1: Burrows's Delta and Its Use in German Literary History

Fotis Jannidis and Gerhard Lauer

Introduction

LITERARY HISTORY OFFERS A guide to the canon of great books. It matters little whether one picks up a history of literature from the nineteenth century or a modern one; they all tend to more or less salvage a small collection of books from the ocean of those published. For the year 1809, for example, any given history of German literature will highlight Johann Wolfgang von Goethe's novel *Die Wahlverwandtschaften* (Elective Affinities), as if no other work of literature had been published that year. This is the case regardless of whether one consults Hermann Hettner's literary history from 1870 or R. H. Stephenson's essay on the novel in Weimar classicism from 2005.[1] A brief perusal of a book catalogue from around 1809, however, suggests an alternative to this canonical picture of literary history. Approximately one hundred German novels were published in 1809. Among the widely read novels included in this list are, for example, the third volume of August Lafontaine's *Die beiden Bräute* (The Two Brides) and August Kotzebue's *Philbert oder die Verhältnisse* (Philbert or the Circumstances). The canon is one story; the cultural history of read books is another. For those who believe that literary history should be more than the history of great books, a corpus-based approach offers one way of dealing with the thousands of books that actually circulated. And we are not using "thousands" metaphorically in speaking of the literary history of the nineteenth century, even when we limit ourselves to the German-speaking countries. A periodical such as *Das belletristische Ausland* (Belles Lettres from Abroad), published from 1843 to 1865, contains 3,618 titles of books translated for the German literary reading audience.[2] Moreover, in the year 1871 alone nearly a thousand new books in the category *belles lettres* were published. And in 1913 more than five thousand books of prose and poetry came out.[3] The corpus of nineteenth-century German books is huge and—as Gregory Crane wrote in his seminal paper, "What do you do with a million books?"—life is too short to read even what was published in a single year of the long nineteenth century.[4]

The number of texts and the ability to access them has always had an influence on philological studies, and new access to sources is a

game-changing event for philologists. When the humanists of the fifteenth and sixteenth century searched the libraries of monasteries for forgotten classical manuscripts, their findings were the basis of a new understanding of history, natural science, and the humanities. When the founding fathers of the new university-based philology, such as Jakob and Wilhelm Grimm, did their research around 1800, access to the medieval sources, which had been mostly neglected until then, became a key factor in forming the foundations of the new discipline. We find ourselves today in a similar situation with the new access to ever-larger collections of literary texts. To be sure, quantitative studies of literary texts have long been undertaken in the context of what we now call digital humanities, but only in the last five years or so has the number of texts available allowed for a quantitatively new form of large-scale research. In the case of German texts, availability is still more recent; only since 2011, with the publishing of the Digitale Bibliothek by TextGrid, has the large collection of texts digitized by Directmedia and later posted online via Zeno.org been encoded with the metadata necessary to make these texts usable for corpus research.[5]

Together with the extraordinary increase in computer processing power, the exploding number of digitally available texts alters the landscape of humanities scholarship in more than one respect. If it had not been the case earlier, then at the very latest when Google digitized about 20 million books from the estimated total number of 130 million unique books in the world, published in 480 languages, books became a corpus.[6] The existence of such large corpora gave rise to a new interest in tools to extract information from them, and Google's widely discussed Ngram Viewer provided a popular example of the kind of research that was now possible. Although the obvious weaknesses of this approach were quickly pointed out, it remains one of the first examples of a tool that allows every humanist—and not just the specialist—to query a vast corpus of texts.[7]

As part of this ongoing transformation, common reading practices have been subjected to renewed scrutiny. It is obvious that millions of books cannot be read or interpreted using existing methodologies of close reading. Coining the term "distant reading" in 2000, Franco Moretti suggested a different approach, claiming that "literary history will quickly become very different from what it is now: it will become 'second hand': a patchwork of other people's research, *without a single direct textual reading*. Still ambitious, and actually even more so than before (world literature!); but the ambition is now directly proportional *to the distance from the text*: the more ambitious the project, the greater must the distance be."[8]

These are, to be sure, polemical words intended to mark a shift in scholarship, whereby distance presents new opportunities for knowledge at the price of a loss of familiarity with individual books. It is thus no

wonder that Stanley Fish and others have attacked Moretti for giving up the intimate relationship with books as friends and exchanging interpretation for mere pattern recognition.[9]

In his response to Fish, Martin Mueller proposed the term "scalable reading" to emphasize that there is no such thing as an opposition between close and distant reading.[10] In his view, these positions differ instead in relation to the amount of text they process; we use corpora or text collections "to perform rough mapping operations that are then followed by a targeted examination of selected examples."[11] As researchers, our position on the continuum that runs from close to distant reading depends on our questions. Obviously, even a literary history concentrating on very few canonical texts has always required a greater distance from a single text than an interpretation of Goethe's *Wahlverwandtschaften*. And this is not the only aspect of the debate on distant reading that is misleading. Although the debates regarding Moretti's term suggest otherwise, there is in fact a long scholarly tradition in literary studies and linguistics that makes use of such techniques as counting words or calculating patterns. Established traditions of lexicometry, research on authorship attribution, stemmatology, concordance, and phylometry have existed for over one hundred years, although they constitute but a very small fraction of humanities scholarship. Technology now allows us to take additional steps in the directions indicated by this earlier research. Using Google's Ngram Viewer, for example, we can explore cultural trends on the basis of a corpus of not less than 4 percent of the books ever printed—an amount of text that no one before the age of digitalization could have handled.[12] And this viewer is not the only new tool for quantitative and corpus-based research. Machine learning, trained on a set of learning examples to work on unseen data; topic modeling, which identifies shared themes through statistical models; and social network analysis for actor pattern recognition—these are just some of the methods available today for answering longstanding questions in the humanities.[13]

Stylometry is another example, and like other quantitative approaches, it too has a long history, predating the digital age. Stylometry has figured prominently in research on authorship attribution and has been used mostly on a small corpus of books. In 1851, the mathematician August de Morgan first proposed using the statistical average of word length, measured in syllables, as a criterion for determining authorship.[14] His test case concerned the epistles of Saint Paul, but others soon followed, such as Thomas Corwin Mendenhall with his word-length studies on Shakespeare/Marlowe.[15] After 1960 linguists expanded the methodology to address larger corpora and examined new features of texts such as average sentence length, vocabulary richness (on the basis of the type-token ratio), homogeneity (based on the number of lemmata within a text), syntactic features, and word classes.[16] Currently, author attribution studies is making use of

a growing range of measures and ever more complex forms of analysis of these factors.[17]

Methods

Since John F. Burrows introduced the Delta measure in his groundbreaking paper of 2002, it has become a common standard in stylometry for establishing the relative stylistic difference between two or more texts.[18] Burrows initially developed his method in the context of author attribution studies in order to enable researchers to create a short list out of a larger group of possible candidates for the authorship of an anonymous text. In the beginning, in other words, it was basically meant to reduce a large set to a small one in order to allow more complex and time-consuming procedures to be applied to this smaller group. The procedure is fairly simple. Like much of Burrows's other work, it is based on counting the most frequent words (MFW), without the use of a stop-word list. Therefore, especially the top of the list consists of words that have almost no semantic importance at all (e.g., "the," "is," "a"). Burrows based his practical demonstration of Delta on longer epic poems; in the years following his publication, studies have duplicated the high success rate of correct attributions with other types of texts (mainly novels)—sometimes in other languages.[19] Today we can say that at least for some languages, such as English, Burrows's Delta has proven to be a remarkably good indicator of stylistic affinity. But we must not be misled: it is nothing like a "stylistic fingerprint," "stylistic DNA," or any other forensic feature to which it might be compared, because Burrows's Delta does not identify individual features but rather describes the relation between a text and other texts in the context of the entire group of texts. If one changes the group, one will also see new relations and maybe new clusterings.

As Burrows himself and follow-up studies by David L. Hoover and others have shown, however, with texts longer than two thousand words, Delta has a high probability of correctly indicating the author—that is, if other texts by the author are included in the comparison. Some differences in the validity of results have been found in the case of prose and poetic texts of shorter length. Due to its brevity, poetry is much more difficult to analyze quantitatively. Applied to texts in Polish or Latin, the results of using Burrows's Delta are less satisfying, which has been attributed to the fact that these are highly inflected languages.[20] Still under discussion is whether (and if so, which) words should be removed (e.g., personal pronouns), which class of words yields the highest accuracy, and what range of the most frequent words leads to the best results—the first hundred or the first eight hundred?[21] In sum, Burrows's Delta has in its short time of existence become well established. However, a great deal of language-specific research remains to be done to understand its strengths and weaknesses in depth.

The use of Delta for research has been greatly facilitated by the implementation of a script written in the open-source statistics language "R." To date, this script—or rather these scripts since the tool has evolved since its first publication to a set of scripts—has been developed by Maciej Eder and Jan Rybicki, who were recently joined by Mike Kestemont.[22] There are other algorithms, scripts, and tools available, but the ease of use of Eder and Rybicki's script makes it our preferred choice: it permits the user to manipulate many different parameters without any need to program.[23] It comfortably allows for a wide range of analytical combinations of style-marker settings, such as "culling" the number of the MFW analyzed, whether or not pronouns will be deleted, and so on. (Note the culling rate specifies the percentage of texts in a corpus in which a given word must be found in order to be included in the analysis.) Given the combination of a relatively simple statistical measurement and the existence of a tool that makes it easy to apply Delta to different texts, there are good reasons for us as nonexperts in the field of language statistics to test it against our knowledge of literary history.

In this chapter we will explore the possibilities that this analytical tool offers those working with larger collections of German texts. Our goal is to evaluate Delta and its use for German literary history. Burrows himself pointed out very early on that stylistic measurements could be used in other fields of literary study outside of authorship attribution, and he has done some research on literary epochs using such measurements.[24] We do not expect any dramatic new insights from this application; 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. By the same token, if we successfully and to a high degree confirm traditional knowledge with this approach, we can then—in cases where the results do not corroborate previously held views—start to investigate and maybe even question traditional knowledge.

We will approach this analysis from two different angles. First, we will use Delta in a series of different tests to group texts from around 1800 according to authorship, epoch, genre, and gender. Second, we will look at a specific problem of individual writing—the position of Heinrich von Kleist in literary history—with the aim of establishing what this instrument can contribute to debates surrounding Kleist's oeuvre.

The Corpus

Our first goal is to evaluate the usefulness of Burrows's Delta for the study of German literary history by applying it to a set of tasks, including author attribution and classification based on genre, time, and gender.

Since we know beforehand what the results of our measurements should be, we can evaluate our tool by comparing the real outcome to the expected results. For a number of reasons, this undertaking can only be a first exploration of this kind of evaluation. The first and most important one is pragmatic—the entire undertaking is based on the validity of our corpus and its metadata. Because the text collection is so huge, we concentrated first on the novels. They provide a larger chunk of text, making the use of Delta more reliable.[25] They are also relatively few in number—only about 430—compared with the thousands of poems.

The texts are part of one of the first large-scale digitization projects in the German language, and most of them are based on scholarly editions used by academics. On the positive side, this means the texts are acceptable for scholarly work, but it is also the case that most of them were published at a time when it was considered good practice to modernize the spelling. For this reason, the corpus is probably not of interest to anyone doing research on the history of spelling or other features dependent on the exact form of the text surface. On the other hand, this modernization of spelling has solved or at least mitigated a problem that heavily burdens all historical language research: the vast variety of spellings in earlier writing systems—including those in Germany before 1800. To put it another way, modernization has in a way done what otherwise would have to be done by the researchers—the normalization of writing.

The corpus is not balanced or representative of the literary production of the period it covers. The main reason for this imbalance is its genesis: it has been compiled from a database that is itself based on a collection of texts produced by a commercial company, Directmedia, over the course of ten years.[26] The company sold digital texts on CDs and DVDs—a total of 165 disks in its main series. The bulk of this collection is copyright free and has been released with a very generous license.[27] One of the disks, the largest collection by far, contained most canonical literary texts up to 1930 and a second one contained literary texts by women. These collections and some other texts are all included now in the section of the database published on Textgridrep.de. Women are thus overrepresented in this corpus relative to their share of the canon and maybe even in relation to their share in literary production.

The collection also contains translations. While it has been shown by Rybicki, for example, that stylometry can be used with interesting results in translation studies, these particular translations represent a very small canon.[28] In addition, because the contents of yet another DVD with erotic texts went into the collection, the canon of erotic literature in the eighteenth century is slightly overrepresented in the collection. We do not mean to imply that only canonized works should be studied but intend instead to point out that the text collection at hand is, especially in some aspects, unlikely to be representative of the totality of published

works. The existing metadata also had to be extended for our research. Dates of publication had to be extracted to be machine readable, and other such potentially significant parameters as literary period had to be added; we also had to add the gender of the author and indicate whether a novel is a translation.

The problems are not confined to the composition of the corpus alone. The generative histories of extensive works such as novels are often long and complicated, and the significant variation among them cannot be easily captured in the Spartan metadata of a corpus. For practical reasons, we decided to use a single field containing the date of the first publication of the first volume or the year of the first number of the first publication in a journal, ignoring the fact that many novels were published over a range of years and that sometimes there was a considerable time gap between the writing and the publication of a novel. We kept to this rule even in the case of posthumously published texts.

Results

With these qualifications in mind, let us start with a simple experiment that looks at all sixty-three German novels in the corpus between 1785 and 1815. We will mainly discuss visualizations of two different procedures: running the algorithm once, which usually results in a dendrogram, or running it repeatedly and adjusting one parameter slightly from run to run, which usually results in a consensus tree. Both the dendrogram and consensus tree show clusters as branches of a tree. The dendrogram contains more information because a greater branch length indicates a greater distance between the two Delta values, which in this case means a greater dissimilarity of style. The consensus tree, on the other hand, is more robust because it is based on repeated runs and it shows the information that is stable across the runs. In figure 1.1, one can see the results from runs starting with the eight hundred MFW up to three thousand, increased by a hundred for each run. Some of the information that can be found in a dendrogram is lost in a consensus tree: a consensus tree does not say anything about the distance of two branches starting from the root. Therefore, in the worst-case scenario, one would see all texts directly connected to the root, which means that there is no reliable information about their similarity at all.[29]

One parameter that can be adjusted is the consensus strength (with values between zero and one). Consensus strength provides a measure of agreement between the original trees; if partitions do not share enough information with those in other trees, they are below this point and will not be shown in the consensus tree. In our example, we used a consensus strength of 0.5.

We can see two sections in the image, the first being the long branch containing mainly Jean Paul's novels and the wheel at the top.

Figure 1.1. Novels between 1785 and 1815: Consensus tree, classic Delta, 800–3000 MFW, consensus 0.5

This wheel shows nine other branches all starting from the center. This configuration basically means that one cannot say anything substantive about the stylistic distance between these branches. One point is rather obvious: almost all the authors are correctly assembled into subbranches. In other words, Delta is indeed a very good indicator for authorship for this group of texts. There seems, however, to be one error: Friederike Helene Unger's novel *Bekenntnisse einer schönen Seele* (Confessions of a Beautiful Soul) from 1806 is not grouped together with her novel *Albert und Albertine* (1804). Yet this configuration may not be an error at all but rather an insight. *Bekenntnisse* was published anonymously, and it is not at all clear whether the novel really is by Unger. Two other writers have been mentioned as possible authors of the book—namely, Paul Ferdinand Buchholz and Charlotte von Ahlefeld.[30] If we focus on a smaller group of texts (classic Delta, three thousand MWF), including another novel by Unger and other female authors and novels by Ludwig Tieck and Goethe, we see that *Bekenntnisse* is still not grouped together with Unger's other novels (fig. 1.2).

In dendrograms like these, only the horizontal lines have a meaning: their length indicates the similarity between two entities—in our case, between texts. Goethe's novels are therefore clustered very closely,
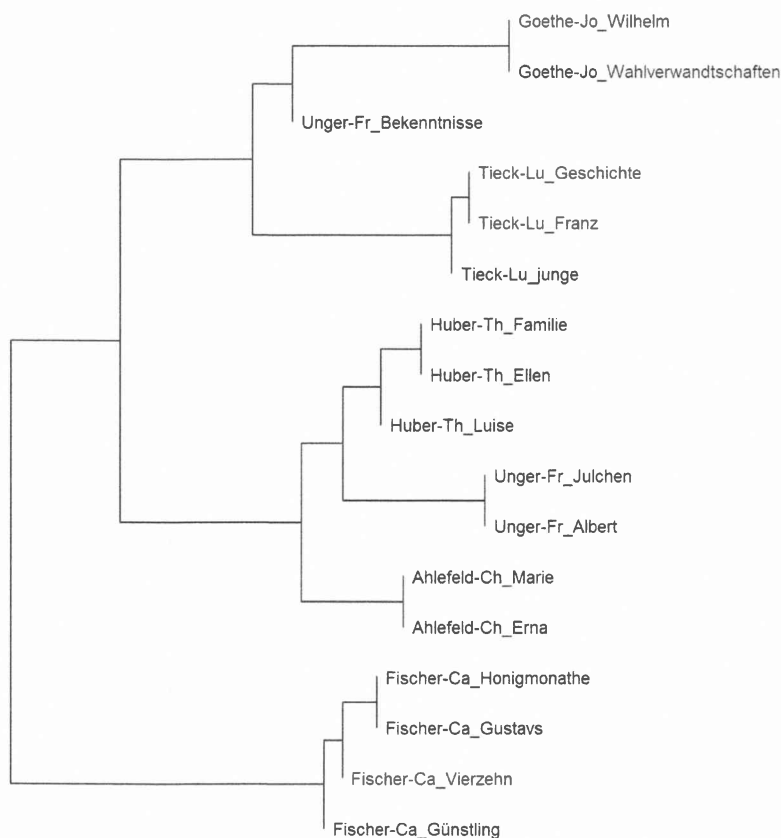
Figure 1.2. Friederike Helene Unger's *Bekenntnisse einer schönen Seele* among the novels: Dendrogram, classic Delta, 3000 MFW

indicating a high similarity, while the novel *Bekenntnisse* is marked as similar to but at a distance from the novels of Goethe. All texts in one branch are more similar to each other than to the other texts in other branches, and this is true as one moves along farther to the left. Thus Goethe's novels are more similar to Tieck's than they are to the group of female writers including Unger, Therese Huber, and Ahlefeld, but all these texts show a greater similarity compared to the outlier group consisting of the novels of Caroline Fischer.

Bekenntnisse does not cluster together with the novels of Ahlefeld or with those of Fischer or Huber. This result seems to be relatively robust, reoccurring in analyses run with different settings and the inclusion of various additional novels by other authors. It therefore would probably be worthwhile to have a closer look at the third candidate, Buchholz. But

as none of his texts are currently available in digital form, for the time being we cannot pursue this line of inquiry any further.

It is a well-established fact in stylometry that genre is one of most important aspects of style. In other words, one can expect that genre will be a good classifying feature. Since we determined that authorship is a very good classifying feature, in the following run we kept this dimension invariant (fig. 1.3). In this test, we used a collection of thirty-nine texts from different genres but all written by the same author—namely, Goethe. For the analysis, we attributed genres to individual works according to conventional scholarly classifications, indicating this information in the file name with the addition of a prefix to the title. As figure 1.3 shows, Goethe's works group nicely, but there are also some interesting unexpected results.

There are two distinct groups: one consists of four subgroups and the other consists of the one group shown at the bottom. The group at the bottom contains all prose texts, the novels, the autobiographical texts, and the smaller novellas. The four groups at the top contain

Figure 1.3. Goethe's work, grouped according to genre: Dendrogram, classic Delta, 1490 MFW, culling 50 percent

Goethe's poetry, both lyric and epic, and plays. Counting down from the top, groups 1 and 2 contain plays (with only one exception), with texts that are stylistically nearer in proximity to each other. Group 3 contains only lyrical texts, and group 4 is a mix of four plays, two epics, and two collections of lyrical texts that group internally along these genre divisions. Obviously time is also a factor here: The plays in the top group were written mostly by the young Goethe or in the early years of his stay in Weimar, while *Tasso* (1790), *Faust* (at least in this 1808/1832 version), and *Iphigenie* (1787) are from his classical period. The poems collected in *Zahme Xenien* (Gentle Xenia) were written in two phases—the first mostly soon after 1815 and the second in the years 1824 to 1827. The *Inschriften, Denk-und Sendeblätter* (Inscriptions and Occasional Pieces) were also written during these late years.[31] Goethe's early novel, *Die Leiden des jungen Werther* (The Sorrows of Young Werther, 1774), is an interesting outlier, because it is the only prose text where the genre attribution did not work. This failure can be easily explained by the fact that it is an epistolary novel consisting mostly of a single voice—that of Werther. But the "false" grouping could also be understood as a signal to look closer at the affinity between the dramatic monologues and this novel. All in all we have seen that, as expected, genre is indeed a critical factor in clustering texts using a stylistic measurement such as Delta. However, it is not as reliable as authorship—a fact that could be seen as an ironic gloss on debates about authorship in the wake of Barthes's and Foucault's famous essays on the death of the author.[32]

After testing Delta as a means to classify novels based on authorship and genre, we were interested to see how well it worked for literary periods. To this end, we compared two test sets of authors who are regarded as typical of their respective epochs (Enlightenment and realism). It is astonishing enough that the twenty-four novels cluster so neatly along the period boundaries. As mentioned previously, a dendrogram the length of a horizontal line indicates the difference between two entities—in our case the difference of styles. In the dendrogram in figure 1.4, it is easy to see that there are two groups of texts that are more similar to each other than any text of the other group and that this clustering happens along the boundaries of the two literary periods. But maybe this clustering is simply an effect of the words used. What happens when we ask the script to use only those words that are common to all texts? The clusters stay the same. Even when we repeated the test with a larger group of texts, we had the same result but with one exception: when we used only those words common to all texts (culling at 100 percent), both novels by Otto Ludwig were grouped with the Enlightenment texts. This grouping was, however, actually a result of the fact that, with culling 100, only 414 MFW were shared by all texts and thus the reliability dropped sharply. As long as 2,000 words were used to compute Delta, the clustering
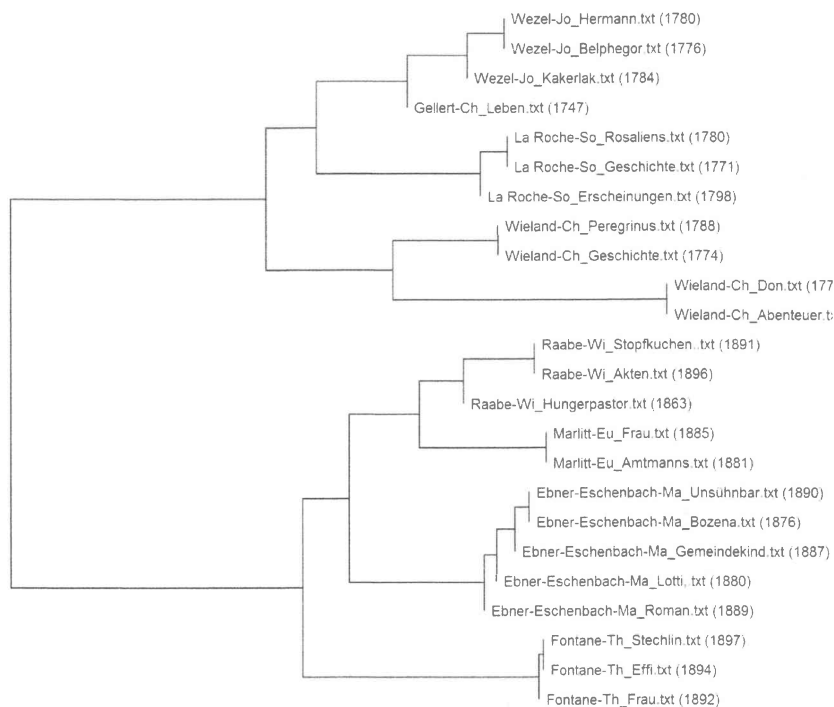
Figure 1.4. Distinction between literary epochs, Enlightenment versus realism: Dendrogram, classic Delta, 2000 MFW

according to epoch remained stable in our test groups. For the most part, then, the stylometric approach reveals a clear-cut division between the texts of the Enlightenment and of realism, although we will discuss examples where the two epochs are not so far apart. For the moment, it is perhaps enough to point out that a simple procedure to classify unknown texts reliably into a specific epoch or time span can be a valuable tool for handling very large corpora.

Additional insights can be gained if we take books into account that have often been omitted by literary history or put into separate categories, such as women's writing. For our next analysis (fig. 1.5), we chose twenty novels written by women and seventeen written by men to see how these German novels from around 1800 group according to their stylistic features. We decided to delete the pronouns in order to eliminate the influence of differences in the gender of the novels' protagonists on our results, although in fact this element itself could be regarded as an important distinguishing feature.

Female and male authors cluster mostly separately from one another, but this separation could simply be an effect of good authorship

Figure 1.5. Male and female authors around 1800: Consensus tree, classic Delta, 200–2500 MFW, consensus 0.5, pronouns deleted

attribution. (And again *Bekenntnisse einer schönen Seele* shows a noticeable distance from Friederike Helene Unger's writings.) Although we did find some evidence for classification based on the gender of the author, we were for the most part unable to separate the texts of all male from all female authors with any real certainty. The results were dependent on the composition of the group of texts as a whole—that is, in some constellations, texts by female authors cluster to a high degree and can be easily distinguished from the cluster(s) of male authors, while in other constellations, using other authors and texts, female and male authors are mixed.

The really interesting thing here is the literary landscape around 1800 that becomes visible through this mapping out of stylistic affinities. Ahlefeld and Huber form a literary province of female writing. But other female writers such as Caroline de la Motte Fouqué, who belongs to the romantic camp, or Benedikte Naubert, best known for her well-received historical novels, cluster near male authors. There seems to be no general stylistic feature of female writing per se but instead a certain

limited number of positions in the literary field around 1800. Some of these positions are indeed occupied exclusively by female authors. But other authors, such as Dorothea Schlegel, write like Goethe, and still others such as Fouqué do not appear to occupy any of the female positions available in the literary field of that time. This result could be used to 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.[33] With such findings, distant reading opens up a view of the structure of a historical literary field that is more than a mere replication of the picture created by the canon. It seems to us that this constellation should be explored in greater detail in the future, especially since we also found that some of the classifications were anything but stable and seemed to be especially dependent on the settings of the parameters.

In another series of tests, we analyzed the historical position of an individual author (fig. 1.6). Literary historians have seen Heinrich von Kleist's work as oscillating between classicism and romanticism and have long debated the classification of his work. He is thus a particularly promising test subject for exploring whether stylometry can contribute to a
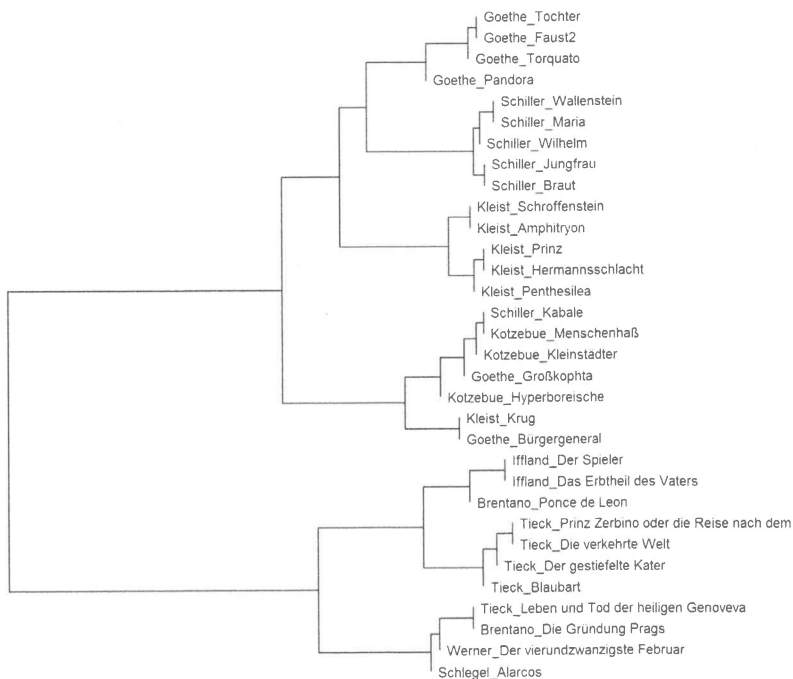

Figure 1.6. Kleist's plays among other plays around 1800: Dendrogram, classic Delta, 2000 MFW

better understanding of how the works of an individual author relate to the broader literary trends of the age. As a pretest, we took a smaller corpus of thirty-two highly canonical works of German dramatic literature around 1800 that also included some popular plays by Iffland and Kotzebue.

In this analysis, as in the previous ones, the texts that constitute the corpus are plain text files, which give rise to a host of problems on the level of textual detail. Plays present the speech of characters, which is not comparable with the voice of a narrator. And especially the longstanding tradition of distinguishing comedy and tragedy might be a hint that drama consists of not one genre but two. We bracketed these concerns for the time being and simply used a two thousand MFW analysis, without culling or deleting pronouns. The distance is measured once again by classic Delta.

As in the previous examples, the distance between the plays is represented by the length of the horizontal lines, while the vertical distance is not of any importance. The preliminary results suggested an answer to our initial question: Kleist is rather close to classical dramatists such as Goethe and Friedrich Schiller and at a marked distance from the group of romantic plays. Interestingly enough, we can see in the group between the classical and romantic plays, but closer to the classical texts, a wild mixture of texts by Goethe, Kotzebue, and Kleist. This clustering can be explained by the fact that these texts are comedies: it seems that the specific language of comedies marks them even more than their author does. The placement of Schiller's *Kabale and Liebe* (Intrigue and Love, 1784) into the context of comedy seems to be a fluke, unless it can be determined that this classification is indeed caused by the more comical parts of this play. On the other hand, Kleist's *Der zerbrochene Krug* (The Broken Jug, 1811) is probably rightly located closer to the domestic tragedy genre than Kleist's other comedy, *Amphitryon* (1807).

These results are only a first indication of how stylometric methods could capture Kleist's characteristic style using only formal style markers. In an attempt to draw more substantive conclusions, we decided to go a step further and take a list of forty-nine plays written or published between 1790 and 1811.

In contrast to the dendrograms, figure 1.7 indicates that Kleist's plays cluster closely together with one exception, his *Käthchen von Heilbronn* (1810). With other Deltas such as that of Eder or Argamon's Delta, *Käthchen* also always appears at a distance from Kleist's other plays and in closer proximity to romantic plays. This result is consistent with the particular subgenre that Kleist chose for this work. He called it "ein großes historisches Ritterschauspiel" (a great historical knights' play).[34] The branch also retraces the development of Kleist's writing of tragedies, starting with his first tragedy, *Die Familie Schroffenstein* (1803), and
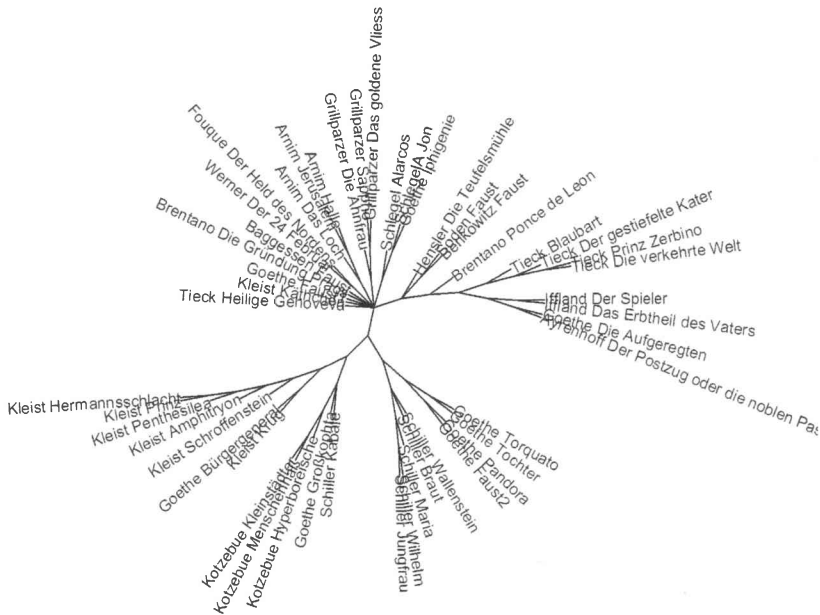
Figure 1.7. Kleist's plays and other plays around 1800: Consensus tree, classic Delta, 200–2000 MFW

ending with the very similar *Hermannsschlacht* (The Battle of Hermann, 1808) and *Prinz von Homburg* (1809). The algorithm maps the dramatic work of Schiller in a parallel fashion.

To avoid any form of cherry picking—that is, in order not to suppress evidence that might contradict the expected results—we have to scrutinize more closely the precise impact of deleting versus not deleting pronouns, different culling rates, and the use of specific Delta algorithms. As can be seen in figure 1.8, with culling, the plays are more similar to each other than without, and they seem to lose some of their specific stylistic features, here rendered more salient by the deletion of pronouns. With culling, the algorithm seems to throw out too many of those words that distinguish the characters of *Käthchen* from Kleist's other characters, insofar as plays consist mainly of the direct speech of characters.

The chosen Delta also yields different results, although only slightly different, because no matter whether we choose classic Delta or Eder's Delta, except in the case of *Käthchen*, Kleist's plays are always close to one another and always at a distance from the romantic plays of his time (fig. 1.9).

Taken together, these consensus trees and the dendrogram from the beginning of our second series of tests yield a fairly robust result. Kleist's

Figure 1.8. Kleist's plays and other plays around 1800: Classic Delta, 200–2000 MFW, pronouns deleted, consensus 0.5: (a) culling 0.5, (b) culling 0.0

Figure 1.9. Kleist's plays and other plays around 1800: 200–2000 MFW, consensus 0: (a) classic Delta, (b) Eder's Delta

style of drama very clearly separates him from the romantic writing of his time. Only his *Käthchen* play clusters closer to the romantic plays. On the other hand, dendrograms and consensus trees generate notably differing results with regard to the closeness of Kleist's plays to the domestic tragedies of this era. Another difference between a 2,000 MFW and the 200–2,000 MFW algorithm is the identification of genre. The dendrograms are better able to divine the differences between comedy and tragedy, while the consensus tree is better at plotting authorial style. Further research is needed to determine whether analysis of drama in general yields more diverse results than that of novels. It is also probable that by around 1800, the German drama had already developed many subgenres, while the novel, an emerging genre, had at that time ramified into fewer subgenres.

Our last arrangement is a bit unusual, even in the context of the rather unfamiliar world that the quantitative analysis of texts still represents. As we pointed out previously, the stylistic affinities between texts can be reliably interpreted as providing an overview of the literary landscape in this epoch, especially if each author is represented by more than one text and, better still, all or at least most of her or his novels. Other instruments and methods are available to reconstruct such a landscape, but usually we use social factors (as in a literary field analysis as modeled by Pierre Bourdieu) or a mix of different factors (as is typical in traditional literary history). One crucial question would thus seem to be how we can relate the clusterings, groupings, and stylistic affinities, which seem to be suggested by the stylistic analysis, to those mapped by these other methods. Is it possible to generalize our findings for specific works to the work of an author in general? One approach we tested (fig. 1.10) was to put all novels by one author into one file and run the Delta analysis on this new configuration.

This approach shows us the stylistic proximity or distance of the authors and can be read as a (incomplete) map of the literary landscape of this period. The map seems to corroborate some well-established views: Romantic authors such as Clemens Brentano, Joseph Freiherr von Eichendorff, Novalis, and Schlegel are clustered together, and they are quite distant from novels by Goethe, Schiller, and Christoph Martin Wieland. On the other hand, the map also raises some new questions. The cluster of female authors (Ahlefeld, Huber, Sophie von La Roche, Caroline von Wolzogen, and Wilhelm Karoline von Wobeser) that comes into view here has to date not really been seen by literary historians as a province of its own. And the proximity of Wilhelm Heinse, Karl Philipp Moritz, and Johann Karl Wezel to the romantic novels deserves closer investigation as well. E. T. A. Hoffmann's position in this field, his proximity to authors of the late Enlightenment and classicism, and especially his proximity to authors associated with
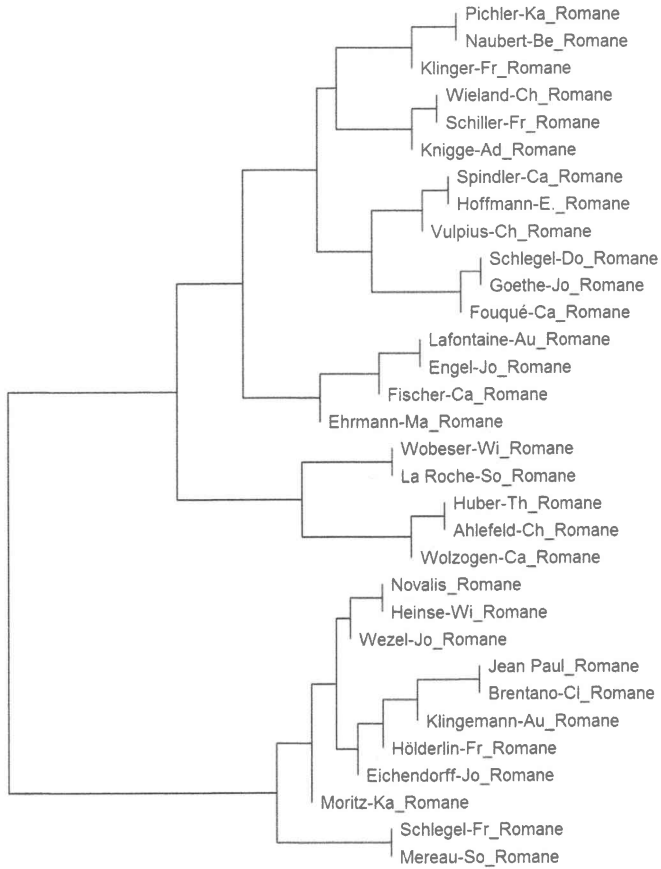
Figure 1.10. Authors of novels as a literary field around 1800: Classic Delta, 2000 MFW

"entertainment literature," could be seen as evidence that classification by period in literary history is based on a bundle of aspects; style, which is modeled here, is only one of these aspects.

Discussion

Stylometry is one example of distant reading. It shows how well quantitative analysis works for research in (German) literary history. It allows us to distinguish authors, such as Jean Paul or Kleist, from other authors around 1800, and it can discriminate among genres and gender, epochs and authors. Eder's and Rybicki's scripts offer style marker settings that transform quantitative text analysis from an arcane knowledge into a practical

tool. For the first time, a quantitative approach can work on large corpora, and we expect that literary historians will appreciate these new research possibilities as soon as they negotiate the well-known two-culture gap.

But stylometry is not a science machine that allows one to pose questions and simply wait for the script to cough up the results. Carefully designed series of tests and some knowledge of the analyzed texts are necessary preconditions for any successful project.[35] Otherwise distant reading runs the risk of predetermining results that confirm the original hypothesis. Furthermore, a better understanding of statistics would improve researchers' ability to do distant reading, and one could make a strong case for the inclusion of statistics courses in programs for literary and cultural studies in the future.

Burrows's Delta is useful for literary history, but it seems to generate results with different degrees of validity with regard to different texts and genres. Authorship attribution worked best in our test series. Genres sometimes could be differentiated, but it depended on the genre and its historical subclassifications. Discrimination of epoch also worked well within the limits of genre in our research with novels. Still, the corpora we used are not sufficient to answer questions on subgenres precisely enough. The effect of missing texts presents a similar problem. The lack of noncanonical but culturally significant texts, including, for example, relevant texts by Buchholz, limited the conclusions we could draw more than once. A major task for further research is therefore the procurement of better and larger corpora.

The test corpora used here do not offer a historically comprehensive map of the written, printed, and read books of their time. Instead they represent a sample of books canonized as a result of the choices made by literary historians, critics, and publishers. A culturally, historically, and socially representative corpus would look different. If we want to find out whether the group of mostly female writers really represents a discrete stylistic position in the literary field around 1800, one that exists alongside the canonized positions of classical and romantic authors, or if we want to decide whether Unger's *Bekenntnisse einer schönen Seele* was written by Buchholz, we simply need more texts—texts by additional male and female writers, including Buchholz himself. A corpus that is representative of the cultural history of books read in German-speaking countries around 1800 is still not available but would be a precondition for further research in literary history beyond the canon and even for the quantitative study of the canon. And even when we have all texts digitized, we have to know more about their distribution, the numbers of editions, and many more factors to make proper use of our tools.

One of the most interesting aspects of Burrows's Delta is the fact that it attends to the entire linguistic universe of a text. In contrast to the many approaches that make use of stop-word lists, Burrows's, Eder's, and

Rybicki's stylometry takes account of every single word. Psycholinguistic research makes good arguments for the significance of these seemingly innocuous words as contributors to individual style. As James Pennebaker and others have shown, function words such as pronouns, articles, prepositions, and auxiliary verbs serve as a good indicator of personal style.[36] Although it does not make exclusive use of such words, Burrowsian stylometry does rely heavily on these MFW. It would be important to know more about the link between psychology, word use, and stylometry. Such knowledge could provide explanations for the deeper mechanism behind the Delta.

The interpretations of the results of quantitative studies using Burrow's Delta are hermeneutic acts of sense making. On the one hand, they will be better if a researcher applies all her or his historical and textual knowledge to the results. On the other hand, the more contextual information we have, the bigger the danger that we try to make sense of everything, even the nonsensical or the random effect. There is no general remedy for this problem, but it helps formulate the question one is trying to answer in the clearest possible terms at the outset of the analysis.

Quantitative text analysis can be seen as a form of scalable reading because its methods allow us to establish the similarity between the styles of authors such as Jean Paul or Heinrich von Kleist and those of their contemporaries' It also enables us to cluster the writing styles of larger groups or literary epochs. It not only allows us to validate existing hypotheses about literary constellations but can also provide us with a more comprehensive snapshot of the literary field. Which works cluster, how many positions are available in the field under specific historical conditions, how large or how small are these positions—questions such as these could be better answered by integrating quantitative research methods into literary history. And MFW is not the only feature one could analyze. Sentiment analysis and genre-specific features such as narrative or dramatic style could serve as further variables in a multivariant quantitative analysis of literary history. Furthermore, comparing such results with analyses of other European, and maybe someday with other non-European texts, presents enormous challenges and opportunities. There are more things in heaven and earth than the canon, and distant reading presents a way of finding them.

Notes

[1] Hermann Hettner, *Literaturgeschichte des 18. Jahrhunderts*, vol. 3 (Braunschweig: Vieweg, 1856–70), 526–41; R. H. Stephenson, "The Novel in Weimar Classicism: Symbolic Form and Symbolic Pregnance," in *The Camden House History of German Literature*, vol. 7, ed. Simon Richter (Rochester, NY: Camden House, 2005), 211–35. It is certainly the case that in most histories of German

literature, from Wilhelm Scherer's *Geschichte der deutschen Literatur* (1883) to Leo Balet's and E. Gerhard's *Die Verbürgerlichung der deutschen Kunst, Literatur und Musik im 18. Jahrhundert* (1936) to Rolf Grimminger's volume *Deutsche Aufklärung bis zur Französischen Revolution 1680–1789* (1980) in *Hansers Sozialgeschichte der deutschen Literatur*, one does discern a wider awareness of the history of books. Nonetheless, areas of study such as positivism, the sociology of literature, book history, and the history of reading merit but a minor niche in current literary studies.

[2] Christine Haug, "Buchserien und Anthologien: Wirkungsmächtige Medien zur Etablierung und Durchsetzung von ausländischen Literaturen in Deutschland im 19. Jahrhundert," IASL Online, last modified October 9, 2005, http://www .iaslonline.lmu.de/index.php?vorgang_id=2122.

[3] Barbara Kastner, "Statistik und Topographie des Verlagswesens," in *Geschichte des deutschen Buchhandels im 19. und 20. Jahrhundert*, ed. Georg Jäger, vol. 1, part 2 (Frankfurt am Main: MVB, 2003), 300–367; also see Ilsedore Rarisch, *Industrialisierung und Literatur. Buchproduktion, Verlagswesen und Buchhandel in Deutschland im 19. Jahrhundert in ihrem statistischen Zusammenhang* (Berlin: Spiess, 1976), and Lynne Tatlock, ed., *Publishing Culture and the "Reading Nation": German Book History in the Long Nineteenth Century* (Rochester, NY: Camden House, 2010).

[4] Gregory Crane, "What Do You Do with a Million Books?" *D-Lib Magazine* 12, no. 3 (2006), doi:10.1045/march2006-crane.

[5] Access to this repository is available at http://www.textgridrep.de.

[6] Leonid Taycher, "Books of the World, Stand Up and Be Counted! All 129,864,880 of You," *Google Books Search* (blog), August 5, 2010, http://booksearch.blogspot .de/2010/08/books-of-world-stand-up-and-be-counted.html.

[7] See, for example, Dan Cohen's posting "Initial Thoughts on the Google Books Ngram Viewer and Datasets," *Dan Cohen* (blog), October 16, 2012, http:// www.dancohen.org/2010/12/19/initial-thoughts-on-the-google-books-ngram -viewer-and-datasets.

[8] Franco Moretti, "Conjectures on World Literature," *New Left Review* 1 (2000): 57.

[9] Stanley Fish, "Mind Your P's and B's: The Digital Humanities and Interpretation," *New York Times*, January 23, 2012, http://opinionator.blogs .nytimes.com/2012/01/23/mind-your-ps-and-bs-the-digital-humanities-and -interpretation.

[10] Martin Mueller, "Stanley Fish and the Digital Humanities," *Center for Scholarly Communication & Digital Curation* (blog), February 8, 2012, http://cscdc .northwestern.edu/blog/?p=332. Also posted on Mueller's blog *Scalable Reading*, http://scalablereading.org.

[11] Ibid.

[12] See Jean-Baptiste Michel et al., "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science* 331, no. 6014 (January 14, 2011): 176–82.

[13] For an application of machine learning, see Russell Horton et al., "Mining Eighteenth Century Ontologies. Machine Learning and Knowledge Classification

in the *Encyclopédie*," *Digital Humanities Quarterly* 3, no. 2 (2009), http://www
.digitalhumanities.org/dhq/vol/3/2/000044/000044.html. A summary of
topic modeling can be found in Clay Templeton, "Topic Modeling in the Human-
ities: An Overview," *Maryland Institute for Technology in the Humanities* (blog),
August 1, 2011, http://mith.umd.edu/topic-modeling-in-the-humanities-an
-overview. A discussion of the value of network analysis for the humanities can be
found in Franco Moretti, "Network Theory, Plot Analysis," *New Left Review* 68
(2011): 80–102.

14 We follow Peter Gryzbek, ed., *Contributions to the Science of Text and Language:
Word Length Studies and Related Issues* (Dordrecht: Springer, 2006), 15–60. See
also Wilhelm Fucks, *Mathematische Analyse von Sprachelementen, Sprachstil und
Sprachen* (Cologne: Westdeutscher Verlag, 1955) and Helmut Kreuzer, Rul Gun-
zenhäuser, eds., *Mathematik und Dichtung. Zur Frage einer exakten Literaturwis-
senschaft* (Munich: Nymphenburger, 1965).

15 Independently from both, the philosopher Wincenty Lutosławski coined the
term "stylometry" with regard to the analysis of peculiarities that determine the
stylistic character of compared texts such as those of Shakespeare and Bacon.
Wincenty Lutosławski, "Principes de stylométrie," *Revue des études grecques* 41
(1890): 61–81; see also Adam Pawłowski, "Wincenty Lutosławski (1863–1954).
A Forgotten Father of Stylometry," *Glottometrics* 8 (2004): 83–90.

16 A good example is the work of the statisticians Frederick Mosteller and David L.
Wallace, *Inference and Disputed Authorship: The Federalist* (Reading, MA: Addison-
Wesley, 1964). They use function words to determine the authorship of the anony-
mously published *Federalist Papers*.

17 See David L. Holmes, "The Evolution of Stylometry in Humanities Scholar-
ship," *Literary & Linguistic Computing* 13, no. 3 (September 1998): 111–17.
For more research in stylometry, see Patrick Juola's survey "Authorship Attri-
bution," *Foundations and Trends in Information Retrieval* 1, no. 3 (2006):
233–334.

18 John Burrows, "Delta: A Measure for Stylistic Difference and a Guide to Likely
Authorship," *Literary & Linguistic Computing* 17, no. 3 (2002): 267–87. A gen-
eral discussion and assessment can be found in Patrick Juola, "Authorship Attri-
bution," section 5.2.1.

19 On novels, see David L. Hoover, "Testing Burrows's Delta," *Literary & Lin-
guistic Computing* 19, no. 4 (2004): 453–75. For different languages, see Karina
van Dalen-Oskam and Joris van Zundert, "Delta for Middle Dutch—Author and
Copyist Distinction in *Walewein*," *Literary & Linguistic Computing* 22, no. 3
(2007): 345–62.

20 Jan Rybicki and Maciej Eder, "Deeper Delta Across Genres and Languages: Do
We Really Need the Most Frequent Words?," *Literary & Linguistic Computing*
26, no. 3 (2011): 315–21; Maciej Eder and Jan Rybicki, "Do Birds of a Feather
Really Flock Together, or How to Choose Training Samples for Authorship Attri-
bution," *Literary & Linguistic Computing* 28, no. 2 (2013): 229–36.

21 For discussions of these issues, see David L. Hoover, "Testing Burrows's Delta";
Marius Popescu and Liviu P. Dinu, "Comparing Statistical Similarity Measures

for Stylistic Multivariate Analysis," in *Proceedings of the International Conference RANLP*, ed. Galia Angelova et al. (Bulgaria: Association for Computational Linguistics, 2009), 349–54; and Peter W. H. Smith and W. Aldridge, "Improving Authorship Attribution: Optimizing Burrows' Delta Method," *Journal of Quantitative Linguistics* 18, no. 1 (2011): 63–88.

22 The script can be downloaded from their website, "Computational Stylistics," https://sites.google.com/site/computationalstylistics. Their application of this tool to a wide range of problems is discussed in Rybicki and Eder, "Deeper Delta," and other papers.

23 For the mathematical background, see Shlomo Argamon, "Interpreting Burrows's Delta: Geometric and Probabilistic Foundations," *Literary & Linguistic Computing* 23, no. 2 (2008): 131–47.

24 John Burrows, "Questions of Authorship: Attribution and Beyond," *Computers and the Humanities* 37, no. 1 (February 2003): 1–26; John Burrows, "Textual Analysis," in *A Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth (Oxford: Blackwell, 2004), 323–47.

25 David L. Hoover, "Testing Burrows's Delta."

26 See "Digitale Bibliothek," Wikipedia, last modified December 31, 2008, http://de.wikipedia.org/wiki/Digitale_Bibliothek_%28Produkt%29.

27 The original data and a version converted to TEI 5 with enriched metadata can be found at "TextGrid: Virtuelle Forschungsumgebung für die Geisteswissenschaften," http://www.textgrid.de/ueber-textgrid/tools-services-ressourcen/digitale-bibliothek.

28 Magda Heydel and Jan Rybicki, "The Stylometry of Collaborative Translation," paper presented at the annual meeting of the Alliance of Digital Humanities Organisations, Hamburg, Germany, July 2012. A detailed abstract can be found at "The Stylometry of Collaborative Translation," http://www.dh2012.uni-hamburg.de/conference/programme/abstracts/the-stylometry-of-collaborative-translation.

29 For a detailed description, see also Maciej Eder, "Computational Stylistics and Biblical Translation: How Reliable Can a Dendrogram Be?" *Philologica Wratislaviensia* 10 (in press).

30 See Susanne Zantop in her afterword to Friederike Helene Unger's *Bekenntnisse einer schönen Seele von ihr selbst geschrieben* (Hildesheim: Olms 1991), 387–88.

31 See the commentary in Johann Wolfgang Goethe, *Gedichte 1800–1832*, ed. Karl Eibl (Frankfurt am Main: Deutscher Klassiker Verlag, 1988), 873–1382.

32 John Burrows was in fact one of the first to point out that Foucault's assertion that the "author" is a mere projection of the reader is contradicted by stylometric findings, but the level of scholarly interest in his fact-based argument was not particularly high. See John F. Burrows, "Computers and the Idea of Authorship," in *Rückkehr des Autors. Zur Erneuerung eines umstrittenen Begriffs*, ed. Fotis Jannidis, Gerhard Lauer, Matias Martinez, and Simone Winko (Tübingen: Niemeyer, 1999), 167–82.

33 See the survey in Renate von Heydebrand and Simone Winko, "Arbeit am Kanon: Geschlechterdifferenz in Rezeption und Wertung von Literatur," in

Genus: Zur Geschlechterdifferenz in den Kulturwissenschaften, ed. Hadumod Buß-mann and Renate Hof (Stuttgart: Kröner, 1995), 206–61.

[34] Heinrich von Kleist, *Das Käthchen von Heilbronn* (Berlin: Realschulbuchhand-lung, 1810).

[35] The list of problems stylometry is facing compiled by Rudman is still valid; cf. Joseph Rudman, "The State of Authorship Attribution Studies: Problems and Solutions," *Computers and the Humanities* 31 (1998): 351–65.

[36] James Pennebaker, *The Secret Life of Pronouns: What Our Words Say about Us* (New York: Bloomsbury, 2011).