

laboratorio dell'immaginario

issn 1826-6118

rivista elettronica

[http://cav.unibg.it/elephant\\_castle](http://cav.unibg.it/elephant_castle)

POSTLUDI. LO STILE TARDO

a cura di Alessandro Baldacci e Amelia Valtolina

novembre 2018

CAV - Centro Arti Visive  
Università degli Studi di Bergamo

SIMONE REBORA, MASSIMO SALGARO

## Is “Late Style” measurable? A stylometric analysis of Johann Wolfgang Goethe’s, Robert Musil’s, and Franz Kafka’s late works

### Late style definition

As many recent publications attest (Leeder 2015; Zanetti 2012; Seidler 2010; Said 2006), “Late Style” has become a fashionable concept in literary theory. The terms “Lateness” and “Late Style” express the idea that the art produced during the final years in the lives of important artists is marked by a profound change of style with respect both to their earlier work and to the work of their contemporaries. Today’s interest in gerontology in the humanities is surely triggered by the aging of contemporary society. This discourse is however, not new, at least not in German culture. In the mid-twentieth century, such important writers and philosophers as Hermann Broch, Gottfried Benn, and T.W. Adorno reflected on the relationship between aging and aesthetic style. An even earlier lecture “On Old Age,” delivered by Jacob Grimm in 1860 (Grimm 2010), may qualify Grimm as the father of this reflection. As Gordon McMullan points out, “‘Late Style,’ invented by romantics, was reinvented by modernists” (2016: 38). In German, we find two concepts that capture the relationship between aging and style: *Spätstil* and *Altersstil*, and historical reasons may also explain why discussion of “Lateness” and aging seems to be a particularly German preoccupation (Leeder 2015: 17ff).

In his “Mythos und Altersstil” (1947), Broch described “radical stylistic change” and “sharp stylistic break” in the creative output of such

geniuses as Titian, Rembrandt, Goya, and Bach (Broch 1995: 213). He also observed these phenomena in Goethe's late writings – the final scenes of *Faust* for instance – in which the language discloses its own mysteries (ibidem). Because style, for Broch, was “the creation of a specific system of conventions for a determinate epoch in history,” the “late stylist” was “an artist who is not happy with the conventional vocabulary provided him by his epoch” (ivi: 214) and, as such, needed to position him- or herself outside that constraint. In his 1954 lecture, “Artists and Old Age,” Benn similarly described the complex relationship between the aging artist and his era (Benn 1989). Although Benn claimed not to consider the physiology of aging, he provided a list of many geniuses who reached old age, including Michelangelo and Titian (ivi: 130). He considered “Late Style” to be “a structural change compared to early work” (ivi: 132). “Late Style” – which for Benn, was a fashionable condition at that time – could assume different faces. For some artists, it was more serene and transcendent, sometimes more autobiographical than their young or mature production. In the case of Michelangelo, “Late Style” was expressed by “a rejection of his peculiar methods and techniques” (ivi: 135) and “a refusal of his role models” (ivi: 136). In a posthumous publication, Edward W. Said (2006) picked up the main ideas developed by Adorno, interpreting “Late Style” as discordance of the artist with his era, making him or her a figure of exile and anachronism. For Adorno, “Late Style” portrayed “in the clearest possible lines, the contradictions and flaws which cut through present-day society” (1957: 391). Said was not interested in late works “that reflect[ed] a special maturity, a new spirit of reconciliation and serenity” (Said 2006: 6) but rather in “the experience of late style that involve[d] a nonharmonious, nonserene tension, and above all, a sort of deliberately unproductive productiveness going *against*...” (ivi: 7). As Said expressed it, “late style is in, but oddly apart from, the present” (ivi: 24).

For McMullan and Smiles, Said's understanding of “Late Style” replicated the romantic understanding of the concept. For them, “the late-style trope takes from romanticism its emphasis on biography, subjectivism, the relationship between creativity and selfhood; from

modernism it derives its interest in tradition, the avant-garde, abstraction, the subordination of self to epoch, the loss of linearity” (McMullan, Smiles: 11). “Late Style” reemerged during modernism at which point the traditional concept of artwork as perfection and totality came to an end (Zanetti 2012: 19-20). The fact that this notion appeared in European culture some decades after the introduction and dissemination of the concepts of childhood (Key 1909) and youth (Pirro, Zenobi 2011) can be no coincidence.

Said's definition of “Late Style” inspired several studies on the late works of artists but also earned negative criticism. Robert Kastenbaum, for example, wrote that “‘Late Style’ is an illusion [...] which ignores the variety of processes and contexts in which creative works are produced late in life” (Kastenbaum 1985: 252; see also Zanetti 2012: 14-15, 204-205). Indeed, Kastenbaum criticized Said's generalization because, in Kastenbaum's eyes, “Late Style” was not a universal phenomenon, and every artist had a different stylistic trajectory. McMullan, similarly, expressed a trenchant judgement of the concept of “Late Style” which, for him, was not “a natural phenomenon,” but a trope, “a critical construct” (McMullan 2016: 36). Given claims regarding the universality of “Late Style,” McMullan considered it regrettable that interdisciplinary studies in literature, music, and the visual arts had rarely been attempted. He also highlighted the collapse of historical difference in the concept of “Late Style” and the lack of distinction between, for example, a specifically modernist and a pre-modernist “Late Style” (ivi: 34).

Later contributors to “Late Style” in literary theory offered more precise outlines of this category. Sandro Zanetti, the author of an insightful monograph on “Late Style,” *Avantgardismus der Greise? Spätwerke und ihre Poetik* (2012), wondered whether generalizable features of late works could be identified. Late works (*Spätwerke*) were, for him, not merely literary texts that came later in life, after other production, but texts that changed perspectives regarding what had been written and published before. He recognized five features of “Late Style”: (1) a work that made reference to another one; (2) a literary text that continued an earlier one; (3) a change in work methods; (4) work produced after half the artist's creative

career had passed; (5) work produced when the end of the career was foreseeable. Not all five criteria needed to be met in order to distinguish a *Spätwerk* from an *Alterswerk* (Zanetti 2013: 55). Rather, what characterized a *Spätwerk* was a double temporal orientation, an internal dialectic that determined its aesthetic quality. For example, it referenced an earlier work and, at the same time, was open to the future, implying the survival of the present moment (Zanetti 2012: 8-9).

Zanetti argued that “Late Style” often appeared when an author produced serial texts or revised a former version of a text, one example of which could be Goethe’s *Faust: Der Tragödie zweiter Theil* (1832), which referenced an 1808 *Faust*. In the late production of an author we could, surprisingly, recognize a sharp break – a “*Greisen-Avantgardismus*” – as Thomas Mann named it. Consequently, late works were a continuation, a survival but also a new beginning (ivi: 8). When he was older, Goethe initiated discussion of his “Late Style” by exposing a rupture between his early and later works in the comments he made about his own work. Ernst Lewy, in his *Zur Sprache des alten Goethe* (1913), described for the first time the main characteristics of Goethe’s “Late Style”: epigrammatic concision, a preference for the unusual and for neologisms, and a tendency towards the symbolic and didactic. From these initial definitions, Zanetti accompanied the reader through the main discussions of “Late Style” in the twentieth and twenty-first centuries. Interestingly, Zanetti placed the concept in dialogue with such important abstractions of literary criticism as Kristeva’s “dialogism” and “intertextuality” and Harold Bloom’s “anxiety of influence” (ivi: 215). For Zanetti, “Late Style” hid exactly this duality: openness to previous works by the same author but, at the same time, resistance to influence and to the anxiety of becoming different (Zanetti 2012: 231). Alongside other concepts born in the time of modernism, such as literariness (Salgaro 2018) and style (Herrmann, van Dalen-Oskam, Schöch 2015), “Late Style” shared a moment of rupture with tradition that was typical of modernist understandings of literature.

## Methods and aim

The scope of our research is to combine literary theories on and analyses of “Late Style” with stylometric analysis of the “Late Style” of three important German authors: Johann Wolfgang Goethe, Robert Musil, and Franz Kafka. Several theoretical and methodological shortcomings are obvious in this approach, of which definition of the concept of style is the most complex. Indeed, it is debatable whether the understanding of style in literary theory and in the digital humanities is equivalent. This, however, is also an empirical question. We propose a truly interdisciplinary approach whose methodology combines the qualitative approach of literary analysis with the quantitative approach of stylometry. This transdisciplinary approach may help resolve controversies regarding the existence of “Late Style” and determine whether it can be detected by quantitative methods.

As part of an attempt to overcome the boundaries between literary stylistics and stylometry, we note that our aim to foster an interdisciplinary dialogue on (late) style is shared by Herrmann, van Dalen-Oskam, and Schöch who, in their 2015 article “Revisiting Style: A Key Concept in Literary Studies,” advocated “work towards a common ground [...] when talking about (literary) style” (26). To do this, they provided a review of the ways in which literary style had been defined since 1945 in Dutch, French, and German language and literary studies. In the German-speaking context, their investigation began with the canonical studies of Wolfgang Kayser (1948) and Emil Staiger (1955) and continued to the founder of literary stylistics, Leo Spitzer (1961), and such important scholars as Peter Szondi (1967).

The methods of Herrmann, van Dalen-Oskam, and Schöch’s archeological research on the concept of style are less important here than are the six definitions of literary style that emerged from their quest: “style as revealing a higher-order aesthetic value, as the holistic ‘gestalt’ of single texts, as an expression of the individuality of an author, as an artifact presupposing choice among alternatives, as a deviation from a norm or reference, or as any formal property of a

text" (Herrmann, van Dalen-Oskam, Schöch 2015: 26).

Many of these features, which belong to the classical definition of style, are also attributed to "Late Style" as previously mentioned. Some of these features are not empirically testable, such as Staiger's fuzzy definition of style as the "Ineffable-Identical" (*Unaussprechlich-Identische*) (1955: 23). More central is the concept of style as deviation from some type of norm, which is important for literary style as well as for "Late Style" (see Bruneau 1951; Riffaterre 1973). So-called "deviance theories of style" match the style of a text against another body of texts which are labelled as "normative" or as background from which the style can express its difference. For Herrmann, van Dalen-Oskam, and Schöch (2015: 43), style as deviation was one of the strongest strands of stylistics across European traditions of style.

This definition of literary style overlaps with the definition of "Late Style" proposed above. For authors like Said or Broch, "Late Style" involves a double deviation – from the style of the early work of the author and from the style of contemporary writers. We call the first deviation an "internal" deviation, because the stylistic break is internal to the style of a single author; and the second "external," because the gap is between the style of an author and that of contemporary works by other authors. For Goethe, Musil, and Kafka, we combined internal and external analyses together with quantitative and qualitative methods. Following Herrmann, van Dalen-Oskam, and Schöch, we consider: (1) qualitative research as the interpretation of literary criticism and (2) quantitative research as the computational analysis of style. In detail:

1) In the 'external' analyses, we compared the style of the author to reference *corpora* of contemporary authors and assessed relative deviations from them. Because we were dealing with huge *corpora*, we used solely quantitative measures based on computed frequencies, relations, and distributions of features and relevant statistics, which are typical of distant reading.

2) In the 'internal' analyses, we compared the early, middle, and late works of the same author. In this case, we also employed an operational description of the "Late Style" of our three authors by a

relevant literary critic. In fact, "Late Style" must be "explicitly defined and clearly identified" (ivi: 44) by a literary critic. This (qualitative) identification may refer to linguistic features at the level of characters, lexicon, syntax, or semantics. Quantitative methods attempt to complement and potentially to confirm, with a different approach, the fine-grained analysis of literary hermeneutics.

Our aim is not to oppose but to combine the two methods and overcome typical misunderstandings and shortcomings of the interdisciplinary dialogue. Even the difference between "evidence" and "interpretation" can be overcome, because, as Jannidis and Lauer assert, in the context of their stylometric analysis of German literary history, the "interpretations of the results of quantitative studies [...] are hermeneutic acts of sense making" (Jannidis, Lauer 2014: 50).

### **Stylometry: history and applications**

The final goal of stylometry is as simple as it is far-reaching. Through statistical analyses of language, stylometry attempts to 'measure' style, thus discerning authors' hidden 'fingerprints.' According to Patrick Juola (2006: 240-243), the origins of stylometry can be traced to the end of the nineteenth century, when Thomas C. Mendenhall (1887) first applied Augustus de Morgan's original intuitions – albeit inconclusively. While the history of stylometry has been marked by groundbreaking successes, such as Mosteller's and Wallace's (1964) analysis of the Federalist Papers, epic failures, such as that of the Cusum technique by Andrew Morton (1978; see Holmes 1998: 114), have also occurred.

The definitive affirmation of this field of research in literary studies, however, dates to the end of the twentieth century, when John F. Burrows proposed a surprisingly effective method for attribution of authorship, known from that moment on as "Delta distance" (Burrows 2002). Its logical implementation was extremely simple. From a group of texts, the occurrences of single words were extracted, and a list was built of the most frequent words in the *corpus*. For each of the texts, the relative frequencies of these words were calculated, generating a vector of numbers. The distance between two

texts was therefore the distance between their two representing vectors, calculated through an *ad hoc* formula. Burrows tested this method on a *corpus* of English Restoration poets, obtaining surprisingly accurate results. In most cases, in fact, the “closest” texts were those written by the same authors.

During the last sixteen years, improvements have been proposed for Delta distance, but the statistical process has remained substantially the same (see Evert et al. 2017). Delta has proved a valid method for attribution of authorship and has been applied to multiple disputes concerning contemporary blockbuster authors, including J. K. Rowling and Elena Ferrante (see Juola 2015; Tuzzi, Cortelazzo 2017), as well as to works attributed to Dante, Shakespeare, and Musil (see Canettieri 2016; Craig, Kinney 2009; Reborá et al. 2018). The success of the Delta distance calculation has frequently been compared to that of the art historian Giovanni Morelli, who attributed numerous works to famous Italian painters based on minute, secondary particulars such as ears, hands, and feet, and not on more typical stylistic choices. In consequence, it has been suggested that the efficacy of this method is determined by the subconscious choices each writer makes when dealing with the most frequent words in his or her vocabulary (such as articles, pronouns, and adverbs, also known as “function words”), which can tell more about an author than conscious stylistic choices (Kestemont 2014).

More recently, Delta has been used for the study of stylistic similarities, influences, and derivations in entire literary canons, including the German one (see Jannidis, Lauer 2014). By combining Delta with clustering techniques derived from fields of research such as phylogenetic and network analysis (see Eder 2017a), stylometry has finally evolved into the branch of digital humanities that provides the most remarkable and reliable expressions of the “distant reading” paradigm (Moretti 2013).

The category of “Late Style” is not new in stylometry. As Jan Rybicki showed (2017), stylometric distances proved to be very sensitive to changes over time, and David J. Hoover (2014) used them to analyze Henry James’ “Late Style.” In a 2018 study, Jonathan P. Reeve tested the category on a wide *corpus*, limiting his focus solely to the inter-

nal perspective (i.e., by comparing the early and later production of individual authors). We will extend the analysis to the external perspective, broadening it beyond distance measures and word frequency. We therefore adopt methodologies that focus on the entire vocabulary and on its semantic aspects: from Zeta analysis, which aims to identify words that are significantly over- or underrepresented in a specific author’s works (Schöch et al. 2018) to word class analysis, by which a text can be mined based on pre-compiled dictionaries to evaluate the semantic areas that dominate it (Tausczik, Pennebaker 2010). For example, the occurrences of a group of words related to the concept of “lateness” (such as “old,” “late,” and “last”) can be counted to evaluate how much a text is dominated by that concept. Of course, word class analysis can adopt much more sophisticated approaches (that involve natural-language processing, statistical estimates, and machine-learning techniques), but the simple “word count” method has also proved quite successful when dealing with issues of literary theory (see Jockers 2014). In addition, this approach can reach aspects of the text that traditional stylometric methods do not fully grasp.

More in general, it is worth noticing that statistical approaches to the study of style cannot be limited to a single calculation or set of techniques because the field remains in flux and may and should be shaped by the needs of traditional stylistic research.

### **Experimental setup and first case study: Goethe**

The decision to start our analysis of “Late Style” with Goethe was inspired by the major theoreticians of “Late Style” who, whether in the case of Benn, Broch, or Said, all cited Goethe’s late works as paradigmatic cases of “Late Style” (Zanetti 2012: 61-70; Sampaoalo 2009: 87-97). Goethe was also chosen to test and refine a methodology that could be expanded to other authors. The large availability of digitized texts by Goethe and of critical studies of his work offered an opportunity to verify the supposed distinctiveness of Goethe’s late production from multiple perspectives, bringing together a variety of approaches and techniques.

The first, indispensable component for any computational textual

analysis, however, is a well-structured *corpus*. Representativeness of *corpora* (Leech 2007) is a quite dated – but still extremely relevant – issue in digital humanities. In particular, testing literary theories requires the definition of a *corpus* that reflects the main characteristics of the literary scene for the period chosen (i.e., it should be sufficiently extended, but also well-balanced in terms of genres, geography, and time). In the case of “Late Style,” the temporal component becomes even more relevant. Samples must be dated with extreme accuracy in order to make both internal and external comparison possible.

As for a *corpus* of Goethe's contemporaries, we followed H.A. and E. Frenzel's indications (1959: 1, 200-295) and selected the most representative works from Sturm und Drang, Classicism, and Romanticism. The plain text versions were downloaded from the online digital libraries *Zeno* and *Project Gutenberg-DE*. We selected thirty-nine of Goethe's works from the *Project Gutenberg-DE* *corpus*. Following indications by Eder (2015), all texts shorter than 5,000 words were excluded from the stylometric analysis.<sup>1</sup> Based on a periodization that is generally shared by Goethe scholars, we split our *corpora*<sup>2</sup> into three sub-selections:

- 1) *Young* selection, corresponding to works published before 1776 (Sturm und Drang);
- 2) *Middle* selection, corresponding to works published between 1777 and 1808 (Classicism);
- 3) *Late/Old* selection, corresponding to works published after 1809 (late works).

### First approach: stylometric (network) analysis

Following the indications of Evert et al. (2017), we chose Cosine

1 Because stylometric analyses are generally based on lists of most frequent words, it is easy to surmise that they may not work well with short texts. Eder (2015) set the text-length threshold to 5,000 words. Even if the discussion is still open and multiple studies propose much lower thresholds (see Eder 2017b), we decided 5,000 words was the most prudent choice.

2 More information about the composition of the *corpora* and the scripts used for their analysis is available in the online repository *LateStyle*.

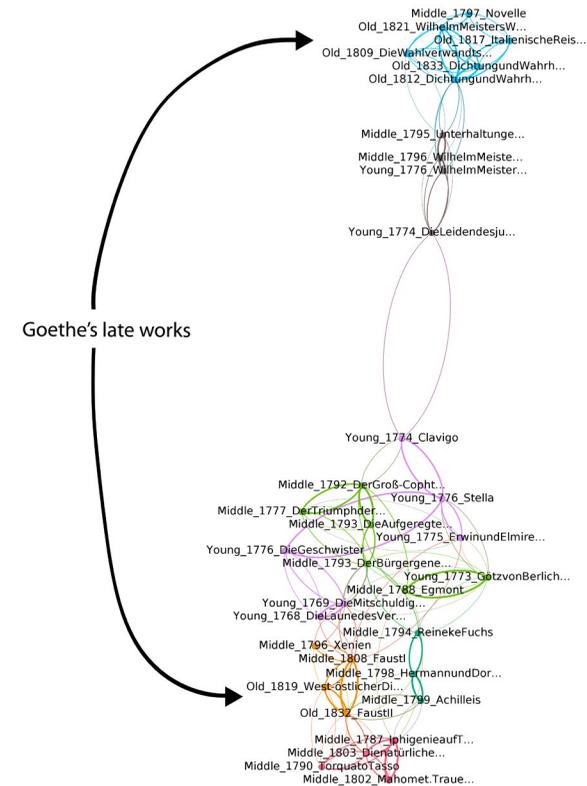


Fig. 1  
Network analysis of Goethe's works.

Delta distance and the 2,000 most frequent words as the best-performing features for stylometric analysis. Results of the internal analysis, obtained by using the *ForceAtlas2* algorithm on the *Gephi* platform (for details, see Eder 2017a), are shown by [Fig. 1].<sup>3</sup> Groupings (evidenced by the different colors of nodes and edges) are

3 For an intuitive interpretation of the graph, all nodes (i.e., Goethe's works) might be considered as planets in a gravitational system, where stronger attractions correspond to shorter stylometric distances. By consequence, the node closest to most of the others will be the most central, while the farthest will be the most peripheral.

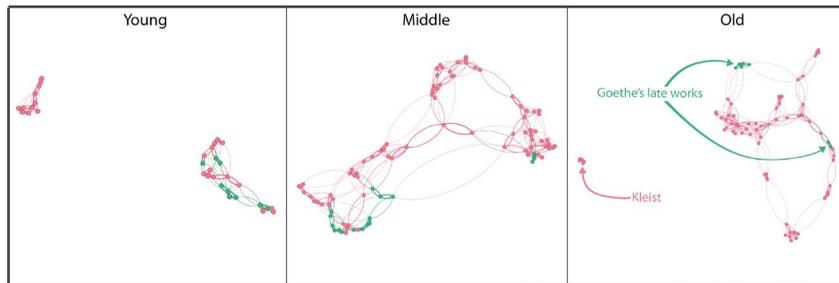


Fig. 2  
Network analyses of Goethe's works (green) and their contemporaries (pink).

automatically determined by the modularity function, as described by Rybicki et al. (2018).

A group of late works (identified by the tag "Old") is isolated on the upper part of the graph. This possible confirmation of the caesura between young and "Late Style," however, is contradicted by at least two factors. First, a work from the "Middle" selection (*Novelle*) is located in this area. Second, two late works (*West-östlicher Divan* and *Faust, zweiter Teil*) are grouped with the "Middle" section.

The results of the external analysis are shown in [Fig. 2]. To reduce complexity, two colors were used, and Goethe's works are always represented by green. As is evident, there is no clear separation between Goethe's late works and those of other writers published contemporaneously. On the contrary, a group of texts by Kleist (the cluster of pink nodes isolated on the left side of the "Old" section of Fig. 2) is most clearly separated from the rest.<sup>4</sup>

## Second approach: Zeta analysis

To confirm these results with a different method, we chose Zeta analysis as described by Craig and Kinney (2009).<sup>5</sup> This method of

<sup>4</sup>This result is in line with the stylometric analyses already performed by Jannidis and Lauer (2014).

<sup>5</sup>The approach is quite straightforward: (1) given two groups of documents, each document is split into a number of segments of equal length; (2) for each

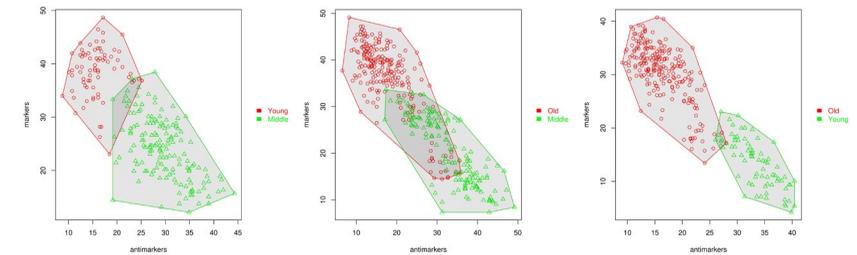


Fig. 3  
Zeta analyses of Goethe's works.

ferred an opportunity to reduce the selection of features to a group of words peculiarly over- or underused by an author; moving beyond the threshold of most frequent words. To generate the final graphs, we adopted the *Markers* method of the *oppose* function in the R package *Stylo* (Eder et al. 2016).<sup>6</sup> All parameters were set to their default values, apart from slice length, which was set to 3,000 words (see Hoover 2013). Results of the internal analysis are shown in [Fig. 3].

The clear separation between "Old" and "Young" areas seems to

word-type in the documents, the proportion of segments in which it appears is calculated (separately for each group); (3) the two values are subtracted. In a real case scenario, suppose that we would like to calculate the Zeta Value for the word "Licht" in Goethe: (1) the first group of documents will be composed of works by Goethe, the second by an ample selection of works by other authors; all documents will be split into 3,000-word-long segments; (2) the proportion of segments in which the word "Licht" appears is calculated both in the Goethe sub-corpus (suppose it will be equal to 0.65) and in the other authors (0.15); (3) the Zeta Value for the word "Licht" in Goethe will thus be  $0.65 - 0.15 = 0.5$  (indicating that it is overrepresented in Goethe). All values will be between  $-1$  (indicating underrepresentation) and  $+1$  (indicating overrepresentation).

<sup>6</sup>The *Markers* method positions each text segment in a two-dimensional space by adding all positive Zeta Values for the words that appear in it (Y coordinate, also defined as "markers") and all negative Zeta Values (X coordinate, "antimarkers"). As a result of following this procedure, the segments tend to appear in two separate areas of the graph (upper-left side for the first group of documents; bottom-right side for the second group). The actual strength of the distinctiveness of words, however, is what determines whether the two groups will be fully separated or partially overlapping.

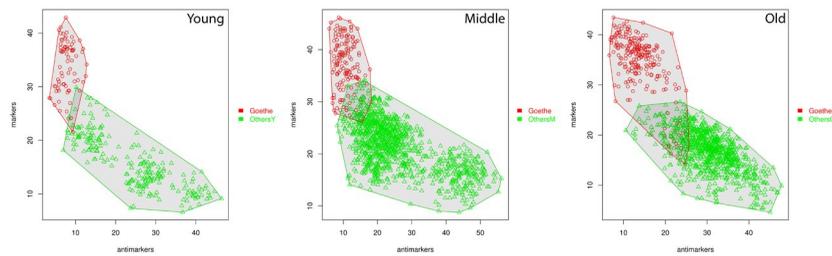


Fig. 4  
Zeta analyses of Goethe's works and works of other authors published contemporaneously.

confirm the peculiarity of "Late Style." The partial intersections between temporally closer selections ("Young" and "Middle"; "Middle" and "Old"), however, suggest that this phenomenon may depend upon the chronological evolution of style, as demonstrated by Eder (2017a) and Rybicki (2017). Quite surprisingly, the external analysis shows that Goethe's late texts are more connected to their contemporaries than the previous ones [Fig. 4],<sup>7</sup> thus contradicting once again the supposed isolation of late style supported by Broch and Said.

The substantial inconclusiveness of these analyses, while hinting at the possible inconsistency of the category of "Late Style" (at least, in the case of Goethe), suggested the need for a different methodology more directly focused on the semantic aspects of the text.

### Third approach: semantic analysis

Erich Trunz's essay, "Goethes Altersstil," first published in 1954, is one of the main studies of Goethe's "Late Style." Trunz distinguished three phases in Goethe's career, the last of which was characterized by strong symbolism and a tendency towards totality (Trunz 1990:

7 See the "Old" section of Fig. 4. The red circles in the lowest positions represent 3,000-word excerpts from *West-östlicher Divan* and *Faust, zweiter Teil* (the same works that were already separated in Fig. 1). These are overlapping with excerpts from Achim von Arnim, Bettina Brentano von Arnim, Jean Paul, and E. T. A. Hoffmann (represented by the green triangles).

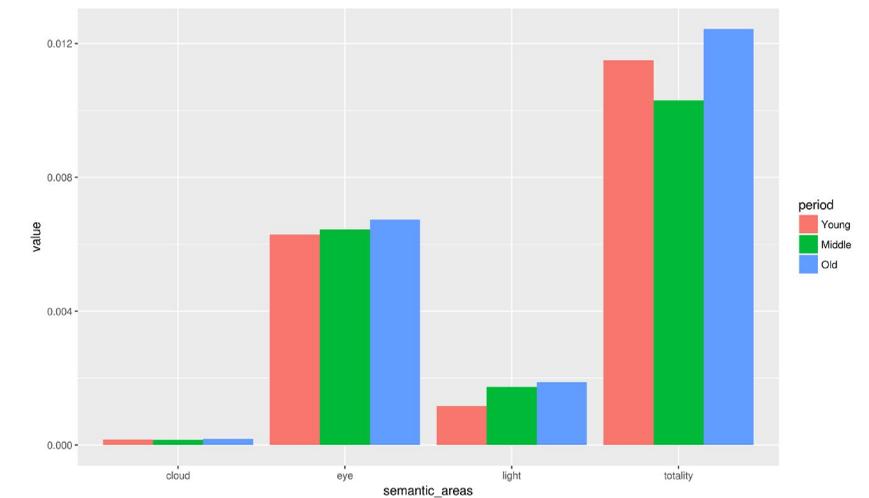


Fig. 5  
Frequency of Trunz's semantic areas in Goethe.

139). Following Trunz, Goethe's "Young" work was inspired by Sturm und Drang and his mature works by the great masters of world literature. In contrast, Goethe's "Late Style" was "completely unique. It doesn't have any relationship to contemporary poetry and doesn't have any model in the wide field of world literature" (ivi: 144). Trunz identified four conceptual symbols as representatives of Goethe's "Late Style": (1) light and (2) totality as symbols of the divine; (3) the eye and (4) the cloud as symbols of humankind's position between finiteness and infinity (ivi: 141). Based on the examples Trunz provided, we extracted and stemmed<sup>8</sup> a series of words connected to these four semantic areas, and we calculated their frequency in the three sub-sections of Goethe's works. Results are shown in [Fig. 5], confirming Trunz's intuition.

To expand this analysis to work contemporaneous with Goethe's, we used a software package that functions with the same logic as the previous experiment but does not depend upon categoriza-

8 Stemming is a procedure that automatically extracts word roots, making it possible to identify (most of) their morphological variations. We stemmed the wordlist using the R package *SnowballC*.

tions of a single literary critic. The LIWC software (Linguistic Inquiry and Word Count, see Tausczik, Pennebaker 2010), more generally adopted in psycholinguistics and sentiment analysis (Liu 2015), offered an opportunity to expand semantic areas to one hundred.<sup>9</sup> For an explanation of these areas, that are represented by sometimes cryptic labels, see the exemplification provided by Wolf et al. (2008):

LIWC Category	Label	Sample Lemmas
3rd person (general)	Other	sie, er, deren
Reference to others	Othref	deine, jemand, uns
Prepositions	Preps	als, bis, von
Present	Present	hilfst, isst, läuft
Pronoun	Pronoun	ich, wir, sie, dein
Social processes	Social	äußern, Begegnung, Kinder
Space	Space	abseits, breit, gegenüber
2nd person (general)	You	du, dein, dir
Article	Article	eine, das, dem
Communication	Comm	ablehnen, sprechen, Verhandlung

<sup>9</sup> LIWC's functional logic is even simpler than that of the other adopted software package. LIWC works on a series of multilingual dictionaries in which each word is connected to various semantic areas (such as "Social Words" and "Cognitive Processes," "Seeing" and "Hearing," "Space," and "Sexuality," to reach a maximum of over 100 categories). LIWC simply counts the words for each semantic area and calculates overall proportions in each text. The German dictionary is introduced and described by Wolf et al. (2008).

Ist Person (singular)	I	ich, mir, mein
Occupation	Occup	Schule, Arbeit, Leistung
Ist Person (general)	Self	ich, wir, mein
Affect Words	Affect	glücklich, hässlich, lächeln
Cognitive Mechanisms	Cogmech	abgrenzen, deshalb, wissen
Past	Past	gestern, hieß, sprach

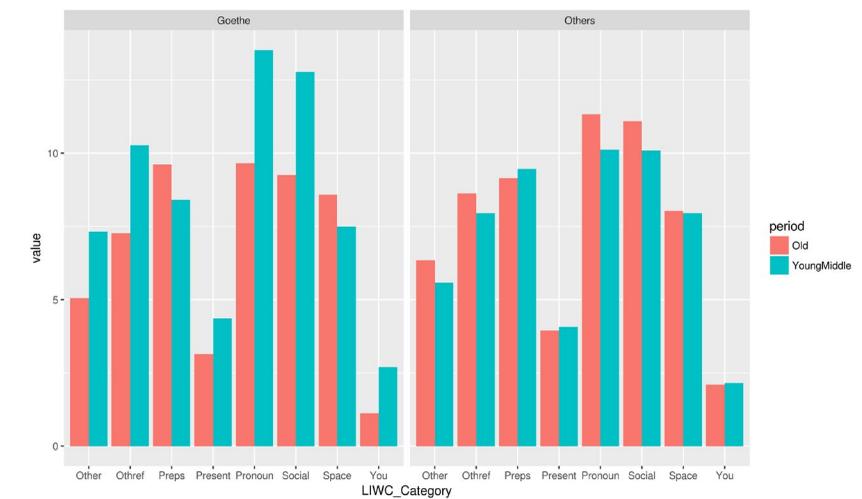


Fig. 6  
Frequency of most distinctive LIWC categories in Goethe's work and in the work of other authors published contemporaneously.

We structured our experiment in two phases. First, we isolated the eight LIWC categories in which the comparison between "Late" and "Young/Middle" Goethe showed the highest discrepancies; second, we calculated the frequencies of the same categories in the

work of other authors. Figure 6 shows that, in five of these categories, Goethe's path diverged diametrically from the approach taken by the authors of other contemporaneously published works. In the other three categories, the difference appears less significant, though the variations are consistent with those found in Goethe's work. Differently from previous methods, our analysis clearly demonstrated the "contradictory, alienated relationship," as Said termed it (2006: 13), between Goethe's "Late Style" and his environment. In addition, the sharp decrease in the "Social," "Other," and "You" areas confirmed the widely held interpretation that Goethe's late works were more abstract and detached.

### Modern case studies: Musil and Kafka

Once the methodology was consolidated, we were able to move towards the twentieth century and extend the analysis to other authors. One important consideration, however, is that we were not able to incorporate truly recent works. Because of copyright issues that interfered with the setup of the external analysis (for which an ample and representative *corpus* should be built), we limited our focus to the first decades of the twentieth century and selected two of the most important representatives of German literary modernism: Robert Musil (1880-1942) and Franz Kafka (1883-1924).

#### Musil

The Musil *corpus* was built thanks to the Klagenfurter digital edition (Amann, Corino, Fanta 2009), which provides a digitized *opera omnia* of the Austrian author. A total of twenty-three text samples were generated by applying a temporal subdivision suggested by the Klagenfurter edition: 1906-1917 for "Young" production; 1918-1927 for "Middle" production; and 1928-1942 for "Late/Old" production. Collections like *Nachlaß zu Lebzeiten (Posthumous Papers of a Living Author)*, which span periods, were segmented into sub-parts using the date of first publication of each text as a reference point. As for Musil's contemporaries, we selected the texts hosted by the KOLI-

MO online database (Herrmann, Lauer 2017). Out of the 42,694 entries in the database, 6,100 provided a date of first publication, 875 of which fell between 1906 and 1942. Once again, copyright limitations interfered with the representativeness of the *corpus*: 470 works are included in KOLIMO for the period 1906-1917, but they decrease to 130 between 1928 and 1942 (a sample that was nonetheless sufficient to run an extensive quantitative analysis).

### Network and Zeta analyses

Internal stylometric analysis (with the same features used for Goethe) produced results that even more strikingly challenged the supposed distinctiveness of Goethe's "Late Style." As is evident in [Fig. 7], the network graph was divided into two distinct clusters, and genre, rather than chronology, played the decisive role. Fiction was clearly separate from essays, regardless of period of publication. These findings were corroborated by external-network and external-Zeta analyses, in which chronology seemed to play an even more contradicting role. In the Zeta analysis, in fact, Goethe's "Early" production stood out most markedly from the work of contemporaneously published authors [Fig. 9]. However, this result was reversed in the network analysis, in which "Early" production was most highly integrated into the system [Fig. 8]. This divergence points to the fact that the two methods highlight different aspects of style, and their combination can actually strengthen results.

### Semantic analysis

Robert Musil is not traditionally mentioned in discussions of "Late Style" (Zanetti 2012: 109ss). Hans Blumenberg described Musil's unfinished novel *Der Mann ohne Eigenschaften (The Man Without Qualities)* as a book that was condemned to be an author's last book (Blumenberg 1997: 165) because it contained obstacles, created by the author himself, that kept it from being completed. Nevertheless, there is a quality in Musil's writings that is typical of "Late Style." The last chapter of the first part of *Der Mann ohne Eigenschaften (1930)*

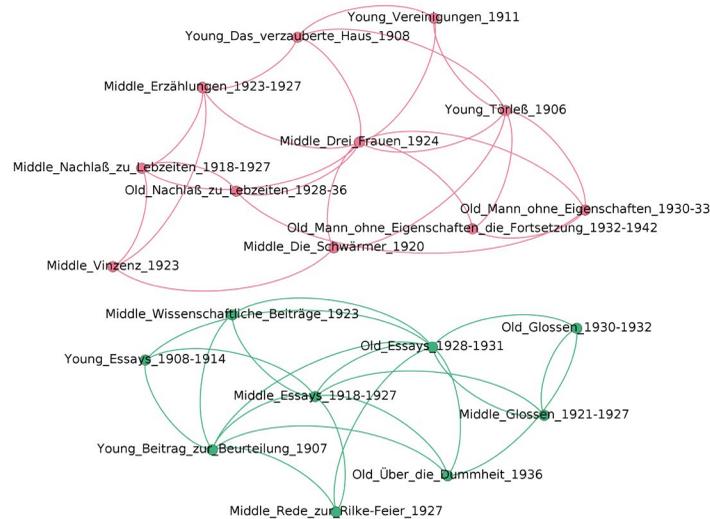


Fig. 7  
Network analysis of Musil's works.

is entitled "Die Umkehrung" (The Reversal), and Walter Fanta, the scholar who contributed fundamental studies on the genesis of the novel, defined the shift from the first to the second book of the novel as "the great passage" (Fanta 2000: 391). Claudia Monti has doubtless studied this passage more in detail than has any other researcher, recognizing two different characteristic styles in the first (1930) and second books (1932).

The first was characterized by what Musil called "the tree of violence," which was supplanted in the second by "the tree of love" (Monti 1995: 71). The "tree of violence," a lifestyle and intellectual approach inspired by Ernst Mach and Friedrich Nietzsche, was expressed through the aggressive intellectual behavior of the protagonist, Ulrich, who criticized such essential constructs of Occidental thinking as identity and causality. In contrast, "the tree of love," inspired by German romanticism, claimed universal analogy and con-

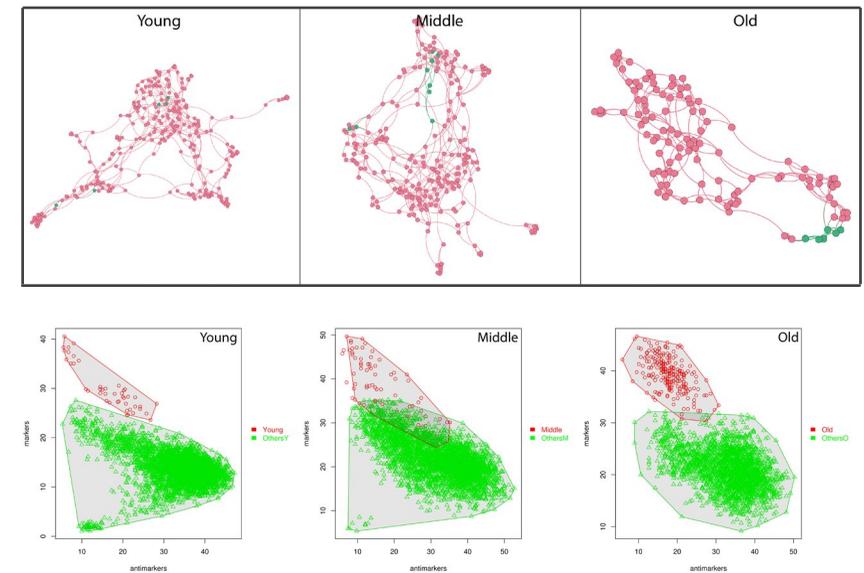


Fig. 8 (above)  
Network analyses of Musil's works (green) and their contemporaries (pink).  
Fig. 9 (below)  
Zeta analyses of Musil's works and their contemporaries.

nectivity and was expressed in the love between Ulrich and his sister Agathe. Both conditions could be seen in Musil's texts through descriptions of Ulrich who, under the sign of the "the tree of violence," was impassible and isolated, but was empathic and affectionate under the sign of "the tree of love." This emotional reversal, from a convex to a concave condition, was inspired by experiments on "optical inversion" by the Gestalt psychologist E. M. v. Hornbostel (Monti 1995).

Monti recognized the same schema in the short story *Die Amsel* (*The Blackbird*) (Monti 2000: 236ff), which led us to date Musil's "Late Style" to 1928, when that story was published. Following Monti's indications and quotations from *Der Mann ohne Eigenschaften* and *Die Amsel*, we created distinct semantic areas for "the tree of violence" and "the tree of love," the latter of which characterizes Musil's "Late Style":

The Tree of Violence	The Tree of Love
Convex	Concave
Violence	Love
Sharp	Soft
Narcissism	Altruism
Lack	Fullness
Apathetic, cold	Welcoming, warm, compassionate

Once again, quantitative analysis confirmed Monti's intuition [Fig. 10]. LIWC analysis did not produce results as significant as it had for Goethe, however. In the eight most distinctive categories in "Late" Musil, only two showed a slight counter-tendency compared to the contemporary context [Fig. 11]. This result has two possible interpretations: either late Musil did not actually fit into the more general category of "Late Style," or the method could not adequately capture the specific nature of "Late Style." To verify which option was most likely, we tested the approach on a third author:

## Kafka

### Semantic analysis

As a sample of qualitative analysis of Kafka's "Late Style," we adopted a 2013 monograph by Malte Kleinwort entitled *Der späte Kafka: Spätstil als Stilsuspension*. Kleinwort noted three main features of Kafka's "Late Style": references to earlier works, a peculiar form of tentativeness, and a poetics of asceticism (Kleinwort 2013: 10). While the majority of Kafka criticism dates the beginning of his "Late Style" to 1917, Kleinwort dated it to 1922 (ivi: 12). The main reason for this forward-dating was that, in 1921, Max Brod published his biography, *Adolf Schreiber: Ein Musikerschicksal*, which Kafka read en-

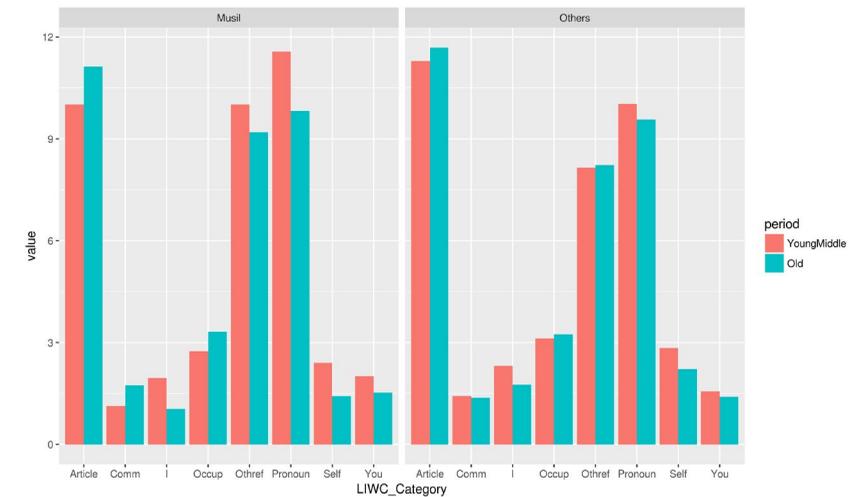
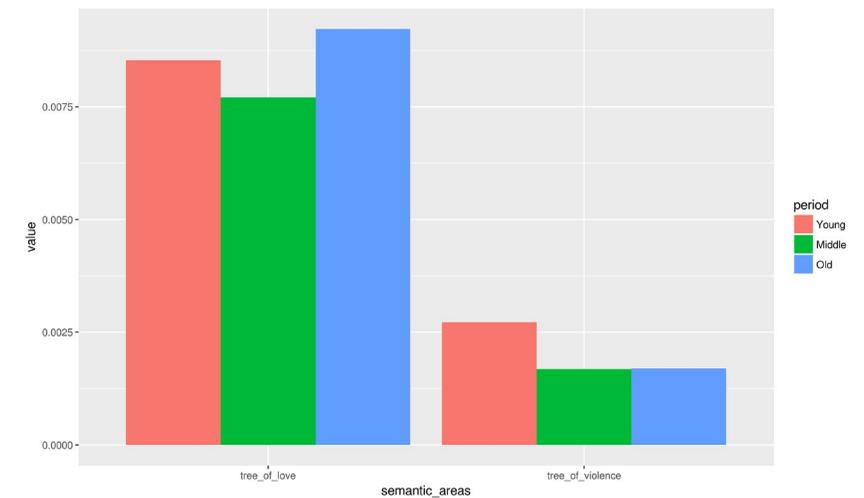


Fig. 10 (above)

Frequency of Monti's semantic areas in Musil.

Fig. 11 (below)

Frequency of most distinctive LIWC categories in Musil's works and in the work of authors published contemporaneously.

thusiastically and which had a pivotal impact on Kafka's late phase (ivi: 59).

In Kleinwort's view, the biography influenced Kafka in four ways, encouraging him to look retrospectively through his writings, to focus on the relationship between artist and impresario, to develop asceticism as a feeling for art, and to adopt reservation as a style of writing (ivi: 13). We can observe the poetics of reservation and asceticism in the loss of subjectivism and dramatization in Kafka's late works, including *Der Hungerkünstler* (*A Hunger Artist*) and *Josefine, die Sängerin oder Das Volk der Mäuse* (*Josephine the Singer, or the Mouse Folk*), in which Adolf Schreiber became the role model for the artist characters (ivi: 59). Not only did music become an important issue in Kafka's "Late Style," but so did its opposite – "the non-musical, the silence and the mere sound, noise or tone" (ivi: 62). Following Kleinwort's qualitative analysis, we determined the following semantic areas as typical of Kafka's "Late Style":

Semantic areas (as mentioned by Kleinwort)	Sample lemmas
The artist	Kunst, Künstlertum
Music	Musik, Konzert
The unmusical	Schweigen, Geräusch, Summen, Zischen, Rascheln, Pfeifen
Disrespect of the audience	Ungeschicklichkeit, Unfertigkeit
The poetics of asceticism	Schlichtheit, Einfachheit
Self-destruction of the artist	Selbstkritik, Selbstzerstörung
Loneliness in the metropolis	Einsamkeit, Fremdheit

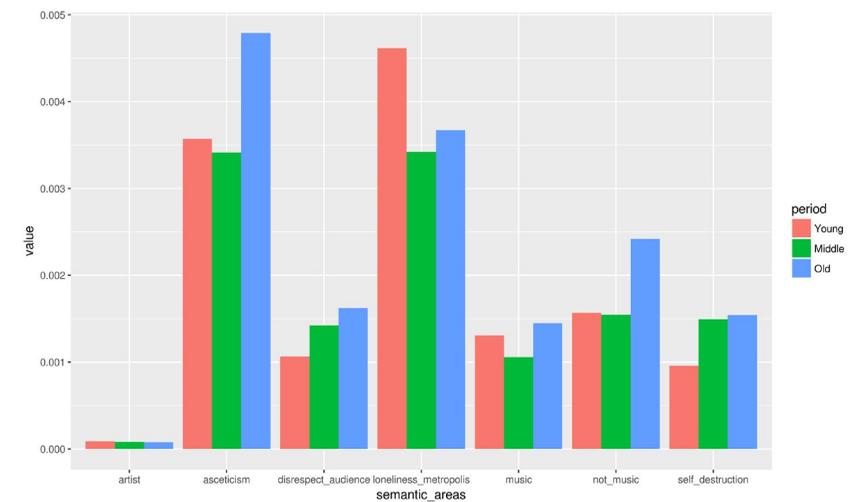


Fig. 12

Frequency of Kleinwort's semantic areas in Kafka's work.

We built the *corpus* for the analysis through the KOLIMO database: twenty texts by Kafka and 564 by other writers that appeared contemporaneously. Following indications by Manfred Engel and Bernd Auerochs (2010), we set the threshold between early and middle production to 1913. Figure 12 confirms, for the most part, Kleinwort's interpretation: the most striking matches were for the semantic areas of "asceticism" and "not music," while only "loneliness in the metropolis" showed a sharp decrease in "Late" production. Moreover, external LIWC analysis proved as efficient as it had for Goethe. Among the eight categories that showed the highest variance in Kafka, six were in counter-tendency to the characteristics of work published contemporaneously [Fig. 13]. Among these was an increase in cognitive activities (indicated by the label "cogmec"), which may be in line with the increased abstractness ("asceticism" in [Fig. 12]) of Kafka's late production. In conclusion, semantic analysis appears to be the most efficient method – though it is not infallible – for measuring "Late Style."

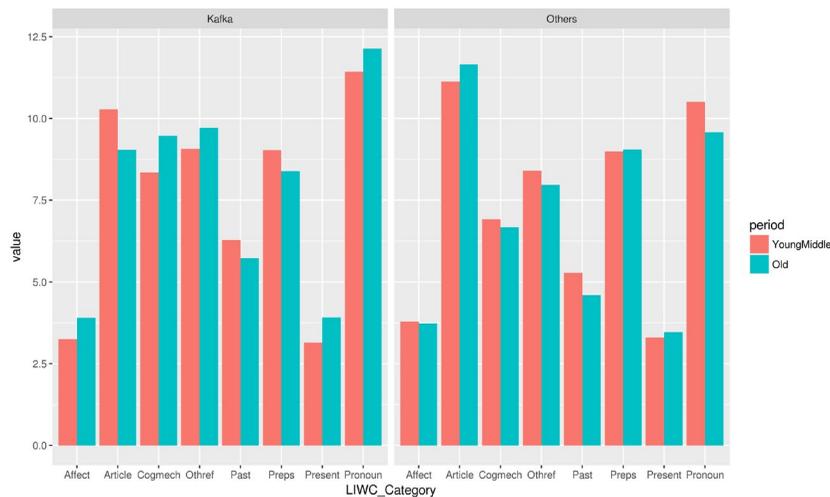


Fig. 13

Frequency of most distinctive LIWC categories in Kafka's work and in the work of other authors published contemporaneously.

## Network and Zeta analyses

The success of semantic analysis is counterbalanced by the failure of the other two, more purely “stylometric” approaches. Internal network analysis shows that the “Early” period – and not the “Late” – distinguished itself most strikingly from the others [Fig. 14]. This discovery was confirmed by external analysis in which Kafka's four early works were clearly isolated from the rest; in addition, while the “Middle” and “Late” works were not as strongly separated, they remained peripheral in the system [Fig. 15].

External Zeta analysis seemed most inconclusive because, in all three periods, no overlap existed between Kafka's work and work published contemporaneously by others; rather, these two categories occupied widely separated areas of the graph [Fig. 16]. This result confirmed the much discussed “uniqueness” of Kafka's style, however, but independent of period of composition, as Herrmann (2017) demonstrated with very similar methods.

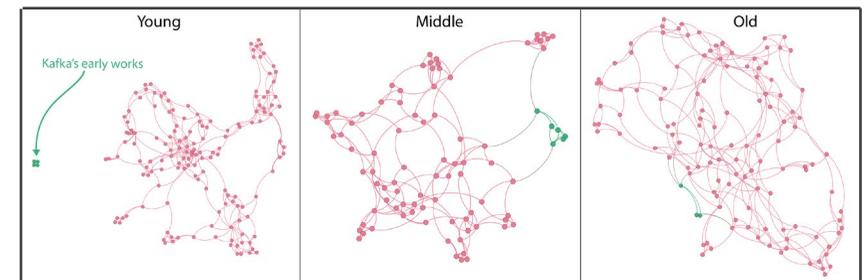
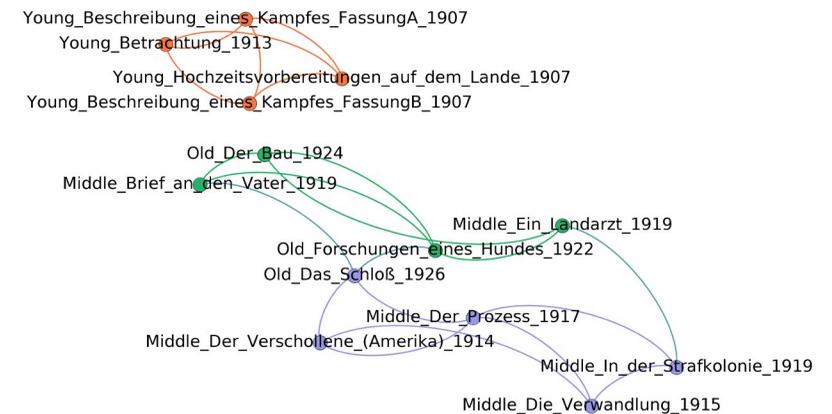


Fig. 14 (above)

Network analysis of Kafka's works.

Fig. 15 (below)

Network analyses of Kafka's works (green) and the work of other authors published contemporaneously (pink).

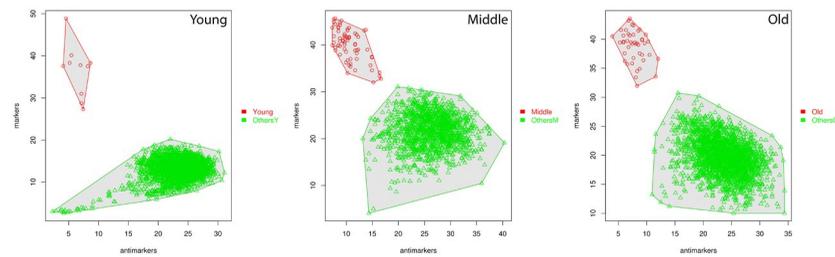


Fig. 16

Zeta analyses of Kafka's works and the work of other authors published contemporaneously.

## Conclusions

The aim of this essay was to combine two research methods, the qualitative analysis of literary criticism and the quantitative analysis of stylometry, to study the phenomenon of "Late Style." We choose three representative and important writers of German literature, Goethe, Musil, and Kafka, and analyzed their late works with different methods while comparing them both to their own early works and to the texts of their contemporaries. The results were as multifaceted as is the concept of style. The reasons for this manifold outcome may be several. On the level of theory, "Late Style" may not be a universal category, as other scholars have suggested (McMullan 2016: 36; Kastenbaum 1985: 252), but rather a tendency that can be observed in the study of single authors (Zanetti 2012: 204).

On the methodological level, the intriguing results of our research underscored the necessity of pairing quantitative and qualitative measures on style. "Late Style" seemed confirmed by our 'internal' studies on semantic areas and through LIWC measurements,<sup>10</sup> but it appeared unsupported by such traditional stylometric methods as network analysis and Zeta analysis:

<sup>10</sup> Note that the LIWC analysis, while adopting a typically 'distant' and 'unsupervised' perspective, necessitates of a confrontation with the analyzed author, in order to identify the distinctive categories. By consequence, it may be considered as a middle way between close and distant approaches in the study of literature.

	Network analysis		Zeta analysis		Semantic analysis	
	Internal	External	Internal	External	Internal	External
Goethe	●	●	●	●	●	●
Musil	●	●	●	●	●	●
Kafka	●	●	●	●	●	●

One reason could be that the phenomenon of "Late Style" is more evident on the semantic level<sup>11</sup> than on the vocabulary/syntactic level. In this perspective, "Late Style" could be the result of an author's decision to take a certain direction and to privilege specific topics more than it derived from the involuntary use of a certain vocabulary. These results may confirm Reeve's claim (2018) that "Late Style" is not measurable – at least, not through most frequent words. However, it should also be noted that many of the words identified by our semantic analyses are actually quite frequent words. What differs from stylometry, is the way in which these words are analyzed (by focusing on more general categorizations, instead of fine-grained frequency patterns). LIWC categories such as "Article", "Prep[osition]s", and "Pronoun" are what Pennebaker termed as "little words", which have a determinant role in identifying psychopathologies and aging (Pennebaker, Chung 2014: 25-26). It may be, in summary, that "Late Style" is more a natural consequence of aging, than the revolutionary stylistic break hypothesized by Broch and Said.

Such multifaceted results, while not allowing to formulate a definitive answer to the question whether "Late Style" actually exists, confirm that the synergy of literary hermeneutics and stylometry can open new and innovative perspectives on traditional research questions.

Another issue concerns the rupture inherent in the category of

<sup>11</sup> The green dots represent the methods that were successful in identifying late style; the red dots represent the ones that failed.

“Late Style.” What does it mean that Kafka differs more from other early-twentieth-century writers than does Musil, as Figures 16 and 9 seem to suggest? And how can we measure the rupture that many deviance theories of literary style conceptualize? (Herrmann, van Dalen-Oskam, Schöch 2015; Salgaro 2018) Once again, the result of our semantic analyses was revealing. Only in relation to a specific vocabulary, suggested in our experiments by literary critics and LIWC categories, we were able to find internal and external differences with reference to style. “Late Style” seemed to emerge only after a reference model was established and not in relation to the vocabulary as a whole. What is amazing, in any case, is that these software programs were able to grasp the “Kafkaesqueness” of the Prague-born author.

All in all, the important questions raised by our approach confirmed the necessity of revisiting the concept of style through stylometry. In this case, we employed a mixed method in which direct confrontation with critical theory and focus on single case studies was merged with DH techniques and distant reading. This approach could be expanded to other authors, including even contemporary ones, if appropriate *corpora* were available.

As for Adorno’s “late works are the catastrophes in the history of art” (Adorno 1937: 567), we can infer that they could have a calamitous effect in literary theory as well. As literary scholars, we want to interpret this “catastrophe” not as a disaster but rather as peripeteia, the final and most exciting part of a beautiful and enigmatic drama.

## BIBLIOGRAPHY

- ADORNO T.W. (1937), “‘Late Style’ in Beethoven”, in LEPPERT R. (Ed.) *Essays on Music by Theodor W. Adorno*, Berkeley, University of California Press, 2002, pp. 564-568.
- Id. (1957), “On the Social Situation of Music”, in LEPPERT R. (Ed.) *Essays on Music, by Theodor W. Adorno*, University of California Press, Berkeley, 2002, pp. 391-436.
- AMANN K., CORINO K., FANTA W. (2009), *Robert Musil, Klagenfurter Ausgabe. Kommentierte digitale Edition sämtlicher Werke, Briefe und nachgelassener Schriften*, Robert Musil-Institut der Universität Klagenfurt, Klagenfurt.
- BENN G. (1989) “Altern als Problem für Künstler”, in SCHUSTER G., HOF H. (eds.), *Gottfried Benn. Sämtliche Werke*, Band 6, Prosa 4, Klett, Stuttgart, pp. 123-151.
- BLUMENBERG H. (1997), “Letzte Bücher”, in *Marbacher Magazin*, LXXX, pp. 165-172.
- BROCH H. (1995), “Mythos und Altersstil”, in LÜTZELER P.M. (ed.) *Schriften zur Literatur 2. Theorie*, Suhrkamp, Frankfurt a. M., pp. 212-234.
- BRUNEAU C. (1951), “La stylistique”, in *Romance Philology* V: 1, pp. 1-14.
- BURROWS J. (2002), “‘Delta’: a Measure of Stylistic Difference and a Guide to Likely Authorship”, in *Literary and Linguistic Computing*, XVII: 3, pp. 267-287.
- CANETTIERI P. (2016), “Chi non ha scritto il Fiore”, in TONELLI N. (Ed.), *Sulle tracce del Fiore*, Le Lettere, Firenze, pp. 121-134.
- CRAIG H., KINNEY A. F. (2009), *Shakespeare, Computers, and the Mystery of Authorship*, Cambridge University Press, New York.
- EDER M. (2015), “Does size matter? Authorship attribution, small samples, big problem”, in *Digital Scholarship in the Humanities*, XXX: 2, pp. 167-182.
- Id. (2017a), “Visualization in stylometry: Cluster analysis using networks”, in *Digital Scholarship in the Humanities*, XXXII: 1, pp. 50-64.

- Id. (2017b), "Short samples in authorship attribution: a new approach", in *Digital Humanities 2017 Conference Abstracts*, ADHO, Montréal, pp. 221-224.
- Id., RYBICKI J., KESTEMONT M. (2016), "Stylometry with R: A Package for Computational Text Analysis", in *R Journal*, VIII: 1, pp. 107-121.
- ENGEL M., AUEROCHS B. (2010), *Kafka-Handbuch: Leben, Werk, Wirkung*, J.B. Metzler, Stuttgart.
- EVERT S., PROISL T., JANNIDIS F., REGER I., PIELSTRÖM S., SCHÖCH C., VITT T. (2017), "Understanding and explaining Delta measures for authorship attribution", in *Digital Scholarship in the Humanities*, XXXII: Supplement 2, pp. ii4-ii16.
- FANTA W. (2000), *Die Entstehungsgeschichte des "Mann ohne Eigenschaften" von Robert Musil*, Böhlau, Wien, Köln, Weimar.
- FRENZEL H. A., FRENZEL E. (1985), *Daten deutscher Dichtung: chronologischer Abriss der deutschen Literaturgeschichte*, Dt. Taschenbuch-Verlag, München.
- GRIMM J. (2010), *Rede über das Alter*, Steidl, Göttingen.
- HERRMANN J. B., LAUER G. (2017), "Das 'Was-bisher-geschah' von KOLIMO. Ein Update zum Korpus der literarischen Moderne", in *DHd 2017 Konferenzabstracts*, pp. 107-111.
- Id., VAN DALEN-OSKAM K., SCHÖCH C. (2015), "Revisiting Style, a Key Concept in Literary Studies", in *Journal of Literary Theory*, IX: 1, pp. 25-52.
- Id. (2017), "In a test bed with Kafka. Introducing a mixed-method approach to digital stylistics", in *Digital Humanities Quarterly*, XI: 4.
- HOLMES D. I. (1998), "The Evolution of Stylometry in Humanities Scholarship," in *Literary and Linguistic Computing*, XIII: 3, pp. 111-117.
- HOOVER D. L. (2013), "Almost all the way through. All at once", in *DH2013 Conference Abstracts*, pp. 223-226.
- Id. (2014), "Modes of composition in Henry James: Dictation, style, and what Maisie knew", in *The Henry James Review*, XXXV: 3, pp. 257-277.
- JANNIDIS F., LAUER G. (2014), "Burrows's Delta and Its Use in German Literary History", in ERLIN M., TATLOCK L. (eds.), *Distant*

- Readings: Topologies of German Culture in the Long Nineteenth Century*, Boydell & Brewer, New York, pp. 29-54.
- JOCKERS M. L. (2014), "A Novel Method for Detecting Plot", in *Matthew L. Jockers's Blog*, <http://www.matthewjockers.net/2014/06/05/a-novel-method-for-detecting-plot/> (last accessed 10 July 2018).
- JUOLA P. (2006), "Authorship attribution. Foundations and Trends", in *Information Retrieval*, I: 3, pp. 233-334.
- Id. (2015), "The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions", in *Digital Scholarship in the Humanities*, XXX, pp. 100-113.
- KASTENBAUM R. (1985), Program of the Thirty-Eighth Annual Scientific Meeting of the Gerontological Society of America, *Gerontologist XXV: special issue*, p. 252.
- KAYSER W. (1948), *Das sprachliche Kunstwerk. Eine Einführung in die Literaturwissenschaft*, Francke, Bern
- 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)*, pp. 59-66.
- KEY E. (1909), *The Century of the Child*, Putnam, New York.
- KLEINWORT M. (2013), *Der späte Kafka. Spätstil als Stilsuspension*, Fink, Paderborn.
- LEECH G. (2007), "New resources, or just better old ones? The Holy Grail of representativeness", in HUNDT M., NESSELHAUF N., BIEWER C. (eds.), *Corpus Linguistics and the Web*, Rodopi, Amsterdam/New York, pp. 133-149.
- LEEDER K. (2015), "Figuring Lateness in Modern German Culture", in *New German Critique*, XLII: 2, pp. 1-29.
- LEWY E. (1913), *Zur Sprache des alten Goethe. Ein Versuch zur Sprache des Einzelnen*, Imberg&Lefson, Berlin.
- LIU B. (2015), *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*, Cambridge University Press, New York.
- MCMULLAN G., SMILES S. (eds.) (2016), *"Late Style" and its Discontents: Essays in art, Literature, and Music*, Oxford University Press, Oxford.

- Id. (2016), "The 'Strangeness' of George Oppen", in MCMULLAN G., SMILES S. (eds.), *"Late Style" and its Discontents: Essays in art, Literature, and Music*, Oxford University Press, Oxford, pp. 31-48.
- MENDENHALL T. C. (1887), "The characteristic curves of composition", in *Science*, IX: 214, pp. 237-249.
- MONTI C. (1995), "Mancanza-pienezza. L'inversione percettiva di Musil", in *Cultura Tedesca*, III, pp. 71-84.
- Id. (2000), "L'altro lato dell'amore, Considerazioni su *Die Amsel* e *Der Mann ohne Eigenschaften*", in BUSCH W., BREUER I. (eds.), *Robert Musils «Die Amsel». Kritische Lektüren. Materialien aus dem Nachlass*, Sturzflüge/Edition Studienverlag, Innsbruck-Wien-Bozen, pp. 225-253.
- MORETTI F. (2013), *Distant reading*, Verso, London.
- MORTON A. Q. (1978), *Literary Detection: How to Prove Authorship and Fraud in Literature and Documents*, Scribner, New York.
- MOSTELLER F., WALLACE D. L. (1964), *Inference and Disputed Authorship: The Federalist*, Addison-Wesley, Reading Mass.
- PENNEBAKER J. W., CHUNG C. K. (2014), "Counting Little Words in Big Data. The Psychology of Individuals, Communities, Culture, and History", in FORGAS J. P., VINCZE O., LASZLO J. (eds.), *Social Cognition and Communication*, Taylor & Francis, New York, pp. 25-42.
- PIRRO M., ZENOBI L. (eds.) (2011), *Rappresentazioni della giovinezza nella letteratura tedesca*, Mimesis, Milano.
- REBORA S., HERRMANN J. B., LAUER G., SALGARO M. (2018), "Robert Musil, a War Journal, and Stylometry: Tackling the Issue of Short Texts in Authorship Attribution", in *Digital Scholarship in the Humanities*, advance article: <https://academic.oup.com/dsh/advance-article/doi/10.1093/llc/fqy055/5140082> (last accessed 22 October 2018).
- REEVE J. P. (2018), "Does 'Late Style' Exist? New Stylometric Approaches to Variation in Single-Author Corpora", in *DH2018 Book of Abstracts*, ADHO, Mexico City, pp. 478-481.
- RIFFATERRE M. (1971), *Essais de stylistique structurale*, Flammarion, Paris.

- RYBICKI J. (2017), "Reading Novels with Statistics: What Numbers of Words Tell Us about Authorship, Genre, or Chronology", in DOBELMAN J. A. (ed.), *Models and Reality: Festschrift For James Robert Thompson*, T&NO Company, Chicago, pp. 207-224.
- RYBICKI J., BIERNACKA-LICZYNAR K., WOŹNIAK M. (2018), "Polysystem Theory and Macroanalysis. A Case Study of Sienkiewicz in Italian", in *DH2018 Book of Abstracts*, ADHO, Mexico City, pp. 490-492.
- SAID E. W. (2006), *On "Late Style": Music and Literature Against the Grain*, Pantheon, New York.
- SALGARO M. (2018), "Historical Introduction to the special issue on Literariness", in *SSOL*, VIII: 1, Special issue on literariness, in press.
- SAMPAOLO G. (2009), "Il tardo Goethe", in COMETA M. (a cura di), *L'età classico-romantica. La cultura letteraria in Germania tra Settecento e Ottocento*, Laterza, Bari, pp. 87-99.
- SCHÖCH C., SCHLÖR D., ZEHE A., GEBHARD H., BECKER M., HOTHO A. (2018), "Burrows' Zeta: Exploring and Evaluating Variants and Parameters", in *DH2018 Book of Abstracts*, ADHO, Mexico City, pp. 274-278.
- SEIDLER M. (2010), *Figurenmodelle des Alters in der deutschsprachigen Gegenwartsliteratur*, Narr, Tübingen.
- SPITZER L. (1961), *Stilstudien. Erster Teil: Sprachstile. Zweiter Teil: Stilsprachen* [1928], Hueber, München.
- STAIGER E. (1955), *Die Kunst der Interpretation. Studien zur deutschen Literaturgeschichte*, Atlantis, Zürich.
- SZONDI P. (1967), "Über philologische Erkenntnis", in Id., *Hölderlin-Studien*, Suhrkamp, Frankfurt a.M., pp. 9-30.
- TAUSCZIK Y. R., PENNEBAKER J. W. (2010), "The psychological meaning of words: LIWC and computerized text analysis methods", in *Journal of Language and Social Psychology*, XXIX, pp. 24-54.
- TRUNZ E. (1990), *Ein Tag aus Goethes Leben*, Beck, München.
- TUZZI A., CORTELAZZO M. A. (2018), "What is Elena Ferrante? A comparative analysis of a secretive bestselling Italian writer", in *Digital Scholarship in the Humanities*, XXXIII: 3, pp. 685-702.

WOLF M., HORN A. B., MEHL M. R., HAUG S., PENNEBAKER J. W., KORDY. H. (2008), "Computergestützte quantitative Textanalyse: Äquivalenz und Robustheit der deutschen Version des Linguistic Inquiry and Word Count", in *Diagnostica* LIV: 2, pp. 85-98.

ZANETTI S. (2013), "Was zu tun bleibt", in *UZH magazin*, IV: 13, pp. 52-54.

Id. (2012), *Avantgardismus der Greise? Spätwerke und ihre Poetik*, Fink, Paderborn.

#### SITOGRAPHY

KOLIMO, <https://kolimo.uni-goettingen.de/index.html> (last visited 10 July 2018).

LateStyle, <https://github.com/SimoneRebora/LateStyle> (last visited 10 July 2018).

LIWC, <https://liwc.wpengine.com/> (last visited 10 July 2018).

Project Gutenberg-DE, <http://gutenberg.spiegel.de/> (last visited 10 July 2018).

SnowballC, <https://cran.r-project.org/web/packages/SnowballC/index.html> (last visited 10 July 2018).

Zeno Library, <http://www.zeno.org/> (last visited 10 July 2018).