

In a test bed with Kafka. Introducing a mixed-method approach to digital stylistics

J. Berenike Herrmann <berenike_dot_herrmann_at_unibas_dot_ch>, Basel University

Abstract

In the emerging field of digital style analysis, scholars have been dwelling on a perceived irreconcilability of “distant reading” [Moretti 2000] and “close reading” [Fish 2012] [Ransom 1937]. This situation is somewhat reminiscent of the “paradigm wars” in social science, where for years, one could either adhere to a quantitative or a qualitative mindset. Taking seriously both the hermeneutic and the empirical traditions of (digital) stylistic research, the present paper proposes “a mixed-methods” coalition of approaches. To establish links between the mixed-methods paradigm from Social Sciences [Creswell and Plano Clark 2007] and stylistic practices within Digital Humanities (DH), the present article discusses common methods of distant, close, and ‘scalable reading’ as well as a flexibly adjustable definition of style [Herrmann et al. 2015]. In the practical part of the paper, I report a ‘mixed-methods digital stylistics’ study on Franz Kafka’s prose. Scaling the degree of abstraction and contextuality of the data according to particular research questions, I combine (1) “quantitative hypothesis testing” (examining Kafka’s stylistic “uniqueness” by means of a stylometric measure); (2) “quantitative exploration” (analyzing the first hundred statistically overrepresented words in Kafka); and (3) “qualitative text analysis” (KWIC and close reading to investigate the functions of a particular style marker in the context of Kafka’s *The Judgment* [*Das Urteil*]). Generally, for digital stylistics, I propose (a) raising epistemological and methodological awareness within the field, and (b) framing the research within the mixed-method paradigm that in fact seems very well suited to DH.

Introduction

A whole new generation of scholars has recently been entering the world of digital stylistics, appreciating the availability of digital resources in the form of data, tools, and infrastructure. The recent proliferation is visible for example in monographs (e.g., [Hoover et al. 2014] [Jockers 2013] [Mahlberg 2013]), easy-to-access analysis tools (e.g., [Anthony 2014] [Sinclair and Rockwell 2016] [Eder et al. 2016]) and guides to text analysis in programming languages such as *R* and *Python* ([Bird et al. 2015] [Jockers 2014]), just to name a few. Contributions to the digital analysis of style abound in journals such as *DH Quarterly*, *Digital Scholarship in the Humanities* (formerly *Literary and Linguistic Computing*), the newly founded *Digital Literary Studies and Cultural Analytics*, the proceedings of the *Workshops on Computational Linguistics for Literature*, as well as in *Language and Literature* and the *International Journal of Corpus Linguistics*. Last, but not least, the Alliance of Digital Humanities Organizations (ADHO) has formed the new special interested group (SIG) *Digital Literary Stylistics* (SIG-DLS). This is an exciting time, with a discipline of digital stylistics in Digital Humanities emerging, expanding, and gradually maturing.

However, despite these recent developments, and despite a substantial tradition of digital analyses of literary style across fields (see [Biber 2011] [Burrows 1987] [Holmes 1998] [Mahlberg 2015] [Semino and Short 2004] [Steen et al. 2010]), “digital stylistics” as a discipline is still in a nascent status. This can be directly linked to its interdisciplinary character, with a kaleidoscope of methods, research paradigms, and world views. This heterogeneity is as much a potential source of confusion as well as a great asset:

On the one hand, the diversity may pose disadvantages. Those digital stylisticians who are trained in hermeneutics see themselves confronted with handling the requirements of an empirical quantitative research paradigm, with issues such as standardized research designs, procedures of data sampling, and analytical strategies, including statistics. Others, who come from fields based around computing, such as natural language processing (NLP), computational linguistics, and corpus linguistics, may be unsure what to do with the deliberate open-endedness of qualitative, hermeneutic research and the important role ascribed to subjectivity and context (cf. [van Zundert 2015]). Yet others may be very skeptical about the possibility of clearly defined ‘methods’ for digital stylistics in the first place, because of a perceived irreconcilability of numerical and hermeneutic approaches. Overall, a number of reasons accounts for a perceived lack of clear methodological standards, procedures and designs.

On the other hand, the juxtaposition of hermeneutic and scientific-empirical traditions provides a great potential for literary style research within DH. Qualitative approaches such as literary criticism and stylistics have developed a rich expertise in describing and understanding the aesthetic effects of literary artifacts. Quantitative approaches such as corpus linguistics, quantitative linguistics, computational linguistics, and data science generally, have concentrated on exploring and explaining large-scale and general language models by linguistic features. In digital stylistics, approaching aspects of style from different angles – be these probabilistically obtained feature patterns, predictive and descriptive accounts of quantitative style distributions, or in-depth interpretations of stylistic phenomena in particular textual contexts – opens up a truly multi-methodological research panorama.

In fact, digital stylistics, with its many branches and predecessors, has seen method-mixing for a long time (see, e.g., [Biber 1988] [Burrows 1987] [Craig 1999] [Leech and Short 2007] [Spitzer 1961]). Yet, we are presently dwelling on a perceived irreconcilability of a form of “distant reading” [Moretti 2000] on the one hand, and “close reading” [Ransom 1937] in a more aesthetically oriented variety [Fish 2012] on the other. Where distant reading applies aggregating measures that are designed to produce quantitative evidence with statistical backup, hermeneutic reading maintains a more qualitative relationship with (whole) texts, explicitly calling for open-ended and subjective dimensions of interpretation. Between these two

“paradigms”, many see close reading as “old-fashioned” for being anecdotal, overly subjective, or uncontrolled, whereas to others, distant reading appears “perilous” for being reductionist, deterministic, or neo-/postpositivist. This situation is reminiscent of the “paradigm wars” in social science, where for years, one could either adhere to a quantitative or a qualitative mindset (cf. [Creswell 2014] [Punch 2014]).

My suggestion for DH is not to fortify an “either-or”-mindset (possibly mistaking digital stylistics for a distant-reading only paradigm, cf. [Gooding 2013]), but to work on a principled coalition of the two views, propelled by a simple and pragmatic view of selecting the best fitting method for any given research question (cf. [Creswell 2014] [Creswell and Plano Clark 2007] [Symonds and Gorard 2010]). In line with the recently emerged mixed-method paradigm of the social sciences, I thus suggest that DH stylistics should move beyond the dichotomies of “close vs. distant”, “qualitative vs. quantitative”, “explanatory vs. exploratory”, “inductive vs. deductive”, “understanding vs. explaining” – and, possibly even “hermeneutic vs. empirical”^[1]. However, “moving beyond” here cannot possibly mean resorting to an “everything-goes” approach that defocuses the nitty-gritty details of difference, but an informed perspective that enables scholars to choose particular research designs by their particular objectives. Seen from the perspective of an emerging discipline of digital stylistics, the challenge is now to provide a common framework for the disparate approaches to stylistic inquiry. This includes a transparent and systematic overview of the different approaches and methods actually employed, as well as a sort of inventory for what kind of method is suitable for what kind of research aim.^[2]

That quantitative style metrics may combine with in-depth studies of style patterns of individual texts has already been proposed from within DH and labeled as “middle-distance reading” ([Craig 2013], for a similar approach, see also [Craig 1999]), or “scalable reading” ([Mueller 2012], see [Jannidis and Lauer 2014, 30–31], [Weitin 2017]). These propositions, however, have just begun to receive broader attention. In fact, many may have mistaken them for the examination of “mid-sized data sets” that are neither “big data” (for a recent discussion of this term, see [Borgman 2015]) nor individual texts, but something in between. In my view, this data-size-driven reading of the term is overly simplistic. Rather, the emphasis should lie not so much on the scale in terms of the *size* of the data volume examined, but on the *research process*. “Scaling” here means complementing one methodology with another, compensating for weaknesses of each approach, and building on the combined strengths. Mueller, in an analogy of using Google earth, describes the essence of scalable reading as to “zoom in and out of things and discover that different properties of phenomena are revealed by looking at them from different distances” [Mueller 2012]. For his account of middle-distance reading, Craig suggests comparing “statistical results [...] with fresh readings of the texts or parts of them, directed by the patterns that are highlighted by the tests” [Craig 2013]. Yet another approach is “rapid shuttling” between quantitatively attained information and hermeneutic close reading (Kirschenbaum 2009), cited in [Hayles 2012, 31]). Accordingly, Hayles states that the tension “between algorithmic analysis and hermeneutic close reading should not be overstated”, since their relation is often “configured not so much as an opposition but as a synergistic interaction” [Hayles 2012, 31].

Ideally, to provide “a better understanding of research problems than either approach alone” [Creswell and Plano Clark 2007, 5], mixed-methods research is a process of combining discrete qualitative and quantitative steps, motivated by some degree of methodological and epistemological awareness. It thus needs to address the role of interpretation^[3] and the degree of open-endedness and subjectivity involved *in either form* of research – issues that have received some attention also from within DH (e.g., [Gius and Jacke 2015] [McCarty 2005] [Ramsay 2011] [van Zundert 2015]). Mixed-method research highlights the hermeneutical dimension of quantitative approaches, not least by emphasizing that quantitative researchers are traditionally “in the background, and their own personal biases and interpretations are seldom discussed” [Creswell and Plano Clark 2007, 9].

“Quantitative”, “qualitative”, and “mixed-methods”

The most basic distinction between qualitative and quantitative methods concerns the nature of the data examined, which is normally numerical vs. non-numerical data [Creswell 2014] [Punch 2014]. For digital stylistics, this may be simplified as ‘*numbers derived from aggregations of words/texts*’ (quantitative) vs. ‘*elements of meaningful textual configurations*’ (qualitative). This discrimination relates to the more common distinction between big and small data sets. One of the most fruitful distinctions is that between *variable-oriented* and *case-oriented* analysis: variable-oriented analysis (e.g., “style variation across author, genre, time in German Modernist literature”) is more often quantitative, and well-suited for finding probabilistic relationships among variables in a large population, whereas case-oriented analysis (e.g., “paradoxical style of Franz Kafka’s *The Judgment*”) is more often qualitative, geared towards finding specific, concrete, historically grounded patterns common to small sets of cases [Punch 2014, 307]. “Variable-approaches” have difficulties with the complexities of causal relationships, while findings of “case-approaches” often remain particularistic, i.e. cannot be easily generalized.

Quantitative, variable-driven research overall strives for a high degree of control, using standardization and tests of reliability to ensure replicability and generalizability. It takes care that its measures and results are not dependent on a particular researcher (“objectivity”). In combination with representative data sampling, it thus aims at overall descriptions of situations and phenomena in a systematic and comparable way. In its stereotypical form, quantitative research uses “close-ended” data collection tools to “settle the discussion”, and to delimit the range of possible interpretation of its results.

By contrast, qualitative research is more sensitive to context and process, and aims for in-depth and holistic understanding, in order to do justice to the complexity of a given phenomenon. Samples are usually small, and sampling is guided by theoretical rather than probabilistic considerations. Methods are less formalized than those in the quantitative approach, less replicable, and have greater flexibility [Punch 2014, 307]. In its stereotypical form, qualitative research uses “open-ended” data collection designed to “opening up the discussion”, i.e. to generate new hypotheses and theories and widen the range of possible perspectives on a phenomenon.

This (simplified) distinction between qualitative and quantitative approaches is enlightening when considering the actual practices in digital stylistics. For example, a prominent approach, *algorithmic criticism* as proposed by Ramsay (2011), distinguishes between “scientific” approaches and what he calls algorithmic hermeneutics. In Ramsay’s view, the “scientific” approach is exceedingly closed-ended, aimed at providing “singular answers to the problem under discussion” [Ramsay 2011, 15], framing “the movement from data to interpretation as fraught with peril” [Ramsay 2011, 19]. By

contrast, an algorithmic criticism approaches a (textual) problem “in order that the matter become richer, deeper, and ever more complicated” [Ramsay 2011, 16]. His suggestion is to apply “hermeneutical freedom” [Ramsay 2011, 2] to the (post-algorithmic) interpretation of quantitative results – while retaining “the commitment to methodological rigor demanded by its tools” [Ramsay 2011, 17]. However, to Ramsay, interpretation is *necessarily* “an insistent subjective manner of engagement” [Ramsay 2011, 8] – he thus appears to deny digital text analysis the epistemological potential that lies in iterative (quantitative) hypothesis-testing. What is more, by putting such a strong emphasis on the role of post-hoc interpretation, he perilously defocuses the implicit assumptions that feed into the algorithmic modeling itself [Sculley and Pasanek 2008].

In my view, the scientific and humanistic paradigms as sketched by Ramsay should *both* contribute to our discipline's multi-methodological setup, establishing a rich interconnected network of possible research questions and methods for answering them. This, of course, has implications for the “nature of discourse in which text analysis bids participation” [Ramsay 2011, 9] – in my view, this discourse should involve deep and open-ended discussions as well as clear-cut conclusions about stylistic facts. Eventually, people engaged in DH stylistics should “generate more concepts, more theoretical constructions”, *and* be able to “test their points” [Van Peer et al. 2012, 7].

The interesting point about mixed-method approaches is that they maintain the basic distinction between quantitative and qualitative methods, while simultaneously transcending it. Thus, they firstly acknowledge the two paradigms, noting their strength and weaknesses, and then propose to:

[c]ombine the methods in a way that achieves complementary strengths and non-overlapping weaknesses. [Johnson and Onwuegbuzie 2004, 18]

With this general rationale at its core, mixed-methods research has been established as a third empirical paradigm^[4], with its own range of research design types (cf. [Creswell 2014] [Creswell and Plano Clark 2007] [Johnson and Onwuegbuzie 2004] [Tashakkori and Teddlie 2010]).

Mixed-methods Digital Style Analysis

The strengths of mixed-methods can be directly applied to digital stylistics: close and distant reading complement each other when adding meaning to numbers, and precision to hermeneutically obtained insight. Also, the generalizability of qualitative findings can be increased, as quantitative findings can be grounded in context-driven analyses. Results from both approaches can be validated through corroboration and convergence. Another great asset of such a coalition of approaches is mutual methods critique [Kelle 2008], which likely increases the validity of data and results as much as that of methods. And finally, a broader range of research questions can be asked, which leads to a varied and flexible generation of knowledge [Johnson and Onwuegbuzie 2004].

In digital stylistics, we are confronted with the fact that “style” is a heterogeneous category – already within literary studies, and ever more so when adopting interdisciplinary and international perspectives (for an overview of distinct philological definitions of style, see [Herrmann et al. 2015]). A definition that allows for a quantitative as well as for a qualitative perspective on style, and mixed approaches, is the following one:

Style is a property of texts constituted by an ensemble of formal features which can be observed quantitatively or qualitatively. [Herrmann et al. 2015, 44]

A quantitative, variable-driven approach may be interested in extracting formal features from a large data set devoid of context, for example gauging the “similarity” of authors by looking at the most frequently used words (the dominant stylometric measure), whereas a “case-driven” approach may be interested in aspects of the holistic gestalt of single texts, or the function of previously extracted features within that context. A third kind of approach may use a quantitative measure for exploring possible new variables (and thus, as a quantitative-heuristic approach may not fit the simplified quantitative-qualitative dichotomy).

The next section starts with a short (and by no means complete) overview of qualitative and quantitative methods relevant for digital stylistics, including a view on their respective strengths and limitations. It is followed by a sketch of my own mixed-method study, which will then be reported in more detail in section (3). Section (4) will subsequently draw conclusions for practices in digital stylistics as an emerging field within DH.

The Method of the Art

In actual stylistic research, the “ensemble of formal features” is as a rule viewed either from a predominantly quantitative or from a predominantly qualitative angle. With their predominant qualitative perspective, literary criticism, (digital) stylistics, aesthetics, and (digital) narratology have traditionally relied on versions of close reading. This mode of scrutinizing analysis and synthesizing interpretation of literary artifacts exploits the kind of situated and intuitive knowledge that human evolution has produced in concert with scholarly expertise (for a discussion of hermeneutics in DH, see [van Zundert 2015]). In stylistics, it has been most typically applied to: (a) explorations of the effect of certain stylistic features in textual interaction, and (b) determining the stylistic character of a piece of art as a coherent unit, including the contributing elements [Herrmann et al. 2015]. Goals of close reading are “linguistic observation” as much as “literary insight”, in the sense of Spitzer's hermeneutic circle [Leech and Short 2007, 11–12].

Applications of close reading to stylistic analysis are documented by a vast literature, with its typical concentration on particular cases (and, alas, sometimes vague generalizations). Variants of close reading, as a form of qualitative research, are applied also in digital text analysis, including computational linguistics/natural language processing (NLP). For example, digital style research employs a range of visualization and manual annotation tools^[5] for close reading, such as KWIC (*Keyword in Context*), which is a method typically applied in corpus stylistics. KWIC displays all instances of a search word vertically, with the words' direct textual context to both sides (also called *concordance*). Running KWIC lists (facilitated by tools such as AntConc and Voyant, as well as particular functions in R and Python) allows the discerning of meaning and possibly the lexicogrammatical function of a particular instance of a lemma in context (for an application in literary criticism, see e.g., [Müller 2013]). *Annotation*, depending on the basic methodological approach,^[6] may also involve some form of close reading. For quantification-oriented annotation paradigms,

this holds true especially for the stage in which annotation (or *tagging*) schemata are developed, with a typical back- and forth-movement between textual context, frequency operations, and theoretical conceptualization. In much the same way, the stages of “close reading”, or qualitative analysis are applied in the development of *machine learning algorithms* and *error analysis* in computational linguistics [Dipper 2008, 86–87].

In order to ensure some standard of compatibility and transparency also for close reading, stylisticians have established basic criteria for “the stylistic method” (e.g., [Simpson 2004]), with *rigorous* (“based on an explicit framework of analysis”), *retrievable* (“organised through explicit terms and criteria, the meanings of which are agreed upon by other students of stylistics”), and *replicable* (“sufficiently transparent as to allow other stylisticians to verify them, either by testing them on the same text or by applying them beyond that text” [Simpson 2004, 4]. Despite its vital role in humanistic inquiry, close reading, as any approach, has its limitations. Mentioned are the limited capacity in terms of data sets and time, the problems with generalization, the prominent role of the scholar’s introspection, and the fact that distinct features and factors are normally intermingled in complex *gestalts* and thus may be hard to isolate. Such issues may be discussed and compensated for in research that combines qualitative phases with quantitative ones.

For quantitative analysis, “style” needs to be formalized and reduced to one or few distinctive and machine-readable features (for a stylistic perspective on style markers, see [Leech and Short 2007]; for a stylometric one, see [Stamatatos 2009]). This applies to a potentially open list of linguistic and literary features (e.g., word frequency, length of words/sentences, vocabulary density, but possibly also more “literary” features, such as metaphor, or speech and thought representation). Quantitative research applies a range of different methods brought in from the fields of computational linguistics/natural language processing (NLP) and corpus linguistics. Simple “counting operations” (e.g., word frequencies, type-token-ratios, word and sentence length) are often followed by statistics such as the chi-square test and t-score, with prominent measures being collocation and keyword analysis [Biber 2011] [Gries 2015] [Hoover et al. 2014]. Statistically more advanced are multivariate methods that examine several variables at the same time (e.g., loglinear regression mainly used in corpus stylistics, as well as ANOVA, principal component analysis, cluster analysis, discriminant analysis, and factor analysis; (cf. [Biber and Reppen 2015] [Diversy et al. 2014] [Gries 2015] [Hoover et al. 2014] [Jockers 2013]; see also [Heyer et al. 2006]). One of the currently most popular techniques of quantitative stylistics is cluster analysis of stylistic similarity (e.g., using Burrows’s Delta; [Burrows 2002]). This kind of analysis operates on the basis of (standardized) measures of the most common features within and across texts [Eder et al. 2016] [Hoover 2004] [Jockers 2013]. A typical objective is to quantitatively scrutinize a particular period for its style, disambiguating an “author effect” from other (hidden) kinds of effects, such as “genre”, and “time.” To this aim, a viable way is to pick this style indicator and generalize over large sets of literary data (see [Eder 2017]; also [Jockers 2013, 70–77]). In stylometry, using supervised machine learning to identify the “strength of the stylistic signal” currently probably comes closest to the core of the prototypical quantitative paradigm, ruling out confounding factors and establishing clear cause-effect relationships (here, the cause would be one or more of the examined factors author, genre, time; the effect is the “stylistic signal” observed in most common words).^[7]

However, quantitative methods (such as factor analysis, multi-dimensional scaling, principal component analysis, and so on) are in fact often used in a more open-ended and exploratory way, something traditionally ascribed to qualitative research. Quantitative stylistic exploration may find new features and relations between known factors, for example between canonicity and writing style. For example, Algee-Hewitt et al. compared what they call the “archive”, that part of ever-published fiction that is now available in digitized form, to canonical sets of novels. Among their findings is that “archive texts” in comparison with canonical ones are stylistically more similar to a written mode of communication, even in their “spoken passages” [Algee-Hewitt et al. 2016, 2, 10]. This finding was not made in answer to a specific hypothesis, but is a valid one. Similarly, the growing number of NLP-adaptations to the literary domain (including neural networks, machine learning, and advanced modes of data mining, cf. [Hoover 2013]; see also the proceedings of the workshops Computational Linguistics for Literature^[8]) is typically used for exploration – hypothesis testing in the strict sense is (still) less common. NLP methods, as well as those of data science^[9] that crunch “big data” to aggregated variables (e.g., [Manovich 2015]), are able to observe new patterns and correlations, allowing for a multitude of possible factors to be examined.

However, regardless of being used exploratorily or explanatorily, with regard to the questions of literary stylistics, at the present point in time, the quantitative types of analysis often appear superficial. Features are necessarily restricted to those than can be reduced to a machine-readable format and used in automatic analysis. Quantitative stylistics, despite a tendency to adapt more complex models, is as a rule decontextualized, operates at the variable level, and is reduced to observing particular types of features. Under certain conditions, precisely these characteristics may foster progress. However, in order to attain a more complete degree of linguistic and literary insight for “style”, quantitative examinations need to be complemented by qualitative ones.

To flesh out these thoughts, in the following I will report on my own mixed-methods study on Franz Kafka’s prose in Newer and Modern German Literature (corpus of about 5.9 million words).

Making the test bed

From the reservoir of digital methods on the qualitative–quantitative scale, I selected three easy-to-access, yet established ones. With style operationalized as an ensemble of formal features (see above), I reconstructed a hypothesis about Kafka’s style from literary studies, which was then examined in a complementary successive analysis:

- *Hypothesis Generation from Literary Criticism.* I distilled from literary criticism converging evidence for a relatively broad assumption about Kafka’s style, the “solitariness”-hypothesis (see below). Finding convincing observations and arguments from the existing literature is thus a necessary step in hypothesis generation. This view has been put forward on the basis of qualitative, hermeneutic interpretations of separate texts (involving the description of linguistic features and literary effects). The kind of evidence given so far offers no firm ground for generalization.
- *Quantitative Hypothesis Testing.* My first empirical step was quantitative hypothesis testing. Extending on prior research [Herrmann

2013a], I examined the “solitariness”-hypothesis on a wider data basis (comparing Kafka’s texts to texts by more than 20 other German authors across time).^[10] For the sake of quantitative breadth, I relinquished the rich complexity of case-driven analysis, using as an indicator of “style” a standardized aggregate measure of most frequent words, *Delta* (see below) and cluster analysis. A problem with this type of analysis is that it does not provide insights into the particular features that may be responsible for Kafka’s (assumedly) distinctive style.

- **Quantitative Exploration.** The second analysis applied another frequency-based measure, *keyness*, to explore the stylistic patterns possibly underlying the results from the cluster analysis. Keeping the time signal relatively constant, I compared Kafka’s word use to four authors belonging to the same literary epoch, Modernism. Style was here identified as a statistically defined “norm” (the numerical reference being the same corpus used in the first analysis), treating word forms that are significantly more frequent than that norm as “stylistically marked”. Building on prior research [Herrmann 2013b], I manually grouped the top 100 overrepresented word forms into (potential) word classes. A category of word forms emerged as particular to the Kafka sample that may be used for subtly managing modality. However, even though this analysis tackles style at the feature level, it is still strongly decontextualized. More context was needed to add value and insight to my observations.
- **Qualitative Text Analysis.** In a qualitative close reading of one of Kafka’s stories (*The Judgment*) taken from the same corpus as before, I focused on just two potential modal particles derived from the previous step. Using *KWIC* (*Key Word in Context*) to systematically analyze their local usage, taking into account the context of Kafka’s story (including the plot), I examined their use and potential functions throughout the unfolding story [Herrmann 2016]. From here, further analysis may take up several possible leads given by the results. These may involve a larger data set and increased statistical rigor – as well as the qualitative detail of other texts and features (modal particles, negations, and adverbs).

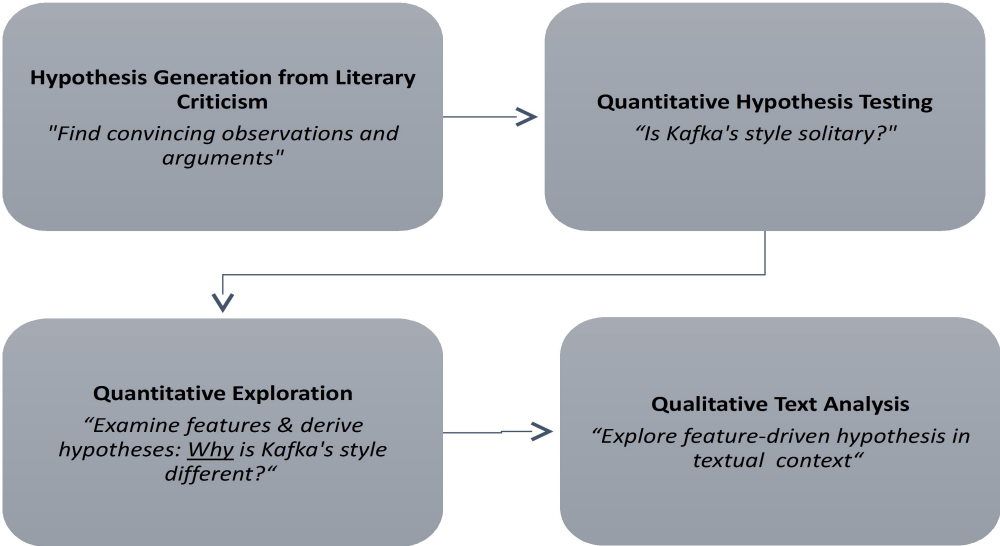


Figure 1. Mixed-Methods Style Analysis of Kafka

Applying Mixed-Methods to Kafka

As described above, a series of interlaced stylistic studies was conducted, applying different methods tailored to distinct questions on aspects of Franz Kafka’s writing style.

26

Hypothesis Generation from Literary Criticism

In Kafka studies, the most basic tenet on Kafka’s style is that it is not only unique, but even “solitary”: Scholars converge in finding that among his generation, nobody else writes like him (cf. e.g., [D’Haen et al. 2012] [Frenzel and Frenzel 1990] [Oschmann 2010] [Trost 2008]). The proposed solitariness has been related to a special lexical precision and “scantiness”, features of “elevated everyday language” and technical language, together with an accumulation of modal adverbs, a hypotactic sentence structure, and a frequent use of inner monologue and especially free indirect discourse. The narrative effect of his language use has been stated as the “objective” registration of *external* events, accompanied by a simultaneous limited *internal* perspective of literary characters (e.g., [Engel 2010] [Oschmann 2010] [Scheffel 2002] [Trost 2008]). The adjective *kafkaesque* generally means “complicated, confusing, and threatening” [kafkaesque], and this label corresponds with a consensus on the irritation that Kafka’s prose instills in readers, be they naïve or professional: “Each sentence says ‘interpret me’, and none will permit it” [Adorno 1981, 245].^[11]

27

Study 1: Quantitative Hypothesis Testing

—

For my first study, I selected an indicator of style that is firmly established within stylometry: the most frequent words (mfw) of a text and text set. The (relative) frequencies of the most common words per text in a given text set are used to compute a distance measure, which determines the stylistic similarity/dissimilarity of each text from each other text. I used a version of Burrows's Delta ([Burrows 2002] [Eder et al. 2016] [Hoover 2004]), which standardizes the mean word frequency (by calculating the z-score) for each text and then computes the "mean of the absolute differences between the z-scores for a set of word-variables in a given text-group and the z-scores for the same set of word-variables in a target text" [Burrows 2002, 271].^[12] In a subsequent cluster-analysis, which employed the ward-linkage algorithm [Eder et al. 2016], the large matrix of distances scores between individual texts was reduced to hierarchically aligned groupings of texts that are visualized, the so-called clusters. I chose a distance measure with cluster analysis because it is a robust measure of text similarity and has been shown to reliably identify signals of author, genre, and time [Eder et al. 2016] [Jockers 2013], also in German literary texts [Jannidis and Lauer 2014].^[13]

My hypotheses were as follows:

- H1: Kafka's texts are *more similar to each other* than they are to other texts, and, as a group, are by comparison more *dissimilar from other (groups of) texts*. For the cluster analysis, this means two things: (a) a similarity-cluster can be observed for Kafka's texts; (b) the observed Kafka-cluster shows distances from other (potential) clusters that are greater than any other distances observed.
- H0: Kafka's texts *do not show a greater similarity to each other* than to any other text. In the cluster analysis, Kafka's texts do not cluster with each other. However, as authorship is a strong predictor of style, clustering by author is generally expected.^[14] Therefore, H0 may be reformulated: Kafka's texts as a potential group of texts related by similarity *do not show greater dissimilarity* to other groups of texts by comparison. In the cluster analysis, any Kafka-cluster of texts thus has a similar distance to other potentially observed author-clusters as these have from each other.

When considering these hypotheses before the background of the chosen method, a caveat is due: in this application, cluster analysis does *not allow statistical hypothesis testing in a strict sense*. In other words, there is no way of determining the probability of falsely rejecting the H_0 (type I error, p -value) or the probability of rejecting a true hypothesis as incorrect (type II error, the failure to reject a false null hypothesis, a “false negative”, i.e., failing to detect an effect that is present). This means I cannot determine causal or even correlational relations between variables (authorship as influencing style as measured by mfw).

I used *cluster analysis* and its iterative version, *bootstrap consensus trees*, applying Eder's Delta as facilitated by the "stylo package" for R [Eder et al. 2016]. My corpus consisted of 64 German prose texts available for the stylo suite of tools^[15], three texts by Robert Walser (to whom Kafka has been traditionally compared), and 27 texts of different length written by Kafka^[16] compiled to seven distinct text samples. The final corpus amounts to altogether 74 text samples (about 5.9 million words). In the analysis, pronouns were deleted to avoid bias for narrative perspective.

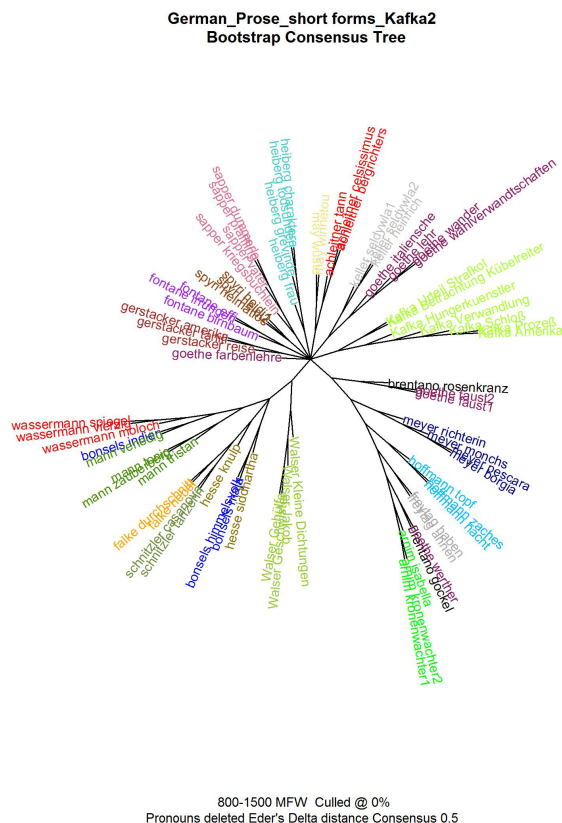


Figure 2. Cluster Analysis using Eder's Delta. Bootstrap Consensus Tree

Figure 2 shows a *bootstrap consensus tree*. It is the result of an iterative cluster analysis, keeping constant all parameter settings but the number of

features [Eder 2017]. Here the selected feature is mfw, the most frequent words shared by the included texts. For the consensus tree, the upper limit of the feature vector was set to 1,500 mfw, the lower one to 800 mfw.^[17] Figure 2 shows that Kafka's texts indeed form a cluster (the light green cluster to the upper right). It also depicts a number of branches that join texts by distinct authors (thereby demonstrating style similarity), most pronouncedly the two main branches in the lower half (both with subbranches of more than seven authors), as well as two smaller branches each joining two authors (Keller/Goethe and May/Achleitner). Most of Kafka's direct contemporaries are thus joined by the big branch at the lower left (Bonsels, Wassermann, Falke, Hesse, Mann, Schnitzler), while the branch at the lower right suggest similarities between older texts (Goethe, Brentano, Arnim, Hoffmann) and the Swiss (historical) realist Meyer.

In opposition to authors in both these groups, Kafka's cluster starts directly from the center, which means that Kafka's texts are more similar to each other than to other authors' texts. However, the same is true for Gerstacker, Heiberg, Sapper, Fontane, and Spyri. Note that the consensus tree is not informative about the distance of any two branches starting from the root. Hence, the first cluster analysis does not falsify the assumption that Kafka writes in a solitary way. However, despite being a relatively robust measure of stylistic similarity (incorporating repeated runs), the consensus tree provides rather limited information.

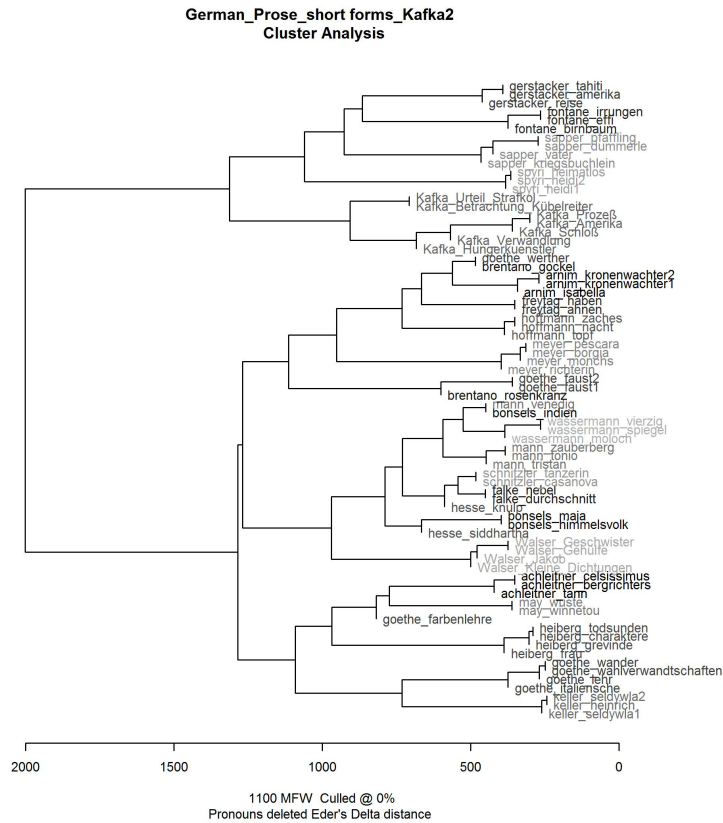


Figure 3. Cluster Analysis using Eder's Delta. Dendrograms (1100 MFW)

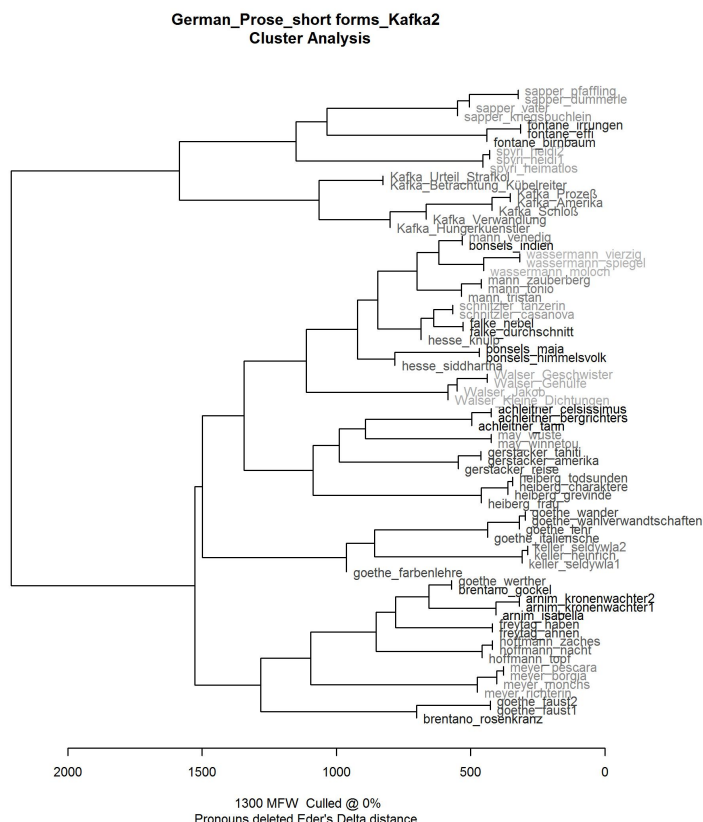


Figure 4. Cluster Analysis using Eder's Delta. Dendrograms (1300 MFW)

Figures 3 and 4 are more informative with regard to the actual similarities between texts than Figure 2. These *dendrograms* (two of the eight separate dendrograms underlying the consensus tree of Fig. 2) are “snapshots” that reflect the particular clustering of texts with feature-parameter set uniquely to 1,100 mfw and to 1,300 mfw, respectively. Both depict a situation in which Kafka’s texts indeed form a cluster, which means that they are more similar to each other than to any other text in the study. Also, Kafka’s texts (as a group) are at a relatively long distance from the nearest neighbors, Spyri, Fontane, Sapper, and Gerstacker (Fig. 3), and Spyri, Fontane, and Sapper (Fig. 4). Additionally, forming a group with the four/three nearest neighbors, Kafka is removed from the rest of the corpus.

Like Fig. 2, despite slight differences, Fig. 3 and Fig. 4 do not falsify the assumption that Kafka writes in a solitary manner. However, we need to take into account the company in this “outlier group”, two children’s books authors (Johanna Spyri, author of *Heidi*, and Agnes Sapper) as well as Germany’s chief proponent of poetic realism (Theodor Fontane), and an author of popular adventure and travel prose (Friedrich Gerstacker). In opposition to Kafka, the other four were immensely popular during their own and Kafka’s lifetime. Without making too strong assumptions on the basis of the two particular dendrograms^[18], these results may be used as a heuristic for furthering the stylistic research into Kafka’s prose: It is common ground in literary studies that Kafka achieved the commonly attested “refusal of making sense” (“Sinnverweigerung”), “with the techniques of the European tradition of narration” (“mit den Mitteln der europaischen Erzahltradition”). In Fig. 3 and 4 this tradition appears to show itself in the form of the nearest neighbors, all of which allegedly use a “realistic” writing style (with a great amount of concrete detail and vividly rendered descriptions, as well as dialogue). Further study needs to establish whether there really is a stylistic similarity with these authors (and among them), and, if so, what features contribute. What is more, Kafka research has shown that he liked reading books for younger readers, as well as (popular) adventure literature and journey descriptions in general (e.g., [Blank 2001]). There is hence some reason to assume that the nearest neighbors may reflect some of Kafka’s own reading habits [Herrmann and Lauer 2016]. This kind of speculation merits further pursuing.

This is where the analysis of this particular quantitative measure stops, letting the aggregated data speak for itself. However, in the present context, it is not really much it can say, despite a relatively weak corroboration of the reported hypothesis, which originates from rich hermeneutic observations. One problem with the measure is here that we do not know which particular textual features are responsible for the dissimilarity (or similarity) of styles: it computes the relative stylistic differences between a whole range of texts, based on a ‘bag-of-words’ approach that has to take *a priori* decisions about the number of mfw, the linkage algorithms, and other parameters. This combination of “black box” and “cherry picking” problems (for a critical overview, see [Eder 2017]) is rather frustrating, especially to the philologist, who would like to pin down the specific textual features that make Kafka unique (for quantitative approaches to features, see for example [Jockers 2013] [Klaussner et al. 2015]). The next step is applying a so-called *keyness* analysis, which allows for an exploration of the features responsible for stylistic differences between distinct texts and text collections.

Study 2: Quantitative Exploration

To pry open the black box of the Delta measure to examine the features that actually make a difference in an author’s style, I selected *keyness* analysis [Herrmann 2013b]. This established measure of corpus stylistics compares the frequencies of single words (or part-of-speech tags, or

semantic tags) included in some text (collection) with those obtained in a (normally larger) reference corpus. It outputs a long list of words that deviate statistically from that reference corpus [Rayson 2012] [Scott and Tribble 2006]. Here, the reference corpus acts as a statistical “norm” against which the word use in the text(s) under scrutiny may be compared. The examined words, depending on whether they deviate positively or negatively, are thus “over-” or “under-represented” with regard to that norm. Keyness analysis traditionally applies a Log-likelihood (LL) measure of difference in word usage (which is here more reliable than chi-squared analysis, cf. [Rayson 2012, 2]).

In order to ensure internal validity across sub-studies, I used the same corpus as in the cluster analysis. The same Kafka samples included in the first study thus formed a “Kafka corpus”, and the remaining 67 texts samples now worked as a reference corpus. In addition to the Kafka corpus, I compiled four new author corpora, extracted from TextGrid: Peter Altenberg, Gustav Meyrink, Arthur Schnitzler, and Georg Trakl (see Table 1 for a summary of the author and reference corpora). In order to control as many factors as possible, the time signal was kept relatively constant: authors were selected from the literary epoch Modernism (roughly 1880-1930). Also, all authors are male, and ideolect was controlled: the four authors share with Kafka a place at a “southern German language continuum” [Becher et al. 2012] [Nekula 2003]. However, given the particularities of Kafka’s (many very short texts and his particular technique of rendering narrative perspective), I deliberately did not control genre, and included drama (Schnitzler’s dramatic work, with its high proportion of fictional dialogue), poetry (Trakl’s poems with their highly subjective and sensual perspective), and (short) experimental prose (Altenberg’s “narrative poems” with a “disengaged” perspective, impressionistic description of sensual perception, and usage of colloquialisms). Through this, I hoped to allow for emerging stylistic patterns particular to Kafka.

38

Epoch	Author	No. of words
Modernism	Kafka	~308,400
	Altenberg	~245,200
	Trakl	~21,300
	Meyrink	~156,500
	Schnitzler	~1,037,000
Newer German Literature	22 authors	~5,873,100

Table 1.

Using the software tool AntConc [Anthony2014] the analysis rendered word lists that are significantly more common in each of the respective target corpora than in the reference corpus “Newer German Literature” (about 5.9 million words).^[19] I decided to examine these lists for word classes (part-of-speech, POS) in order to observe not content (as a tendency indicated by proper names and common nouns), but rather stylistic patterns. According to authorship attribution studies, the latter may be observed especially among functional word classes. My research questions were:

39

- What differences between the five authors are striking among the top 100 keywords?
- Which word classes are particularly common in Kafka when compared to the other authors?

The distribution of word classes among the top 100 keywords in each author-corpus rendered a number of interesting findings (see also [Herrmann 2013b]). For the current purpose, however, the most interesting one is this: Among the words most strongly overrepresented in Kafka’s prose, a variety falls into a category of potential “modal words”, which I operationally defined as word forms that (a) possess few semantic features when decontextualized; and (b) are typically used as adverbs, negations, and modal particles (but may have functions as conjunctions or other types of particles). For reasons of focus, I excluded (modal) verbs from the analysis. The observed types (e.g., *aber*, *nicht*, *allerdings*, *vielleicht*, *schon*, *ja*, *auch*) are semantically very flexible, potentially exerting distinct functions across different co(n)texts, e.g., to manage doubt/certainty, limit/broaden scope of assertions, and for negative/positive evaluation.

40

Proportion of words indicating possible modality

among top 100 key words (types)

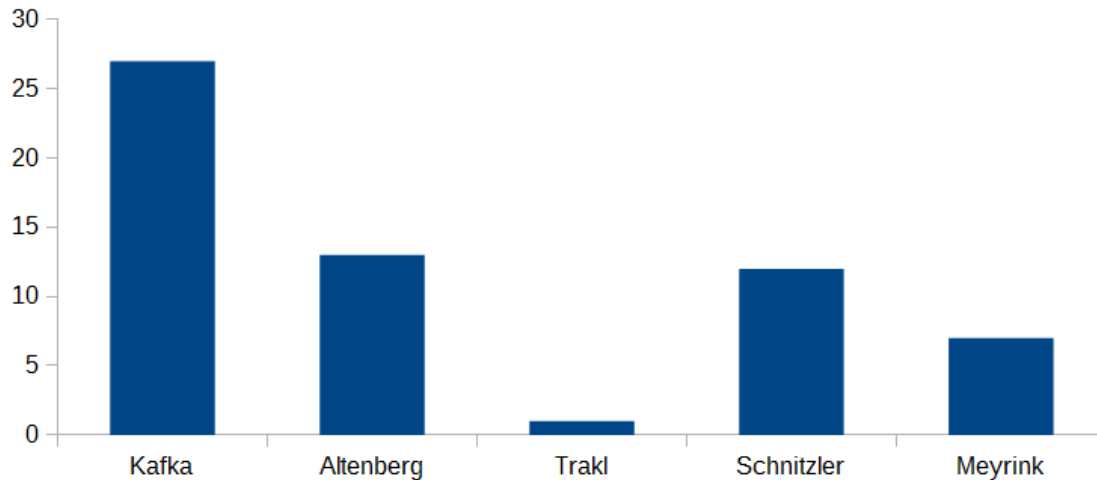


Figure 5. Number of “modal words” identified in the keyness analysis (Kafka, Altenberg, Trakl, Schnitzler, Meyrink)

Figure 5 shows the number of “modal words” identified in the keyness analysis for each of the five authors: The distinct types show up among the 100 statistically overrepresented word-forms (e.g., the types *ja*, *doch*, *schon* amounting to three counts). Fig. 5 reports that Kafka’s usage of this category is more substantial (N=27) than Altenberg’s (N=13)^[20], Schnitzler’s (N=12), Meyrink’s (N=7), and Trakl’s (N=1). Given the reference corpus, Kafka’s prose is hence marked by a more varied use of “modal words”.^[21] Specifically in their use as modal particles, words like *ja*, *schon*, *doch*, and *nur* may exert very subtle, but important, pragmatic functions of negotiating shared knowledge and beliefs. Note that they belong among the most frequent words of the German language generally, but are still overrepresented in Kafka’s prose – at least when compared with our reference corpus of literary texts. They may thus be a key to explaining (aspects of) Kafka’s subtly disconcerting style. However, more analysis is needed to further test this assumption.

To sum up, the application of keyness analysis to Kafka and other Modernist authors has pointed to a quantitative style pattern in Kafka’s prose, a high frequency of lemmas that may perform “modal” functions in the discourse. While it is clear that more research is required – for example varying reference corpora with principled attention to the variable “genre” (the used corpora contain narrative texts as well as poems and dramatic texts), and using more author-corpora – the analysis has shown that an exploration of quantitative patterns in terms of single words is quite useful. Indeed, it provides a first take at answering the question of what underlies the clustering observed in the *Quantitative Hypothesis Testing*. It could not falsify the philological finding that Kafka’s style is defined by a particular use of “modal adverbs” (e.g., [Trost 2008]). However, the identified category of lemmas needs disambiguation as to their grammatical categories and local functions in textual passages. We thus need more context.

Study 3: Qualitative Text Analysis

The third step of my project breaks down the complexity of the broad list of “modal”-type words run up by the keyword list, by zooming in on just two lemmas, namely, *ja* and *doch* [Herrmann 2016]. As modal particles, both have been well described in the linguistic literature and belong to the most frequent German modal particles [Hentschel 1986]. Generally, modal particles cater to intentionally attaching an utterance to the conversational frame, structure the common knowledge base of the interlocutors by indicating common ground, give prompts as to the felicitous interpretation of utterances in the given context, indicate subjective stance of speakers, and establish anaphoric reference with prior discourse.

chaftlichen Beziehungen wiederaufzunehmen – wofür **ja** kein Hindernis bestand – und im übrigen auf überhaupt nach Hause zu bringen – er sagte **ja** selbst, daß er die Verhältnisse in der zurückfahren. Allein – weiß du, was das ist?« »**Ja**, kann er denn von unserer Heirat nicht ättest du dich überhaupt nicht verloben sollen.« »**Ja**, das ist unser beider Schuld; aber ich Riese«, sagte sich Georg. »Hier ist es **ja** unerträglich dunkel«, sagte er dann. »Ja, dunke es ja unerträglich dunkel«, sagte er dann. »**Ja**, dunkel ist es schon«, antwortete der Vater. » ?« »Ich habe es lieber so.« »Es ist **ja** ganz warm draußen«, sagte Georg, wie im und die Arme über der Brust kreuzt. »**Ja**, Deinem Freunde«, sagte der Vater mit Betonung. Brille, die er mit der Hand bedeckte. »**Ja**, jetzt habe ich es mir wieder überlegt. mir im Geschäft unentbehrlich, das weißt du **ja** sehr genau, aber wenn das Geschäft deine bei mir im Zimmer saß. Ich konnte **ja** deine Abneigung gegen ihn ganz gut verstehen, erhob und die Menge anrief. Du hast **ja** selbst diese Geschichte hier und da wieder seinen künftigen Haushalt mitzunehmen. Es schien **ja** fast, wenn man genauer zusah, daß die seine Zunge, daß er vor Schmerz einknickte. »**Ja**, freilich habe ich Komödie gespielt! Komödie! G er schon seit Jahren nicht, er weiß **ja** alles hundertmal besser als du selbst, deine gelb zum Wegwerfen, und ich, du siehst **ja**, wie es mit mir steht. Dafür hast du wahrscheinlich früher sagen. Jetzt paßt es **ja** gar nicht mehr.« Und lauter: »Jetzt weißt von dir! Ein unschuldiges Kind warst du **ja** eigentlich, aber noch eigentlicher warst du ein

Figure 6. KWIC list of all instances of *ja* in Kafka's *Das Urteil* (*The Judgment*)

In answer to the research question “What functions do the modal particles *ja* and *doch* exert in Kafka's short narration *The Judgment* (1912)?”, the method of choice was a digital type of close reading of the text. – I used a digital KWIC (*Keyword in Context*) tool for a starting point.^[22] Figure 6 shows that of the N=18 instances of *ja*, n=12 may be identified as modal particles; and a similar picture can be observed for *doch*, where of N=19 retrieved instances, n=9 are clear modal particles. This frequency roughly equals the proportion of each word found in natural *spoken* discourse in German [Hentschel 1986].

The combined linguistic/narratological examination of the usage of *ja* and *doch* throughout the story revealed that the two lemmas mostly do not accord with the causality structure of the narrated world (for details, see [Herrmann 2016]: The two characters, a father and his son, do not communicate on the basis of a reliable common ground. Instead, there are constant shifts in basic beliefs, and thus, in part due to the use of modal particles (or accompanied by them as symptoms of underlying communicative shortcomings) the management of information fails epically, resulting in the death of the son. In particular, *ja* is normally used to signal that interlocutor A assumes that interlocutor B maintains a shared knowledge/belief at the present point in time; *doch* is normally used to signal that interlocutor A assumes that interlocutor B is presently not aware of an allegedly shared knowledge/belief. However, in *The Judgment*, the characters often use them in an infelicitous way, not respecting conventional pragmatic principles. The fact that in terms of frequency, the modal particles in Kafka's fictional narrative roughly equal the natural distribution in spoken language may explain the overuse identified in the keyness analysis, especially when assuming that modal particles are overall not as frequent in fictional writing. Speculatively, one may adopt the idea that the use of *ja* and *doch* in *The Judgment* is typical for Kafka's use of modal particles in general, especially taking into account that modal particles are indicators of free indirect discourse, which is one of the properties frequently assigned to Kafka's prose.

Yet, more (quantitative and qualitative) analysis is needed to further flesh out and test these hypotheses. For now, it may be stated that the use of two specific modal particles indeed seems to contribute to the “elevated everyday language” described as typical of Kafka's style (in *Metamorphosis*, cf. [Trost 2008]). The examination of two modal particles in one particular text used the local lexical/syntactical context as well as that of the story world with its complex social-communicative relations. Since modal particles are heavily dependent on contextual analysis, both levels are essential when striving to make well-informed statements about their function in Kafka's style.

To sum up, this article has presented new findings on Kafka's prose (compared to other authors writing in German) gathered in complementary successive analysis. It tested philological findings by means of a method of data aggregation operating on a comparably large data set, explored author's style by looking at the probabilities of word frequencies, and eventually zoomed into the possible functions of two modal particles within the text world of one story. To ensure compatibility, my study kept the reference corpus constant, and varied the style measure only subtly (both quantitative analyses work with raw quantities of inflected word forms, and two of these word forms were queried through KWIC in the particular story *The Judgment*).

Yet, my main objective was to raise awareness of the methodological issues of doing digital stylistics. The different steps varied in the way in which they approached Kafka's style – quantitatively and qualitatively, explanatory and exploratory, deductively and inductively, and also in the degree of speculation and interpretative freedom licensed. With regard to quantitative analysis, I have highlighted that hypothesis testing in the strict sense of the term is still a desideratum in my work. The same holds for analysis of statistical correlations and interactions on a broader data basis. My future research will take care of this: We are currently building a “Corpus of Literary Modernism” (KOLIMO, see footnote above) for examining the possible relations between factors such as author, genre, time of publication etc. and style variables such as part of speech, lexical density, and readability, but

also more traditionally literary features such as metaphor [Herrmann2018].

Conclusions

Above, I first proposed that the interdisciplinary setup of DH stylistics has advantages and disadvantages, depending on how we handle the issue, and that in fact, DH style studies may be a natural environment for the mixed-methods paradigm. To contextualize this, I have reported on my own research, assuming that in terms of its general character, it may be quite representative of digital text analysis, even though the range of methods and questions found in DH stylistics is naturally much wider. For reasons of focus, further crucial dimensions of doing DH stylistics that relate to research design have not been mentioned, such as programming, sample selection, edition and publication.

I pointed out that in some moments, in line with a 'quantitative paradigm', we may strive for open questions to be settled, and possibilities for interpretation of results to be delimited in order to achieve significant and replicable results. At other points of the research process, we may tap into the "qualitative paradigm", in line with our hermeneutic tradition, allowing higher degrees of (aesthetic) subjectivity, generating ideas, and opening discussion of alleged securities. Pointing towards a "mixed-methods paradigm", I would now like to conclude by considering Moretti's take on distant reading:

If we want to understand the system in its entirety, we must accept losing something. We always pay a price for theoretical knowledge: reality is infinitely rich; concepts are abstract, are poor. But it's precisely this 'poverty' that makes it possible to handle them, and therefore to know. This is why less is actually more. [Moretti 2000, 578–8]

My point in this paper has been to embrace distant reading's variable-driven principle, but to complement it by a paradigm that is case-driven, looks at smaller samples, and may indulge in open-ended interpretation about connections between different variables. Both general approaches have their merit, and tapping into the methodological flexibility offered by a mixed-methods paradigm will allow for the interaction of different mindsets in a structured way.

By mutual methods critique, we may more vigorously utilize the meaning potential of crunched numbers, as well as sharpen hermeneutically obtained insight to a new degree of precision. Here, the digital turn is possibly what really accelerates the growth of a third culture in the sense of Snow [Snow 2012] – one that has explanatory as well as exploratory moments, with "thick" understanding as well as the "poverty" of aggregation mentioned by Moretti. A mixed-method paradigm includes approaches for 'opening up the discussion' as much as those that 'settle it'.

Author's Note

Many thanks to the editorial team, led by Joris van Zundert, for invoking and guiding the volume, as well as for their feedback at multiple stages of the publication process. Thanks also to the three anonymous reviewers who provided me not only with detailed and constructive feedback, but who also were generous in pointing out the things they valued. A big thank you to Andrea Hense, an ally from the *Sociological Research Institute* Göttingen (SOFI) and to Steffen Kühnel of the *Center of Methods in Social Sciences* at Göttingen University, both of whom made time for discussing my research. And finally, many thanks to Camilla di Biase-Dyson, who proof-read an earlier version of the paper, and to Zsófia Demjen, for proof-reading the final version. Any remaining errors are my own.

Notes

[1] One should note that mixed-methods is essentially an empirical research paradigm. This is compatible with a digital stylistics that understands "empirical" in the wide sense – as the "acquisition of knowledge through observation" ("Wissensgewinn durch Beobachtung") [Eibl 2013, 23].

[2] This may possibly even entail striving for some unifying epistemological perspective. Some mixed-methods scholars advocate pragmatism, such as Creswell who assumes a strong link between "worldviews" (e.g., postpositivist, social constructivist, participatory, pragmatic) and research methods [Creswell 2014]. Others, such as Kelle describe weaker links between particular world views and methods [Kelle 2008].

[3] Within mixed-method approaches, interpretation is being recognized as a vital operation, linked to the tradition of hermeneutic inquiry: "data analysis, including quantitative data analysis, is a process of interpretation involving basic hermeneutic principles" [Ness et al. 2011, 294]. Good mixed-methods research in DH stylistics should fulfill criteria that allow for complementarity (cf. [Miles et al. 2014, 311–314]) – and methodological awareness. If such criteria are met, subjectivity and phases of open-ended interpretation are highly justified. These become problematic only where introspection and impressionistic observations are used in an opaque way, which however is true for any quantitative-scientific study as well [Rudman 2016]. Problematic is also a dogmatic rejection of (aggregated) "data". Indispensable practical factors of sound mixed-method research are clear language use and the definition of concepts and terms.

[4] Recently, scholars like Symonds & Gorard have emphasized that a distinction between "quantitative" and "qualitative" research is a stereotypical idealization [Symonds and Gorard 2010]. As such, it indeed helps structuring our understanding of practices. However, it may be misleading to use a dichotomy as a basis for invoking a mixed-methods paradigm, since research processes 'in the wild' are not necessarily paradigmatic. Rather, they lie on a continuum between "open-endedness" vs. "closed-endedness" (data collection tools), "non-numerical" vs. "numerical" (data types), and "interpretation of natural occurring data" vs. "counting/aggregation/statistics" (analytical techniques). A gradual character of quantification/qualification is actually evident in digital stylistics: there are many quantitative studies that aim at exploration, rather than at explanation through hypothesis testing (e.g., data science in the sense of [Manovich 2015]). At the same time, close reading can be used to test hypotheses derived from statistical quantitative analyses (e.g., revealing what mode of thought representation is actually invoked by a certain verb type that contributes significantly to a factor analysis of genre style). However, I believe that working with research paradigms has a useful structuring function and therefore shall be maintained for the time being.

[5] For a number of easily accessible text analysis visualization tools, see the web-based tool Voyant [Sinclair and Rockwell 2016].

[6] CATMA is a tool for manual hermeneutic digital text annotation that features a comprehensive narratological tag set [Gius and Jacke 2015].

49

50

51

52

[7] See [Underwood 2017], whose article on the genealogy of distant reading was not yet available at the time of writing and will thus not be discussed in depth. His key argument is that the basic distinction of what he calls “distant reading” from other forms of literary criticism is “the practice of framing historical inquiry as an experiment, using hypotheses and samples (of texts or other social evidence) that are defined before the writer settles on a conclusion.” (5). He thus draws a basic distinction between exploratory (“other forms of literary criticism”) and explanatory (“distant reading”) inquiry. Although, as explained above, I see a great value in mapping the distinction of “distant vs. other forms of reading” onto the basic “quantitative vs. qualitative” paradigms, I do not fully agree with Underwood’s general characterization of “distant reading” as an “experimental” hypothesis-testing endeavor. There is, after all, a great number of “distant reading”-studies that uses data aggregation methods and big data sets to explore (emerging) patterns. Such studies, which are fully valid, as I will show below, do not apply hypothesis-testing in the strict sense.

[8] Retrievable under <http://www.aclweb.org/anthology/W/W15/#0700>

[9] Scholars like Manovich (2015) and Anderson (2008) take a relatively extreme stance, proclaiming that with the large available repositories of data, the nature of numerically transformable data is fundamentally changing: instead of “long data”, which is (traditionally) organized in few variables with many pertaining cases, the new type is “wide data”, clustered in many variables with a huge amount of individual cases [Manovich 2015]. In “data mining”, data exploration techniques incl. dimension reduction are most important for predictive analysis and machine learning based on probabilities of correlations. While correlative/association analyses have long been established in computational and corpus linguistics (as in social sciences), “big data science” appears in a new quality. In its most prominent formulation it forsakes the quest for causality (and “theory”) altogether [Anderson 2008]. Seen from the perspective of mixed-methods, however, there is no reason for why one should not include stages of zooming in on causal relationships.

[10] As will be seen below, this is a somewhat loose understanding of hypothesis testing. Explanatory hypothesis testing in the strict quantitative sense would have selected a combination of hypothesis and method that leaves much less room for confounding variables. The particular analysis is thus a typical example for the current state of the field, pending more rigorous and wide-spread applications of hypothesis testing.

[11] It is an open (empirical) question how much of this effect of Kafka’s prose may in fact be attributed to his ‘style’, and in how much it is a matter of “plot”. Since both dimensions tend to feed into each other, further research is needed to creatively establish ways of description and measurement.

[12] Z-scores represent the distance between the individual raw frequencies and an observed group mean, in units of the standard deviation.

[13] Stylometric measures, first developed for authorship attribution, and lately adopted by DH for mostly literary history studies, have in fact been used widely for exploratory purposes [Jannidis and Lauer 2014]. This shows that a quantitative-explanatory mapping is too facile; the way in which a measure is applied essentially depends on the research question.

[14] None of the texts used in the analysis are of disputed authorship, and the genre signal was kept relatively constant, with all texts labeled as ‘prose’. The genre signal is a major confounding factor in authorship attribution [Rudman 2016, 318].

[15] See <https://sites.google.com/site/computationalstylistics/corpora>

[16] Texts by Walser were extracted from Gutenberg.org. Those texts, as well as ones from many more authors (currently more than 2,000) extracted from Gutenberg-DE, TextGrid Repository, and the German Text Archive (DTA) have been compiled to form a unified corpus of narrative modernism (KOLIMO), together with added metadata and a few style features. Visit the ongoing project under <https://kolimo.uni-goettingen.de/about.html>

[17] The lower margin was set to 800 mfw in accordance with Jannidis & Lauer, who however set the upper limit to 3,000 mfw [Jannidis and Lauer 2014]. I set the upper limit to 1,500 mfw, thus capturing the range of values where Eder’s Delta’s appears to perform best (between 1,000 and 1,500 mfw; cf. [Jannidis et al. 2015]). However, Evert et al. (2015), who replicated Jannidis et al. (2015) with more feature samples, state that the number of features (mfw) remains “a critical factor for which no good strategy is available” [Evert et al. 2015, 87]. The results thus need to be further tested.

[18] Stylometric methods such as cluster analysis applied to literary data typically involve a series of tests with different parameter settings, including varying the distance measure, the number of mfw, the linkage algorithm, and others (cf. [Eder 2017] [Evert et al. 2015] [Jannidis and Lauer 2014], for German literature). It has been pointed out that this kind of scenario is prone to “cherry picking”, i.e. omitting results that do not fit the expected/desired outcome [Rudman 2003]. The overall results, specifically the bootstrap consensus tree, however, give some reason to assume that Kafka is indeed an author with a distinct – if not solitary – style. Yet, the single dendrograms vary with regard to Kaka’s nearest neighbors. Therefore, all obtained results need to be further evaluated, and possibly sided by experiments with other parameter settings.

[19] Comparable to the “oppose-function” of the stylo-package [Eder et al. 2016], which has been further developed since the time of writing.

[20] Altenberg was newly introduced for the present analysis. Because of his colloquial style, I extrapolated that he might use adverbs, negations, and modal particles in a similar way as Kafka. However, the result is very similar to that of Schnitzler, and the number of types among the top 100 keywords is only half of Kafka’s.

[21] The present analysis has good external validity, as shown by prior analysis [Herrmann 2013b] based on a much larger and more varied reference corpus (~ 93,000 texts, ~ 600 authors, ~ 134 million words, diverse genres). Comparison shows very similar results for Kafka (N=24 (big reference corpus) vs. N=27 (present corpus)) and Schnitzler (N=10 (big reference corpus) vs. N=12 (present corpus)), and numerically identical ones for Meyrink (N=7) and Trakl (N=1). Since the larger corpus is closer to an assumed “true population” of newer German literature, there is good reason to assume that the findings obtained can be generalized by taking as a stylistic “norm” a larger population (of texts, genres, and authors across time).

[22] There are many KWIC tools available online, e.g. in Voyant, AntConc, as well as functions in R and Python.

Works Cited

Adorno 1981 Adorno, T. W. (1981). “Notes on Kafka.” In S. Weber & S. Weber (Trans.), *Prisms* (pp. 243–271). Cambridge Mass.: MIT Press.

Algee-Hewitt et al. 2016 Algee-Hewitt, M., Allison, S., Gemma, M., Heuser, R., Moretti, F., & Walser, H. (2016). “Canon/Archive. Large-scale dynamics in the Literary Field.” *Stanford Literary Lab Pamphlet*, 11. Retrieved from <http://litlab.stanford.edu/LiteraryLabPamphlet11.pdf>.

Anderson 2008 Anderson, C. (2008). “The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.” Retrieved July 24, 2013, from http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

- Anthony2014** Anthony, L. (2014). AntConc (Version 3.4.4). Tokyo, Japan: Waseda University. Retrieved from <http://www.laurenceanthony.net/>
- Becher et al. 2012** Becher, P., Höhne, S., & Nekula, M. (Eds.). (2012). *Kafka und Prag: Literatur-, kultur-, sozial- und sprachhistorische Kontexte*. Köln: Böhlau.
- Biber 1988** Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber 2011** Biber, D. (2011). "Corpus linguistics and the study of literature: Back to the future?" *Scientific Study of Literature*, 1(1), 15–23. 10.1075/ssol.1.1.02bib.
- Biber and Reppen 2015** Biber, D., & Reppen, R. (2015). *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press.
- Bird et al. 2015** Bird, S., Klein, E., & Loper, E. (2015). *Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit*. Retrieved from <http://www.nltk.org/book/>
- Blank 2001** Blank, H. (2001). In *Kafkas Bibliothek: Werke der Weltliteratur und Geschichte in der Edition, wie sie Kafka besaß oder kannte; kommentiert mit Zitaten aus seinen Briefen und Tagebüchern*. Stuttgart: Blank.
- Borgman 2015** Borgman, C. L. (2015). *Big data, little data, no data: scholarship in the networked world*. Cambridge: MIT Press.
- Brenner 2011** Brenner, P. J. (2011). *Neue deutsche Literaturgeschichte: vom "Ackermann" zu Günter Grass* (3., rev. and extended edition). Berlin: de Gruyter.
- Burrows 1987** Burrows, J. F. (1987). *Computation into criticism. A study of Jane Austen's novels and an experiment in method*. Oxford: Clarendon.
- Burrows 2002** Burrows, J. (2002). "'Delta': a measure of stylistic difference and a guide to likely authorship." *Literary and Linguistic Computing*, 17(3), 267–287. <https://doi.org/10.1093/lc/17.3.267>.
- Craig 1999** Craig, H. (1999). "Authorial attribution and computational stylistics: if you can tell authors apart, have you learned anything about them?" *Literary and Linguistic Computing*, 14(1), 103–113. <https://doi.org/10.1093/lc/14.1.103>.
- Craig 2013** Craig, H. (2013, September). *Middle-distance reading with information-theory metrics*. (Talk abstract). Retrieved from <http://www.gcdh.de/en/events/calendar-view/prof.-hugh-craig-middle-distance-reading-with-information-theory-metrics1/>
- Creswell 2014** Creswell, J. W. (2014). *Research design: qualitative, quantitative, and mixed methods approaches* (4th ed.). Los Angeles: Sage.
- Creswell and Plano Clark 2007** Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks: Sage.
- D'Haen et al. 2012** D'haen, T., Damrosch, D., & Kadir, D. (2012). *The Routledge companion to world literature*. London: Routledge.
- Dipper 2008** Dipper, S. (2008). "Theory-driven and Corpus-driven Computational Linguistics, and the Use of Corpora." In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics. An International Handbook* (pp. 68–96). Berlin: De Gruyter.
- Diversy et al. 2014** Diversy, S., Evert, S., & Neumann, S. (2014). "A semi-supervised multivariate approach to the study of language variation." In B. Szmrecsanyi & B. Wälchli (Eds.), *Aggregating Dialectology, Typology, and Register Analysis: Linguistic Variation in Text and Speech* (pp. 174–204). Berlin: De Gruyter.
- Eder 2017** Eder, M. (2017). "Visualization in Stylometry: Cluster Analysis Using Networks." *Digital Scholarship in the Humanities*, Volume 32, Issue 1(1), 50–64. <http://doi.org/10.1093/lc/fqv061>
- Eder et al. 2016** Eder, M., Kestemont, M., & Rybicki, J. (2016). "Stylometry with R: A package for computational text analysis." *R Journal*, 16(1). Retrieved from <https://journal.r-project.org/archive/accepted/>
- Eibl 2013** Eibl, K. (2013). "Ist Literaturwissenschaft als Erfahrungswissenschaft möglich? Mit einigen Anmerkungen zur Wissenschaftsphilosophie des Wiener Kreises." In P. Ajouri, K. Mellmann, & C. Rauen (Eds.), *Empirie in der Literaturwissenschaft* (pp. 19–45). Münster: Mentis.
- Engel 2010** Engel, M. (2010). "Drei Werkphasen." In M. Engel & B. Auerochs (Eds.), *Kafka-Handbuch. Leben, Werk, Wirkung* (pp. 81–90). Stuttgart: Metzler.
- Evert et al. 2015** Evert, S., Proisl, T., Jannidis, F., Pielström, S., Schöch, C., & Vitt, T. (2015). "Towards a better understanding of Burrows's Delta in literary authorship attribution." *Proceedings of NAACL-HLT Fourth Workshop on Computational Linguistics for Literature*, 79–88.
- Fish 2012** Fish, S. (2012, January 23). "Mind your p's and b's: The digital humanities and interpretation." *Opinionator, New York Times*. Retrieved from <http://opinionator.blogs.nytimes.com/2012/01/23/mind-your-ps-and-bs-the-digital-humanities-and-interpretation/>
- Frenzel and Frenzel 1990** Frenzel, H. A., & Frenzel, E. (1990). *Daten deutscher Dichtung. Band II: Vom Realismus bis zur Gegenwart*. München: Dt. Taschenbuch-Verl.
- Gius and Jacke 2015** Gius, E., & Jacke, J. (2015). "Informatik und Hermeneutik. Zum Mehrwert interdisziplinärer Textanalyse." In C. Baum & T. Stäcker (Eds.), *Grenzen und Möglichkeiten der Digital Humanities. Sonderband der Zeitschrift für digitale Geisteswissenschaften* (Vol. 1). Retrieved from <http://www.zfdg.de/informatik-und-hermeneutik-zum-mehrwert-interdisziplin%C3%A4rer-textanalyse>
- Gooding 2013** Gooding, P. (2013). "Mass digitization and the garbage dump: The conflicting needs of quantitative and qualitative methods." *Literary and Linguistic Computing*, 28(3), 425–431. <http://doi.org/10.1093/lc/fqs054>
- Gries 2015** Gries, S. Th. (2015). "Some Current Quantitative Problems in Corpus Linguistics and a Sketch of Some Solutions." *Language and Linguistics* 16(1), 93–117. <https://doi.org/10.1177/1606822X14556606>.
- Hayles 2012** Hayles, N. K. (2012). *How we think: digital media and contemporary technogenesis*. Chicago; London: The University of Chicago Press.
- Hentschel 1986** Hentschel, E. (1986). *Funktion und Geschichte deutscher Partikeln: Ja, doch, halt und eben*. Berlin: De Gruyter.
- Herrmann 2013a** Herrmann, J. B. (2013a). *Kafka among the authors. Stylometric analyses*. (Paper). Presented at the Expert Workshop, Stylometry@Kraków. Krakow.

- Herrmann 2013b** Herrmann, J. B. (2013b). *Computing Kafka - How keyness and collocation analysis help explain paradoxical style*. Talk and poster presented at the International Herrenchausen Conference "(Digital) Humanities Revisited – Challenges and Opportunities in the Digital Age." Herrenchausen.
- Herrmann 2016** Herrmann, J. B. (2016). "Läuse im Pelz der Sprache?' Zu Funktionen von Modalpartikeln in narrativen (De-)Motivierungsstrategien bei Franz Kafka." In M. Horváth & K. Mellmann (Eds.), *Die biologisch-kognitiven Grundlagen narrativer Motivierung*. Münster: Mentis.
- Herrmann and Lauer 2016** Herrmann, J. B., & Lauer, G. (2016). *Aufbau und Annotation des Kafka/Referenzkorpus*. Paper presented at the Conference Digital Humanities im deutschsprachigen Raum (DhD), Leipzig. <http://www.dhd2016.de/abstracts/vortr%C3%A4ge-011.html>
- Herrmann et al. 2015** Herrmann, J. B., van Dalen-Oskam, K., & Schöch, C. (2015). "Revisiting style, a key concept in literary studies." *Journal of Literary Theory*, 9(1), 25-52. 10.1515/jlt-2015-0003.
- Herrmann2018** Herrmann, J. B. (2018). "Anschaulichkeit messen. Eine quantitative Metaphernanalyse an deutschsprachigen Erzählanfängen zwischen 1880 und 1926." ["Measuring Stylistic Vividness. A quantitative metaphor analysis of the beginning sections of German narrative fiction 1880 -1926"]. In T. Köppe & R. Singer (Eds.), *Show, don't tell: Konzepte und Strategien anschaulichen Erzählens*. Bielefeld: Aisthesis.
- Heyer et al. 2006** Heyer, G., Quasthoff, U., & Wittig, T. (2006). *Text Mining: Wissensrohstoff Text: Konzepte, Algorithmen, Ergebnisse*. Herdecke : W3L-Verl.
- Holmes 1998** Holmes, D. I. (1998). "The Evolution of Stylometry in Humanities Scholarship." *Literary and Linguistic Computing*, 13(3), 111–117. <http://doi.org/10.1093/lilc/13.3.111>
- Hoover 2004** Hoover, D. L. (2004). "Testing Burrows's Delta." *Literary and Linguistic Computing*, 19(4), 453–475. <http://doi.org/10.1093/lilc/19.4.453>
- Hoover 2013** Hoover, D. L. (2013). "Quantitative Analysis and Literary Studies." In R. Siemens & S. Schreibman (Eds.), *A Companion to Digital Literary Studies* (pp. 517–533). John Wiley & Sons, Ltd. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/9781405177504.ch28/summary>
- Hoover et al. 2014** Hoover, D. L., Culpeper, J., & O'Halloran, K. (2014). *Digital literary studies: Corpus approaches to poetry, prose, and drama*. New York: Routledge, Taylor & Francis Group.
- Jannidis and Lauer 2014** Jannidis, F., & Lauer, G. (2014). "Burrows's Delta and its use in German literary history." In M. Erlin & L. Tatlock (Eds.), *Distant Readings. Topologies of German Culture in the Long Nineteenth Century* (pp. 29–54). Rochester, New York: Camden House.
- Jannidis et al. 2015** Jannidis, F., Steffen Pielström, Christoph Schöch, and Thorsten Vitt (2015). "Improving Burrows Delta. An empirical evaluation of text distance measures." In *Abstracts for the Digital Humanities 2015*, Sidney. Retrieved from http://dh2015.org/abstracts/xml/JANNIDIS_Fotis_Improving_Burrows_Delta__An_empirical/JANNIDIS_Fotis_Improving_Burrows_Delta__An_empirical_.html
- Jockers 2013** Jockers, M. L. (2013). *Macroanalysis: digital methods and literary history*. Urbana, Ill.: Univ. of Illinois Press.
- Jockers 2014** Jockers, M. L. (2014). *Text analysis with R for students of literature*. Cham: Springer.
- Johnson and Onwuegbuzie 2004** Johnson, R. B., & Onwuegbuzie, A. J. (2004). "Mixed Methods Research: A Research Paradigm Whose Time Has Come." *Educational Researcher*, 33(7), 14–26. <https://doi.org/10.3102/0013189X033007014>.
- Kelle 2008** Kelle, U. (2008). *Die Integration qualitativer und quantitativer Methoden in der empirischen Sozialforschung: theoretische Grundlagen und methodologische Konzepte* (2nd Ed.). Wiesbaden: VS, Verlag für Sozialwissenschaften.
- Klaussner et al. 2015** Klaussner, C., Nerbonne, J., & Çöltekin, Ç. (2015). Finding Characteristic Features in Stylometric Analysis. *Digital Scholarship in the Humanities*, 30(suppl 1), i114–i129. <http://doi.org/10.1093/lilc/fqv048>
- Leech and Short 2007** Leech, G. N., & Short, M. (2007). *Style in fiction. A linguistic introduction to English fictional prose* (2nd ed). New York: Pearson Longman.
- Mahlberg 2013** Mahlberg, M. (2013). *Corpus Stylistics and Dickens's Fiction*. London: Routledge.
- Mahlberg 2015** Mahlberg, M. (2015). "Literary style and literary texts." In *The Cambridge Handbook of English Corpus Linguistics*. Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139764377.020>
- Manovich 2015** Manovich, L. (2015). "Data Science and Digital Art History." *International Journal for Digital Art History*, 0(1). <http://dx.doi.org/10.11588/dah.2015.1.21631>.
- McCarty 2005** McCarty, W. (2005). *Humanities computing*. Basingstoke & New York: Palgrave Macmillan.
- Meindl 2011** Meindl, C. (2011). *Methodik für Linguisten. Eine Einführung in Statistik und Versuchsplanung*. Tübingen: Narr.
- Miles et al. 2014** Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: a methods sourcebook* (3rd ed.). Los Angeles, Calif.: Sage Publ.
- Moretti 2000** Moretti, F. (2000). "Conjectures on world literature". *New Left Review*, (1). Retrieved from <https://newleftreview.org/II/1/franco-moretti-conjectures-on-world-literature>.
- Mueller 2012** Mueller, M. (2012, May 29). "Scalable Reading." (Blog-entry). Retrieved from <https://scalablereading.northwestern.edu/scalable-reading/>
- Müller 2013** Müller, R. (2013). "Parallelstellenmethode – digital. Wie computer-gestützte Korpus-Analysen die Hermeneutik empirisieren." In P. Ajouri, K. Mellmann, & C. Rauen (Eds.), *Empirie in der Literaturwissenschaft*. Mentis: Münster.
- Nekula 2003** Nekula, M. (2003). "Franz Kafkas Deutsch." *Linguistik Online*, 13(1). Retrieved from <https://bop.unibe.ch/linguistik-online/article/view/879/1533>
- Ness et al. 2011** Ness, P. H. V., Fried, T. R., & Gill, T. M. (2011). "Mixed methods for the interpretation of longitudinal gerontologic data. Insights from philosophical hermeneutics." *Journal of Mixed Methods Research*, 5(4), 293–308.
- Oschmann 2010** Oschmann, D. (2010). "Kafka als Erzähler". In M. Engel & B. Auerochs (Eds.), *Kafka-Handbuch* (pp. 438–449). Stuttgart: J. B. Metzler.

- Punch 2014** Punch, K. F. (2014). *Introduction to social research: Quantitative and qualitative approaches* (3., [rev.] ed.). Los Angeles: Sage.
- Ramsay 2011** Ramsay, S. (2011). *Reading Machines: Toward an Algorithmic Criticism*. Chicago: University of Illinois Press.
- Ransom 1937** Ransom, J. C. (1937). "Criticism, Inc." *VQR Online*, 13(4). Retrieved from <http://www.vqronline.org/essay/criticism-inc-0>
- Rayson 2012** Rayson, P. (2012). "Corpus analysis of key words." In *The Encyclopedia of Applied Linguistics*. Blackwell. Retrieved from <http://dx.doi.org/10.1002/9781405198431.wbeal0247>
- Rudman 2003** Rudman, J. (2003). "Cherry Picking in Nontraditional Authorship Attribution Studies." *CHANCE*, 16(2), 26–32. <https://doi.org/10.1080/09332480.2003.10554845>.
- Rudman 2016** Rudman, J. (2016). "Non-Traditional Authorship Attribution Studies of William Shakespeare's Canon: Some Caveats." *Journal of Early Modern Studies*, 5(0), 307–328. <https://doi.org/10.13128/JEMS-2279-7149-18094>.
- Scheffel 2002** Scheffel, M. (2002). "Das Urteil" - Eine Erzählung ohne 'geraden zusammenhängenden, verfolgaren Sinn'? *Strukturalismus mit strukturaler Erzähltheorie*. In O. Jahraus & S. Neuhaus (Eds.), *Kafkas "Urteil" und die Literaturtheorie. Zehn Modellanalysen* (pp. 59–77). Stuttgart: Reclam.
- Scott and Tribble 2006** Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in Language Education*. Amsterdam, Philadelphia: John Benjamins.
- Sculley and Pasanek 2008** Sculley, D., & Pasanek, B. M. (2008). "Meaning and mining: the impact of implicit assumptions in data mining for the humanities." *Literary and Linguistic Computing*, 23(4), 409–424. <http://doi.org/10.1093/lc/fqn019>
- Semino and Short 2004** Semino, E., & Short, M. (2004). *Corpus stylistics. Speech, writing and thought presentation in a corpus of English writing*. London: Routledge.
- Simpson 2004** Simpson, P. (2004). *Stylistics. A resource book for students*. London: Routledge.
- Sinclair and Rockwell 2016** Sinclair, S., & Rockwell, G. (2016). *Voyant Tools*. Web. Retrieved from <http://voyant-tools.org/>
- Snow 2012** Snow, C. P. (2012). *The Two Cultures* [1959]. Cambridge: Cambridge University Press.
- Spitzer 1961** Spitzer, L. (1961). *Stilstudien* [1928]. *Erster Teil: Sprachstile. Zweiter Teil: Stilsprachen*, München: Max Hueber.
- Stamatatos 2009** Stamatatos, E. (2009). "A Survey of Modern Authorship Attribution Methods." *Journal of the American Society for Information Science and Technology*, 60(3), 538–556. <http://dx.doi.org/10.1002/asi.21001>.
- Steen et al. 2010** Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., & Krennmayr, T. (2010). "Metaphor in Usage." *Cognitive Linguistics*, 21(4). <https://doi.org/10.1515/cogl.2010.024>.
- Symonds and Gorard 2010** Symonds, J. E., & Gorard, S. (2010). "Death of mixed methods? Or the rebirth of research as a craft." *Evaluation & Research in Education*, 23(2), 121–136. <https://doi.org/10.1080/09500790.2010.483514>.
- Tashakkori and Teddlie 2010** Tashakkori, A., & Teddlie, C. (Eds.). (2010). *Sage handbook of mixed methods in social & behavioral research* (2nd ed.). Los Angeles: Sage.
- Trost 2008** Trost, I. (2008). "Erzählen und Besprechen: zum Stil von Franz Kafkas Erzählung 'Die Verwandlung'." In: Fritz, Th. A., Koch, G. & Trost, I., *Literaturstil - sprachwissenschaftlich. Festschrift für Hans-Werner Eroms zum 70. Geburtstag* (pp. 144–168). Heidelberg: Universitätsverlag.
- Underwood 2017** Underwood, T. (2017). "A Genealogy of Distant Reading," *DH Quarterly*, 11(2). Retrieved from <http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html>.
- van Zundert 2015** van Zundert, J. J. (2015). "Screwmenetics and Hermenumericals." In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A New Companion to Digital Humanities* (pp. 331–347). Wiley. 10.1002/9781118680605.ch23.
- Van Peer et al. 2012** Van Peer, W., Hakemulder, J., & Zyngier, S. (2012). *Scientific methods for the humanities*. Amsterdam, Philadelphia: John Benjamins.
- Weitin 2017** Weitin, T. (2017). "Scalable Reading." *Zeitschrift für Literaturwissenschaft und Linguistik*, 47(1), 1–6. <https://doi.org/10.1007/s41244-017-0048-4>.
- kafkaesque** Macmillan Publishers (Ed.). (2015). *kafkaesque. Macmillan Dictionary and Thesaurus Online*. Retrieved from <http://www.macmillandictionary.com/dictionary/british/kafkaesque>



This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.