentation-bei-arte-17646920.html> (Accessed: October 25, 2024).

19. Arnold, H. L. Thomas Mann. München: Edition Text u. Kritik, 1976. ISBN: 9783921402221. 226 S.

20. M. Travers, Thomas Mann. London: Macmillan Education, 1992. Pp. vii + 146. ISBN : 978-0333517079.

21. Thomas Mann-Handbuch: Leben – Werk – Wirkung, A. Blödmann, F. Marx (Eds.), DOI: <https://doi.org/10.1007/978-3-476-05341-1>, Verlag J.B. Metzler Stuttgart, Springer-Verlag Berlin Heidelberg 2015. ISBN 978-3-476-02456-5. IX + 425 pages.

22. Thomas Mann: neue kulturwissenschaftliche Lektüren. S. Börnchen, G. Mein, G. Schmidt (Eds.), Wilhelm Fink Verlag, 2012. ISBN 9783846753897. 457 Seiten.

23. C. Grawe, Sprache im Prosawerk. Beispiele von Goethe, Fontane, Thomas Mann, Bergengruen, Kleist und Johnson. Bonn: Bouvier Verlag Herbert Grundmann, 1987. ISBN: 9783416009584. 111 Seiten.

24. Dowden, S. D.: Sympathy for the abyss: a study in the novel of German modernism: Kafka, Broch, Musil, and Thomas Mann. Tübingen: Niemeyer, 1986. ISBN 3-484-18090-0. 195 p.

25. Nübel, B. Robert Musil – Essayismus als Selbstreflexion der Moderne, Berlin, New York: De Gruyter, 2006. URL: <https://doi.org/10.1515/9783110201857>. 548 S.

26. Nübel, B. and Wolf, N. Ch. Robert-Musil-Handbuch, Berlin, Boston: De Gruyter, 2016. URL: <https://doi.org/10.1515/9783110255577>. 1064 S.

27. Boelderl, A. R. and Neymeyr, B. Robert Musil im Spannungsfeld zwischen Psychologie und Phänomenologie, Berlin, Boston: De Gruyter, 2024. URL: <https://doi.org/10.1515/9783110988352>. 366 S.

28. H. Bloom, Robert Musil's the Man Without Qualities. Chelsea House Publishers, 2005. ISBN 9780791081228. 211 pages.

29. J. Bouveresse, La Voix de l'âme et les Chemins de l'esprit. Dix études sur Robert Musil. Éditions du Seuil, 2001, ISBN: 9782020362894. 462 p.

30. A Companion to the Works of Robert Musil, P. Payne, G. Bartram, and G. Tihanov (Eds.). Camden House, Rochester, New York. 2007. ISBN: 978-1-57113-110-2. 472 p.
31. P. Payne. Robert Musil's 'The Man Without Qualities': A Critical Study. Cambridge University Press, 1988. ISBN: 978-0-521-11060-0. 271 p.
32. Th. Sebastian, The Intersection of Science and Literature in Musil's The Man Without Qualities. Camden House, an imprint of Boydell & Brewer Inc., Rochester, 2005. ISBN: 1-57113-116-7. 159 p.
33. F. Schwarzwälder, Der Weltanschauungsroman 2. Ordnung: Probleme literarischer Modellbildung bei Hermann Broch und Robert Musil. transcript Verlag, Bielefeld, 2019, 372 Seiten. ISBN: 978-3-8376-4996-3.
34. A Companion to the Works of Hermann Broch, G. Bartram, S. McGaughey and G. Tihanov (Eds.), 2019. Camden House, an imprint of Boydell & Brewer Inc., Rochester, ISBN: 9781571135414, 290 p.
35. Hermann-Broch-Handbuch: Zeit – Werk – Forschung, M. Kessler, P. M. Lützeler (Eds.), De Gruyter, 2015. ISBN: 978-3110200713. 685 S.
36. Wohlleben, D. and Lützeler, P. M. (Eds.). Hermann Broch und die Romantik, Berlin, Boston: De Gruyter, 2014. <https://doi.org/10.1515/9783110351958>. 235 S.
37. Hermann Broch, Visionary in Exile, The 2001 Yale Symposium, P. M. Lützeler, M. Konzett and W. Riemer (Eds.). Camden House, an imprint of Boydell & Brewer Inc., Rochester. ISBN: 9781571132727. 280 p.
38. W. C. Donahue, The End of Modernism: Elias Canetti's Auto-da-Fé. The University of North Carolina Press, 2001. ISBN: 978-1-4696-5742-4. 302 p.
39. J. P. Arnason and D. Roberts, Elias Canetti's Counter-Image of Society: Crowds, Power, Transformation. Camden House, an imprint of Boydell & Brewer Inc., Rochester. 2004. ISBN: 9781571131607. 174 p.
40. A Companion to the Works of Elias Canetti, D. C. G. Lorenz (Ed.). Camden House, an imprint of Boydell & Brewer Inc., Rochester. 2004. ISBN: 9781571134080. 364 p.
41. J S McClelland, The Crowd and the Mob: From Plato to Canetti. Unwin Hyman Ltd, 2011. ISBN 9780415602495. 356 Pages.
42. B. Neumann, G. Wimmer, Elias Canetti in seiner Zeit: Kulturelle, wissenschaftliche und politische Deskriptionen. J.B. Metzler, ein Imprint des Springer-Verlages, 2020. ISBN 978-3-476-05649-8. 264 S.
43. Radaelli, G. Literarische Mehrsprachigkeit: Sprachwechsel bei Elias Canetti und Ingeborg Bachmann, Berlin: Akademie Verlag, 2011. <https://doi.org/10.1524/9783050053592>. 304 S.
44. Th. Mann, Die Erzählungen, Zweiter Band. Fischer Taschenbuch Verlag GmbH, Frankfurt am Main. 1979.
45. H. Broch, Gedichte. Kommentierte Werkausgabe, Band 8. Suhrkamp, Frankfurt, 1980. ISBN: 978-3518370728. 244 S.
46. Moisl H. Cluster Analysis for Corpus Linguistics. De Gruyter Mouton, 2015. – 381 p. ISBN:9783110350258.
47. Gan G., Ma C., Wu J., Data Clustering: Theory, Algorithms, and Applications. Society for Industrial and Applied Mathematics, 2007. – 466 p. DOI: 10.1137/1.9780898718348.

Результаты процедуры рецензирования статьи

В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.

Со списком рецензентов издательства можно ознакомиться [здесь](#).

Рецензируемая статья посвящена стилометрическому сопоставлению немецкоязычных модернистских текстов. Актуальность предмета исследования не вызывает сомнения и

обусловлена необходимостью осмысления использования стилометрического метода анализа авторских текстов. Как отмечается в работе, стилометрия до сих пор не имеет подходящего универсального рабочего метода: «в некоторых исследованиях учитывается частота встречаемости в текстах знаменательных частей речи и служебных слов (предлоги, союзы), средние длины слов и предложений; в паре анализируемых текстов сравниваются самые часто встречающиеся слова и даже буквосочетания. Однако разные методы часто приводят к противоречивым выводам, поэтому надёжнее использовать несколько методов одновременно».

Теоретической основой исследования выступили многочисленные труды российских и зарубежных ученых на русском, чешском и английском языках, изучение которых показало, что «манера использования числительных относится к авторской особенности, позволяющей решать проблему авторства текста, изучать жанровые и стилистические особенности, а, следовательно, результаты анализа допускают содержательную филологическую интерпретацию».

Методологическую базу настоящего исследования составил авторский стилометрический метод анализа литературных текстов, основанный на учёте использования числительных, так как «среди знаменательных частей речи именно числительные по своей природе наиболее легко поддаются количественному учёту». Авторами была разработана компьютерная программа, распознающая числительные в немецкоязычных текстах, с помощью которой в данной работе анализируются оригинальные художественные тексты Томаса Манна, Германа Броха, Роберта Музиля и Элиаса Канетти с точки зрения использования данной знаменательной части речи. Полученные результаты представлены в таблице «Частотность числительных в исследуемых текстах», а также на 4 рисунках, отражающих результат применения иерархического кластерного анализа к немецкоязычным текстам Т. Манна, Х. Броха, Р. Музиля и Э. Канетти.

В ходе проведенного исследования было выявлено, что авторский подход к решению задач стилометрии, основанный на анализе статистики числительных в текстах, несмотря на свою простоту, показал высокую эффективность и чувствительность. Тексты Т. Манна, Х. Броха, Р. Музиля и Э. Канетти, которые до сих пор анализировались только традиционными лингвистическими методами, впервые были подвергнуты формальному стилометрическому анализу, который позволил установить некоторые особенности стиля авторов и характерные отличия в употреблении ими числительных. Следовательно, представленный метод подходит для атрибуции текста.

Теоретическая значимость исследования связана с его вкладом в развитие количественной лингвистики: «предложенный новый стилометрический метод анализа авторских текстов, основанный на анализе частотности числительных в (авторских) литературных текстах, способен успешно решать задачи стилометрии, в том числе связанные с атрибуцией текстов». Практическая значимость работы заключается в том, что данный стилометрический метод может применяться наряду с другими методами в исследованиях литературных текстов и при идентификации автора текста (в случае необходимости).

Библиография статьи насчитывает 47 источников, в том числе на иностранных языках, что представляется достаточным для обобщения и анализа теоретического аспекта изучаемой проблематики. Рекомендуем автору(ам) обратить внимание на оформление библиографического списка (см. требования редакции).

Стиль статьи отвечает требованиям научного описания, содержание представленной работы соответствует названию, логика исследования четкая. Статья имеет завершённый вид; она вполне самостоятельна, оригинальна, будет интересна и полезна широкому кругу лиц и может быть рекомендована к публикации в научном журнале «Филология: научные исследования».

