

Robert Musil, a war journal, and stylometry: Tackling the issue of short texts in authorship attribution

Simone Rebori 

Foreign Languages and Literatures, University of Verona, Italy

J. Berenike Herrmann and Gerhard Lauer

University of Basel, Switzerland

Massimo Salgaro

Foreign Languages and Literatures, University of Verona, Italy

Abstract

During World War I (WWI), between 1916 and 1917, Robert Musil was the chief editor of the *Tiroler Soldaten-Zeitung* in Bozen. This activity probably also involved authorship of articles and has posed a philological problem to scholars, who have not been able to attribute with certainty a range of relatively short texts to Musil. With this article, we present a new approach that combines philological research with stylometric methods. Exploration of WWI archives and digitization of historical documents were paired with application of authorship attribution techniques, following extensive evaluation. To build the training set, we adapted the ‘impostors method’ by grouping three ‘distractor authors’ (similar to Musil in terms of style) and three actual candidates for authorship. In the test set, we developed two designs for tackling the issue of text length: a combinatory design, where longer chunks were composed by the juxtaposition of short texts; a simplified design, where the texts for attribution were merged with already attributed texts. Results of our experiment suggest that Musil attribution may be disproved with a high level of confidence for ten texts that were more probably written by a less well-known author, Albert Ritter. We carried out a keyness analysis on the specific words preferred or avoided by the two authors, which not only corroborated the results of the quantitative analysis but also findings from Musil philology. Our study showcases the potentialities of using mixed methods in stylometry.

Correspondence:

Simone Rebori, Foreign Languages and Literatures, University of Verona, Verona, Italy.

E-mail:

simone.rebori@univr.it

1 The Case Study: Robert Musil and the *Tiroler Soldaten-Zeitung*

The evolution of stylometry—the computational study of writing style—is tightly connected to the issue of authorship attribution. Pioneering studies like those by Mosteller and Wallace (1964), Forsyth

et al. (1999), and Craig and Kinney (2009), have successfully applied stylometric methods to determine authorship of the *Federalist Papers*, Cicero’s spurious *Consolatio*, and Shakespeare’s disputed plays. Although extensive research has recently been dedicated to the improvement of stylometric methods (Jockers and Witten, 2010; Evert *et al.*,

2015), empirical applications to attributive problems in literary history are still quite scarce (a recent one being Tuzzi and Cortelazzo's (2018) work on Elena Ferrante). With our study, we aim to contribute to the integration of computational methods and literary criticism for resolving empirical questions of authorship.

Robert Musil, one of the most important authors of twentieth-century literature written in German, fought in the Austrian army at the Italian front. During the First World War, between 1916 and 1917, he was the chief editor ('Schriftleiter') of the propagandistic journal *Tiroler Soldaten-Zeitung* (TSZ) in Bozen. While his role as editor is undisputed, it is an open question whether Musil also authored articles, and if so, how many. The first issue of the TSZ was published in June 1915. In 1916, the publishing was entrusted to the Bozen-based *Heeresgruppenkommando Erzherzog Eugen*, to which lieutenant Musil was assigned during the same year. At the beginning of October, Musil became the chief editor of the newspaper. After his arrival, major changes in the magazine's layout were introduced, the issue numbering was restarted, and the name was changed to *Soldaten-Zeitung*. These changes indicate Musil's high ambitions as an editor. The articles being published were now much more elaborate (Urbaner, 2005), and the TSZ became soon a highly regarded journal. Due to the repositioning of the commands as well as technical problems, the magazine publication ended in April 1917. In November 1917, Robert Musil's work as the chief editor was honored by his commander in chief explicitly for his highly influential propaganda work (Dinklage, 1960, pp. 230–31).

But how can one make sense of Musil's success as the chief editor—and possibly writer—of the TSZ? After all, Musil's involvement in the propagandistic war writings seems in clear opposition to his much more tentative, essayistic, and open-minded work in general. This issue is closely connected with a philological problem. As the TSZ-texts appeared anonymously, authorship attribution has been a challenge, which was further increased by the problem of providing a clear account of Musil's writing style in the first place. For example, Arntzen (1980,

p. 178) states that 'in comparison with Thomas Mann and Franz Kafka, [Musil] does not have a work-immanent personal style, but a functional style'. Regina Schaunig, the author of the only monograph on Musil's activity in the TSZ, defines his editorial style as 'stylistically neutral' (Schaunig, 2014, p. 100), although she also stated that the TSZ articles become more subtle and psychological after Musil being in charge of the journal (Schaunig, 2009/2010). Others, such as Gschwandtner (2013/2014, 2015/2016, 2016), conceive of the TSZ-texts as products of an 'anonymous' or 'collective' authorship, being written under the pressure of war propaganda and censorship and not by a single autonomous author. A similar argument is raised by Krause (2012), when he talks of Musil's 'Doppelleben' (double life) as a propaganda journalist and as an artist—much the same was already argued by Giovannini (1987/1988). In sum, Musil scholarship concludes that no obvious stylistic similarities between both roles are perceivable.

Despite this fact, different samples of TSZ articles (from a minimum of three to a maximum of thirty-eight) have been attributed to Musil over the past 60 years (for a detailed list, see Schaunig, 2014, pp. 356–57). The surprising aspect of these attributions is the lack of factual or formal evidence. For example, Roth (1972, p. 528) lists nineteen texts from TSZ introducing them with the quite cryptic phrase: 'anonymous texts which have not yet been identified with certainty'. Subsequent studies, such as the one by Arntzen (1980), build on Roth's assertions and propose new attributions, again without clear evidence-based corroboration. The Italian edition of TSZ articles attributed to Musil (Fontanari and Libardi, 1987) simply includes any texts previously indicated as possibly authored by Musil. Schaunig (2009/2010) and Gschwandtner (2013/2014), in line with Corino (2003, p. 1759), attribute the editorials to Musil on two grounds: his role as the chief editor and the articles' content and style. Language and (anti-irredentistic) politics, book reviews, and feuilletonistic topics were among Musil's major interests, and the articles are simply better written than many other war time writings. Even the prestigious digital edition of Musil's works (*Klagenfurter Ausgabe*) is no more accurate,

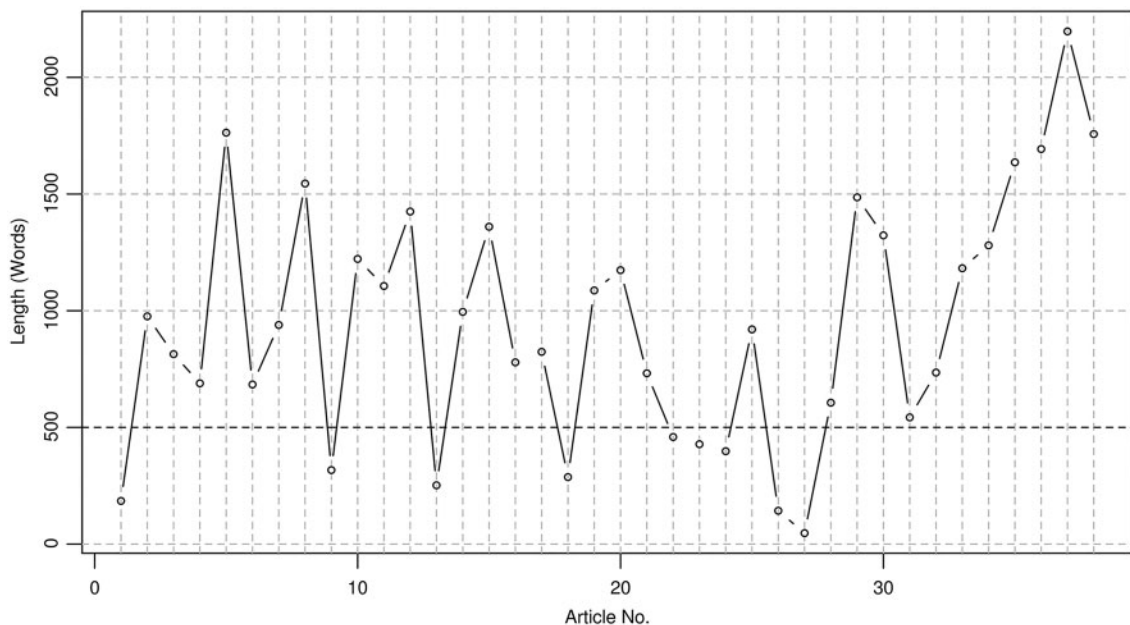


Fig. 1 Length of Musil's articles in the TSZ

resorting to defining the problem as one of a 'work in progress' (Amann *et al.*, 2009). It is thus quite clear that Musil scholarship provides good clues for further inquiry, but as each contribution focuses on different aspects of language use in context, they are approximations that do not permit formal grounds for authorship attribution. To provide an evidence-based approach for further enquiry, the present study evaluates how some of the most recent stylistic methods may offer a solution to the problem.

2 Methodological Issues: Not Only Text Length

The first issue to be tackled when dealing with TSZ articles is that of text length. In his first experiments with authorship attribution, Burrows (2007, p. 30) stated that 'around 10,000 word-tokens seem to suffice as a reliable minimum for an authorial set, 500 (but preferably many more) for an independent text'. The authorial set (now more generally known as 'training set') does not constitute a problem for our analysis, as we dispose of a much wider

corpus of Musil writings. More problematic is the situation with the 'independent texts' to be attributed (known as 'test set'), for which Burrows gives a very provisional limit. Luyckx and Daelemans (2011) were the first to analyze systematically the effects of text length on authorship attribution. However, the results of their nuanced experiment are not easily applicable to our case study. First, unlike ours, their focus was on the dimensions of the training set—because it has 'more influence on performance than the amount of test material' (Luyckx and Daelemans, 2011, p. 38). Second, their experiment featured a varying number of candidate authors—whereas our goal is simply to test the probability of Musil's authorship. More immediately adaptable to our needs is the study by Eder (2015), who tested different stylistic methods on multilingual corpora, proposing that the minimum dimension for an attributable text chunk is about 5,000 words, independently of language—while an ideal 'comfort zone' (meaning an attribution accuracy of about 80%) is around 8,000 words. Figure 1 shows that the average length of Musil's disputed articles in the TSZ is slightly below 1,000 words,

with a minimum of only forty-seven words. Quantitative authorship attribution of the individual texts thus poses a challenge, which we tackled by adopting a combinatorial design (see Section 3.3).

Another complication is brought by the actual availability of the digitized texts. In fact, Schaunig (2014) suggests that Musil may have authored many more than the thirty-eight texts disputed by the critics, providing a list of 165 possible titles. However, an analysis of the entire corpus of the TSZ is presently not possible, for a number of reasons. The journal has been digitized by the Austrian National Library,¹ but the quality of digitization is still too low. The plain text versions of the scanned pages comprise about 20% errors at the level of character recognition, which is an error margin sufficient to invalidate a stylometric analysis of short texts (Eder, 2013, p. 608). The problem derives from the fact that the TSZ was printed using the font Fraktur, which still poses a challenge for Optical Character Recognition software—for the most advanced approach, see (Breuel *et al.*, 2013). Given this limitation, we decided to restrict the test set to the selection of thirty-eight articles for which Musil's authorship has been proposed (for details, see Table 2). A good part of these texts is available in good quality, thanks to the *Klagenfurter Ausgabe*. The remaining part was typed by us. Next to the issues of text length and digitization, the chief challenge for our study lies in the nature of the intended stylometric task.

Following the definition by Koppel *et al.* (2008, p. 9), Musil's relation to the TSZ is a typical 'verification problem', where there is a no closed candidate list, but a suspect—with the challenge of determining whether that suspect is the author. An approach that is now generally adopted for coping with this issue is known as the 'impostors method'. It has already contributed to some of the most celebrated successes in authorship attribution—such as that of J. K. Rowling's *The Cuckoo's Calling* (Juola, 2015). As originally introduced by Koppel and Winter (2014, p. 178):

Suppose we are asked to determine if the documents *X* and *Y* were written by the same author. We systematically produce a set of 'impostor' documents and—in a matter reminiscent of a police

lineup—ask if *X* is sufficiently more similar to *Y* than to any of the generated impostors. The trick is using the proper methods to select the impostors and, more important, to measure document similarity.

While the idea is convincing in its simplicity, it needs to be carefully applied to prevent misinterpretation. Koppel *et al.* (2008, p. 19) note that the perspective taken on the test is crucial:

While it is indeed reasonable to conclude that *A* is not the author [of a text *X*] if most chunks are attributed to not-*A*, the converse is not true. Any author who is neither *A* nor represented in the sample not-*A*, but who happens to have a style more similar to *A* than to not-*A*, will be falsely determined by this method to be *A*.

In other words, the impostors method can be effective in disproving an attribution, but it is less reliable for confirming authorship. Another complication, however an interesting one, is that Musil—in his role as chief editor of the journal—may have corrected and adapted a sizeable number of the published articles, thus intermixing his style with those of others. By consequence, traces of his authorial signal may be found also in texts that were not originally written by him. Thus, both the setup of our case study and methodological considerations point out that it is disapproval of authorship, not its confirmation, which should be sought.

3 Experimental Design

3.1 Training set composition

In our study, we adapted the impostors method to include true impostors but also such authors that actually run a slight chance of being true candidates for authorship. Juola (2015, p. 106) notes that 'the exact number of distractor authors [i.e. impostors] is open, but three to seven seems a reasonable range'. Depending on the case study, then, the impostors should be 'matched for time period, language, region, genre, and gender' (Juola, 2015, p. 106). Following these instructions, we conducted a survey of early twentieth-century authors, which rendered a potential list of ten authors matching

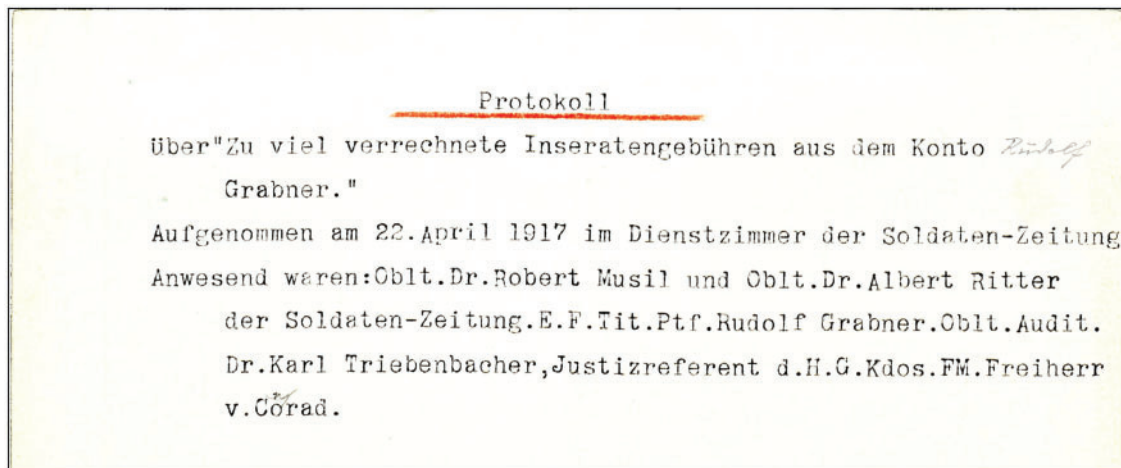


Fig. 2 Minutes of the Austrian Army, 22 April 1917 (*Kriegsarchiv*, Wien)

Musil's literary style. Taking into account copyright issues, three authors could be singled out as close matches that are also digitally available: Franz Kafka, Franz Blei, and Stefan Zweig. Alfred Döblin, who according to the pioneering quantitative study by Thöming (1970) would be an ideal impostor, could not be included for copyright reasons. Subsequently, we expanded the list by three authors introduced by Urbaner (2001) as collaborators of the TSZ: two (Hugo Salus and Marie delle Grazie) identified with certainty, and one (Albert Ritter) presented as a possible collaborator (see also Schaunig, 2014). The inclusion of Salus and delle Grazie added a slight inconsistency to the training set, particularly because only fictional writings of theirs were available, but the choice seemed acceptable for a preliminary testing of the approach. Taking up on the debate on Ritter's status at the TSZ, further research at the *Kriegsarchiv* in Wien demonstrated that Ritter was actually a member of the journal's editorial board (see Fig. 2).

Albert Ritter was a journalist and writer whose book entitled *Autonomie? Zur Frage der Neugestaltung Österreichs* was also advertised in the TSZ on 28 January 1917. During his career he made use of several pseudonyms, such as Karl v. Winterstetten. We digitized *Autonomie* together with another political pamphlet entitled *Grossdeutschland und die Weltpolitik*. To sum up,

our criteria for selection of impostors and candidates were (1) scholarly assertions of stylistic similarity to Musil (the impostors Blei, Kafka, and Zweig) and (2) membership in the journalistic team of the TSZ (the candidates Ritter, Salus, and delle Grazie).

The training set-texts authored by Musil were taken from the 'Essays' section of the *Klagenfurter* edition, selecting all articles published between 1911 and 1919. The training set was thus composed by seven authors: Musil, and the six candidates/impostors. To give it flexibility and to test its reliability, we structured the training set according to an 'instance-based' architecture (Stamatatos, 2009, p. 548), where each author is represented by more than one textual sample. Fitting in the 'comfort zone' described by Eder (2015), we therefore composed twenty-one text chunks with a length comprised between 6,000 and 8,000 words, three for each author (for details, see Table 1). To compensate for major fluctuations in German orthography at the beginning of twentieth century, all Umlauts and the Eszett graphemes were transformed into digraphs (*ä* -> *ae*; *ö* -> *oe*; *ü* -> *ue*; *ß* -> *ss*).

3.2 Validation and feature selection

Both validation and experimentation were implemented through the *Stylo* package,² developed by Eder et al. (2016) in the statistical programming

Table 1 Training set composition

Author	Original source	Genre	Digital version	Text chunks	Length
Franz Blei	Extracts from <i>Das große Bestiarium der modernen Literatur</i> (1922), <i>Formen der Liebe</i> (1930), and <i>Männer und Masken</i> (1930)	Essay/Fictional	Project Gutenberg-DE	<i>blei_1</i> <i>blei_2</i> <i>blei_3</i>	7,906 W. 6,822 W. 6,583 W.
Marie delle Grazie	Extracts from the novel <i>Das Buch der Liebe</i> (1916)	Novels	Project Gutenberg-DE	<i>dellegrazie_1</i> <i>dellegrazie_2</i> <i>dellegrazie_3</i>	7,426 W. 7,870 W. 7,201 W.
Franz Kafka	Letters written between 1909 and 1919	Letters	Project Gutenberg-DE	<i>kafka_1</i> <i>kafka_2</i> <i>kafka_3</i>	7,499 W. 7,799 W. 7,117 W.
Robert Musil	Journal articles published between 1911 and 1919	Essay/Journalistic	Klagenfurter Ausgabe	<i>musil_1</i> <i>musil_2</i> <i>musil_3</i>	6,548 W. 7,988 W. 6,291 W.
Albert Ritter	Extracts from the pamphlets <i>Autonomie?</i> (1916) and <i>Grossdeutschland und die Weltpolitik</i> (1919)	Essay/Political	Self-built	<i>ritter_1</i> <i>ritter_2</i> <i>ritter_3</i>	7,819 W. 7,047 W. 6,272 W.
Hugo Salus	The short stories <i>Pietà</i> , <i>Der Spiegel</i> , and <i>Das Meerweibchen</i> (1906)	Short stories	Project Gutenberg-DE	<i>salus_1</i> <i>salus_2</i> <i>salus_3</i>	7,129 W. 7,805 W. 7,643 W.
Stefan Zweig	Reviews published between 1902 and 1939	Reviews	Project Gutenberg-DE	<i>zweig_1</i> <i>zweig_2</i> <i>zweig_3</i>	6,798 W. 6,831 W. 6,818 W.

environment R. The structure of the training set allowed an extensive validation procedure, in which we adapted the canonical ‘ten-fold cross-validation’ (Weiss and Kulikowski, 1991) to develop a twenty-one-fold cross-validation: each text chunk was tested against the remaining twenty.³ From the areas of machine learning and distance measures, we selected as most canonical procedures Support Vector Machines, *k*-Nearest Neighbors (*k*-NNs), and Nearest Shrunken Centroids for machine learning (see Jockers and Witten, 2010), and Burrows’s Delta, Eder’s Delta, Canberra, Cosine, and Cosine Delta for distance measures (see Evert *et al.*, 2017). Results of the evaluation are shown in Fig. 3.

As the mean values at the bottom line show, distance measures work slightly better than machine learning algorithms on this specific corpus. Therefore, and since machine learning requires more computational power, we decided to only include the distance measures in the experiment. On the whole, percentages of correctly guessed attributions across methods and texts were high, with the exception of the text chunk entitled *blei_1* (from *Das große Bestiarium* by Franz Blei),

where both Cosine and *k*-NN failed. For this reason, to keep the *blei_1* text in the experimental design, we excluded the Cosine distance from our experiment. Our set of measures thus included Burrows’s Delta, Eder’s Delta, Cosine Delta, and Canberra—with a mean accuracy of 89% for *blei_1*.

Results were further corroborated by the stratified cross-validation protocols included in the *classify()* function of the Stylo package (see Eder *et al.*, 2016, p. 117): once again, Cosine distance and machine learning algorithms scored worse than the selected methods. Two parameter settings were also evaluated, namely, merging of training and test sets, and the analyzed features (most frequent words (MFWs) versus character *n*-grams). The practice of merging the two sets, while not possible with machine learning approaches, is widely adopted with distance measures and does not produce significant changes in accuracy (Eder, 2015, p. 182). We verified this by running three rounds of stratified cross-validation on our corpus. Training and test sets were prepared through random selection over three different subdivisions

90	83.5	93.5	0	89	92	0	55	blei_1
97.5	99.5	100	100	93.5	100	100	94	blei_2
98	97	99	98	98.5	98	97	100	blei_3
100	100	100	100	100	100	100	100	dellegrazie_1
100	100	100	100	100	100	100	100	dellegrazie_2
100	100	100	100	100	100	100	100	dellegrazie_3
100	100	100	100	100	100	100	100	kafka_1
100	100	100	100	100	99	100	100	kafka_2
100	100	100	100	100	99	100	100	kafka_3
100	100	100	100	99.5	100	100	100	musil_1
99.5	99	99.5	99.5	99.5	99	100	98	musil_2
100	100	99.5	100	100	55	100	97	musil_3
100	100	99	100	100	99	100	100	ritter_1
100	99.5	100	100	100	100	100	100	ritter_2
100	100	100	100	100	100	100	100	ritter_3
100	100	100	100	100	100	100	100	salus_1
100	100	100	100	100	100	100	100	salus_2
100	100	100	100	100	99	100	100	salus_3
100	100	100	100	100	100	100	100	zweig_1
100	100	100	100	100	100	100	100	zweig_2
99.5	100	99.5	100	99.5	90	100	50	zweig_3
99.26	98.98	99.52	95.12	99.02	96.67	95.1	94.95	mean_values
Burrows	Eder	Canberra	Cosine	CosineDelta	SVM	KNN	NSC	

Fig. 3 Percentages of correctly guessed attributions (200 iterations, 10–2,000 MFWs)

of the corpus. Mean accuracy scores were slightly higher for the merged sets (see Fig. 4). For this reason, we decided to work with merged sets, which is standard procedure in the Stylo package.

Second, we used as the unit of analysis the level of words, not that of character *n*-grams. The use of characters instead of words is widely adopted in authorship attribution and has shown a better efficiency in some specific cases (Grieve, 2007). A stratified cross-validation using the most frequent character 10-grams (the best performing feature for German; Halvani *et al.*, 2016, p. 39), however, showed that performance became substantially lower (see Fig. 4).

Finally, we tested the effect of parameter settings of MFWs used in the analysis (Fig. 5). Accuracy is overall high (already over 85% for twenty MFWs). It increases quite steadily with the number of MFWs but fluctuates below 500 MFWs for all distance measures. Therefore, it should suffice to include 500 MFWs, as this is the range where we can assure that the procedure captures predominantly ‘function words’, which are indicators of style that are normally unconsciously produced (Chung and Pennebaker, 2007; Kestemont, 2014).

As Evert *et al.* (2017) have recently suggested 2,000 MFWs as an ideal parameter setting, we also included higher settings for MFWs, comparing the results of

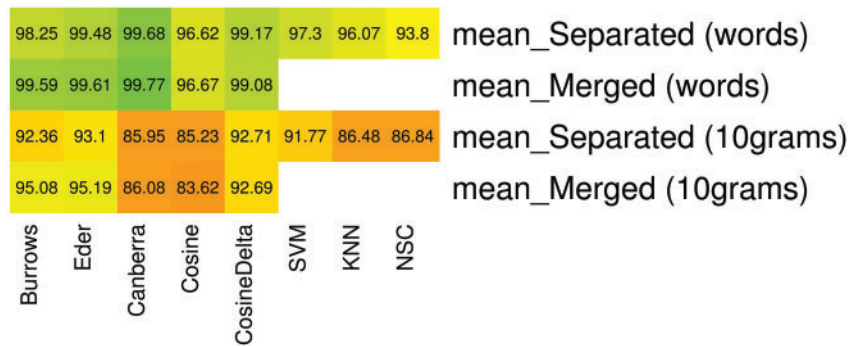


Fig. 4 Twenty-fold stratified cross-validation results (200 iterations, 10–2,000 MFWs)

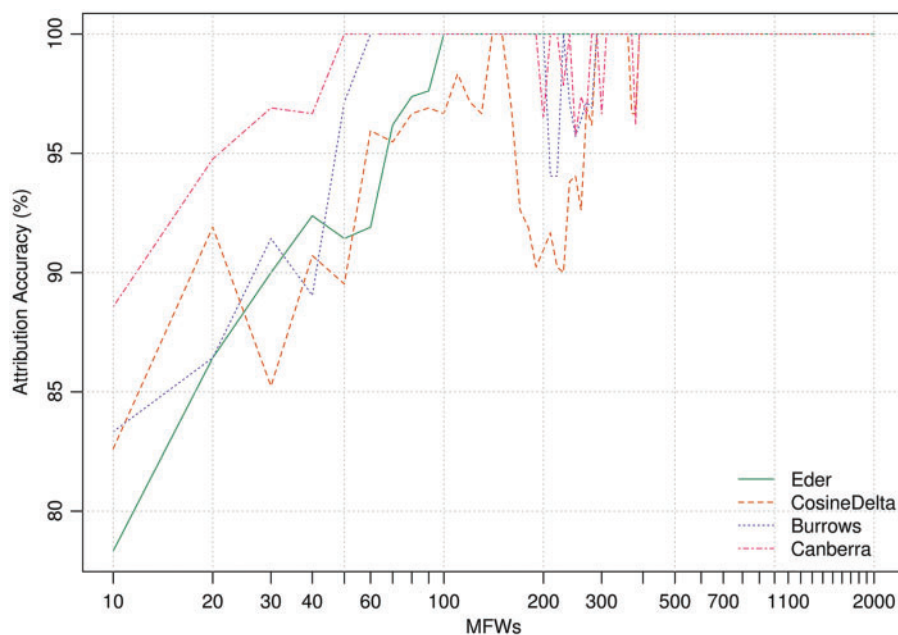


Fig. 5 Mean attribution accuracy for different MFWs parameter settings (three rounds of twenty-fold stratified cross-validation)

multiple combinations of distance measures and MFW settings. As genre may work as a possible confounding signal, we examined the use of culling—the removal of words that are too characteristic for individual texts (Hoover, 2002) which thus helps with fine-tuning the authorial signal. However, as our study deals with very short texts, we eventually decided against culling, since it reduces the number of features substantially below 2,000 MFWs.

3.3 Test set composition: The combinatory design

Tackling the issue of short text length in a test set, Eder (2015, p. 175) suggests that if individual texts are too short, ‘concatenated samples would display a very good performance’. In other words, texts that are too short to be meaningfully used in the procedure are combined. As Eder suggests that the ‘comfort zone’ for these text chunks lies between

5,000 and 8,000 words, concatenating chunks of eight TSZ-texts (with their mean length of about 1,000 words each) seemed to be sufficient. However, the potential combinations of eight out of thirty-eight texts produce a very complex combinatory design, with a total of $N = 48,903,492$ iterations.⁴ Counting with one iteration per second, the computing of the results would take 1 year and eight months to be completed. We thus needed to reduce the complexity of the combinatory design.

Our first solution was to reduce the number of very short texts. As shown by Fig. 1, some texts fall clearly below the 1,000 word-mean. We decided to exclude such texts from our experiment, because they are less certainly attributable and because they sharply decrease the length of possible concatenated samples. Following Burrows's (2007) suggestion, we reasoned that a sensible limit of text length may be at 500 words, thus cutting off nine texts from the experiment. Another text that was excluded is *Aus der Geschichte eines Regiments* (TSZ, 26 July 1916), as it has been attributed to Musil on the basis of philological proof (Corino, 1973). The combinatory design was thus repeated on this simplified corpus shown in Table 2, comprising a total of twenty-eight texts. Through this, the number of texts combined to concatenated chunks could be reduced to six.

The text chunks had an average length of $N = 6,963$ words with an $SD = 909$ words, and only 0.99% of the texts under the limit of 5,000 words. The test set⁵ was composed using the combinations of twenty-eight texts. However, this design still is highly demanding in computational terms (with $N = 376,740$ iterations, counting one iteration per second, it would have required 104 h). Therefore, an *ad hoc* simplification was introduced during the experiment, using just a tenth of the iterations (see Section 4.2).

4 The Experiment

4.1 Implementation

The stylometric analysis was carried out using the *stylo()* function in the Stylo package. In addition, we used other functionalities of the R programming language, implementing the combinatory design

through a series of simple scripts. As a first step, we tested the design on a selection of test sets, obtaining some promising results (Herrmann *et al.*, 2017). In fact, our studies using Cosine Delta on 200–2,000 MFWs and Bootstrap Consensus trees confirmed our intuition.

As noted by Eder (2017a, p. 56), the consensus tree method 'is aimed at producing a number of virtual dendrograms, and then at evaluating robustness of groupings across these dendrograms'. This procedure appears more stable than those based on unique feature selection, because it 'captures the average behavior of a corpus for a given frequency strata'. For example, there are ten virtual dendrograms behind Figs 6 and 7, produced by shifting MFWs from 200 to 2,000, with a regular interval of 200 words. This approach compensates for local aberrant behaviors, like those evidenced in by the fluctuations in Fig. 5. However, a shortfall of consensus trees—as already remarked by Jannidis and Lauer (2014, p. 35)—is that they do not show the distances between texts. In authorship attribution, these distances can be highly significant, because they can be used to estimate the reliability of an attribution (e.g. if one candidate surpasses another by a significantly wide—or dangerously thin—interval). By consequence, we decided to modify the procedure slightly.

Dendrograms—and consensus trees—are visualizations of text similarity. They are created by two ingredients: a matrix containing all the distances between the text samples and an algorithm for reducing the matrix to a hierarchy of clusters. The most widely used algorithm was designed by Ward (1963) and tested in stylometry by Hoover (2003), but some criticism regarding its application has emerged more recently (see Eder, 2012, 2017a). Although cluster algorithms are widely used in stylometry, authorship attribution does not depend on it. The visual imagery of cluster analysis has a clear descriptive power in showing complex mutual relationships between all the texts in a corpus, but it is not necessary when a study—as ours—focuses on the relationship between a corpus as a whole and just one disputable text. In other words, a multi-perspective problem becomes a single-perspective problem, and the distance matrix becomes a

Table 2 Test set comprising TSZ articles previously attributed to Musil; derived from (Schaunig, 2014); D = (Dinklage, 1960); R = (Roth, 1972); C = (Corino, 1973, 2003, 2010); A = (Arntzen, 1980); FL = (Fontanari and Libardi, 1987)

Text #	Title	Date of publication	Attributed by	Length
Excl_1	Der Weg zu den Sternen	8 July 1916	C, FL	185 W.
Excl_2	Aus der Geschichte eines Regiments	26 July 1916	C, FL	976 W.
1	Kameraden arbeitet mit!	6 August 1916	A, FL	814 W.
2	Bin ich ein Österreicher?	20 August 1916	A, FL	689 W.
3	Herr Tüchtig und Herr Wichtig	27 August 1916	C, FL	1,763 W.
4	Das Schlagwort	27 August 1916	A, FL	684 W.
5	Die Erziehung zum Staat	3 September 1916	A, FL	939 W.
6	Bauernleben	1 October 1916	C	1,545 W.
Excl_3	Kunst hinter der Front	8 October 1916	C	317 W.
7	Sonderbare Patrioten	15 October 1916	A, FL	1,222 W.
8	Noch einmal Bauernleben	29 October 1916	C	1,106 W.
9	Opportunität	12 November 1916	FL	1,425 W.
Excl_4	Kannst Du deutsch [III]	12 November 1916	A, FL	252 W.
10	Eine gute persönliche Beziehung	26 November 1916	A, FL	995 W.
11	Eine österreichische Kultur	10 December 1916	R, A, FL	1,360 W.
12	Der Nörgler und der neue Österreicher	17 December 1916	A, FL	779 W.
13	Das Kompromiß	24 December 1916	A, FL	824 W.
Excl_5	Der Augenzeuge	24 December 1916	C	287 W.
14	Heilige Zeit	31 December 1916	A, FL	1,087 W.
15	Zentralismus und Föderalismus	7 January 1917	FL	1,174 W.
16	Föderalismus oder Zentralismus	14 January 1917	FL	732 W.
Excl_6	Kannst Du Deutsch [V]	21 January 1917	A, FL	459 W.
Excl_7	Vorpolitische Reinigung	4 February 1917	A, FL	428 W.
Excl_8	Kannst Du Deutsch [VI]	4 February 1917	A, FL	398 W.
17	Zu Milde und zu Wilde	11 February 1917	A, FL	920 W.
Excl_9	Aus einer öffentlichen Schwulstfabrik	18 February 1917	A, FL	143 W.
Excl_10	Schnucki in der "großen Zeit"	18 February 1917	A, FL	47 W.
18	Neu-Altösterreichisches	25 February 1917	A, FL	606 W.
19	Ist die "österreichische Frage" schwierig?	4 March 1917	FL	1486 W.
20	Seiner Hochwohlgeboren!	4 March 1917	D, A, FL	1323 W.
21	Luxussteuern	4 March 1917	A, FL	543 W.
22	Positive Ziele	11 March 1917	FL	735 W.
23	Der Frieden versprochen!	18 March 1917	FL	1,182 W.
24	Das Staatsprogramm der Deutschen	18 March 1917	A, FL	1,280 W.
25	Wehe dem Staatsmann!	25 March 1917	FL	1,636 W.
26	Der Frieden und die Zukunft	1 April 1917	FL	1,693 W.
27	Presse und Krieg	8 April 1917	FL	2,197 W.
28	Vermächtnis	15 April 1917	D, R, C, A, FL	1,757 W.

distance vector. In our experiment, by consequence, the only part of interest in the distance matrix is the row (or column) containing the information on the test set, while the remaining twenty-one rows (or columns) can be ignored.

When foregoing cluster analysis and its linkage algorithms, we can thus fundamentally simplify our experiment and simply calculate the mean values of the distance vector in all the different configurations. Keeping the distance measure constant,

mean distances can be calculated in different frequency strata (10–100 MFWs, 100–1,000 MFWs, etc.). The major issue with this approach is that some distance measures depend on the number of MFWs used (e.g. the distances calculated with ten MFWs are relatively lower than those calculated with 1,000 MFWs), so it is necessary to normalize the values at each step. In our experiment, we divided all values by their maximum, to obtain normalized distances between 0 and 1. As this



Fig. 6 Preliminary testing. Test set grouping with Musil

procedure renders a large output (see Section 3.3; each of the twenty-eight texts in the test set appears $n=80,730$ times throughout the $N=376,740$ iterations; each combination of an individual text with one of the other five texts produces a unique value), the number of distance values needed to be reduced. We thus calculated the mean value for each text, which was subsequently used for determining the probability of authorship for impostors and candidate authors.

4.2 Combinatory design results

The experiment used a ‘virtual machine’ provided by the *Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen*, with eight Cores and 16 GB RAM. For a preliminary testing of the procedure, we ran an

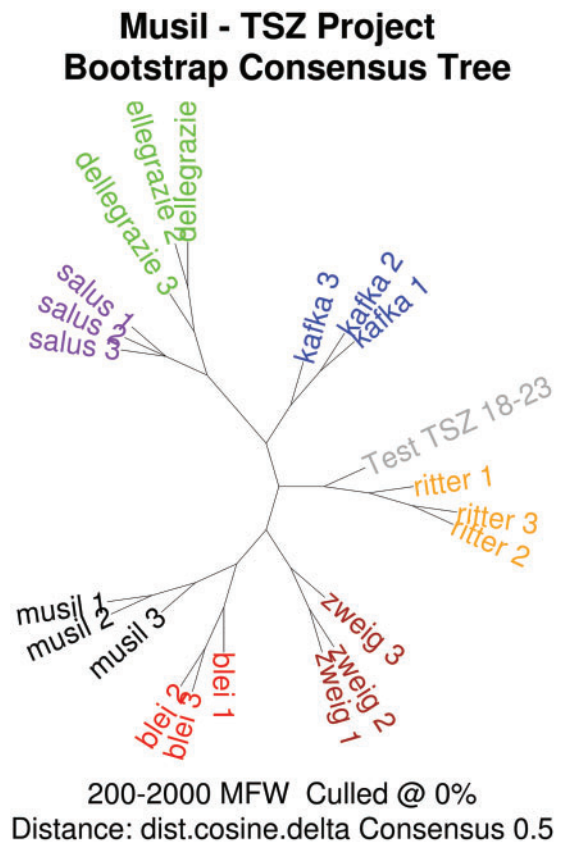


Fig. 7 Preliminary testing. Test set grouping with Ritter

analysis using Burrows’s Delta and 50–500 MFWs, with about 1.5 s for processing each iteration. To reduce the total running time to a manageable size, we implemented an *ad hoc* simplification. After having processed all 376,740 combinations of texts in the test set (which took about 6 days to be completed), we examined whether the results may be reproduced by processing just a fraction of the texts. An analysis of a randomly selected fraction that comprised a tenth of the combinations took about 14 h and produced sufficiently similar results (for details, see Salgaro *et al.*, 2018). The subsequent analyses were just produced on the simplified design.

Figure 8 shows the mean distances between each text in the test set (listed on the x axis) and the texts in the training set (grouped by author). The lower a curve is, the closer it is to the test set, thus suggesting the most plausible author for each individual

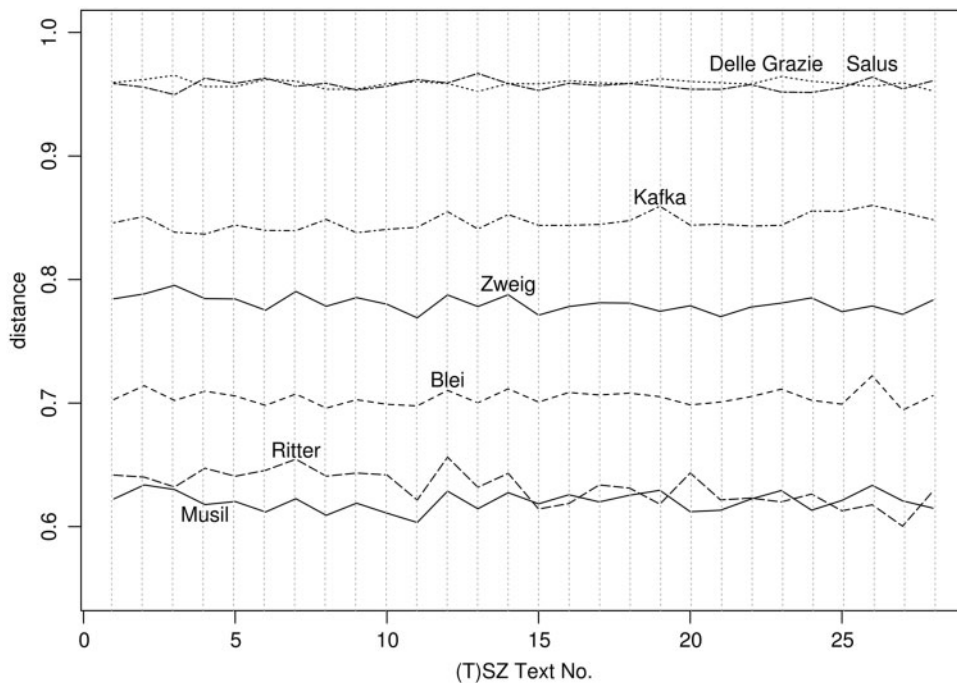


Fig. 8 Mean distances from test set (Burrows's Delta 50–500 MFWs, 37,674 iterations)

text. In the configuration analyzed here, it is clear that the only strong candidates are Musil and Ritter.

In the last step of the experiment, we expanded the design to a total of sixteen different configurations: for details, see Table 3.

Note that the frequency strata are all partially overlapping: we avoided complete separation to test the orders of magnitude, instead of specific values. Results show how different parameter settings cause significant variation in the relative distances, while some tendencies are consistent overall. When considering the candidate authors, Salus (short stories) and delle Grazie (novels) are always furthest from the test set and close to each other. Kafka (letters), Zweig (reviews), and Blei (relatively essayistic fictional texts) each are substantially removed from the test set as well. The impostors thus serve the purpose of building a backdrop to test Musil's authorship of the TSZ articles—at least to some extent. However, their positioning also suggests dominance of the genre signal, as

Table 3 Parameter settings (distance measures + MFWs)

Burrows's Delta	Eder's Delta	Cosine Delta	Canberra
20–200 MFWs	20–200 MFWs	20–200 MFWs	20–200 MFWs
50–500 MFWs	50–500 MFWs	50–500 MFWs	50–500 MFWs
100–1,000 MFWs	100–1,000 MFWs	100–1,000 MFWs	100–1,000 MFWs
200–2,000 MFWs	200–2,000 MFWs	200–2,000 MFWs	200–2,000 MFWs

Note: Each MFWs stratum comprises ten steps.

Salus and delle Grazie are furthest from the test set also in terms of genre.

The most important result is however that the only other author close to Musil's signal is Ritter, who contests Musil in terms of authorship in particular with the highest numbers of MFWs, but not with the lower ones. A synthesized representation of Musil and Ritter in relation to the test set is shown in Fig. 9a.

Figure 9a was created in the following steps. First, considering that in all sixteen different configurations the only authors contending authorship are Musil and Ritter, the graph represents only the

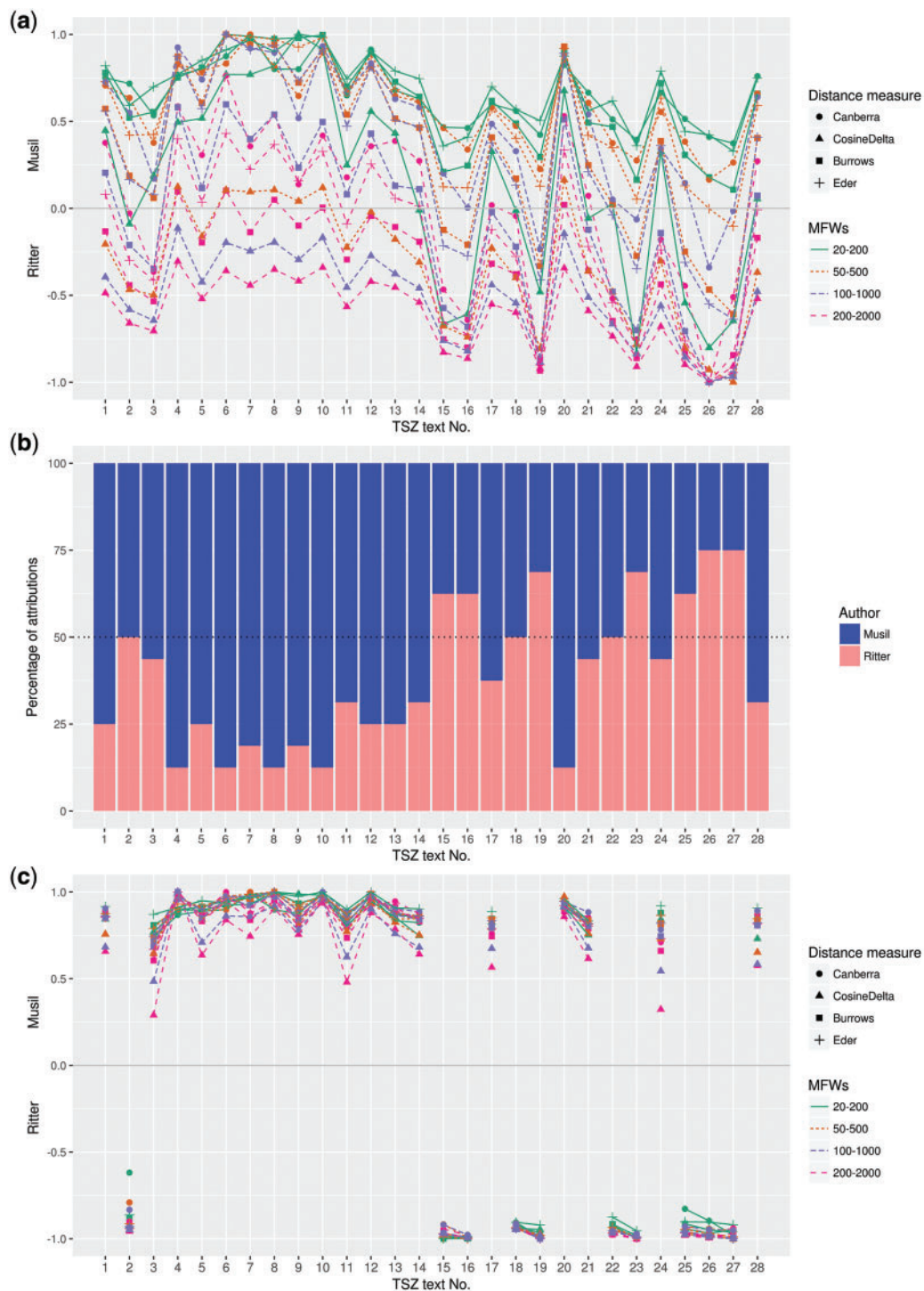


Fig. 9 (a) Dominant authorial signal in the test set (Musil versus Ritter); (b) percentages of attributions; (c) dominant authorial signal in the test set (two groups, ten + eighteen texts)

distances between their respective signals. Second, for better highlighting the distinctions, the measures were normalized to a range between -1 and $+1$. Where a point in the curve is below 0, Ritter's signal is stronger than Musil's; where it is above 0, Musil is dominant. The figure shows some general indications: first, with a few exceptions, Musil's authorial fingerprint is clearly dominant in the first part of the corpus (roughly corresponding with the year 1916), while with the beginning of the year 1917 (text numbers fifteen and sixteen), some more substantial doubts start to emerge. In particular, texts numbers twenty-six and twenty-seven show negative peaks. However, these results are to be considered as no more than a general indication. In fact, they highly depend on the characteristics of the sixteen classifiers, with their particular merits and defects. For example, it is evident that most of the failed Musil-attributions derive from the high MFWs strata. These show better performance overall (see Fig. 5), but they also drive the analysis outside the area of 'function words', which are an important factor for theoretical validity (Kestemont, 2014). One distance measure that attributes all texts from the year 1916 to Ritter (between 100 MFWs and 2,000 MFWs) is Cosine Delta. It has been evaluated as most reliable among stylometric measures (Evert *et al.*, 2017) but appears to be highly dependent on the choice of MFWs—which brings up the already mentioned caveat in terms of attributional validity.

The obtained results thus need to be interpreted with caution. In fact, our method, being based on concatenations of short texts, may have partly masked the absence of Musil's signal, through a saturation effect. In this regard, it calls our attention that the two texts showing the lowest levels of attributability are also among the longest in the corpus. We thus repeated the experiment on a reduced selection of texts, first using the ten texts that were not attributed (NA) to Musil by at least 50% of the classifiers (see Fig. 9b), and then on the remaining eighteen texts.

Figure 9c shows that in this experiment, Ritter's signal clearly overrides Musil's for the first sample (texts numbers two, fifteen, sixteen, eighteen, nineteen, etc.). This holds regardless of the distance

measure and of the frequency stratus. Conversely, Musil's signal is equally dominant in the second sample (texts numbers one, three, four, etc.), with all values between 0 and $+1$.

This result enlarges an effect that was more latent in the first experiment. Using a combinatory design for short texts thus appears as a promising way of indicating attributability. Even though it possibly suffers from a saturation effect, and precisely because it shows how the two signals cannot be entirely separated, it points the way to a more refined analysis.

4.3 Simplified design and final results

To validate these results, we devised a more simplified design, by profiting of the fact that the training set could be reduced to Musil and Ritter alone. In a 'profile-based approach' (Stamatatos, 2009, p. 546), the three text chunks by each author were concatenated into two single chunks, both normalized to a length of 20,000 words. From each of these two vectors, 2,000 words were randomly extracted and combined with the TSZ texts, thus building a series of twenty-eight test sets with a length comprised between $n=4,543$ and $n=7,197$ words, within which Musil's and Ritter's authorial signals were equally represented, together with the disputed texts. The distances between these test sets and the two remaining text chunks (18,000 words for each author) were calculated using the sixteen previously tested classifiers. The procedure was repeated ten times (in a design comparable to a ten-fold cross-validation), thus generating a total of 160 measurements for each one of the TSZ texts.

Results are synthesized in Fig. 10, where the median in the box plot indicates who the majority of the classifiers votes for as the most plausible author of the TSZ texts. *T*-tests on each distribution of measurements confirm that just two texts are NA with a significant margin of certainty. With a few exceptions, results are consistent with those obtained with the combinatory design and, most importantly, all the texts attributed to Ritter with the first method are confirmed by the second one.

Table 4 is an overview of the different attributions. The final attribution combines the two computational methods: in case of disagreement, the

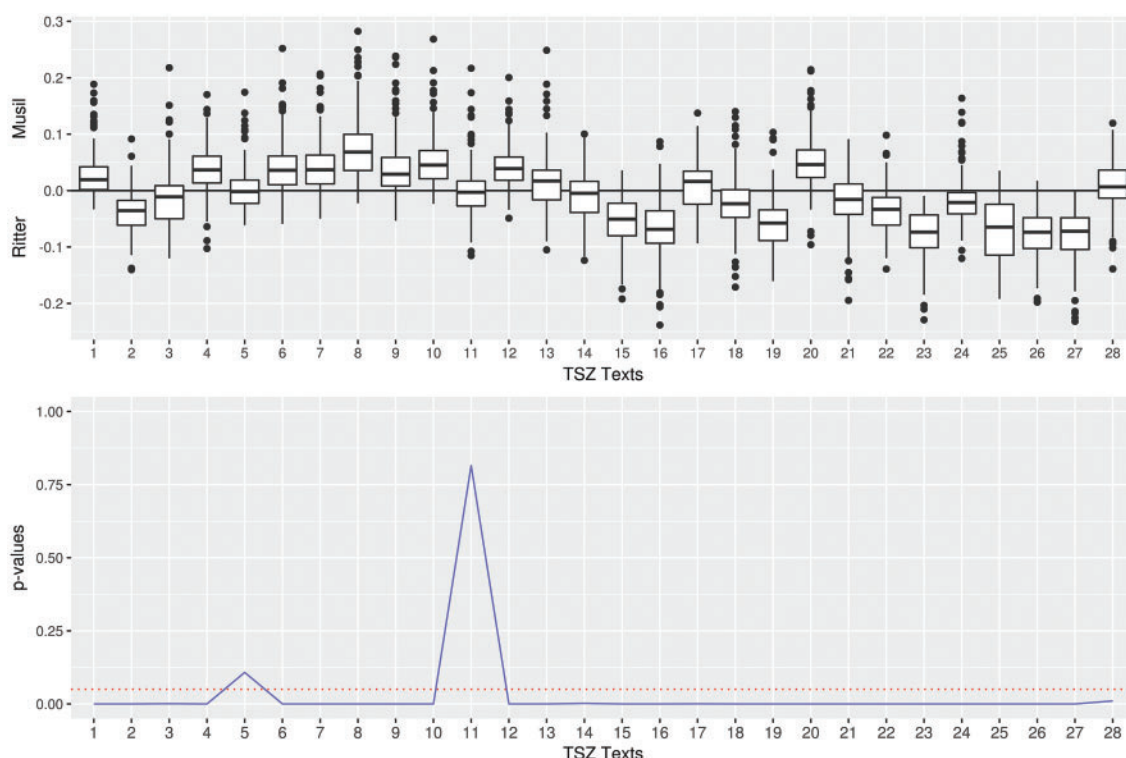


Fig. 10 Attributions with the simplified design (Musil versus Ritter)

text is considered as NA. Providing a still somewhat preliminary answer to our research question, Musil-attribution can be disproved with a high level of confidence for texts numbers two, fifteen, sixteen, eighteen, nineteen, twenty-two, twenty-three, twenty-five, twenty-six, and twenty-seven (see Table 4). Our statistical analysis gives reason to assume that Ritter may be the author of these ten articles. Authorship cannot be disproved at a significant level of confidence for fourteen TSZ-texts, which may cautiously suggest Musil as the probable author.

5 Keyness Analysis

As has been repeatedly pointed out, there is no scientific consensus available for authorship attribution (Stamatatos, 2009; Luyckx, 2010), and the measures we have applied above are still black

boxes when asking the following questions: given that both are candidates for authorship of the TSZ-articles, what do Musil and Ritter have in common when compared with the rest of the training set? What lexico-grammatical features are distinctive of each author's language use? How do the attributed sets of articles of the test set behave when compared to each other? Applying a mixed-methods approach (Herrmann, 2017), the following study supplements the abstract approach of the distance measures with a more contextualized view on the lexico-grammatical features themselves. Specifically, we investigate key words, defined as words that occur 'with unusual frequency in a given text. This does not mean high frequency but unusual frequency, by comparison with a reference corpus of some kind' (Scott, 1997, p. 236).

Experimentally applying the option for measuring keyness available from the Stylo package, we used the *oppose()*-function, which performs a

Table 4 TSZ articles previously attributed to Musil confronted with our attributions

Text #	Title	Combinatory design	Simplified design	Final attribution
1	Kameraden arbeitet mit!	Musil	Musil	Musil
2	Bin ich ein Österreicher?	Ritter	Ritter	Ritter
3	Herr Tüchtig und Herr Wichtig	Musil	Ritter	NA
4	Das Schlagwort	Musil	Musil	Musil
5	Die Erziehung zum Staat	Musil	NA	Musil
6	Bauernleben	Musil	Musil	Musil
7	Sonderbare Patrioten	Musil	Musil	Musil
8	Noch einmal Bauernleben	Musil	Musil	Musil
9	Opportunität	Musil	Musil	Musil
10	Eine gute persönliche Beziehung	Musil	Musil	Musil
11	Eine österreichische Kultur	Musil	NA	Musil
12	Der Nörgler und der neue Österreicher	Musil	Musil	Musil
13	Das Kompromiß	Musil	Musil	Musil
14	Heilige Zeit	Musil	Ritter	NA
15	Zentralismus und Föderalismus	Ritter	Ritter	Ritter
16	Föderalismus oder Zentralismus	Ritter	Ritter	Ritter
17	Zu Milde und zu Wilde	Musil	Musil	Musil
18	Neu-Altösterreichisches	Ritter	Ritter	Ritter
19	Ist die "österreichische Frage" schwierig?	Ritter	Ritter	Ritter
20	Seiner Hochwohlgeboren!	Musil	Musil	Musil
21	Luxussteuern	Musil	Ritter	NA
22	Positive Ziele	Ritter	Ritter	Ritter
23	Der Frieden versprochen!	Ritter	Ritter	Ritter
24	Das Staatsprogramm der Deutschen	Musil	Ritter	NA
25	Wehe dem Staatsmann!	Ritter	Ritter	Ritter
26	Der Frieden und die Zukunft	Ritter	Ritter	Ritter
27	Presse und Krieg	Ritter	Ritter	Ritter
28	Vermächtnis	Musil	Musil	Musil

contrastive analysis between two given sets of texts. On the basis of Craig's (Craig and Kinney, 2009) and Eder's zeta (Eder *et al.*, 2016), *oppose()* generates a vector of words significantly preferred or avoided by an author when compared to a set of texts. In our analysis, we combined these two measures (by simply multiplying the values obtained for each word) and ran two types of analysis: to analyze similarities between Musil and Ritter, we contrasted their texts of certified authorship to the remaining training set (Fig. 11a and b). To analyze differences between the two prime candidates for authorship, we contrasted Ritter to Musil (Fig. 12) on the basis of the texts that we attributed to them (see Table 4). In comparison with other contrastive measures, zeta can reveal how consistently a feature is used across each of the compared text sets (for a most recent discussion of zeta, see Schöch, 2018).

Figure 11 shows the words that were significantly preferred by the two authors (right-hand side) when compared to the remaining training set (left-hand side). The two partitions used for the first analysis were (1) journal articles authored by Musil (20,827 words) and Ritter (21,238 words), respectively; (2) the texts by the remaining six authors (130,485 words for Musil; 130,074 words for Ritter, see Table 1). For both partitions, all texts were merged into single text chunks. Parameters were set to 5,000 words for text slice length, with a moving window of 100 (text slice overlap was set at 4,900). The rare occurrences threshold was set at 2 (default value). A clear genre signal ('journalistic prose') was expected, since the training set includes fictional texts, letters, and essayistic literature. Also, it was hypothesized that Musil and Ritter would show structural and possibly thematic overlap.

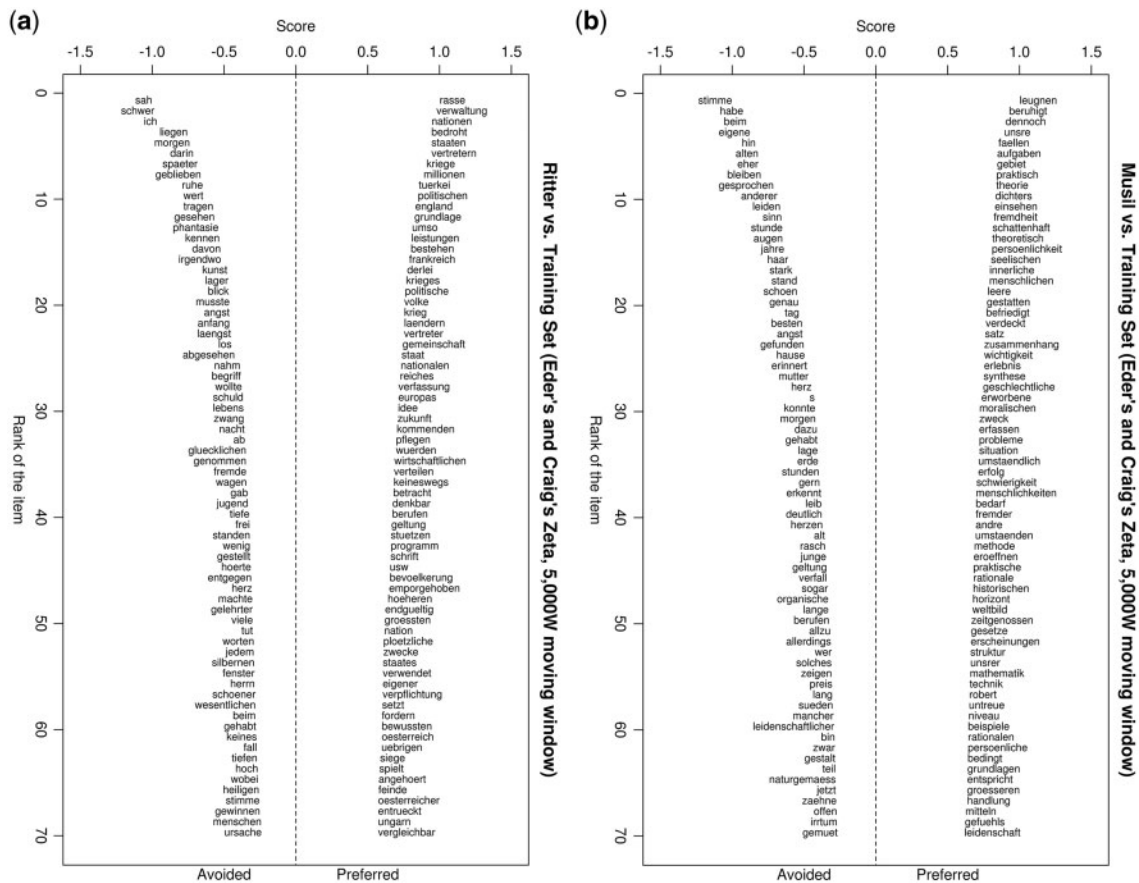


Fig. 11 Zeta analysis: words significantly preferred against training set. (a) Ritter; (b) Musil

The two vectors in Fig. 11 comprise words with a zeta between -1 and $+1$, with positive values reflecting an overuse by Ritter (a), and Musil (b), respectively. With rank order of distinctiveness from top to bottom (rank 1 has a zeta of ± 1), Fig. 11 clearly shows a comparatively formal, complex, and abstract style for both Ritter and Musil. While the training set prefers items indicating an involved style in the sense described by Biber and colleagues (Biber and Conrad, 2009)—frequent use of verbs and pronouns, with basic concrete meanings—both Ritter and Musil show a high use of nouns, which indicates a more formal style, and use comparatively longer and more abstract words than the training set, indicating complexity, and an

informational character (see Biber, 1988). There is thus a clear similarity, which may in part be explained by true stylistic overlap, in part by genre. However, differences between the two contenders are also visible: In contrast to Ritter, Musil's preferred words establish reference not to matters of ideology, war (Ritter: *rasse*, *krieg*, and *programm*), and administration (Ritter: *verwaltung*, *nationen*, and *vertretern*), but to matters of the *conditio humana* and its deliberation, with words such as *gebiet*, *erlebnis*, *zusammenhang*, *weltbild*, and *untrue*. At the same time, Ritter uses formalistic function words (*usw*, *umso*, and *derlei*) which are absent from Musil's vector. As a tendency, Ritter appears to write more pronouncedly political than

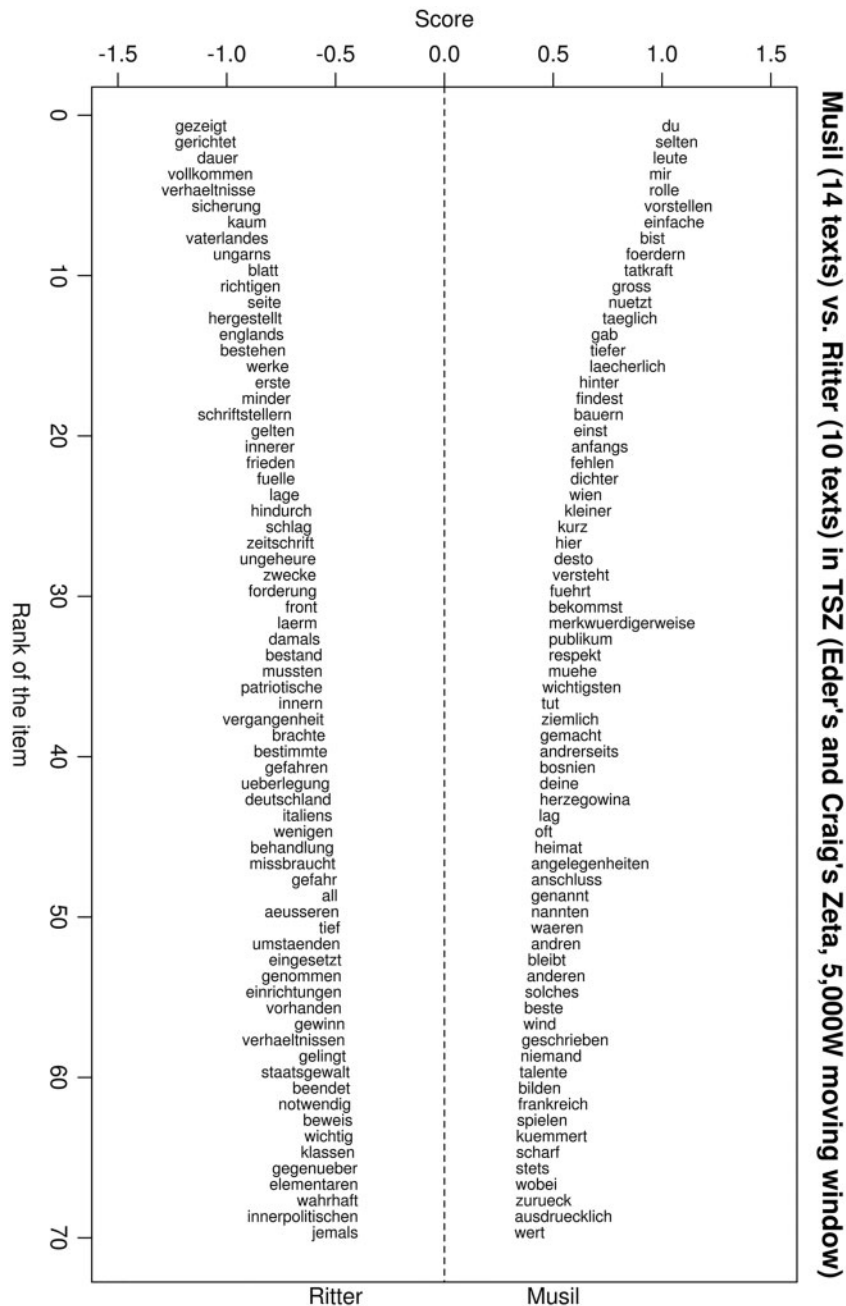


Fig. 12 Zeta analysis: words significantly preferred by texts attributed to Ritter and Musil (test set)

Musil, possibly in an inflammatory 'pamphlet style'. While similarities are clear, the noted differences in style were explored in a second keyness analysis, contrasting the previously attributed articles.

Figure 12 shows the words that were significantly preferred in the text sets attributed to Musil and Ritter when directly compared to each other. The two partitions were (1) the fourteen texts

attributed to Musil (15,743 words); (2) the ten texts attributed to Ritter (12,164 of words). All parameters were set as in the previous analysis.

Figure 12 depicts positive zetas for words overused by the articles attributed to Musil (right-hand side), and negative ones overused by the articles attributed to Ritter (left-hand side). Interestingly, the difference is not just in the content but largely structural. The articles attributed to Ritter are more formal, featuring way more nouns (more than thirty) than verbs (around fourteen), and no pronouns. In terms of content, nouns (*verhältnisse*, *sicherung*, and *vaterlandes*) are abstract and categorical, clearly referring to (European and German) politics, as well as writing, time and space, and reasoning. His adjectives also suggest a formal, categorical style used for political purposes, with evaluative (*vollkommen* and *richtigen*) and political reference (*innerer* and *patriotische*).⁶ The verbs are mostly used as past participles (*gezeigt*), thus as past tense and/or passive voice—a feature that is consistent with language use in a pamphlet style intended as inflammatory but needs to be cautious to not reveal author's identity (see Salgaro (2018), who introduces Ritter as a writer of the TSZ, earmarking his political orientation). The determiners (*bestimmte* and *wenigen*) may be used to establish explicit reference, which is also consistent with an inflammatory political style of pamphlets.

In contrast, Musil's vector suggests a more situated, informal, and much less politically oriented style. It features pronouns (*du* and *mir*), which are absent from Ritter's preferred words and has a much higher proportion of verbs (more than twenty), adverbs (more than ten), and comparably fewer nouns (around fifteen). The ratio of verbs/nouns is thus reversed. His verbs are mostly used in present tense, including second person (*bist* and *findest*). In terms of content, they mostly establish general reference to everyday affairs, including mental, physical actions, and more abstract relations (*vorstellen*, *fördern*, *nutzt*, and *gab*). His nouns refer to local and national culture (*leute*, *bauern*, *dichter*, *wien*, and *heimat*), with items more indicative of culture than politics (*rolle*, *tatkraft*, and *publikum*). His adjectives are basic (*einfache*, *klein*, and *kurz*), and his determiners (*einige*, *jeder*, and *andren*)⁷ may

be used to establish less explicit reference, again consistent with the more informal and involved style. The identified language use as comparatively more abstract, and cultural, is consistent with findings from the Musil philology, which suggest that Musil kept a distance to the strong ideological positions of his contemporaries (Amann, 2007). In contrast to Ritter, the ideological and political thinker dealing with present conditions, Musil appears more as philosophical essayist, addressing the possibilities of human experience (Nübel, 2016). It is remarkable how this stance is not only reflected in content but also in structure, in a less categorical and more involved style.

Overall, only few function words appear on the list, which however would be expected, as important discriminative features of style on theoretical grounds (Kestemont, 2014). Ritter's pronoun *all* and adverb *kaum* are candidates for unconscious style use, as well as Musil's abovementioned use of pronouns and his conjunction *desto*. While not function words in the strict sense, Musil also shows a conspicuously frequent use of primary and delexicalized verbs such as *bist*, *gab*, *tut*, and *bleibt*. Taken together, the two contrastive analyses thus support the stylometric findings—first, Ritter and Musil use quite similar structural and thematic features when opposed to the other authors. This corroborates their joint position at the bottom in Fig. 8. More importantly, the direct comparison of the attributed texts reveals non-trivial differences that are plausibly explained by attested knowledge about Ritter as a political activist and Musil as an essayist and life philosopher. The authorial profile and intentions here seem to correspond with idiosyncratic stylistic choice, both in terms of lexical and grammatical dimensions.

As was shown, keyness analysis is a corpus-driven method that requires a post hoc assessment of the results—which we conducted from the perspective of semantic, lexical, and grammatical features, relating the concepts of involved and informational prose (Biber and Conrad, 2009) to 'inflammatory', 'formalistic', or 'pamphlet-like' style. Logically, future work should complement this perspective, operationalizing these intuitions for quantitative hypothesis testing, by means of resources such as

part-of-speech and named entity tagging, measures of syntactic complexity, but also (semi-)automatic recognition of rhetorical constructions and narrative structure. In addition to complementary methodology, finally, more data are needed to further corroborate and extend our initial findings. As genre and topic appear to play a key role, both parameters should ideally be kept constant.

6 Discussion and Future Perspectives

Our experiments have shown how stylometry, on the level of distance measures and feature exploration, may be utilized for determining authorship attributability of short texts. Our results contribute a new perspective to research on the TSZ: by introducing Ritter next to Musil as a probable candidate, and by offering data-driven and statistical discriminations of twenty-eight debated TSZ articles that have not yet been described in this pattern. Yet, these results should be considered as preliminary. One of the biggest limitations of our experiments is their dependence on the specific training set comprised by an initial set of impostors and candidate authors. Our results should thus be replicated by more experiments, run on a carefully sampled and possibly more extensive training set.

From this point of view, our ‘alpha experiment’ was useful for defining a methodology that, from now on, can be re-applied on different corpora. In addition, it has given some significant indications about the possible composition of these corpora. The fact that the candidate authors delle Grazie and Salus are much further removed from the test set in comparison with the impostors suggests that future studies should sample training sets in which genre is explicitly modeled. It should not be underestimated, in fact, that the pamphlets by Ritter (together with most of the articles published by Musil between 1911 and 1919) were the only texts in the training set that match the politically committed tones—and overtones—of a journal like the TSZ, whose addressees were the soldiers at the front. Further research should also consider other authors active in the TSZ: there were at least four other

potential members of the editorial board and many other occasional collaborators (Urbaner, 2001), but documentation about their activity is scarce if not totally absent, as well as texts of certain attribution. Therefore, the ‘impostors method’ is a relevant method that should be further refined. An ideal selection of training material may for example be Austrian propagandistic writings published during the First World War. Although this material is not readily available, the substantial efforts of its collection and digitization should provide much better results than our impostors Blei, Kafka, and Zweig. An expansion of the training set may also solve some of the parameter-issues of the experiment. As for the choice of MFWs, Evert *et al.* (2017) suggest that topics may become dominant only above 2,000 MFWs, but inspection of the word lists generated by the procedure shows that nouns such as *autonomie*, *politik*, and *krieg/kriege* are present around the 500th position (words that denote the subjects preferred by Ritter). Their presence gives reason to question attribution obtained with high MFWs settings, precisely because content here seems to overshadow more structural aspects of style. However, it is possible that this is an ‘anomaly’ primarily created by the limited dimensions of our training set, and that more function words will appear at high ranks in delta when working with bigger corpora. Using bigger corpora also offers the advantage of enabling culling, disposing over a sufficient number of features, while eliminating the words that are too characteristic of particular texts.

In addition, it should be noted that the methods here applied, especially in the combinatory design, still lack systematicity. The empirical procedure adopted for stabilizing the graphs in Fig. 9 can be ‘bootstrapped’ from the data itself (i.e. by developing an algorithm that automatically stabilizes distances to their maximums) but still depends too much on the intrinsic characteristics of the results—and, especially, on the level of dominance of one candidate over the others.⁸

It should also be noted that 5,000 words are not at all an insuperable limit of text length in authorship attribution. As recently demonstrated by Eder (2017b), even 2,000 words often work fine. Our decision to set the limit to 500 words in our

test set was also motivated by the PAN-framework, which reported an accuracy of 95.7% for Dutch essays with an average of only 412 words (Stamatatos *et al.*, 2014), and an accuracy of 77% for an average of 354 words (Stamatatos *et al.*, 2015). However, the procedure still needs to be validated on German. For a just slightly different procedure, Halvani *et al.* (2016) report an accuracy for German of 78%; yet, while they offer an extremely promising approach to impostor research, specific indications on text length are still missing. Musil's writing should thus be an interesting case for those methods.

Authorship attribution is a heterogeneous and active research field, and there are many more methodologies that may be applied to our problem (see also Hirst and Feiguina, 2007; Basile *et al.*, 2008; Koppel *et al.*, 2008). In addition, as Gschwandtner (2015/2016) noted that many articles in the TSZ—although not the ones taken into consideration here—were actually plagiarisms of previously published material, also text reuse software (Büchler *et al.*, 2018) may be added to the equation. This often deplored methodological multiplicity of authorship attribution actually can be turned into an asset when being used to cross-check results: 'If K different and independent techniques are each $X\%$ accurate, the chance of them all being wrong is $(1 - X)^K$, which becomes arbitrarily small as K increases' (Juola, 2015, p. 107). The combination of only three solid methods (with an 80% of accuracy each) can thus jointly provide results with a validity of more than 99%. However, this approach has its limitations too: good methodologies, even if generally independent, cannot avoid significant overlapping (see Juola, 2012, p. 281). Other dimensions of the research, such as careful sampling of new corpora, will thence be just as crucial.

When the specifics of an author's style, as in the case of Musil, are an open question, digital humanities scholars of different backgrounds normally become specifically intrigued. Here, procedures aimed at the contrastive description of discriminatory features such as key words (Scott, 1997) become specifically valuable resources: they allow

scalability and a closer grip at the words in context, as well as the type of heuristic research that is indispensable for true progress within a larger frame, especially when followed up by genre-sensitive hypothesis-driven analysis of specific style markers such as sentential complexity. Thus, further development and application of zeta, log-likelihood (Rayson and Garside, 2000), and term frequency-inverse document frequency (Robertson, 2004) shall advance digital humanities combinations of explanatory and exploratory research. Our study has geared up new instruments for authorship attribution and style exploration. At the same time, it has hopefully provided food for thought for Musil scholars. May readers of all kinds take it as an invitation for doing the next couple of steps together.

Acknowledgements

The authors would like to thank Maciej Eder and Fotis Jannidis for their support and insightful comments. This work also benefits substantially from the suggestions of our anonymous reviewers.

References

- Amann, K. (2007). *Musil - Literatur und Politik: mit einer Neuedition ausgewählter politischer Schriften aus dem Nachlass*. Reinbek bei Hamburg: Rowohlt.
- Amann, K., Corino, K., and Fanta, W. (2009). *Robert Musil, Klagenfurter Ausgabe. Kommentierte digitale Edition sämtlicher Werke, Briefe und nachgelassener Schriften*. Klagenfurt: Robert Musil-Institut der Universität Klagenfurt.
- Arntzen, H. (1980). *Musil-Kommentar sämtlicher zu Lebzeiten erschienener Schriften außer dem Roman "Der Mann ohne Eigenschaften"*. München: Winkler.
- Basile, C., Benedetto, D., Caglioti, E., and Degli Esposti, M. (2008). An example of mathematical authorship attribution. *Journal of Mathematical Physics*, 49: 125211.
- Biber, D. (1988). *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, D. and Conrad, S. (2009). *Register, Genre, and Style*. Cambridge: Cambridge University Press.

- Breuel, T. M., Ul-Hasan, A., Al Azawi, M., and Shafait, F. (2013). High-performance OCR for printed English and Fraktur using LSTM networks. In *Proceedings of the 12th International Conference on Document Analysis and Recognition*. Washington, DC, USA: IEEE Computer Society Conference Publishing Services (CPS), pp. 683–87.
- Büchler, M., Franzini, G., Franzini, E., Moritz, M., and Bulert, K. (2018). TRACER - a multilevel framework for historical text reuse detection. *Journal of Data Mining and Digital Humanities*, in press.
- Burrows, J. (2007). All the way through: testing for authorship in different frequency strata. *Literary and Linguistic Computing*, 22(1): 27–48.
- Chung, C. and Pennebaker, J. W. (2007). The psychological functions of function words. In Fiedler, K. (ed.), *Social Communication*. New York, NY: Psychology Press, pp. 343–59.
- Corino, K. (1973). Robert Musil, Aus der Geschichte eines Regiments. *Studi Germanici*, 11: 109–15.
- Corino, K. (2003). *Robert Musil: eine Biographie*. Reinbek bei Hamburg: Rowohlt.
- Corino, K. (2010). Klaviersonnen über Schluchten des Gemüts. Robert Musil und die Musik. *Das Plateau*, 120: 4–21.
- Craig, H. and Kinney, A. F. (2009). *Shakespeare, Computers, and the Mystery of Authorship*. Cambridge: Cambridge University Press.
- Dinklage, K. (1960). Musils Herkunft und Lebensgeschichte. In Dinklage, K. (ed.), *Leben – Werk – Wirkung*. Reinbek: Rowohlt, pp. 187–264.
- Eder, M. (2012). Computational stylistics and Biblical translation: how reliable can a dendrogram be? In Piotrowski, T. and Grabowski, Ł. (eds), *The Translator and the Computer*. Wrocław: WSF Press, pp. 155–70.
- Eder, M. (2013). Mind your corpus: systematic errors in authorship attribution. *Literary and Linguistic Computing*, 28(4): 603–14.
- Eder, M. (2015). Does size matter? Authorship attribution, small samples, big problem. *Digital Scholarship in the Humanities*, 30(2): 167–82.
- Eder, M. (2017a). Visualization in stylometry: cluster analysis using networks. *Digital Scholarship in the Humanities*, 32(1): 50–64.
- Eder, M. (2017b). Short samples in authorship attribution: a new approach. In *Digital Humanities 2017 Conference Abstracts*. Montréal: ADHO, pp. 221–4.
- Eder, M., Kestemont, M., and Rybicki, J. (2016). Stylometry with R: a package for computational text analysis. *The R Journal*, 8(1): 107–21. <https://journal.r-project.org/archive/2016/RJ-2016-007/index.html> (accessed 31 May 2018).
- Evert, S., Proisl, T., Jannidis, F., Pielström, S., Schöch, C., and Vitt, T. (2015). Towards a better understanding of Burrows's Delta in literary authorship attribution. In *Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature*. Denver, Colorado, USA: Association for Computational Linguistics, pp. 79–88.
- Evert, S., Proisl, T., Jannidis, F., Reger, I., Pielström, S., Schöch, C., and Vitt, T. (2017). Understanding and explaining Delta measures for authorship attribution. *Digital Scholarship in the Humanities*, 32 (suppl. 2): ii4–16.
- Fontanari, A. and Libardi, M. (1987). *La guerra parallela*. Trento: Reverdito.
- Forsyth, R. S., Holmes, D. I., and Tse E. K. (1999). Cicero, Sigonio, and Burrows: investigating the authenticity of the Consolatio. *Literary and Linguistic Computing*, 14(3): 375–400.
- Giovannini, E. (1987/1988). Der Parallelkrieg. Zu Musils Arbeit in der "Soldatenzeitung". *Musil-Forum*, 13(14): 88–99.
- Grieve, J. W. (2007). Quantitative authorship attribution: an evaluation of techniques. *Literary and Linguistic Computing*, 22(3): 251–70.
- Gschwandtner, H. (2013/2014). Dienst und Autorschaft im Krieg. Robert Musil als Redakteur der Zeitschrift "Heimat". *Musil-Forum*, 33: 101–24.
- Gschwandtner, H. (2015/2016). In der Sperrgewalt der Fackel? Karl Kraus, Robert Musil und die "Tiroler Soldaten-Zeitung". *Musil-Forum*, 34: 157–76.
- Gschwandtner, H. (2016). Kriegspublizistik. In Nübel, B. and Wolf, N. C. (eds), *Robert-Musil-Handbuch*. Berlin: De Gruyter, pp. 434–40.
- Halvani, O., Winter, C., and Pflug, A. (2016). Authorship verification for different languages, genres and topics. *Digital Investigation*, 16: 33–43.
- Herrmann, J. B. (2017). In a test bed with Kafka. Introducing a mixed-method approach to digital stylistics. *Digital Humanities Quarterly*, 11(4). <http://www.digitalhumanities.org/dhq/vol/11/4/000341/000341.html> (accessed 31 May 2018).
- Herrmann, J. B., Lauer, G., Rebora, S., and Salgaro, M. (2017). Short texts in authorship attribution. The Case

- of Robert Musil's war articles. In *AIUCD 2017 Book of Abstracts*. Roma: AIUCD, pp. 57–61.
- Hirst, G. and Feiguina, O.** (2007). Bigrams of syntactic labels for authorship discrimination of short texts. *Literary and Linguistic Computing*, **22**(4): 405–17.
- Hoover, D.** (2002). Frequent word sequences and statistical stylistics. *Literary and Linguistic Computing*, **17**(2): 157–80.
- Hoover, D.** (2003). Frequent collocations and authorial style. *Literary and Linguistic Computing*, **18**(3): 261–86.
- Jannidis, F. and Lauer, G.** (2014). Burrows's delta and its use in German literary history. In Erlin, M. and Tatlock, L. (eds), *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*. Rochester: Camden House, pp. 29–54.
- Jockers, M. L. and Witten, D. M.** (2010). A comparative study of machine learning methods for authorship attribution. *Literary and Linguistic Computing*, **25**(2): 215–23.
- Juola, P.** (2012). Large-scale experiments in authorship attribution. *English Studies*, **93**(3): 275–83.
- Juola, P.** (2015). The Rowling case: a proposed standard analytic protocol for authorship questions. *Digital Scholarship in the Humanities*, **30**: 100–113.
- Kestemont, M.** (2014). Function words in authorship attribution. From black magic to theory? In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*. Gothenburg, Sweden: Association for Computational Linguistics, pp. 59–66. <http://aclweb.org/anthology/W/W14/W14-0908.pdf> (accessed 31 May 2018).
- Koppel, M. and Winter, Y.** (2014). Determining if two documents are by the same author. *JASIST*, **65**(1): 178–87.
- Koppel, M., Schler, J., and Argamon, S.** (2008). Computational methods in authorship attribution. *Journal of the American Society for Information Science and Technology*, **60**(1): 9–26.
- Krause, M.** (2012). Beitrag zur Beurteilung der Autorschaften Musils. *Sprache und Literatur*, **43**(110): 66–80.
- Luyckx, K.** (2010). Scalability Issues in Authorship Attribution. Ph.D. thesis, University of Antwerp.
- Luyckx, K. and Daelemans W.** (2011). The effect of author set size and data size in authorship attribution. *Literary and Linguistic Computing*, **26**(1): 35–55.
- Mosteller, F. and Wallace, D. L.** (1964). *Inference and Disputed Authorship: The Federalist*. Reading, MA: Addison-Wesley.
- Nübel, B.** (2016). Möglichkeitssinn und Essayismus. In Nübel, B. and Wolf, N. C. (eds), *Robert-Musil-Handbuch*. Berlin, Boston: De Gruyter: 719–25.
- Rayson, P. and Garside, R.** (2000). Comparing corpora using frequency profiling. In *Proceedings of the workshop on Comparing Corpora*. Hong Kong: Association for Computational Linguistics, pp. 1–6.
- Robertson, Stephen** (2004). Understanding inverse document frequency: on theoretical arguments for IDF. *Journal of Documentation*, **60**(5): 503–20.
- Roth, M.-L.** (1972). Robert Musil. *Ethik und Ästhetik*. München: List.
- Salgaro, M.** (2018). Albert Ritter, der ghostwriter in der Redaktion der Tiroler Soldaten Zeitung – ein biographisches Profil. In Dimino, M., Locher, E., and Salgaro, M. (eds), *Landsturm-Oberleutnant Dr. Robert Musil als Redakteur der Tiroler Soldatenzeitung*. Paderborn: Fink Verlag, in press.
- Salgaro, M., Rebora, S., Lauer, G., and Herrmann, J. B.** (2018). The 'Tiroler Soldaten-Zeitung' and its authors. A computer-aided search for Robert Musil. In *DHD 2018 Konferenzabstracts*. Köln: DHd, pp. 315–20.
- Schaunig, R.** (2009/2010). Viribus unitis. Robert Musils Schreiben in kollektiver Anonymität. *Musil-Forum*, **31**: 202–23.
- Schaunig, R.** (2014). *Der Dichter im Dienst des Generals. Robert Musils Propagandaschriften im ersten Weltkrieg*. Klagenfurt: Kitab.
- Schöch, C.** (2018). Zeta für die kontrastive Analyse literarischer Texte. Theorie, Implementierung, Fallstudie. In *Quantitative Ansätze in den Literatur- und Geisteswissenschaften. Systematische und historische Perspektiven*. Berlin, Boston: De Gruyter, pp. 77–94.
- Scott, M.** (1997). PC analysis of key words and key key words. *System*, **25**(2): 233–45.
- Stamatatos, E.** (2009). A survey of modern authorship attribution methods. *Journal of the Association for Information Science and Technology*, **60**(3): 538–56.
- Stamatatos, E., Daelemans, W., Verhoeven, B., Potthast, M., Stein, B., Juola, P., Sanchez-Perez, M. A., and Barrón-Cedeño, A.** (2014). Overview of the author identification task at PAN 2014. In *Proceedings of PAN/CLEF 2014*, Sheffield, UK, pp. 877–97.
- Stamatatos, E., Daelemans, W., Verhoeven, B., Juola, P., López-López, A., Potthast, M., and Stein, B.** (2015). Overview of the author identification task at PAN 2015. In *Working Notes Papers of the CLEF 2015 Evaluation*

- Labs. <http://ceur-ws.org/Vol-1391/inv-pap3-CR.pdf> (accessed 31 May 2018).
- Thöming, J.** (1970). Wie erkennt man einen anonym-veröffentlichten Musil Text? *Études Germaniques*, **25**: 170–83.
- Tuzzi, A., and Cortelazzo, M. A.** (2018). What is Elena Ferrante? A comparative analysis of a secretive bestselling Italian writer. *Digital Scholarship in the Humanities*. <https://doi.org/10.1093/llc/fqx066> (accessed 31 May 2018).
- Urbaner, R.** (2001). "... daran zugrunde gegangen, dass sie Tagespolitik treiben wollte"? Die "(Tiroler) Soldaten-Zeitung" 1915-1917. *eForum zeitGeschichte*, **3**(4). http://www.eforum-zeitgeschichte.at/3_01a8.html (accessed 31 May 2018).
- Urbaner, R.** (2005). Schriftführer Musil. Der Jahrhundertschriftsteller als Chefredakteur der Soldatenzeitung. *Quart-Heft für Kultur Tirol*, **5**: 54–67.
- Wagner-Egelhaaf, M.** (2016). Anderer Zustand. In Nübel, B. and Wolf, N. C. (eds), *Robert-Musil-Handbuch*. Berlin, Boston: De Gruyter, pp. 710–12.
- Ward, J. H.** (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, **58**: 236–44.
- Weiss, S. and Kulikowski, C.** (1991). *Computer Systems that Learn*. San Mateo, CA: Morgan Kaufmann.
- 2 <https://github.com/computationalstylistics/stylo> (accessed 31 May 2018).
- 3 All scripts and corpora for this and for the following experiments are available at: <https://github.com/SimoneRebora/MusilStylometry> (accessed 31 July 2018).
- 4 Thus, strictly speaking, $N = 48,903,492$ test sets.
- 5 More properly put, the $N = 376,740$ different test sets.
- 6 Interestingly, the articles attributed to Ritter appear more typically in line with what philological study has revealed about the journal's central topics (see Gschwandtner, 2016, p. 435). It is also interesting that Gschwandtner mentions the article 'Bin ich Österreicher' ('Am I an Austrian') as 'einschlägig' ('pertinent', Gschwandtner, 2016, p. 435). Our attribution suggests it was authored by Ritter.
- 7 Note that the determiner/pronoun *andere* ('different/other(s)') also appears among the Musil keywords in Fig. 11b. It may be related to his concept of 'Anderer Zustand' ('different/other condition'), one of the most important references in Musil's thought. This condition involves an experience ('erlebnis', a keyword in Fig. 11b) of a uniqueness beyond the present conditions, such as in love, mystical experiences, and primitive thinking (see Wagner-Egelhaaf, 2016).
- 8 Also, computing time can be substantially reduced through some accurate modifications in the algorithm. Embedding Stylo functions, in fact, resulted in a significant simplification of the script, but it also imposed a series of useless operations to the virtual machine. For example, as already noted, matrix distances can be reduced to simple vectors and the generation of dendrograms/bootstrap trees for each iteration can be entirely skipped.

Notes

- 1 <http://anno.onb.ac.at/cgi-content/anno?aid=tsz> (accessed 31 May 2018).