Massimo Salgaro[*]

THE DIGITAL HUMANITIES AS A TOOLKIT FOR LITERARY THEORY: THREE CASE STUDIES OF THE OPERATIONALIZATION OF THE CONCEPTS OF "LATE STYLE," "AUTHORSHIP ATTRIBUTION," AND "LITERARY MOVEMENT"

A *curiosum* on the final pages of Umberto Eco's 1979 *Lector in Fabula*, demands further interpretation. After discussing the role of his model reader, Eco engaged his students at the University of Bologna and the University of Urbino in a reading experiment, asking them to read, summarize and interpret Alphonse Allais's novel *Un drame bien parisien*. This experiment produced quantitative data: "Only 4% were unable to capture the contradictions of the story, 40% tried to identify a semiotic mechanism, 20% tried to rationalize in various manners, less than 20% declared themselves totally lost. The rest provided imprecise and incomplete reports" (Eco 1979: 230). Despite the limitations of this experiment— the small number of subjects and the amount of missing information regarding methodology—the attempt is relevant to literary criticism because it shows one of its major representatives attempting to combine his own theoretical claims with empirical data. Some decades later, within a completely different academic and intellectual context, Franco Moretti offered this ground-breaking statement in his book, *Distant Reading*:

> In the last few years, literary studies have experienced what we could call the rise of quantitative evidence. This had happened before of course, without producing lasting effects, but this time it is probably going to be different, because this time we have digital databases and automated data retrieval.... When it comes to phenomena of language and style, we can do things [with them] that previous generations could only dream of. (Moretti 2013a, 212)

This mass of data demands that "close reading" be complemented by "distant reading," allowing "patterns among billions of sentences" of digitized texts to be identified (Moretti 2013a, 164). The problem of close reading is that it takes into account only a small group of canonical texts and neglects the thousands beyond the canon. Moretti was aware that the goal of the distant-reading approach was to capture large systems like the "Western European Novel" (49) on a theoretical level.

The exploitation of big data requires a methodological revolution by literary scholars who generally tend to "listen" to literary texts. Digitized archives "are not messages that were meant to address us, and so they say absolutely nothing until one asks the right question," Moretti adds (165). This "encounter of the formal and the quantitative" (164) fascinated him.

Even though such important literary scholars as Eco and Moretti argued for the combined use of quantitative and qualitative research in literary criticism, such a method encounters skepticism among humanities scholars. In 2015 Gerhard Lauer, editor of a special issue of the *Journal of Literary Theory* on "Empirical Methods in Literary Studies," made the alarming claim that "[a]lmost universally in literary theory, a skeptical perspective on empirical methods prevails" (Lauer 2015, 1). And yet empirical methods are widely used in literary criticism (Ajouri & al. 2013) to search for concordances and word frequencies, for example. Consider, too, that the International Society for the Empirical Study of Literature (Internationale Gesellschaft für Empirische Literaturwissenschaft) was founded as early as 1987 and that, since 2013, a Max Planck Center

---
[*] *Massimo Salgaro works at the intersection of theoretical and empirical research in the fields of literary theory, Empirical Aesthetics and Digital Humanities. He works as Assistant professor of German Studies at the University of Verona. Among others, he has held research fellowships at the Tongij University of Shanghai (China) and at the University of Göttingen (Germany, Alexander von Humboldt-fellow) in 2015, at the University of Alberta, Edmonton (Canada) in 2011, at the Peter-Szondi-Institut of the Freie Universität Berlin (DAAD fellow) in 2010 and at Columbia University, New York (USA) in 2008. He is Vice-President of the Internationale-Robert-Musil Gesellschaft, Treasurer of IGEL (International Society for the Empirical Study of Litterature) and in the board of the COST-network E-READ.His current research project is "Sympathy for the devil" (at Max Planck Institute for Empirical Aesthetics, Institut for Advanced Study, Paris) on the moral value of literary empathy.*

for Empirical Aesthetics has existed with a section devoted specifically to the study of language and literature. Even if "empirical methods undoubtedly have their history and assured place in literary theory" (Lauer, 2) only a few literary scholars have made use of them, often because of lack of competence. The mass of data produced by digitization pushes us to revise the *status quo* and to rethink the difference between empirical and theoretical approaches to literary studies.

Digital Humanities (DH) offers contemporary literary criticism, as well as other fields, a unique and constantly expandable set of "big data" (digital libraries and collections, for example) that can be investigated through statistical analysis. It's merely impossible to give an exhaustive definition of the DH and of its applications, as the bibliography on "What is Digital Humanities" on the website of the European Association for Digital Humanities shows (https://eadh.org/publications/what-digital-humanities). To give an example, the DH includes text encoding and digital scholarly editing (Ciotti 2016, Pierazzo&Mancinelli 2018). Computational text analysis and stylometry are only part of the "toolkit" DH can offer to the Humanities but, for the sake of concision, in this article we are only referring to these tools.

The digitization of literary texts and the use of software for their analysis is producing a growing mass of data that changes the balance between quantitative and qualitative research in literary studies. At the same time, it is becoming increasingly clear that a partnership between these two approaches can lead to new information and innovative perspectives.

1. Quantitative and qualitative research

To grasp the interaction between quantitative and qualitative research, an examination of the social sciences, which have been dealing with this issue for a long time, can be of service. When sociology was born In the middle of the nineteenth century, its founders shared a naive belief in the natural sciences. The foundations of sociology, therefore, appear under the sign of positivism. Only post-positivistic thinking after 1960 replaced the concept of science, based on a mechanistic model of reality, of the certainty of scientific laws, and of faith in the progress of mankind. Probability and uncertainty entered the scientific paradigm. Although the social sciences lost their certainties as a result, they didn't abandon empiricism (Corbetta 2014, 28). In the 1870s, the German sociologist Max Weber took an interpretative approach to social phenomena and rejected the parallelism between sociology and the natural sciences. For him, the core of sociological understanding was *Verstehen*—that is, the attempt to capture the uniqueness and non-replicability of social phenomena. Following in these two traditions, we can observe two different perspectives in the social sciences: the quantitative and the qualitative methods of research (Corbetta 2014, 39ff). The difference between these approaches is summarized by Corbetta (51) in a table partially reconstructed here for the concepts relevant to literary criticism:

	QUANTITATIVE RESEARCH	QUALITATIVE RESEARCH
Relation theory-research	Structured, logically following phases, deduction (theory precedes observation)	Open, interactive
Concepts	Operational	Orientated, open, in progress
Psychological Interaction between researcher and observed phenomenon	Scientific observation, detached, neutral	Empathic identification in the perspective of the studied object
Representativeness	Statistical representative sample	Single cases, not statistically significant
Mathematical and statistical techniques	Intense use	No use
Implications of the results	Generalizable	Specific

As Francesco Ronzon noted, ethnographic fieldwork is another example of qualitative research. Ethnographers engaged in "participant observation" adopt an ecological approach, involving themselves with the objects of study in their natural environment (Ronzon 2008, 15). The participant observer doesn't manipulate the context, as happens in experimental work in the laboratory. Rather, she or he observes the

research object, which may be an event, a person, or a phenomenon, in the "real world," capturing its uniqueness and originality. Ethnographic observation does not exclude the observer's subjectivity; instead, it recognizes the individuality of the researcher, who belongs to a specific culture, society, and group with specific beliefs, motivations, and interests. In this view, the interpretation of the observed data assumes high importance, and the endeavor to find the correct method for each case leads to a certain "methodological pluralism" (Ronzon 2008, 20) in ethnographic research.

Undoubtedly, qualitative and quantitative approaches are separated by a barrier in the social sciences. In recent years, a third paradigm has begun to emerge that rejects methodological orthodoxy in favor of methodological appropriateness (Patton 1990, Bryman 1988). For these scholars, the legitimization of quantitative or qualitative analysis depends upon the research object (Corbetta 2014, 70).

The concept of "operationalization" serves as a bridge in the dialogue between quantitative and qualitative research. For Corbetta, this means "translation from theoretical language to empirical language" (Corbetta 2014, 89) or, in other words, the transformation of a concept into empirically observable properties. The operationalized property of a research object thus constitutes a variable. Two examples: As Corbetta (2014, 89) points out, the concept of "cultural level" can be operationalized as a property known as "diploma degree," which can further be measured as: "without graduation, high school diploma, PhD, etc." Similarly, the concept of "weight" can be expressed in the property "weight of a book" and operationalized with the measure of a specific weight of .7 kilograms (Corbetta 2014, 93). As these examples show, such properties allow the quantification, measurement, and classification of a phenomenon.

"Operationalization" is not limited to the social sciences. Recently, it has reached the field of literary theory as well. Here is Franco Moretti's pithy definition:

> [Operationalization] describes the process whereby concepts are transformed into a series of operations—which, in turn, allow all manner of phenomena to be measured. Operationalizing means building a bridge from concept to measurement, and then to the world. In our case, [the bridge] is from the concepts of literary theory to some form of quantification and then to literary texts. (Moretti 2013b, 3)

What follows are three case studies[1] of computational analysis of "Late Style," "authorship attribution," and "literary movement." This analysis represents an attempt to demonstrate that the quantitative approaches of DH allow the "operationalization" of concepts in literary theory—that is, their translation from a theoretical level to an empirical one. Once these concepts are operationalized through quantitative analysis, data are produced that can inform qualitative research; in this way, a circular form of interdisciplinary research is created.

1.1 Example 1: Late style

One example of "operationalization" is presented in a recent study (Rebora & Salgaro 2018) that examined Edward Said's concept of Late Style (Said 2006). Following Said's definition, Late Style expresses the idea that the art produced during the final years 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. In the mid-twentieth century, such important writers and philosophers as Hermann Broch, Gottfried Benn, and T. W. Adorno had already described Late Style as "radical stylistic change" and a "sharp stylistic break" in the creative output of such geniuses as Titian, Rembrandt, Goya, and Bach (see, e.g., Broch 1995, 213).

Said's definition of Late Style inspired several studies on artists' late work but was also criticized. 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). Indeed, in Kastenbaum's eyes, Late Style was not a universal phenomenon, and the style of each artist followed a unique trajectory. Gordon 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 & Smiles 2016, 36).

---

[1] The first two case studies are taken from papers that I have co-authored and that are currently "accepted for publication." The graphs and tables used here were prepared by Simone Rebora and were published in the studies cited (Rebora&Salgaro 2018; Rebora&al. 2018).

The research reported in this paper combined literary theories on and analyses of Late Style with stylometric analyses of the Late Style of three important German authors: Johann Wolfgang Goethe, Robert Musil, and Franz Kafka. 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." It has already contributed to some of the most celebrated successes in attribution of authorship—such as that of J. K. Rowling's *The Cuckoo's Calling* (Juola 2012).

To ask whether Late Style is measurable, however, implies another, unspoken question: "Does Late Style actually exist?" To answer this question, we combined qualitative research, such as Edward Said's, with quantitative analysis of the Digital Humanities. Corbetta posited five different phases of quantitative research (Corbetta 2014, 79).

1. Theory
2. Hypothesis
3. Data collection
4. Data analysis
5. Results

We adopted schema based upon Said's and Broch's theories (1) that Late Style involves a double deviation—from the style of the early work of the author and from the style of other writers working at the same time. The hypothesis (2) to be tested was whether Late Style actually existed and could be measured through quantitative methods. To find evidence for the hypothesis, we designed a three-phase experiment that involved collecting (3) and analyzing (4) data. Between steps (3) and (4) the hypothesis was operationalized. This step was key to setting the mixed-method approach into motion. Because the results (5) only partially confirm the Late Style theory and hypothesis, they show the potential of a mixed method to integrate, confirm and challenge qualitative research.

We undertook two different analyses:[2] an "internal" one, in which we measured stylistic breaks between different working phases of an author's life (early work, late work, etc.); and an "external" one, in which we compared the style of the late works of each of the three authors to a reference corpus, assembled from authors who wrote at the same time as each study subject, and assessed deviations from that corpus. We then adopted methodologies that focused on the entire vocabulary: from "Delta distance" analysis, which is based upon the relative frequencies (per book and per author) of the most frequent words in the corpus (Burrows 2002) to Zeta analysis, which aims to identify words that are significantly over- or underrepresented in a specific author's works (Schöch et al. 2018).

The graphs below describe the deviation in the Zeta Analysis between the late works of Goethe, Musil, and Kafka and those of their contemporaries:
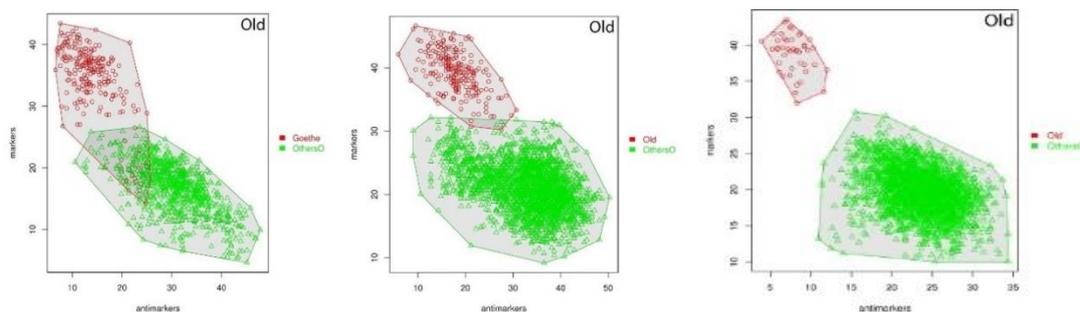

Fig. 1: Results of Zeta analysis of the late works of Goethe, Musil, and Kafka (red) compared to that of their contemporaries (green)

The diagram shows that external analysis does not confirm the existence of Late Style (a break between

---

[2] Because this paper is intended to show the intersection between quantitative and qualitative research, our research will not be described in depth. The methodology and data from the three case studies appear in the published papers.

their style and the style of their contemporaries) except in the case of Kafka, whose style seems to diverge from that of other writers who were his contemporaries. But the Kafka analysis also showed that his early work period—and not the "Late"—stood out most strikingly from the others.

Because the two methods based on word frequencies did not confirm the existence of Late Style, we attempted to analyze the same corpus from a semantic point of view using Linguistic Inquiry and Word Count software (LIWC). LIWC allows a text to be mined based on pre-compiled dictionaries in order to evaluate its dominant semantic areas (Tausczik & Pennebaker 2010). LIWC works on a series of multilingual dictionaries in which each word is connected to various semantic areas ("Social Words," "Cognitive Processes," "Seeing," "Hearing," and "Space," for example).

As a second method, we built our quantitative approach by starting with a hermeneutic analysis of the late works of Goethe, Musil, and Kafka. As an example, we adopted a 2013 monograph by Malte Kleinwort entitled *Der späte Kafka: Spätstil als Stilsuspension.* Kleinwort noted three main characteristics of Kafka's Late Style: references to earlier works, a peculiar form of tentativeness, and a poetics of asceticism (Kleinwort 2013, 10). 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

The quantitative approach based on semantic difference yielded much better results. It showed that the three semantic areas (as identified by such literary scholars as Kleinwort) linked to the Late Style of Goethe, Musil, and Kafka did, in fact, increase in the authors' late works. LIWC analysis showed that their late works were in counter-tendency to the characteristics of the work of other writers published contemporaneously. The success of semantic analysis is counterbalanced by the failure of the other two, more purely "stylometric" approaches. The following recapitulates successful (green) and unsuccessful (red) methods:

	Network analysis		Zeta analysis		Semantic analysis	
	Internal	*External*	*Internal*	*External*	*Internal*	*External*
Goethe	🟥	🟥	🟥	🟥	🟩	🟩
Musil	🟥	🟥	🟥	🟥	🟩	🟥

Kafka							

Our results were multifaceted and intriguing on the methodological level to the extent that Late Style seemed confirmed by our "internal" studies on semantic areas and through LIWC measurements, though the concept of Late Style appeared unsupported by such traditional stylometric methods as network analysis and Zeta analysis. It may be that the phenomenon of Late Style is more evident on a semantic level than on a vocabulary/syntactic level. Such multifaceted results, while not permitting a definitive answer to the question of whether Late Style actually exists, do confirm that the synergy of literary hermeneutics and stylometry can open new and innovative perspectives on traditional research questions.

Far from "solving" problems in a definite way, this combined qualitative/quantitative approach was circular in nature. It began with qualitative theory, moved through Said's and Broch's theories on Late Style, and continued to different quantitative analyses that produced results that again necessitated qualitative interpretation. Only this circular motion allows the potential of the combined approaches to be fully exploited.
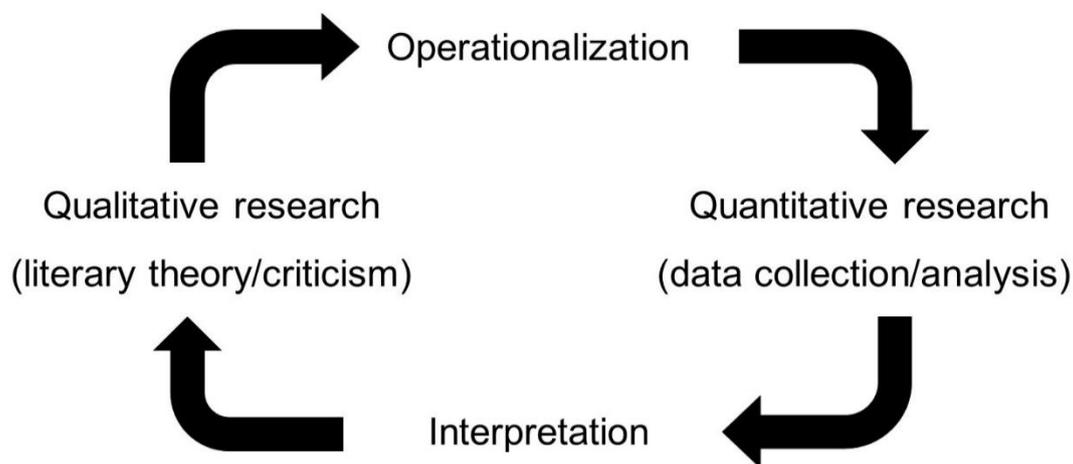

Fig. 2: Operationalization of literary theory

1.2 Example 2: Authorship Attribution

In the second case study (Rebora, Herrmann, Lauer, Salgaro 2018), we dealt with a problem of attribution of authorship. Robert Musil, one of the most important authors of twentieth-century Austrian literature, fought in the Austrian army on the Italian front during the First World War. Between 1916 and 1917, he was chief editor of the propagandistic journal *Tiroler Soldaten-Zeitung* in Bozen. While his editorship is undisputed, it is an open question whether Musil also wrote articles in the *Tiroler Soldaten-Zeitung*, and if so, how many. Because articles in the *Tiroler Soldaten-Zeitung* appeared anonymously, Musil's activity as a writer has posed a philological problem for scholars, though the major Musil scholars have attributed various articles from the *Tiroler Soldaten-Zeitung* to Musil over the last sixty years (for a detailed list, see Schaunig 2014). The surprising aspect of these attributions is the lack of factual or formal evidence for them. Marie-Louise Roth (1972, 528), for example, lists nineteen texts from the *Tiroler Soldaten-Zeitung* as Musil's, though she introduces them with the cryptic phrase "anonymous texts which have not yet been identified with certainty." Subsequent studies, such as the one by Arntzen (1980), built on Roth's assertions and proposed new attributions, again without evidence. The following table lists the scholars who have attributed *Tiroler Soldaten-Zeitung* articles to Musil, the year of publication, and the number of the attributed texts.

Name of the Musil-expert attributing texts appeared in the *Tiroler Soldaten-Zeitung*	Year of the publication of the paper including the attribution of authorship	Number of texts attributed to Musil

Karl Dinklage	1960	3 texts
Marie-Louise Roth	1972	19 texts
Karl Corino	1973, 2003, 2010	8 texts
Helmut Arntzen	1980	22 texts
Fernando Orlandi	1987, 2003, 2011	36 texts
Amann et al., *Klagenfurter Ausgabe*	2009	36 texts
Regina Schaunig	2011, 2014	38 texts ("a further 165 texts can be offered for discussion" Schaunig 2014, 358-361)

As is obvious, the attributions range from three to thirty-eight, a considerable gap. During our research we made another discovery. In the Vienna military archives,[3] we found that another author was present in the editorial office of the *Tiroler Soldaten-Zeitung*: Albert Ritter. He was a journalist, fiction writer, and politician who plead for the unification of German speaking countries (*Alldeutsche Verband*). During his career he made use of several pseudonyms, including Karl v. Winterstetten and R. W. Conrad. We digitized Ritter's texts and compared them to digitized texts by Musil that are found in the *Klagenfurter Ausgabe*. We used stylometry software to perform a series of complex experiments intended to determine the authorship of the articles published in the *Tiroler Soldaten-Zeitung*.

Our statistical analysis provides evidence for the proposition that Ritter may be the author of ten articles attributed by Musil scholars to Robert Musil. The graph below shows the results of our authorship attribution for twenty-eight texts published in the *Tiroler Soldaten-Zeitung*:
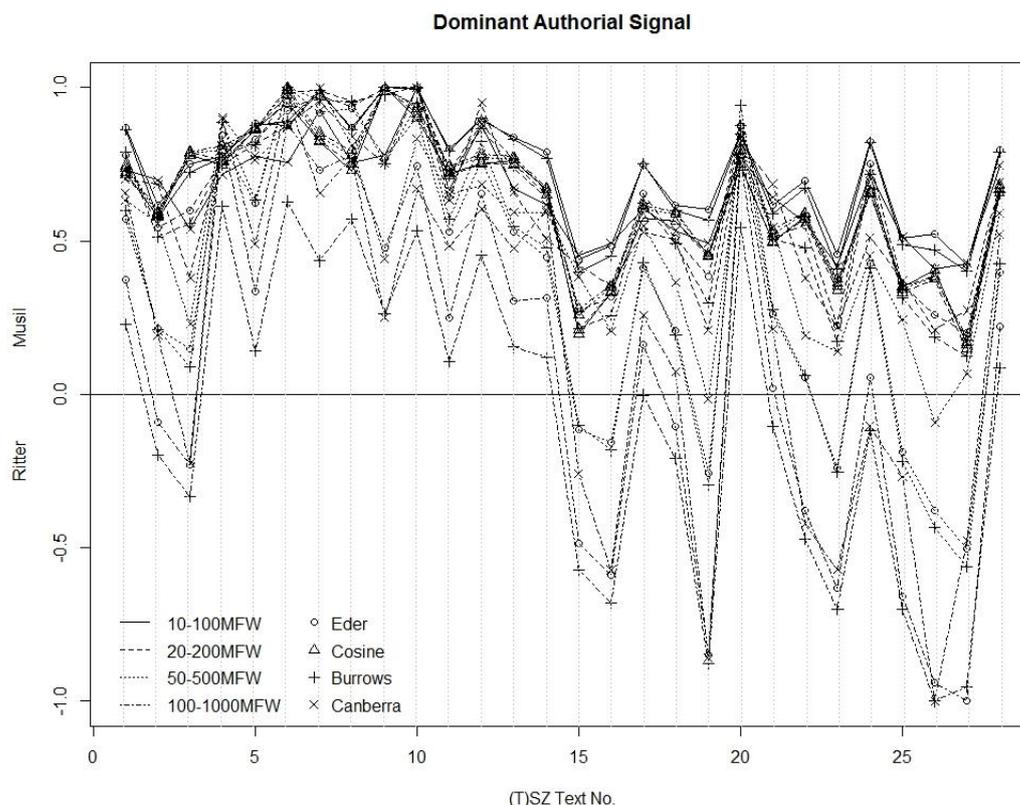

Fig. 3 Comparison of the statistical occurrence of most frequent words between Musil and Ritter through the use of various stylometric measurements (distances and most frequent words (MFW)).

---

[3] The document in question is cataloged as AT-OeStA/KA FA NFA HHK HFK/HGK Conrad (Erzherzog Joseph) Akten 1852 in the Kriegsarchiv of Vienna.

Figure 3 combines a total of sixteen different stylometric measurements for each analyzed text. If the majority of the measurements are below zero, Ritter is the most probable author of the text; if they fall above, the author is most probably Musil.

Our study has identified new instruments for attribution of authorship and style exploration and has helped solve a problem of attribution that has kept Musil researchers busy for some sixty years. But the research is not concluded; rather, it has provided food for thought for Musil scholars and DH experts. The texts attributed with more certainty to Musil and Ritter can now be compared on a more fine-grained level, comparing, for example, ideological views or political statements expressed in the *Tiroler Soldaten-Zeitung* articles. In addition, though Albert Ritter was unknown as an author prior to this analysis (Salgaro in press), the discovery of his writings through attribution analysis and his "15 minutes of fame" among Musil experts may make him the object of greater attention by German literary scholars.

1.3 Example 3: Literary history

A seminal paper published in 2014 offers a third example of the use of stylometric analysis in literary history (Jannidis and Lauer 2014). Literary history generally concerns itself solely with canonical works such as those of Goethe and Schiller. In 1809, however, not only was Goethe's canonical novel *Die Wahlverwandtschaften* published, but so were a hundred other German novels. Jannidis and Lauer therefore conclude the "[t]he canon is one story; the cultural history of read books is another" (Jannidis and Lauer 2014, 29). No one would be able to read all these texts using existing methodologies of close reading, but a distant-reading method based on digitized texts permits a consideration of non-canonical literary texts.

Jannidis and Lauer used the Delta Measure, introduced by Burrows in 2002, for their stylometric analyses. Burrows initially developed his method as a tool for author-attribution studies—that is, to narrow a large group of possibilities for the authorship of an anonymous text down to a short list of candidates. Their 2014 study sought to explore the possibilities of the Delta Measure for larger collections of German texts (33). Their first experiment was based upon a corpus of sixty-three German novels published between 1785 and 1815, and the result was a tree diagram that grouped all novels written by the same author. There was one exception, however. Friederike Helene Unger's 1806 novel *Bekenntnisse einer schönen Seele* ("Confessions of a Beautiful Soul") was not grouped close to her other work *Albert und Albertine* from 1804. Because *Bekenntnisse* was published anonymously, it has remained unclear whether Unger was the novel's true author. When the texts were grouped on the basis of gender, *Bekenntnisse einer schönen Seele* once again stood noticeably apart from Unger's other writings. As a result, Jannidis and Lauer suggested that *Bekenntnisse* might have been written by a male writer named Paul Ferdinand Buchholz or, alternatively, could "corroborate the findings of scholars in gender studies who have pointed out the bias in constructing the canon, on the one hand, and the continuous disregard of female traditions on the other (Jannidis and Lauer 2014, 42).

In another series of tests, Jannidis and Lauer analyzed the historical position of an author who wrote between the last decade of the eighteenth century and the first decade of the nineteenth, Heinrich von Kleist. It is a commonplace of literary historians that von Kleist's works oscillate between Classicism and Romanticism. In an attempt to draw more substantive conclusions, the researchers compiled a corpus of forty-nine plays in the Classicist or Romantic tradition, written or published between 1790 and 1811, and analyzed them with Kleist's works. The results suggested that Kleist's style clearly separated him from other Romantic writers. Only Kleist's *Käthchen von Heilbronn* (1810) showed similarities with Romantic plays by other writers of the same period, perhaps due to the specific nature of the work, which Kleist himself defined as "a great historical knights' play" (Jannidis and Lauer 2014, 43).

The Jannidis and Lauer example shows how DH studies can contribute to the discussion of literary texts. The researchers were able to attribute a text published anonymously and to disentangle an enigma that surrounded Kleist's literary production. Their operationalization of existing literary theories also gave rise to the typical circular form of mixed-method studies that begin with qualitative research, move on to quantitative analysis, and then return to qualitative research (Fig. 2).

2. Discussion

Taken together, the three examples show the potential of "distant reading" in literary criticism. "Distant

reading" allows scholars to go beyond the close reading of canonical texts, which offers only a limited perspective on literary history, and to exploit the "big data" of literature in order to formulate new theories on literary reception and the production of texts. Case studies 2 and 3 showed that unknown authors can be discovered through distant-reading methods (Albert Ritter and Paul Ferdinand Buchholz) and encourage literary critics to deepen their knowledge of the texts of formerly unknown authors through hermeneutic analysis.

As all three examples show, quantitative analysis offers an opportunity to resolve areas of doubt in literary theory and to answer such questions as those that implicitly guided the studies mentioned above: Does Late Style exist? Is Robert Musil the author of the texts published during the First World War? Does Kleist, as an author, belong to Classicism or Romanticism? Quantitative analyses are based on statistics that can test whether a hypothesis is valid (Corbetta 2014, 570). The best cases for such combined research are those that present a binary option—that is, which can test whether a given hypothesis is correct or not. Consequently, the use of computers in the Humanities does not lead to the abandonment of theory. As Ganascia, wrote: "On the contrary, programs need to refer to well-defined theoretical frameworks on which they can bring pieces of material evidence to bear" (2015, 4).

The use of statistics is not common in the Humanities, but the case studies examined herein show some of the ways that statistical analysis can be used as a tool for literary theory through the operationalization of literary concepts. Even when it does no more than confirm an existing theory, the attempt is not tautological because it gives rise to interdisciplinary research, as Jannidis and Lauer convincingly show:

> We do not expect any dramatic new insights from this application [of stylometry]; instead we seek to evaluate the method in terms of the knowledge we already possess. If, however, we can corroborate the existing scholarly consensus with this new computational method, then we will have succeeded in providing a firmer foundation for this knowledge, because we will have achieved the same result or similar results by two independent research methods. (Jannidis and Lauer 2014, 33)

Clearly, not every literary theory can be approached through computational analysis. As Uri Margolin argued (Margolin 1981, 18), concepts in literary theory are sometimes too vague or imprecise (aesthetic predicates such as "elegant," "graceful," and "subtle," for example). In the second case study we examined, this kind of vagueness appeared in descriptions of Musil's "functional style" (Arntzen 1980, 177) or stylistic "neutrality" (Schaunig 2014, 100). Herrmann and her colleagues encountered a similar problem when they attempted to combine traditional stylistics with stylometric approaches, some of which were not empirically testable (Emil Staiger's fuzzy definition of style as the "Ineffable-Identical," for example—the *Unaussprechlich-Identische*—was too vague to be operationalized.) (Herrmann, Van Dalen-Oskam & Schöch 2015, 31).

Sometimes the findings of a computational analysis can't be transferred to literary studies because the balance between the knowledge in the source discipline (DH) and the state of research in the receiving discipline (literary theory) is unbalanced. It can be the case that (a) the current state of research in DH, (b) the current state of research in the study of literature, are too different to permit beneficial transfers between the two disciplines.[4] The result of an analysis could be either too complex or too simplistic to be transferred to the other discipline. As a consequence, some papers on computational studies seem to literary scholars more a way "of showing that his tools work rather than using them to uncover new readings" (Hammond 2017, 5). Additional obstacles include the fact that OCR programs may be unable to recognize medieval texts written in ancient characters and that contemporary authors cannot be studied because their texts are protected by copyright and are thus not available for digitalization.

Despite these limitations, a mixed-method approach opens up a considerable field of research and demands that prejudices regarding quantitative and qualitative research, as well as traditional concepts such as "evidence" and "interpretation," be overcome. Moving beyond these barriers, Jannidis and Lauer commented that "interpretations of the results of quantitative studies ... are hermeneutic acts of sense making" (2014:

---

[4] The ideas of Killian Koepsell and Carlos Spoerhase (2008), who reflected interestingly on the limits and potential of a transfer of knowledge from the cognitive sciences to literary theory, are relevant to our discussion.

50). This paper therefore adopts the view of Jean-Gabriel Ganascia who, in his 2015 article "The Logic of the Big Data Turn in Digital Literary Studies," proposes to overturn fruitless discussions regarding whether DH should be classified as a "natural science" or as "cultural studies." Indeed, in Ganascia's view, computer-aided methods are "a continuation of traditional humanistic approaches" (Ganascia 2015, 6).

In the renewal of literary theory and literary history initiated by the DH, the "operationalization" of literary and critical concepts plays a central role. As Moretti points out:

> Digital humanities may not yet have changed the territory of the literary historian, or the reading of individual texts; but operationalizing has certainly changed, and radicalized, our relationship to concepts: it has raised our expectations by turning concepts into magic spells that can call into being a whole world of empirical data and it has sharpened our skepticism because, if the data revolt against their creator, then the concept is really in trouble. A theory-driven, data-rich research program has become imaginable, [whose goal is] testing, and, when needed, falsifying the received knowledge of literary study. Of this enterprise, operationalizing will be the central ingredient. (Moretti 2013b, 15)

The digitization of texts offers new potentialities for the study of literature even as literary scholars confront the need to adapt their approaches to the demands of literature in the digital age, a literature that often is produced and consumed on digital devices. We are experiencing the third "reading revolution." After the invention of writing, 6.000 years ago, and of the Gutenberg printing press in the 15th century, the introduction of digital texts and the arrival of the e-books is changing our reading minds and the Western culture which is based on the "Order of the Book" (Van der Wheel, 2). Our cultural heritage is being digitized very soon as of in 2014 Project Gutenberg offered free access to over 46.000 books.

In conclusion, the DH offer two very different set of tools to literary studies: to preserve texts through digitization and to study them through computational analysis. In our research we focused exclusively on the latter ones. The more such tools can be adapted to the present, the more a precious legacy of literary history can be salvaged for future generations.

Works Cited

Amann, Klaus, Corino, Karl, and Fanta, Walter. *Robert Musil, Klagenfurter Ausgabe: Kommentierte digitale Edition sämtlicher Werke, Briefe und nachgelassener Schriften*. Klagenfurt, Austria: Robert Musil-Institut der Universität Klagenfurt, 2009.

Arntzen, Helmut. *Musil-Kommentar sämtlicher zu Lebzeiten erschienener Schriften außer dem Roman "Der Mann ohne Eigenschaften."* Munich: Winkler, 1980.

Ajouri, Philip, Mellmann Katia, and Rauen Christoph, eds. *Empirie in der Literaturwissenschaft*. Münster, Germany: Mentis, 2013.

Broch, Herrmann. "Mythos und Altersstil." *Schriften zur Literatur 2: Theorie*, Ed. P.M. Lützeler. Frankfurt: Suhrkamp, 1995. 212-234.

Bryman, Alan. *Quantity and Quality in Social Research*. London: Uniwin Hyman, 1988.

Burrows, John. "Delta': A Measure of Stylistic Difference and a Guide to Likely Authorship." *Literary and Linguistic Computing*, 17.3 (2007): 267–287.

Ciotti, Fabio. "Toward a Formal Ontology for Narrative." *MATLIT: Materialidades da Literatura*, IV, 1, 29-44, 2016, doi: 10.14195/2182-8830.

Corbetta, Piergiorgio. *Metodologia e Tecniche della Ricerca Sociale*. Bologna: Il Mulino, 2014.

Corino, Karl. "Robert Musil, Aus der Geschichte eines Regiments." *Studi Germanici* 11 (1973): 109-115.

Corino, Karl. *Robert Musil: Eine Biographie*. Reinbek bei Hamburg, Germany: Rowohlt, 2003.

Corino, Karl. "Klaviersonnen über Schluchten des Gemüts: Robert Musil und die Musik." *Das Plateau*, 120 (2010): 4-21.

Dinklage, Karl. *Robert Musil: Leben, Werk, Wirkung*. Zürich: Rowohlt, 1960.

Eco, Umberto. *Lector in Fabula: La Cooperazione Interpretativa nei Testi Narrativi*. Milan: Bompiani, 1979.

Eder, Maciej, Kestemont, Mark, and Rybicki, Jan. "Stylometry with R: A Package for Computational Text Analysis." *The R Journal* 8.1 (2016): 107-121. https://journal.r-project.org/archive/2016/RJ-2016-

007/index.html. Last visited 31 May 2018.

Ganascia, Jean-Gabriel. "The Logic of the Big Data Turn in Digital Literary Studies." *Frontiers in Digital Humanities*, 2015. https://doi.org/10.3389/fdigh.2015.00007. Last visited 10/10/2018.

Hammond, Adam. "The Double Bind of Validation: Distant Reading and the Digital Humanities' "Trough of Disillusionment." *Literature Compass* 14:8 (2017): 1-13.

Herrmann, Julia Berenike, Van Dalen-Oskam, Katrin, and Schöch, Christoph. "Revisiting Style, a Key Concept in Literary Studies." *Journal of Literary Theory* 9.1 (2015): 25-52.

Jannidis, Fotis, and Lauer, Gerhard. "Burrows's Delta and its Use in German Literary History." *Distant Readings: Topologies of German Culture in the Long Nineteenth Century*. Eds. M. Erlin and L. Tatlock. Rochester, NY: Camden House, 2014, 29-54.

Juola, Patrick. "Large-Scale Experiments in Attribution of Authorship." *English Studies* 93.3 (2012): 275-283.

Kastenbaum, Robert. "Program of the Thirty-Eighth Annual Scientific Meeting of the Gerontological Society of America." In Nasar, Jack and Farokhpay Mitra, "Assessment of Activity Priorities and Design Preferences of Elderly Residents in Public Housing: A Case Study," *The Gerontologist* 25.3 (1985): 251-257.

Kleinwort, Malte. *Der späte Kafka: Spätstil als Stilsuspension*. Paderborn, Germany: Fink, 2013.

Koepsell, Killian and Spoerhase, Carlos. "Neuroscience and the Study of Literature: Some Thoughts on the Possibility of Transferring Knowledge." *Journal of Literary Theory* 2:2 (2008): 363-374.

Lauer, Gerhard. "Introduction: Empirical Methods in Literary Studies." *Journal of Literary Theory* 9.1 (2015): 1-4.

Margolin, Uri. "On the Vagueness of Critical Concepts." *Poetics* 10:1 (1981): 15-31.

McMullan, Gordon and Sam Smiles, eds. *Late Style and its Discontents: Essays in art, Literature, and Music*. Oxford: Oxford University Press, 2016.

Moretti, Franco. *Distant Reading*. London, New York: Verso, 2013a.

Moretti, Franco. "Operationalizing": or, the Function of Measurement in Modern Literary Theory, Pamphlet of the Stanford Literary Lab, December 2013b, 1-15. https://litlab.stanford.edu/LiteraryLabPamphlet6.pdf.

Orlandi, Fernando. "Nota Editoriale." *Robert Musil: La Guerra Parallela*. Ed. Fernando Orlandi. Rovereto, Italy: Reverdito, 1987.

Patton, Michael Quinn. *Qualitative Evaluation and Research Methods*, Newbury Park, CA: Sage, 1990.

Pierazzo Elena, Tiziana Mancinelli. *Che cos'è un'edizione scientifica digitale?*. Rome: Carocci, 2018.

Rebora, Simone and Salgaro, Massimo. "Is Late Style Measurable? A Stylometric Analysis of Johann Wolfgang Goethe's, Robert Musil's, and Franz Kafka's Late Works." *Elephant&Castle*, 18, 2018, 4-39.

Rebora, Simone, Herrmann, Berenike, Lauer, Gerhard, and Salgaro, Massimo. "Robert Musil, a War Journal, and Stylometry: Tackling the Issue of Short Texts in Attribution of Authorship." *Digital Scholarship in the Humanities* fqy055, https://doi.org/10.1093/llc/fqy055, 2018.

Ronzon, Francesco. *Sul campo: Breve Guida alla Ricerca Etnografica*. Rome: Meltemi, 2008.

Roth, Marie-Louise. *Robert Musil: Ethik und Ästhetik*. Munich: List, 1972.

Said, Edward W. *On Late Style: Music and Literature Against the Grain*. Pantheon: New York, 2006.

Salgaro, Massimo. "Albert Ritter, der Ghostwriter in der Redaktion der Tiroler Soldatenzeitung—ein biographisches Profil." Oberleutnant Robert Musil als Redakteur der Tiroler Soldaten Zeitung. Eds. Di Mino, M., E. Locher, and M. Salgaro. München: Fink Verlag, in press.

Schaunig, Regina. *Der Dichter im Dienst des Generals: Robert Musils Propagandaschriften im ersten Weltkrieg*. Klagenfurt, Austria: Kitab, 2014

Schöch, Christof, Schlör, Daniel, Zehe Albin, Gebhard, Henning, Becker Martin, and Hotho Andreas. "Burrows' Zeta: Exploring and Evaluating Variants and Parameters." *DH2018 Book of Abstracts*. Mexico City: ADHO, 2018: 274-278.

Tauszik, Yla R. and Pennebaker James W. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods." *Journal of Language and Social Psychology* 29 (2010): 24-54.

Van der Weel, Adriaan. *Changing Our Textual Minds: Towards a Digital Order of Knowledge* Manchester: Manchester University Press, 2011.