TexTiles: Exploring Patterns in Historical Discourse

Robert Roessler*
Harvard University, USA

Caiseen Kelly†
Harvard University, USA

Michael Behrisch‡
Tufts University, USA

Johanna Beyer§
Harvard University, USA

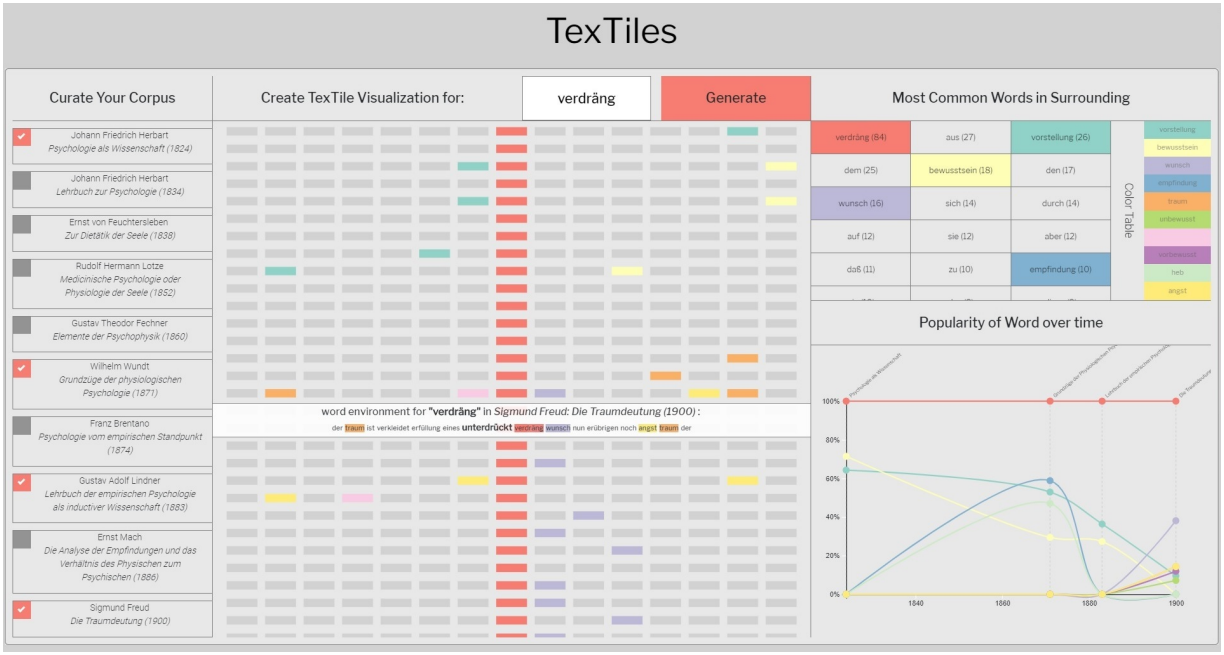

Figure 1: Initial TexTile view: Domain experts can search for keywords in an individually curated corpus and explore text segments horizontally and vertically, i.e., syntagmatically and paradigmatically. A complementary word frequency visualization of words in keyword surroundings over time and through different works (line graph) allows for diachronic analysis and intuitive pattern recognition.

Abstract

Discourse analysis is a well-established method in the humanities to analyze historical trends and knowledge figurations that are inscribed in the texts of an era. While these trends are manifested as linguistic nuances in a variety of ways, the actual discourse remains a rather abstract concept. The goal of our paper is to develop a visual representation of such a discourse. We present TexTiles, a Visual Analytics framework that allows domain specialists to visually trace historical discourses in large text corpora. By allowing scholars to curate a customizable corpus, analyze keywords in context, and ultimately explore the network of these context words within the corpus, TexTiles serves as an exploratory research tool for domain experts. Allowing for a hybrid approach between close and distant reading practices, we demonstrate the utility of our application based on a case study on the discovery of the unconscious and the mechanics of repression in the long 19th century. TexTiles allowed users to trace knowledge formations that occurred already around 1800 leading to Sigmund Freud's *Interpretation of Dreams* in 1900.

Keywords: Digital humanities, computational discourse analysis, computational literary studies, distant reading

*e-mail: robertroessler@g.harvard.edu
†e-mail: caiseen_kelley@college.harvard.edu
‡e-mail: michael.behrisch@tufts.edu
§e-mail: jbeyer@g.harvard.edu

1 Introduction

Discourse analysis is a popular method in the humanities to examine complex and multilayered historical developments. Prominent discourses, such as the ideas of "love", "sexuality", "race", and "gender", have been subject to changes and shifts over the last few centuries. These changes are reflected in the institutionalized language of the time, which bears witness to ongoing ideological trends. While the impact of a discourse may be explicit, the idea of the discourse itself remains a rather abstract concept and eludes itself from being unfolded, unveiled, or measured. This general academic consensus served as initial motivation for the paper to develop a framework that would help to visually trace a historical discourse. Another motivation for the paper was that in the past, scholars examined archival materials only manually and were thus restricted in their archaeological work. It is the goal of this paper to use computational techniques to make use of the rich materials made available through recent digitization efforts and visualize historical discourses inscribed in these documents in a much more comprehensible manner. For this purpose, we developed TexTiles, a Visual Analytics framework that enables scholars to recognize trends in large amounts of texts, exceeding the capacity of established analog analyses. With TexTiles, users can explore changes within the contexts of keywords in depth for nuanced insights and to further strengthen their arguments regarding historical developments.

The contribution of this project is twofold. The first contribution is the visual framework TexTiles itself, which allows scholars to

curate their own corpus, search for keywords, and most importantly explore the environment of these keywords in depth. We structure the TexTiles system into three intertwined parts: (1) The *TexTile view* lets the user understand syntactic contexts (Fig. 1, left); (2) The *Word-Net view* lets users understand the discoursive network of nuanced meanings of a key term at one point in time (Fig. 2); and (3) The *Comparative Word-Matrix view* lets users compare and contrast two distinctive discoursive networks and explore temporal changes (Fig. 3). By example of TexTiles, the paper, furthermore, aims to contribute to the theory of discourse analysis in a computational era by revisiting structuralist theory and incorporating a carefully conducted hybrid approach between close and distant reading practices.

We demonstrate the usefulness of TexTiles and underlying theory in a case study tracing the formation of the unconscious and the mechanics of repression in the long 19th century. TexTiles allows to facilitate this kind of analysis in hours, while the typical work-flow of experts in the field would have taken weeks. We verify this claim and investigate potential future works through multiple expert study interviews.

2 Background

The beginning of discourse analysis is marked by Michel Foucault's monograph *Archaeology of Knowledge*, in which he argues that "discourses", i.e., historical knowledge formations and narratives, emerge according to a vast and complex set of discursive and institutional relationships that manifest themselves in language [8]. By defining a discourse as a system of relations, the discourse reveals itself as a formal structure of statements that can be analyzed linguistically. For this purpose, Ferdinand de Saussure's model of language as a system of signs that are arranged both horizontally and vertically (i.e., syntagmatically and paradigmatically) [5], as well as Roman Jakobsons model of an axis of combination and an axis of selection [12] served as a theoretical framework. It was these structural models that gave us the idea to introduce natural language processing together with visualization techniques in an effort to more fully grasp a historical discourse and render it visible.

However, favoring a quantitative approach over a former qualitative approach also has its downside, as it does not allow scholars to fully encode all the meanings that are inscribed in a statement. In his *Archaeology*, Foucault shows that the semantic and syntactic structures do not suffice in determining the discursive meaning of an expression, as e.g., the combination of the letters 'qwerty' produce discoursive but no semantic meaning. As a result, the meaning of an expression depends on the conditions in which it emerges and exists within a field of discourse. This understanding influenced TexTiles' design as an exploratory framework, specifically for domain specialists who want to focus their analysis on large scale patterns (distant reading) while still maintaining an awareness of the specific utterances and words (close reading) that make up the word environments. Rather than providing a black-box tool, TexTiles enables users to scale their textual, archaeological work and allows them to examine vast amounts of textual data in order to find patterns in previously overlooked or marginalized archival materials, which have been made more and more accessible by digitization initiatives in recent years.

3 Requirement Analysis

TexTiles has been developed in close collaboration with domain experts in the humanities and the visualization community. Based on their expertise and alongside well-established methodological ideas, we collected a list of requirements tailored towards a visual encoding of historical discourses. Since every discourse produces and uses specific vocabulary, we allow users to focus on key terms and enable scholarly users to analyze change and frequency of surroundings words. Similar to linguistic and structuralist theory [5,12] we provide a two-dimensional model that allows for analysis along two vectors:

R1: Synchronic Analysis. Users should be able to explore a system of statements that construct the discourse at *one* moment in time without taking its history into account. For example, what nuances define the meaning of a word at one point of time?

R2: Diachronic Analysis. Complementary to the synchronic analysis, users should also be able to analyze discoursive statements *over time*, i.e., compare how the meaning or the understanding of a term might have shifted over the course of several years.

Diachronic and synchronic analysis allow to conceive the broader discourse in its temporal dimensions. However, users also need the possibility to explore individual sentences - again, along two vectors:

R3: Syntagmatic Analysis. Users should be able to explore text segments horizontally, i.e., identifying combinations and connections of words surrounding a previously defined keyword.

R4: Paradigmatic Analysis. Complementary, users should also be able to explore text segments vertically and to identify paradigmatic patterns and rules such as interchangeability in certain word slots following a keyword.

4 Related Work

Visualizing Text Patterns, Word Relationships, and Word Co-Occurrence. A range of visualization systems for text analysis have been developed, see Jänicke et al. [13] for a comprehensive overview. The comparative exploration of document collections has been the focus of a series of works in the past, such as in [17, 23, 24]. For example, Keim and Oelke present a visualization for text patterns in (temporal) document collections, where arbitrary text features (syntactic, semantic, higher-level) are mapped to pixel colors [15]. Pixels are arranged in series of glyphs, depicting paragraph, chapter, and book relationships. ThemeRiver-inspired approaches [11], such as Tiara [19] or Textflow [4], focus more on the development of topics over time in order to depict stability, or branching and merging behaviors. Similar to our work, co-occurrence visualizations are demonstrated, e.g., in [27] (matrix view) or [16] (arc diagram) with different visual means in order to support named entity resolution or social network analysis. TexTiles' core contribution is a holistic framework for analyzing historical discourse. As such, we follow a pragmatic approach and adapt and borrow successful counterparts from the related works in the field of visualization, rather than designing entirely novel visual metaphors.

Computational Discourse Analysis. Computational discourse analysis is centered around analyzing specific language usage patterns by extrapolating beyond clause level analysis [2]. Current models focus on the conversational, argumentative, and debate-like aspects of discourse to illustrate a discoursive structure at one point in time. [18]. While such synchronic approaches have been subject to extensive research and elaborated frameworks and models have been developed [10, 20], scholars have paid less attention to diachronic aspects to analyze how a discourse might have changed over time. By allowing users to compare and contrast a keyword's contexts at different points in time, TexTiles, on the other hand, seeks to provide users with the tools they need to explore trends in historical discourses.

Corpus Linguistics. With its goal to analyze contexts of keywords over time, TexTiles draws inspiration from well established techniques and methods in the field of corpus linguistics, most and foremost concordances and keywords in context (KWiC). Prominent tools such as Voyant [26], AntConc [1], or Trameur [7] have been developed to visualize contexts of keywords. While these tools offer a variety of features in addition to analyzing concordances, they are limited in their capacity to thoroughly compare and contrast all concordance lines. With its Word-Net and Word-Matrix view, TexTiles adds another dimension by allowing users to examine all

concordance lines in greater depth and analyze discoursive trends.

Distant Reading. In contrast to close reading, distant reading is a method of engaging with texts on a more abstract, quantitative level. The practice was introduced by Franco Moretti [21, 22] as a strategy to deal with the ever growing amounts of (digitized) textual data on a macro-level [14]. Critics, however, argue that purely quantitative practices reduce literary complexity and thus violate the individuality of the text [28]. Mixed forms such as mid-range reading [3] have since been introduced, trying to bridge close and distant reading practices. TexTiles' architecture draws inspiration from these non-binary approaches. It is theoretically guided by Andrew Piper's circular model of targeted close readings validating patterns in larger distant readings [25].

5 Historical Discourse Analysis with TexTiles

Based on our requirement analysis, we derived and specified a set of core features that guide functionality, design, and implementation of our framework. Allowing for close and distant reading practices [6, 21], our initial, eponymous, and most central design choice was to abstract text into tiles, allowing for better pattern recognition when dealing with vast amounts of text, while still enabling users to explore on the textual level. With this in mind, we derived the following feature list, with features F3-F5 forming a triad of complementary visualizations.

F1: Customizable Corpus Curation. In order to increase clarity and usefulness of the visualization, users need to be able to select individual texts of the corpus for subsequent analysis.

F2: Keyword-Based Initiation. All visualizations are initiated and updated based on the keyword a user chooses to explore.

F3: TexTile View. After keyword input, an initial visualization displays all relevant text segments. This visualization allows for horizontal and vertical exploration, i.e., synchronic and diachronic analysis.

F4: Word-Net View. For deeper insights on the syntagmatic but also the paradigmatic level, users can curate their own network visualization based on selected words in the initial keyword's surrounding. The Word-Net view enables users to model a representation of the state of a discourse at one point in time.

F5: Word-Matrix View. This matrix-based view supports diachronic analysis, i.e., the temporal comparison of two distinctively curated network views.

5.1 TexTile View

The first of the three triadic visualizations in our application is the TexTile view itself (Fig. 1, center), which aims to let users explore specific keywords, their word environments, and their syntactic contexts in a clean and uncluttered view.

Users start their exploration by curating their individual corpus (Fig. 1, left) and searching for a specific keyword within that corpus. In order for users to still get a sense of the original text, we used a customized stemming library and dropped only a few stop words. As soon as a keyword has been selected, the TexTile visualization and its associated word frequency graph (Fig. 1, right) are initiated. We encode each instance of the keyword within the corpus as a separate row (concordance line), and display the keyword as an abstract red tile in the center of that row. Additionally, we display the surrounding words of the keyword as grey tiles, allowing users to examine the environment of each keyword occurrence in more detail.

Once a user hovers over a grey tile, the system visually highlights all tiles representing the same word. This allows users to quickly get an overview of the frequency of certain words. In addition, when hovering over a tile, we show a tool tip of the original text behind the abstract tiles in the selected row. This enables users to go back to their original data on demand, without cluttering the view. By clicking on a tile, a user can assign a fixed color to a word which

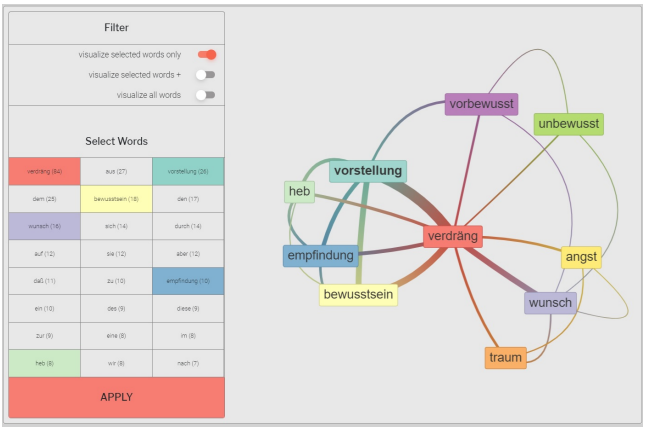

Figure 2: Word-Net view, showing co-occurring words in the word surroundings of the stemmed keyword "verdräng" (repression) within previously selected corpus texts (see Fig. 1)

updates the word frequency line graph. The line graph displays the relative popularity of the selected word in the environment of the initially chosen keyword through all works of the corpus. As a result of these side-by-side visualizations, users can see how term importance changes over time.

5.2 Word-Net View

The second visualization of TexTiles is the Word-Net view (Fig. 2), a relational network showing the strength of connections between co-occurring words within the environments of a keyword. These co-occurrences shed light on the keyword's different nuances in meaning and it is this synchronic snapshot that renders a discourse visible.

The Word-Net view allows users to continue their initial analysis and visualize previously selected words, but also allows to redefine and adjust their selection. We then compute the connections between co-occurring words and encode the strength of their connection (i.e., the frequency of their co-occurrence) in the distance between nodes and their edge thickness. The width of an edge is a direct linear mapping to the frequency of co-occurrence. Edge lengths are computed in a force-directed layouting algorithm with the frequency of co-occurrence between nodes as additional input parameter.

5.3 Comparative Word-Matrix View

The final visualization of TexTiles is the Comparative Word-Matrix view (Fig. 3), which is initially only a transformation of the relational network visualization into an adjacency matrix using the same data and data structure. The adjacency matrix initially only shows the discoursive composition of the keyword at one point in time. However, in a next step, our Comparative Word-Matrix view allows users to compare synchronic snapshots in a single visualization. By curating two contrasting corpora, users generate two distinct data sets with texts from different authors and/or from different periods. In order to display both data sets in the same matrix, rather than using rectangles as representations of connections we use triangles of different color.

As in the Word-Net view, users can filter the underlying word basis. In case of the Word-Matrix view, data is gathered from both data sets, resulting in a different clustering, as seen in Fig. 3. These patterns indicate not only the different nuances the keyword has in the two distinct data sets - they also allow users to trace shifts in meaning and thus analyze discoursive trends. However, it is crucial that users are well aware of the underlying texts as it is very easy to curate highly diverging clusters. Users need to carefully select

Figure 3: Comparative Word-Matrix view, displaying the co-occurrences of selected words in two different corpora, enabling scholars to detect trends in discourse through time.

the words used in the visualization, especially words with diverging meaning. With that in mind, domain experts are able to analyze and contrast corpora diachronically and synchronically and compare patterns within multiple works, scholars, and time periods.

6 Evaluation

In this section, we first describe a brief case study tracing the formation of the unconscious and the mechanics of repression between 1800-1900 and, next, report on expert feedback from scholars at the German Studies Department at Harvard University.

6.1 Case Study

In our case study, we focus on the idea and the usage of the word "Verdrängung" during the 19th century - the German term for repression. TexTiles allows us to rephrase our research question to ask: First, in what contexts was the stem of the term "Verdrängung", i.e., "verdräng" used within the different texts of our corpus, and secondly, how did its context change? Shifts in context would consequently signify a change of the discourse.

For this purpose, we first selected four texts as our corpus: 1. Johann Friedrich Herbart's *Lehrbuch der Psychologie als Wissenschaft (1824)*, one of the key texts for the emerging field of empirical psychology that founded an entire school of thought, the so-called Herbartianismus, 2. Wilhelm Wundt's *Grundzüge der physiologischen Psychologie (1871)*, one of the most widely read textbooks on empirical psychology in Germany with Wundt being regarded as the founder of psychology as a science, 3. Gustav Adolf Lindner's *Lehrbuch der empirischen Psychologie (1882)*, the most popular psychology textbook in Austrian schools, and 4. Sigmund Freud's *Interpretation of Dreams (1900)*, which marks the beginning of modern Psychoanalysis.

Having assembled these texts, we start our analysis in TexTiles by searching within this text corpus for the stemmed term "verdräng". The keyword search initializes the TexTiles view., i.e., computes the syntagmatic word surroundings for all instances of the stem in all corpus texts as seen in Fig. 1. By exploring words that occur in the surroundings (using the TexTiles view), we are able to make a variety of significant findings: The terms "Vorstellung" (notion, imagination, or mental representation) and "Bewusstsein" (consciousness, awareness) were in frequent use and thus appear to be highly important in the context of repression for Herbart, Wundt, and Lindner, while Freud used these terms less often when thinking about repression. Instead, he developed a new terminology around the mechanics of repression composed of the triadic model of "Vorbewusst"-"Bewusst"-"Unbewusst" (unconscious, conscious, preconscious) that describes the state of mental images ("Vorstellungen") more accurately. In addition, Freud expanded the idea of repression by putting it into the context of "Angst", (fear), "Wunsch" (wish), and "Traum" (dream). The curves in the line chart in Fig. 1 signify this shift in discourse clearly.

This initial analysis encouraged us to explore the original text behind the TexTile view to reexamine how Freud situated wish and dream within the mechanics of repression. TexTiles allows us to find an answer immediately - the tooltip provides the sentence behind the hovered-on tiles: "Der Traum ist verkleidete Erfüllung eines unterdrückten verdrängten Wunsches" (The dream is the disguised fulfilling of a suppressed repressed wish), as seen in Fig. 1, tooltip. Being a scientific textbook, some of its sentences read like definitions, substance that makes for a structural linguistic analysis. Regarding the psyche, it was this paradigmatic attempt to transform formulas into formulations that resulted in the scientification of the psyche. The process was initiated by Herbart who suggested the application of mathematics and physics to the psyche. However, he still put psychology on a metaphysical foundation that led to the popularity of the terms "Vorstellungen" and "Bewusstsein" in the context of "verdräng" as can be seen in the line graph of Fig. 1. In the mid-19th century, however, a purely empirical approach gained more favor. Wilhelm Wundt, who was at the forefront of this movement, introduced the measurable sensation ("Empfindungen", see Fig. 1) as a physical and bodily counterpart to the mental image of the "Vorstellungen". Alongside physical terms like e.g., "heben" (to lift), Wundt put "Verdrängung" into a more scientific context and amended the discourse by strengthening its mechanical aspects. While Freud did not totally abandon this path and adopted many of the mechanical principles on a metaphorical level, he not only introduced specific terminology but also expanded the earlier narrow idea of "Verdrängung".

While the TexTile view helps us understand the significance of these shifts in context and discourse, the Word-Net view (Fig. 2) reveals the characteristics of the pre- and the post-Freudian discourse in one view. In Fig. 2 we can see on the left side the narrow idea of repression in pre-Freudian texts with "Vorstellung" and "Bewusstsein" as two strong points of reference while on the right side the network of word co-occurrences signifies how Freud expands and widens the context for the notion of repression in 1900 by including the nuances of the wish and the dream. The Comparative Word-Matrix view lets us analyze this structure closely and provides a clear, unambiguous pattern. Fig. 3 shows that Freud's notion of "Verdrängung" (blue) does not overlap with the ones of Herbart, Wundt, or Lindner (red), which suggests that with regard to "Verdrängung", Freud is indeed a founder of discourse, as Foucault claimed [9]. However, the notion of "Verdrängung" makes up only a small part of the larger psychological discourse in the 19th century and a TexTiles analysis of other key terms such as "unconscious" suggests that Freud too did not reinvent the wheel entirely but built it on ideas that had been circulating earlier.

6.2 Expert Feedback

For collecting expert feedback, we used the same text corpus as in the case study and asked a total of 5 experts to explore the word surroundings of the key term "repression". The goal of our evaluation was to test whether TexTiles was effective in two ways: First, we wanted our framework to be able to reproduce the results typically gained with traditional discourse analysis. Secondly, we wanted to test whether TexTiles provides scholars with a way to gain additional insights and explore overlooked nuances. In sum, we wanted our tool to enable scholars to better support their arguments or find hidden aspects of the discourse.

The experts were intrigued by the possibility of exploring thousands of pages of text by tracing key terms and their surroundings in visual patterns. Overall, four aspects stood out to them: 1) The

possibility to examine a previously unreadable amount of text; 2) To analyze the text in an elaborate fashion and beyond a simple keyword search; 3) To get a quick overview, and; 4) To ultimately explore text and discourse as patterns. On the other hand, some scholars expressed a potential limitation of TexTiles: Rhetorical strategies in other text genres (e.g., metaphors or ellipses, or irony in general) might distort TexTiles' results and patterns. In addition, scholars also pointed out the neighborhood size of the visualized word's surroundings. Currently, we display seven words to the left and seven words to the right of the selected keyword in the TexTile view. Our experts, however, expressed the need to revisit the theoretical underpinning of this setting.

7 Discussion & Conclusion

TexTiles allows scholars to curate a corpus of different texts and provides them with three exploratory visualizations for historical discourse analysis. With the help of the TexTile, the Word-Net, and the Word-Matrix view, scholars can perform historical discourse analysis digitally. TexTiles enables them to explore large amounts of texts quickly and thoroughly, to create unique visualizations for word constellations, and to search these patterns to proof old hypotheses as well as for finding previously overlooked details that might change the debate in the field. Currently, TexTiles is designed to create meaningful patterns of semi-scientific texts and thus allow users to draw conclusions regarding the discourse of the time. In the months to come, we plan to incorporate the valuable feedback we received from experts who tested TexTiles to make it an effective tool for examining literary texts. For instance, it may be of great interest to a scholar to analyze all the poems of a particular author over time and examine how his/her textual universe expanded and what nuances the author consciously added to certain terms in later poems. Ultimately, TexTiles could also prove highly beneficial for online archives, the Hathi-Trust currently being the most important one for scholars in the Humanities. These archives allow scholars to download lists of copyright-free texts. A plugin that would allow scholars to export and examine their pre-curated list of texts with TexTiles would be a tremendous improvement over current practices.

References

[1] L. Anthony. Developing antconc for a new generation of corpus linguists. *Proceedings of the Corpus Linguistics Conference*, CL 2013:14–16, 2013.

[2] W. Benzon. Computational linguistics and discourse analysis. *SSRN Electronic Journal*, March 1979. doi: 10.2139/ssrn.2508667

[3] A. Booth. Mid-range reading: Not a manifesto. *PMLA*, 132(3):620–627, 2017. doi: 10.1632/pmla.2017.132.3.620

[4] W. Cui, S. Liu, L. Tan, C. Shi, Y. Song, Z. Gao, H. Qu, and X. Tong. Textflow: Towards better understanding of evolving topics in text. *IEEE Trans. Vis. Comput. Graph.*, 17(12):2412–2421, 2011. doi: 10.1109/TVCG.2011.239

[5] F. de Saussure. *Course in General Linguistics (originally: Cours de linguistique gnrale (1916))*. Fontana/Collins, Glasgow, 1977.

[6] M. Erlin, ed. *Distant Readings. Topologies of German Culture in the Long Nineteenth Century*. Camden House, New York, Rochester, 2014.

[7] S. Fleury and M. Zimina. Trameur: A framework for annotated text corpora exploration. *Proceedings of COLING (International Conference on Computational Linguistics), System Demonstrations*, pp. 57–61, 2014.

[8] M. Foucault. *The Archaeology of Knowledge*. Pantheon Books, New York City, 1972.

[9] M. Foucault. *What is an Auhtor*, vol. Textual Strategies. Cornell University Press, Ithaca, N.Y., 1979.

[10] J. Fuhse, O. Stuhler, J. Riebling, and J. L. Martin. Relating social and symbolic relations in quantitative text analysis. a study of parliamentary discourse in the weimar republic. *Poetics*, 2019. doi: 10.1016/j.poetic.2019.04.004

[11] S. Havre, B. Hetzler, and L. Nowell. Themeriver: visualizing theme changes over time. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings*, pp. 115–123, Oct 2000. doi: 10.1109/INFVIS.2000.885098

[12] R. Jakobson. *Linguistics and Poetics*, vol. Style in Language. MIT Press, Cambridge, Massachusetts, 1960.

[13] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. On close and distant reading in digital humanities: A survey and future challenges. In *Eurographics Conference on Visualization, EuroVis 2015 - State of the Art Reports, STARs, Cagliari, Italy, May 25-29, 2015*, pp. 83–103, 2015. doi: 10.2312/eurovisstar.20151113

[14] M. L. Jockers. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, Champaign, IL, USA, 1st ed., 2013.

[15] D. A. Keim and D. Oelke. Literature fingerprinting: A new method for visual literary analysis. In *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, IEEE VAST 2007, Sacramento, California, USA, October 30-November 1, 2007*, pp. 115–122, 2007. doi: 10.1109/VAST.2007.4389004

[16] L. F. Klein. Social network analysis and visualization in 'the papers of thomas jefferson'. In *DH*, pp. 254–255, 2012.

[17] K. Lagus, S. Kaski, and T. Kohonen. Mining massive document collections by the WEBSOM method. *Inf. Sci.*, 163(1-3):135–156, 2004. doi: 10.1016/j.ins.2003.03.017

[18] C. Lipizzi, D. G. Dessavre, L. Iandoli, and J. E. R. Marquez. Towards computational discourse analysis: A methodology for mining twitter backchanneling conversations. *Computers in Human Behavior*, 64:782 – 792, 2016. doi: 10.1016/j.chb.2016.07.030

[19] S. Liu, M. X. Zhou, S. Pan, Y. Song, W. Qian, W. Cai, and X. Lian. Tiara: Interactive, topic-based visual text summarization and analysis. *ACM Trans. Intell. Syst. Technol.*, 3(2):25:1–25:28, Feb. 2012. doi: 10.1145/2089094.2089101

[20] O. Loyola-Gonzlez, A. Lpez-Cuevas, M. A. Medina-Prez, B. Camia, J. E. Ramrez-Mrquez, and R. Monroy. Fusing pattern discovery and visual analytics approaches in tweet propagation. *Information Fusion*, 46:91 – 101, 2019. doi: 10.1016/j.inffus.2018.05.004

[21] F. Moretti. *Distant Reading*. Verso, London, New York, 2013.

[22] F. Moretti and A. Piazza. *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, 2005.

[23] D. Oelke, H. Strobelt, C. Rohrdantz, I. Gurevych, and O. Deussen. Comparative exploration of document collections: a visual analytics approach. *Comput. Graph. Forum*, 33(3):201–210, 2014. doi: 10.1111/cgf.12376

[24] F. V. Paulovich and R. Minghim. Hipp: A novel hierarchical point placement strategy and its application to the exploration of document collections. *IEEE Trans. Vis. Comput. Graph.*, 14(6):1229–1236, 2008. doi: 10.1109/TVCG.2008.138

[25] A. Piper. *Enumerations: Data and Literary Study*. University of Chicago Press, 2018.

[26] S. Sinclair and G. Rockwell. Voyant tools - https://voyant-tools.org/.

[27] F. Stoffel, W. Jentner, M. Behrisch, J. Fuchs, and D. A. Keim. Interactive ambiguity resolution of named entities in fictional literature. *Comput. Graph. Forum*, 36(3):189–200, 2017. doi: 10.1111/cgf.13179

[28] K. Trumpener. Critical response i. paratext and genre system: A response to franco moretti. *Critical Inquiry*, 36(1):159–171, 2009. doi: 10.1016/j.ins.2003.03.017