ARTICLE

Conceptual Forays: A Corpus-based Study of "Theory" in Digital Humanities Journals

Rabea Kleymann[1] , Andreas Niekler[2] , Manuel Burghardt[2]

[1] Leibniz Center for Literary and Cultural Research, [2] University of Leipzig

The status of theory in the Digital Humanities (DH) has been the subject of much debate. As a result, we find different theory narratives competing and entangled with each other. If at all, these narratives can only be grasped and examined from a somewhat detached perspective. Here, we attempt to investigate these elusive narratives by means of a conceptual history approach. In doing so, we define different theory dimensions, ranging from specific cultural and literary theory frameworks to more generic uses of the concept of *theory*. We examine the use and semantic changes of these theory notions in a large corpus of DH journals. Using a mixture of heuristic methods and approaches from the field of distributional semantics, we aim to create tellable conceptual stories of DH theory.

1. Introduction

In his provocative 2008 text on the "end of theory," Chris Anderson asked whether the ubiquity of data could make theory-building obsolete altogether. This question seems to be particularly relevant for the Digital Humanities (DH), where metaphors such as "distant reading" (Moretti), "macroanalysis" (Jockers), and "culturomics" (Michel et al.) indicate a strong focus on empirical and data-driven approaches. Indeed, Anderson's thesis is also echoed in many DH debates; for instance, in the common narrative of a post-theoretical era that entails a "lack of theory" (Kleymann) in favor of an overly "positivist methodological fetishism" (Arnold).

A closer look, however, clearly shows that the issue of DH theorizing has played and continues to play a central role in the community (Warwick). Typically, theories and theorizing are called into question in order to reflect epistemological stances within DH research, with Matthew K. Gold even asking "does DH need theory?" Johanna Drucker, however, suggests that be reformulated, as "the question is not, does digital humanities need theory? But rather, how will digital scholarship be humanistic without it?" According to Drucker, theoretical constructs could be regarded as humanistic safeguards, as "humanistic theory provides ways of thinking differently, otherwise, specific to the problems and precepts of interpretative knowing – partial, situated, enunciative, subjective, and performative." A similar view was later expressed by Rafael C. Alvarado, who speaks of theories as a unique feature of DH scholarship, suggesting that "digital humanists may reconnect with the production of theory, an area where the humanities and interpretive social sciences have developed expertise." As Alvarado notes, theoretical

underpinnings are an essential feature distinguishing DH from data science. Demanding more theoretical interventions from DH scholarship seems to be one trajectory within the debate.

In addition, DH theories are discussed with regard to timing issues. Following wider post-theoretical narratives since the 1980s, a post-theoretical era has also been heralded within DH scholarship, with Tom Scheinfeldt arguing that "the time for theory is over, in the sense it is now the time for methodology" (qtd. in Hall). Ted Underwood, however, claims that DH simply missed the right moment for theoretical ventures (66). Contrary to appeals for a post-theoretical state, there are also arguments for a pre-theoretical state. Julia Flanders and Fotis Jannidis state that "a theory of digital humanities cannot simply coincide with its praxis. It can [...] very probably learn a lot from older theories [...] but first of all it must be founded in a very close look at the activities of digital humanists" (3). Moreover, Gary Hall remarks that the argument that "critical and self-reflexive theoretical questions about the use of digital tools and data-led methodologies should be deferred for the time being" has become prominent within DH.

What now lies before or after theory formation is similarly determined by focusing on praxeological perspectives. In this context, further dichotomies such as "saying and doing" and "building" (Endres) versus writing are introduced. Particularly prominent is the phrase "more hack, less yack" (Warwick 538), which highlights another realm of the debate, namely the textual form or linguistic condition of theories in the humanities. In other words, the entanglement of theory and textual practices seems to be outdated, while new forms, such as "materialist epistemology" (Ramsay and Rockwell) or "materialized contemplative knowledge" (El Khatib et al. 2), have started to appear.

Oscillating between celebration, regrets, and hesitation, theories continue to diversify within DH research (Elliott and Attridge 2). Given these ambivalent views of the role and function of theory in DH, we would like to attempt a form of analysis that has received little attention so far. Specifically, in this article, we will use the framework of conceptual history to investigate narratives concerning DH theory. Beyond literary framings, we define *narrative* as a form of ordering pattern for knowledge production within scientific discourses. The suggestion that there could be a science narrative, as Marie-Laure Ryan puts it, "carries the implication that scientific discourse does not reflect, but covertly constructs reality, does not discover truths, but fabricates them according to the rules of its own game" (344). Such an understanding ties in with Science and Technology research since the 1980s, which has focused on narrative structures within epistemic cultures (Knorr Cetina; Latour and Woolgar). The narrative of an "end of theory," for example, not only gains relevance in mediating knowledge and practices within DH's epistemic cultures, rather, such *grand narratives* or *master narratives*, following Jean-François Lyotard's

conception, provide specific epistemological and social settings for knowledge production. Within this framework, storytelling, on the one hand, can be regarded as organizing and mediating knowledge in research (Brandt 215). On the other hand, researchers, as Rom Harré points out, become storytellers sketching out storylines to contextualize their research (81–89).

Accordingly, we address two research questions in this article: 1) What kind of narratives are linked to the concept of theory in DH's epistemic cultures? and 2) What kind of epistemological struggles and semantic paradoxes are entangled with theory in DH? Our approach is founded on two premises. On the one hand, we proceed from the premise that theory can be regarded as a concept in terms of conceptual history approaches. Moreover, we assume that the theory discourse in DH can be addressed as a research problem of conceptual history approaches. As Ernst Müller and Falko Schmieder remark, conceptual history—grasped here as a history of science—assumes that single concepts within a scientific community are not only strongly shaping research environments (Müller and Schmieder, "Begriffsgeschichte und Wissenschaftsgeschichte" 89), but that expectations, practices, and interpretations are manifested in their usage (*Begriffsgeschichte und Historische Semantik* 604). On the other hand, applying a conceptual history approach to DH is premised on the assumption that DH can already be historicized. Here we follow Müller and Schmieder's argument that theory could be regarded as a "cipher or abbreviation for heterogeneous meanings and argumentations" (*Begriffsgeschichte* 74),[1] applying it to DH research. Therefore, this article deploys the argument that the concept *theory* brings into sharper relief characteristics of DH as an epistemic culture (Malazita et al.).

Conceptual history (or *begriffsgeschichte*) falls under the umbrella term of historical semantics, but it is also a method particularly associated with the work of Reinhart Koselleck, among others. Neighboring methods are discourse analysis, metaphorology, and the history of ideas (Müller and Schmieder, *Begriffsgeschichte* 122). One research focus of conceptual history is to investigate how concepts are formed, perceived, and incorporated in time. Concepts reflect and address social structures, while, as Kai Vogelsang notes, they themselves influence reality, "shaping the way it perceives itself and constituting patterns by providing models for action and increasing the likelihood of their usage" (16).

Recently, conceptual history approaches have been discussed in the natural language processing (NLP) community, leading to the idea of an approach called *digital begriffsgeschichte*. Our methodological framework for the investigation of theory takes up this thread, as it combines conceptual history with computational approaches from distributional semantics. The aim of this

---

1 All German references are translated by the authors of this paper. The authors of the paper are entirely responsible for translation errors.

paper is therefore twofold. On a discursive level, it presents and reflects (partial) perspectives on the theory discourse in order to uncover latent epistemological settings within the DH. On a methodological level, however, it brings to the fore the premises, implications, and pitfalls of linking conceptual history approaches with computational methods that are inspired by frequency analysis and distributional semantics (Wevers and Koolen 226).

The paper is largely organized around the presentation of two case studies, which serve as the first steps into a conceptual-historical inquiry of theory in DH research. Before embarking on this main task, however, in Section 2, we first discuss prior attempts to operationalize conceptual history in a computational way. The two case studies "Theory frameworks in DH research" and "Semantic spaces of theory and related concepts" are presented in Section 3. In closing, we reflect on the results and provide some concluding remarks.

2. Computational approaches to conceptual history

In this section, we provide an overview of approaches that use conceptual history in combination with computational techniques. As noted above, the emergence of conceptual history in general was coined and significantly developed by Koselleck. Müller and Schmieder, Kathrin Kollmeier, and Frank van Vree et al. have also provided concise introductions to the topic. Here, however, we show that an increasing number of approaches can be designated *digital begriffsgeschichte*.

Alexander Friedrich and Chris Biemann are among the first to have enhanced the conceptual history approach with computational techniques. They explore quantitative, semi-automatic approaches to digital conceptual history, analyzing the concepts *net*, *network*, and *networking*. One issue that Friedrich and Biemann already raise is the operationalization of semantic ambiguity, especially in abstract concepts and metaphors. Methodologically, the authors propose a prior, knowledge-free meaning induction (in German: *vorwissensfreie Bedeutungsinduktion*).

A similar approach has been provided by Silke Schwandt, whose paper seeks to highlight "the relevance of digital methods for historical semantics, using the Latin term *virtus* and its medieval use as an example" (107). In her conceptual-historical study, Schwandt presents a computational semasiological approach that relies on cooccurrence analyses using Voyant Tools (Sinclair and Rockwell), among others. Furthermore, she also proposes onomasiological procedures that enhance discursive addressing of keywords. A comparable approach can be found in Daniel Burckhardt et al.'s study, where diachronic collocation information is used to induce semantic change of words. There the authors employ the DiaCollo tool to investigate historical semantics used in the GDR's (German Democratic Republic) press language.

In another study, Peter De Bolla et al. propose an interesting computational approach that relies on cooccurrence information of words to construct a measure for "conceptual coherence" (75). This approach allows them to identify more complex verbal constellations that might comprise "concepts." Their proposed measure builds on *pointwise mutual information* (PMI), which is a popular measure of association, to which they add a smoothing exponent in the denominator.

Another computational approach that very much focuses on the history of ideas is presented by Arianna Betti and Hein van den Berg, who introduce a new methodological approach which they describe as a "model approach to the history of ideas" (812). Christian Haase et al. provide an interesting approach to investigate the lexical change of words, as they cluster the senses of different words in a diachronic way, using a network approach called neighborhood-graph over time. In this way, they can visualize the formation and change of the meaning of words using their exploratory *SCoT* (*Sense Clustering over Time*) tool.

Yet another approach, this time focusing on state-of-the-art word embeddings and their application for historical research questions, can be found in an article by Melvin Wevers and Marijn Koolen. There, the authors not only provide a general insight into the function and significance of word-embedding models; rather, they also discuss conceptual-historical example analyses for the terms *democracy* and *abortion*. In addition to word embeddings as analytical tools for conceptual history, as Wevers and Koolen suggest, they can also address questions of semantic change.

This review of related work shows that there have been attempts to apply a computational conceptual history and that these can mostly be found in the areas of digital history and computational linguistics (approaches to lexical and semantic change). Tools like *SCoT* allow for a very detailed and exploratory examination of how a word's meaning changes in detail, but unfortunately this seems inappropriate as an operationalization to our context of discovery; namely, the quantitative temporal mapping of a change in meaning. Since we are interested in determining larger-scale conceptual changes, we believe the word embeddings approach described by Wevers and Koolen appears to be the most promising for an investigation of DH epistemic culture and, more specifically, the role of *theory* in DH.

3. Tellable conceptual stories of DH theories

In this section, we present two tellable conceptual stories of *theory* in DH. While the terms *narrative* and *story* are often synonymously used in everyday language, we regard our case studies only as seeds for possible storylines.[2]

---

2 Contrary to narrative theory, Seymour Chatman distinguishes between narrative, story, and discourse. In his definition, story (histoire or fabula) and discourse (or récit or syuzhet) are two dimensions of narratives (9).

Contrary to *metanarratives* in science, which are "presumed to uniquely comprehend or in practice govern a culture" (Plotnitsky 516), our conceptual stories are locally situated and merely suggest possible theoretical formations. In other words, these case studies provide an opportunity to shift attention away from dominant narratives, such as "the end of theory," to potentially more differentiated and nuanced stories within DH's research. In doing so, we ask to what extent stories are entangled with *theory* and whether these conceptual stories are worth telling (Baroni)?

Because *theory* as a concept is simultaneously ambiguous, indispensable, and controversial, but also marks intersections between traditional disciplines and has already undergone semantic changes in the context of DH ("theory in practice"), we believe that it is highly appropriate for our inquiry. As a concept, *theory* cannot be easily defined or translated in DH research. Therefore, we argue that *theory* in DH can only be narrated, if at all.

Case study 1 offers a semasiological perspective, while case study 2 offers a more onomasiological perspective, which deals with semantic shifts of the term *theory* and related concepts. More precisely, case study 1 focuses on the semantic and discursive scope of *theory*, which we trace through frequencies and cooccurrences of theory references. In case study 2, we are concerned with a temporal internal structure (or propositional system) of the term *theory*. We create contextual word embeddings of the term and examine how they have developed or transformed in DH research, as well as exploring which concepts can be identified as nearest neighbors. The methodology of our two case studies is inspired by a related project called "The Trace of Theory", in which Geoffrey Rockwell et al. investigate keywords from the domain of literary theory in large text collections by means of a (1) dictionary approach as well as (2) a machine learning approach.

Our corpus-based study relies on academic publications in the field of DH. Most of the scholarly communication in DH takes the form of conference abstracts and articles in dedicated journals. While there is a growing collection of abstracts from various past conferences being added to "The Index of Digital Humanities Conferences" (Weingart et al.), this resource until now has many blind spots, as indexing is still in progress. For this reason, we have chosen instead to rely on journals (see Table 1).

The journals studied here are all well-established in DH[3] and cover a time span from 1966 to 2020 (see Figure 1). The journal that goes furthest back in time is *Computers and the Humanities* (*CHum*), which was renamed *Language*

Table 1. Overview of journals and the overall corpus composition.

Journal	Time span	Articles	Tokens
Computers and the Humanities (*ComHum*) https://link.springer.com/journal/10579/	1966–2004	1,560	approx. 6.8 million
Digital Humanities Quarterly (*DHQ*) http://digitalhumanities.org/dhq/	2007–2019	418	approx. 3.4 million
Literary and Linguistic Computing (*LLC*)	1986–2014	1,454	approx. 2.0 million
Digital Scholarship in the Humanities (*DSH*) https://academic.oup.com/dsh	2014–2020	305	approx. 6.9 million
Total		3,737	approx. 19.1 million

Resources and Evaluation (*LRE*) in 2005. However, we decided to exclude *LRE* from the corpus, as it has an explicit focus on linguistics and therefore is not representative for DH as a whole. To compensate for this gap, we decided to add *Digital Humanities Quarterly* (*DHQ*), which has been around since 2007. The final journal studied is *Literary and Linguistic Computing* (*LLC*), which was renamed in 2015 to *Digital Scholarship in the Humanities* (*DSH*), at which point its thematic focus became rather broader, explaining why we kept it in the corpus.

3.1. Theory frameworks in DH (case study 1)

Although DH is a highly ambiguous term with many different definitions (Terras), all DH approaches share a basic humanities perspective. *Theory* within DH research often refers to specific theoretical frameworks (e.g., poststructuralism) or their representatives (e.g., Michel Foucault), which are the focus of our analysis in this first case study. More concretely, we seek to answer the following questions: In how many articles does *theory* or any reference to a humanities approach to theory appear? How often is a specific theoretical framework referenced within a document? Which theoretical concepts cooccur and thus indicate theory clusters within DH research?

DICTIONARIES OF HUMANISTIC THEORY FRAMEWORKS

Two limitations arose in the context of our experimental design. First, we have used an edited list of theory frameworks and representatives, which we then systematically searched in our corpus. We decided to use a manually edited list of theory frameworks, as we encountered various problems when using only high-level concepts (such as structuralism or postcolonialism), because many of the articles instead mention typical representatives of the specific theoretical currents—for example, one tends to find Eichenbaum rather than formalism. Second, our corpus of DH journals is certainly not meant to be representative

173–200; Chris Alen Sula and Heather V. Hill, "The Early History of Digital Humanities: An Analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004)," Digital Scholarship in the Humanities 34, no. 1 (2019): 190–206, https://doi.org/10.1093/llc/fqz072.
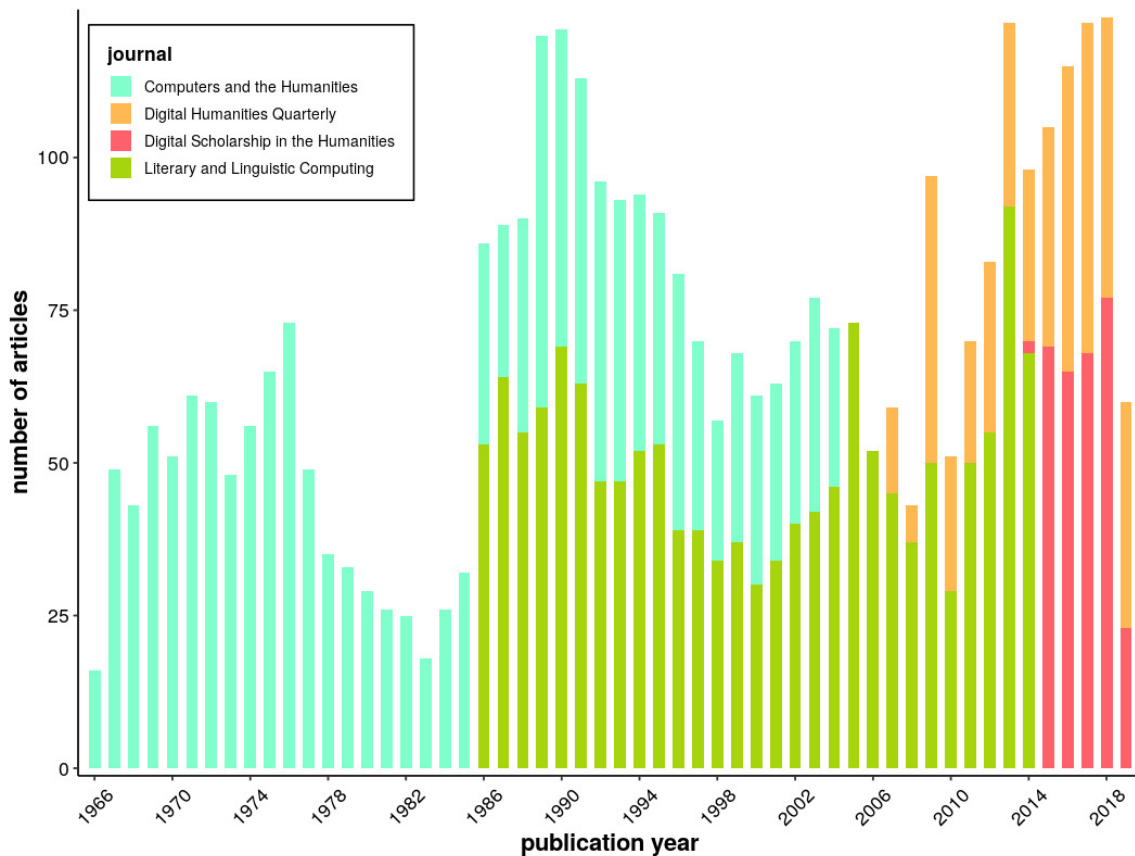
Figure 1. Overview of the temporal distribution of journal articles in our corpus.

of all DH scholarly communication. Since it is impossible to survey the entirety of theory frameworks in the humanities, we have chosen literary and cultural theories as a representative subfield.

To investigate this subfield we created dictionaries, which are based on three widely used introductory works to literary theory: Selden et al., Rivkin and Ryan, and Castle. Selection criteria included an assumed degree of familiarity and dissemination of the introductory works (such as the number of editions and citations). Our selection focused less on a diversification of the theoretical canon and thus shows a rather strong gender and diversity imbalance (Risam 17). The dictionaries are structured as follows: each dictionary contains an umbrella term for the theoretical approach, typical representatives ("name and surname" as well as "surname" only), and common multi-word combinations.[4]

Our dictionaries provide heuristic tools that allow us to address potential discursive intersections of *theory*. Please note that we do not intend to accurately reproduce individual branches of theory by means of representative authors, but rather to generate a generic inventory of relevant terms that are suitable to represent the typical use of theory in the humanities. Furthermore,

---

4 For the full dictionaries see "Dictionaries (Markdown)" at https://theory-in-dh.github.io/conceptual_forays/JoCA2022/conceptual_forays_supplementary.html

we do not claim that these dictionaries are exhaustive or representative of all theory frameworks that might be used in DH. Our compilation of traditional textbooks is rather a first attempt to address the question of what DH scholars probably mean when they talk about *theory*. We believe that our dictionaries provide first insights into the nature of theory in DH research, because they serve as syntheses of high-level concepts and representatives for literary and cultural theories.

In total, 13 unordered dictionaries were aggregated on the basis of their tables of contents, including names of theoretical approaches and schools as well as some of their representatives. Furthermore, we added hermeneutics, which was not explicitly part of the three mentioned introductory books, but which we frequently observed in our corpus during the first experiments. The final list of 14 theory frameworks, along with one representative given as an exemplar of each, runs as follows:

1. *Formalism and New Criticism* (Boris Eichenbaum, …)

2. *Structuralism* (Ferdinand de Saussure, …)

3. *Phenomenology, Rhetoric, and Reader-oriented Theories* (Edmund Husserl, …)

4. *Marxist Theory* (Georg Lukács, …)

5. *Poststructuralism* (Michel Foucault, …)

6. *Critical Race Theory and Ethnic Studies* (Lisa Lowe, …)

7. *Postcolonial Studies* (Gayatri Chakravorty Spivak, …)

8. *Psychoanalysis* (Julia Kristeva, …)

9. *Political Criticism* (Antonio Gramsci, …)

10. *Gender and LGBTQ+ Studies* (Judith Butler, …)

11. *Feminist Theory* (Coppélia Kahn, …)

12. *Cultural Studies and Critical Theory* (Theodor Adorno, …)

13. *Historicism* (Stephen Greenblatt, …)

14. *Hermeneutics* (this was used as an additional meta category; no specific representative authors were defined for this dictionary)

Finally, ambiguity was addressed in several ways. First, although in Selden et al.'s *Reader* theoretical constructs or other umbrella terms, such as metaphor or sexual politics, appear alongside names of theories and representatives, we elected not to include these constructs in our dictionaries because of their semantic ambiguity. Second, we also removed some names that were highly

ambiguous; for instance, (*Walter*) *Benjamin*, whose last name is also a first name that appears in multiple articles. Third, classifications of representatives are not always clear-cut as they may appear in two or more dictionaries at the same time (e.g., Michel Foucault). In such cases, we manually selected what we assumed to be the most representative dictionary. Lastly, we split Rivkin's "Feminism" and "Gender studies" and Selden et al.'s "Feminst theories" and "Gay, lesbian, queer theories" categories into two dictionaries called "Gender and LGBTQ+ Studies" and "Feminist Theory."[5]

FREQUENCY ANALYSES

Searching for the items from our dictionaries, we found that at least one of the specific cultural and literary theory terms appears in 793 articles of a total of 3,737 articles in the corpus. We found additional verbatim occurrences for *theory* and *theories* (1,037) that were not part of our dictionaries (see [Figure 2](#)). In order to get a rough overview of the share of further theory frameworks, we searched for all instances of "theory + of," "noun + theory/ies," and "adjective + theory/ies," which, for the most part, brings to the fore further theory frameworks that go far beyond the scope of our dictionaries for cultural and literary theory. Not only did we find numerous other humanistic theories (e.g., "theory of meaning," "theory of textuality," "theory of genres," "theory of metaphor," "theory of lexical diffusion"), we also found theories from other domains and disciplines (e.g., "information theory," "graph theory," "evolution theory," "game theory," "chaos theory"). In follow-up studies, we will systematically extract other theory frameworks and produce further dictionaries and augment them with representative theoreticians from *Wikipedia* and *Wikidata* (Gutiérrez de la Torre et al.).

These simple figures are by themselves already telling, as it shows that almost half the articles (48.97%) mention theory in one way or another, thereby already challenging the popular narrative that DH may lack theory (Cecire). That said, it is obvious that the mere mention of *theory* does not automatically entail an actual application or development of a theory. This is why we wanted to take a closer look at the use of specific theory frameworks from traditional humanities disciplines.

TYPE-TOKEN-RATIO (TTR)

After some corpus-wide frequency analysis, we took a closer look at the frequency of theory references in single documents. More concretely, we were interested in how often a dictionary item is used within one article. As a measure, we adopted the type-token-ratio (TTR), which is popular in quantitative linguistics to analyze the complexity of language by means of its vocabulary performance (Hess et al.). TTR distinguishes *types*, which are the

---

5 We are aware that the language to describe spectrums of gender and sexuality is still changing (Thelwall et al.; Cameron and Kulick). In addition to this, we would like to point out that such categorizations, like our dictionaries, are starting points for bias in data capturing (D'Ignazio and Klein 97).
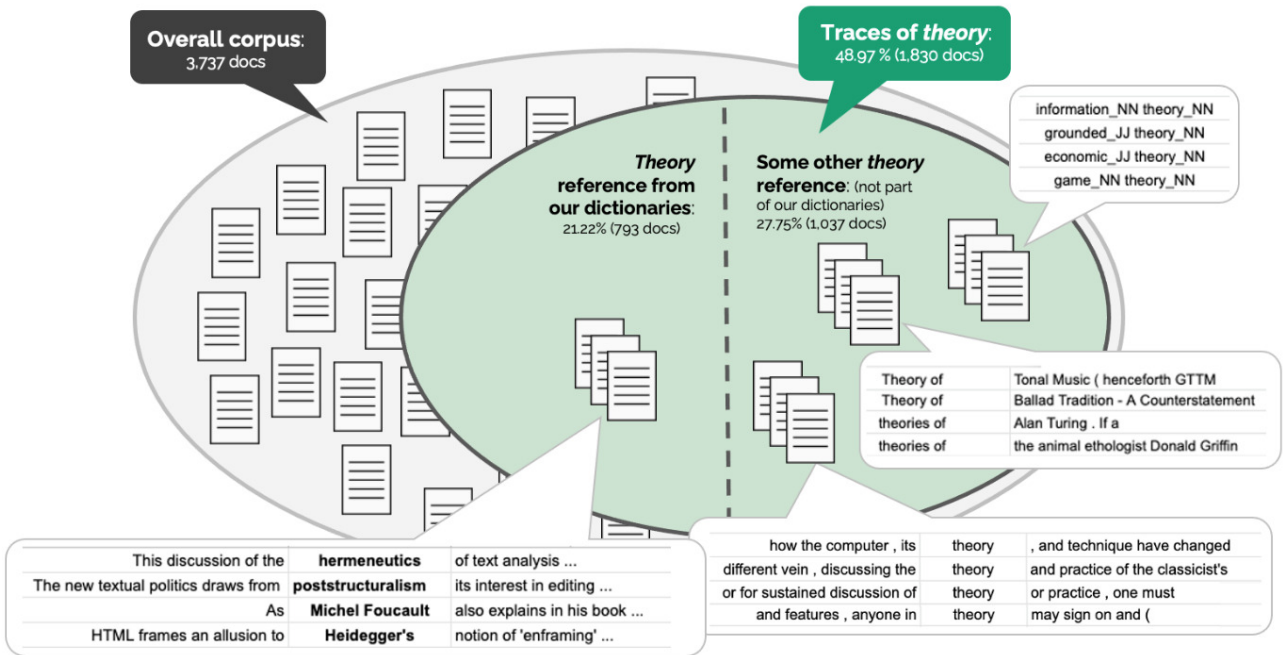
Figure 2. Overview of theory references in our corpus of DH journal articles.

number of unique words in a text, and *tokens*, which are the actual realizations of one type in a text. The TTR is calculated by dividing the number of types by the number of tokens. We obtained results in the range of 0 to 1, where values toward 1 can be interpreted as having high lexical variety. An actual score of 1 would mean that every type of a text is realized by exactly one token; that is, every word used in the text is unique.

For our case of investigating the frequency of theory references within different documents, we calculated TTR exclusively for the theory items of our dictionary, not for the whole document texts. A TTR of exactly 1 here would mean that each theory *type* is realized by exactly one token, which could be interpreted as a rather shallow reference to the theory framework, as one would assume that an article that heavily relies on theoretical references to, for instance "Roland Barthes," would mention him more than just once in the paper. Interestingly, this happens to be the case for 34.8% of articles that have at least one specific theory reference. This shows that most of the documents with a TTR=1 reference exactly one specific theory item, one single time (222). In a few cases, we found two (40), three (11), or four (2) theory references being mentioned exactly one time each. As only a comparatively small number of articles (222) picks up one specific theory reference from our dictionaries only once within the whole article, this might lead to the conclusion that most DH articles do indeed address the topic of theory in more than just a cursory way. This assumption is also reflected by the document frequencies (type count) and total occurrences (token count) of the most frequent theory items in our

Table 2. Overview of the ten-most-frequent dictionary terms with regard to their document frequency.

theory framework / representative scholar	document frequency (type count)	token count	avg. token count	standard deviation avg. token count
hermeneutics	126	600	4.76	1.08
linguistic theory	87	339	3.90	0.46
cultural studies	86	465	5.41	1.35
roland barthes	78	161	2.06	0.49
new critics	66	271	4.11	2.29
critical theory	62	104	1.68	0.31
michel foucault	59	127	2.15	0.58
jacques derrida	54	135	2.50	0.80
phenomenology	34	58	1.71	0.72
sigmund freud	33	60	1.82	0.58

corpus (full list available online),[6] which indicate that these items are heavily referenced in many different articles and also with a rather high density within individual articles.

As the ranking of the top-ten document frequencies (df) by type count (see [Table 2](#)) shows, hermeneutic references (df=126) are particularly dominant, which is probably due to the wide semantic range of hermeneutics. According to Joris van Zundert, hermeneutics "turned from a theory of the interpretation of text into an ontological theory of understanding. It can now be understood broadly as the theory of the processes that turn information into knowledge" (333). As humanities are implicitly indicated as being hermeneutic, DH is also often located within a hermeneutical tradition. Moreover, DH is certainly deeply rooted in textual scholarship and philology. This could also explain the dominance of linguistic theory (df=87).

It is also noteworthy that *close reading* is one of the core methods of "new criticism" (Ransom), which was "an early to mid-twentieth-century literary movement that subordinated the historical [...] concerns of previous scholarship to the text itself" (Bode 92). Therefore, the high document frequency of new critics (df=66) might point to theoretical endeavors framing and conceptualizing *close* and *distant reading*. Alan Liu even argues that DH—the catch-all term *distant reading* in particular—has disturbed the truce between new criticism and cultural-critical readings post 1968: "An unspoken demilitarized zone thus intervened between close and cultural-critical reading. The digital humanities break this détente."

Against this background, the document frequency of critical theory (df=62) can be further commented on. Our dictionary of cultural studies contains critical theory as one type. Critical theory includes, in a broader sense, many

---

6 For a complete frequency list see the "Supplementary Table" spreadsheet (1st tab: frequencies dict_items) at https://theory-in-dh.github.io/conceptual_forays/JoCA2022/conceptual_forays_supplementary.html

theoretical approaches, which "have emerged in connection with the many social movements that identify varied dimensions of the domination of human beings in modern societies" (Bohman). In a narrower sense, critical theory designates the Frankfurt School. Within the theory debate, David M. Berry and other scholars invoke critical theories to strengthen the role of criticism within DH research (Berry 140; Burdick et al. 76).

With regard to Table 2, it is also noteworthy that while poststructuralism is not mentioned explicitly as a theoretical framework, three of its most popular representatives—Roland Barthes (df=78), Michel Foucault (df=59), and Jacques Derrida (df=54)—have surprisingly high document frequencies. We will discuss the role of poststructuralism in more detail in the section on cooccurrence analysis, where we will encounter the names of these three French philosophers once again.

As opposed to the single mentions of theory items that were discussed previously, there are also terms that are referenced extensively within single articles. In our corpus, the terms with the highest average token count in one document are not umbrella terms, such as hermeneutics or poststructuralism, but rather specific theorists (full list available online).[7] Jacques Lacan (dict_psychoanalysis) has an average token count per document of 11.6 and appears in 16 different papers. Lacan is closely followed by Vladimir Propp (dict_structuralism) with an average of 10.2 tokens per document and a document frequency of 20. Frank Raymond Leavis (dict_formalism_new criticism) is another example, as he is mentioned in a total of 8 papers with an average token count of 6.6.

Approaching these single articles via close reading reveals for Lacan that his theoretical constructs are discussed under a computational paradigm. More concretely, Terry Harpold deals with Lacan's four discourses, while Tamise van Pelt's article refers to a Lacanian notion of subjectivity. The case for Propp is a bit different. His formalistic approach seems highly adaptable to the demand for discrete categories that are readable by the computer. Journal articles referencing Propp are then concerned, for example, with the reproducibility of text annotations as well as machine learning (Fisseni et al.; Finlayson).

Leavis's references mostly appear in the context of Charles P. Snow's "two cultures" dichotomy. It is worth highlighting that the Leavis references seem to be intertwined with the narrative "bridging the gap" (Porsdam), which plays an important role in defining an epistemic culture of DH. Leavis (and Snow) are thus representatives of a larger humanities discourse, which is also echoed in DH.

---

7 For a complete frequency list see the "Supplementary Table" spreadsheet (1st tab: frequencies dict_items) at https://theory-in-dh.github.io/conceptual_forays/JoCA2022/conceptual_forays_supplementary.html

Table 3. Overview of the ten-most-frequent dictionary term cooccurrences.

cooccurrence count	type_1	type_2
18	roland barthes (dict_poststructuralism)	michel foucault (dict_poststructuralism)
18	jacques derrida (dict_poststructuralism)	michel foucault
15	roland barthes	jacques derrida
14	cultural studies	critical theory (dict_cultural studies)
13	roland barthes	hermeneutics
13	cultural studies	hermeneutics
13	critical theory	michel foucault
12	jacques derrida	critical theory
12	hermeneutics	critical theory
10	roland barthes	poststructuralism

For row 1, a cooccurrence count of 18 means that Roland Barthes and Michel Foucault cooccurred—each at least one time—in 18 different documents.

COUNTING COOCCURRENCES OF DICTIONARY TERMS

Having discussed the frequencies of single theory items, we will now outline in more detail the results of our investigation into the relations between those items by means of cooccurrence analysis of theory items (full list of cooccurrences available online).[8] These results (see Table 3) largely aligned with the results of the previous frequency analysis (see Table 2), with unexpected theory cooccurrences being rather rare. This might well be an effect of our dictionaries, which are limited to cultural and literary theories in a broader sense. We also did not use any significance weights like *Pointwise Mutual Information* at this point. Rather, this evaluation is intended to explore frequent patterns in the texts and to serve as a first plausibility check.

The highest ranks are again taken by renowned poststructuralists, namely Foucault, Barthes, and Derrida. As might be expected, cultural studies and hermeneutics frequently cooccur with these poststructuralist representatives. The pairing of Barthes and Foucault appears in a total of 18 journal articles, 9 of which are dedicated to the broader topic of authorship theories. Another explanation may be that cooccurrences of Foucault and Derrida (as well as Barthes) could be traced back to text encoding initiatives, which shaped, in particular, early DH projects. Schreibman points out that critics

> saw the possibilities afforded by HTML as the realisation of theories by Barthes, Foucault, Bakhtin and Derrida who wrote of textual openness, nonlinearity and intertextuality [...]. Indeed,

> many first-generation electronic editions conceived in Hypertext Markup Language (HTML) were viewed by their creators as embodiments of post-structuralist theory (285).

In this regard, "hypertext" also appears in over half of the mentions of Derrida and Foucault (10 of 18 cooccurrences), while George P. Landow, who adapted Barthes' ideas for his hypertext theory, is mentioned in these contexts, too.

A close reading of these passages suggests that poststructuralist approaches in DH are mostly being used to underpin aspects of modeling and textual representation. This anecdotal close read demonstrates how the shared uses of our previously defined dictionary items can be utilized to identify possible patterns and qualitatively evaluate them in more detail. The example shown here only highlights the theoretical embeddedness of a particular topic; however, with expanded term lists, more such forays would be possible to explain theory usage in DH publications.

CONCLUSIONS

Our semasiological investigations have revealed how usage, functions, and semantics of *theory* are interfering in DH research. The ranking of the document frequencies by type count provides a rather expected result of *theory* within DH. Hermeneutics, cultural studies (critical theory), and new criticism are frequently brought up as theoretical frameworks considering our dictionaries. What our case study indicates is that the empirical basis of the "end of theory"-narrative in DH is weak, insofar as there are diverse significant references to canonical authors in the tradition of theory in the humanities. This continuity of theoretical reflection might be worth telling, because it is contrary to the idea of a disruptive break in DH's knowledge production. Jean Bauer has already stated in 2010 that "I am sick and tired of people saying that my friends, my colleagues, and I do not understand or care about theory. Every digital humanities project I have ever worked on or heard about is steeped in theoretical implications AND THEIR CREATORS KNOW IT" (Bauer's emphasis). Moreover, it is noteworthy that our case study hints at an uneventfulness of theory in DH. This leads us to suspect that the narrative of "theorylessness" must play a different role for DH's epistemic cultures. M. Beatrice Fazi explains that "the prospect of the end of theory is also reflected in popular concerns about the end of cognitive work due to algorithmic automation, and in related worries about the shrinking of human intellectual faculties in a society where rational decision is increasingly delegated to machines" (107).

The single mentions of theory items as well as its cooccurrences give further insights into the different ways *theory* is used within DH research. We also conducted some scalable readings by oscillating between quantitative explorations and context-sensitive references in the respective articles. While referencing Lacan, for example, seems to go hand in hand with reflecting on philosophical ideas such as subjectivity for DH's specific issues, Leavis and

Snow ("two culture"-dichotomy) show that DH's theories are located within a larger story of theory in sciences. Propp's formalistic approach, however, serves more as a demonstrative example of "theory in practice." Cooccurrences can be interpreted in a similar vein. Theoretical references are made here in order to weave them into practices, tools, and (digital) representations. Although this operationalization of theory has been recently subjected to criticism (Alvarado), it is nevertheless reasonable to assume that theory references are thus linked to a certain expectation; namely, to address humanities claims under a computational paradigm. Future consideration could be given to systematically expanding the dictionaries. In particular, theories from media studies, linguistics, sociology, and computer science should be included.

3.2. Semantic spaces of theory and related concepts (case study 2)

In our second case study, we explored the semantic spaces of *theory* and related concepts, utilizing state-of-the-art neural embedding models (Wevers and Koolen 232). We were thus less concerned with the scope, function, and usage of specific theory frameworks, and instead focused on *theory* itself. The focus therefore shifts to an onomasiological investigation, complementing the semasiological approach in the previous case study. The guiding questions of this second case study are: What other terms have similar contexts like theory? Which terms are used "instead of" or "complementary" to theory in the same or similar contexts? When comparing semantic contexts between related terms or counter concepts, what similarities and differences emerge?

THEORY EMBEDDINGS

One conceptual foundation of this case study can be found in the idea of distributional semantics. The distributional hypothesis suggests that words with similar distributions of context—that is, similar surrounding words—have similar meanings (Harris). For example, lion and tiger have other context words (e.g., teeth and claws) in common than car and bus (e.g., wheels and street). Word embeddings model such cooccurrences of words as vector representations in a multidimensional space, which then can be compared to each other using similarity metrics, such as the cosine distance (Mikolov et al.). We followed existing approaches which assume that semantic representations and their stability can be represented by means of word embeddings and that these embeddings can be compared between diachronic time periods in corpora (Martinc et al.; Giulianelli et al.; Kahmann et al.; Hamilton et al.; Jatowt and Duh).

As a first step, we fine-tuned a pre-trained BERT language model for domain adaptation on our DH journal corpus. Following the approach outlined by Matej Martinc et al., we did not conduct any diachronic fine-tuning. Since embeddings in the BERT language model are contextual, which means they are dependent on the time-specific context, we used this as our input to access the diachronic semantic stability. We used the English BERT-base-uncased model

with 12 attention layers and a hidden layer size of 768. Much like Mario Giulianelli et al. and Martinc et al., we created sequences of byte-pair encoding tokens.

For each of these sequences, we generated a sequence embedding by summing the last four encoder output layers. The resulting sequence embedding represents a concatenation of contextual embeddings for the tokens in the input sequence. We sliced those concatenations and acquired a representation (i.e., a contextual token embedding) for each word usage in our journal articles. These representations differ depending on the context in which the token is embedded. As a consequence, the same word has a different representation in each context. Finally, we could aggregate the embeddings on the token level and were able to compare different time spans in the corpus with regard to the semantic representation of the theory-related vocabulary in our study.

In a first experiment, we extracted different word usage contexts of *theory* via its local embeddings, hoping to reveal different senses of *theory*. A k-means clustering approach was chosen to explore different embeddings in the texts (see Figure 3). We used a silhouette analysis to identify the optimal number of clusters (*k*), which in this case appears to be 6 clusters. Interestingly, the clusters mostly reflect different forms of syntactic embeddings of the word form *theory*, for instance "*in* theory," "theory *of*," "*noun* + theory," and the plural form "theories," as well as a cluster with concrete instances of theories and a cluster with rather conceptual aspects of theory.

The mere clustering of the local embeddings of the word theory is obviously limited in its expressive power regarding the role of theory in DH, however, as this simple approach does not reveal distinct senses and usage patterns of *theory* as a concept. We therefore also took a diachronic perspective on *theory*, aggregating the concept's local embeddings for different time slices. To gain better insights into different senses of *theory*, we also chose to analyze semantically related concepts, to see how those relate to *theory* in the course of time.

COMPARISON OF *THEORY* AND OTHER DH CONCEPTS THROUGH TIME

Because we sought to examine relations of *theory* to contextual word embeddings of further concepts that play a role in DH's epistemic cultures, we took a closer look at *model*, *method*, *experiment*, and *tool*, as they seem essential for characterizing DH research and are frequently encountered in corresponding discourses. Above all, we were interested not only in semantic ambiguity but also in questions of controversiality and indispensability of *theory*. *Modeling*, for instance, is described as a core DH activity (Flanders and Jannidis, *Knowledge Organization and Data Modeling in the Humanities*). At the same time, DH research is considered to expand the *methodological*
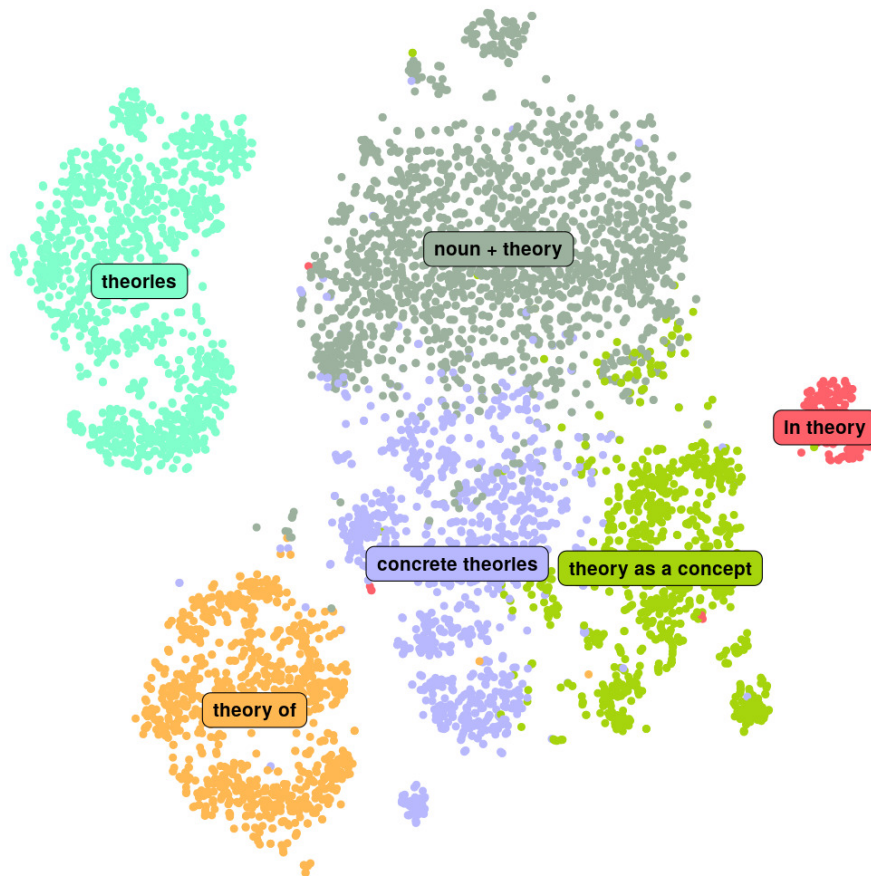
Figure 3. K-means clustering of the embeddings with k = 6.

repertoire of the humanities by using *tools* and other infrastructural settings. Finally, the term *experiment* will be discussed against the background of the emerging DH laboratories (Pawlicka-Deger).

To provide an overview of the basic contexts of the epistemic concepts *theory*, *model*, *method*, *tool*, and *experiment*, Table 4 shows the top-10 terms that have the most similar embedding vectors. We will revisit Table 4 in the following discussion on the relationship between these five concepts. Figure 4 provides an overview of how the concepts *tool*, *model*, *experiment*, and *method* evolve with regard to their semantic similarity to *theory*. Following the suggestions of Shen Dinghan et al. and Vitalii Zhelezniak et al., we used max pooling of the contextualized embeddings of each term within a 3-year slice, as it typically takes out the influence of syntactical embedding information.

Interestingly, *theory* is almost a straight line, as its self-similarity is rather high, even from the beginning of the period surveyed. As time goes by, some more nuanced contexts are added to the *theory* vector. The vectors of the other concepts are stacked for every 3-year-slice, which means as time goes by, more contextual meaning is also added to the other vectors. The graph also shows

Table 4. Top-10 highest-ranked terms for theory, model, method, tool, and experiment.

Rank	theory	model	method	tool	experiment
1	ideology	framework	technique	instrument	trial
2	theorization	representation	procedure	technique	test
3	principle	theory	approach	method	laboratory
4	hypothesis	approach	algorithm	resource	investigation
5	doctrine	prototype	strategy	implement	exercise
6	methodology	paradigms	tool	software	simulation
7	idea	simulator	tactic	facility	research
8	philosophy	idealization	mode	platform	observation
9	conceptualization	conceptualization	mechanism	device	study
10	paradigms	system	process	weapon	project

The ranking is based on cosine similarities, which is mostly between 0.6 and 0.7 for the above terms.
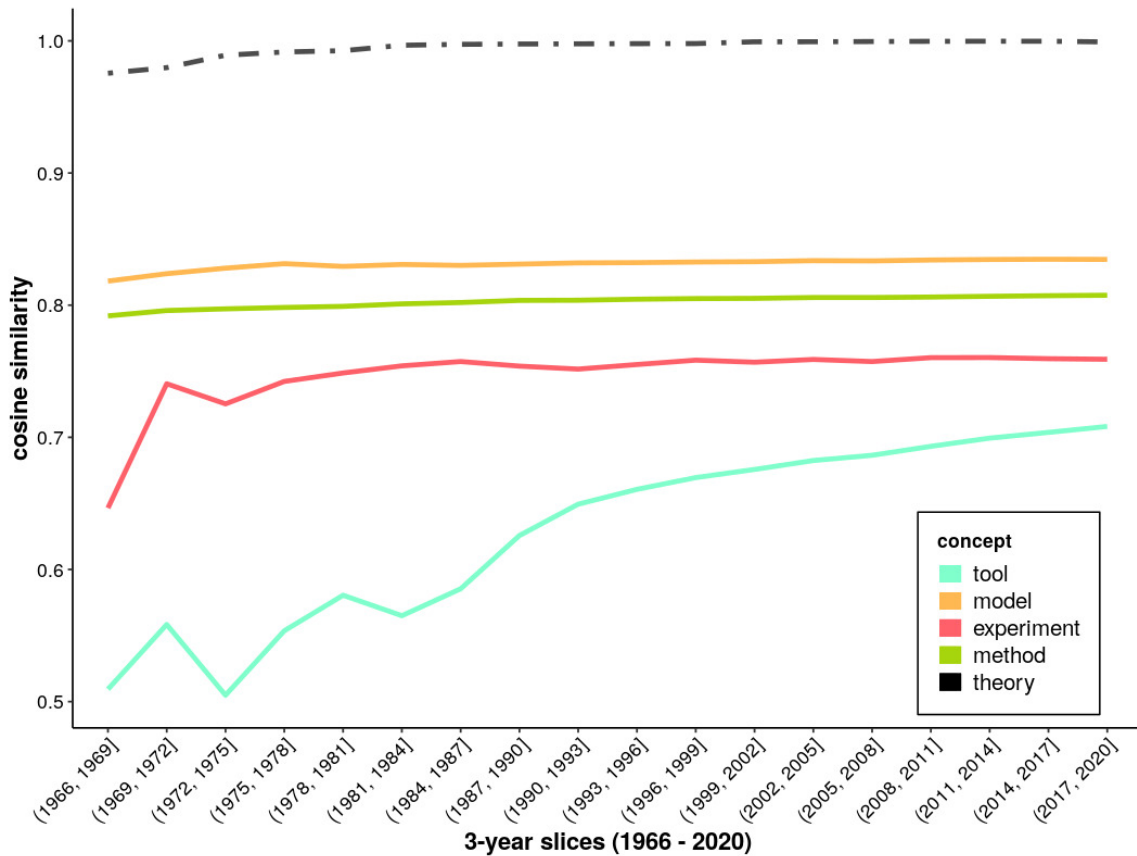
Figure 4. Overview of different concepts and their accumulated cosine similarity to the theory vector through time.

The cosine similarity is based on max-pooling of the contextualized embeddings of each term within a 3-year slice.

that the ranking of similar concepts is stable over time, meaning *model* is always the concept that is most similar to *theory*, then comes *method*, *experiment*, and, finally, *tool*, which has fairly low numbers in the beginning.

As [Figure 4](#) shows, we initially observe a gradual increase of contextual meaning of *theory*, while from the 1990s on, contextual meanings of *theory* became established. As with *model*, between the 1966 and 1990, an increase of contextual meanings can be observed, but this quickly levels off. According

to our experimental design, it seems worth mentioning that *theory* and *model* were neither used synonymously nor did they gradually diverge. Rather, both concepts converged quite soon to their final contextual meaning. Fotis Jannidis and Julia Flanders mention that within DH research views on formal modeling among others are advocated which "tend [...] to focus on how the concept of the 'model' is itself embedded in more general concepts like theory and how they, theories and models determine or at least interact with the formal model" (28). Comparing both context vectors may broaden the perspective of how *model* and *theory* are related.

As [Table 4](#) shows, different semantic fields of *theory* and *model* can be described over the entire period covered by our corpus. The terms *principle*, *hypothesis*, *doctrine*, and *paradigm* indicate that *theory* rather covers contexts of regularity as well as (axiomatic) programmatic and forms of intersubjective knowledge, while *model* has more contextual similarity with terms such as *representation*, *prototype*, and *simulation*, which focus on different kinds of mapping as well as formal systematizations. This could be one reason why *theory* and *model* have the same contextual distance to each other (see [Figure 4](#)): they consistently seem to cover different semantic fields. It could be assumed that these two semantic fields might complement each other. Thus, in the top-10 ranking of terms with the most similar context words as *model*, *theory* takes the third highest rank. This is not so with *model*, which does not appear in the top-10 ranking of *theory* at all.

These semantic frictions between *theory* and *model* seem to represent the ambiguity of these concepts within DH research. Neither the concepts nor their semantic fields can be transferred—rather, they intertwine. In comparison, the relation between *theory* and *method* seems to be more explicit. Not only do they cover distinguishable semantic fields, as [Table 4](#) clearly shows, but the top terms of *methods* include *technique*, *procedure*, *approach*, and *strategy*, while the distance of the context-vectors remains constant (see [Figure 4](#)).

A slightly different picture emerged when we compared the semantic contexts of *theory* and *experiment* (see [Figure 4](#)). While the vector for *experiment* seems to have been established in the first six years, it then slowly converges to *theory*, adding further meanings. The top-10 highest-ranked terms for *experiment*, however, point to an (oppositional) semantic field; namely, the more practical side of DH research. Moreover, the highest ranking terms for *experiment* seem to break into two semantic fields: 1) terms such as *trial*, *test*, *laboratory*, *investigation*, and *observation* suggest contexts that could be associated with empirical science research settings. The fact that the first place is occupied by *trial* is particularly noteworthy, as it introduces notions of failure within DH research. 2) The terms *research*, *study*, and *project* seem to bring into sharper relief ongoing transformations in DH in general. Strictly speaking, the most similar context to research and project has the term *experiment* with regard

to our five selected terms. According to Ian Hacking, experiments are not subordinate to theory; rather, "experimentation has a life of its own" (Hacking 150). With regard to the decoupling of experiment and theory building, Willard McCarty concludes that for DH,

> we can infer that humanities computing likewise need not wait on the emergence of a theoretical framework, that its semidirected, semicoherent activities are no discredit, rather the norm for an experimental field. Furthermore, we may find deep kinship in the complex, constructivist idea that, to put the matter crudely, scientific knowledge is both found and made (1133).

Interestingly, however, the semantic similarity between *theory* and *experiment* does not change within our corpus (see [Figure 4](#)). Thus, no meanings are added that would further distance the two concepts. This situation is somewhat different for the concept *tool*. Between 1966 and 1985 the *tool* vector fluctuated in its movement as it drew closer and moved further away from the *theory* vector. From the mid-1980s on, there is a steady approximation. It could be hypothesized that one reason for this constant convergence of these two vectors might be the emerging idea of a theory-driven development of tools.

CHANGE OF NEAREST NEIGHBORS

Next, we took a closer look at the *theory* vector and how it changed its contexts over time. We created a matrix that contains all the ranks for all the other terms for each of the 3-year-slices (full list of ranks over time available online).[9] For the purpose of a more consistent comparison that is not distorted by grammatical or syntactic effects, we compared *theory* only to other nouns.

[Figure 5](#) provides an overview of the top-25 terms that are best ranked by means of their cosine similarity to the concept *theory* throughout the entire corpus. As some of these concepts have quite some fluctuations in their ranks through time, it is impossible to visualize the whole spectrum of ranks for multiple terms in one plot. Therefore, we decided to only visualize the rank movement within the first 25 ranks. *Theorist*, for example, steadily increases its ranks (1969–1972: rank 1,030 → 1981–1983: rank 472 → 2002–2004: rank 49), but it is only in the time slice of 2011 to 2013 that it hits the top 25 ranks and thus appears in the graph.

Strikingly, the terms *principle*, *idea*, and *concept* have the most similar context vectors to *theory,* indicating a rather fundamental and conceptual use of *theory* in the context of DH research articles. These terms are closely followed by the term *model*, reaffirming the observations about the rather stable and semantically distinct relation between *theory* and *model* that were already

---

9 For a complete list of ranks over time see the "Supplementary Table" spreadsheet (3rd tab: ranks over time) at https://theory-in-dh.github.io/conceptual_forays/JoCA2022/conceptual_forays_supplementary.html
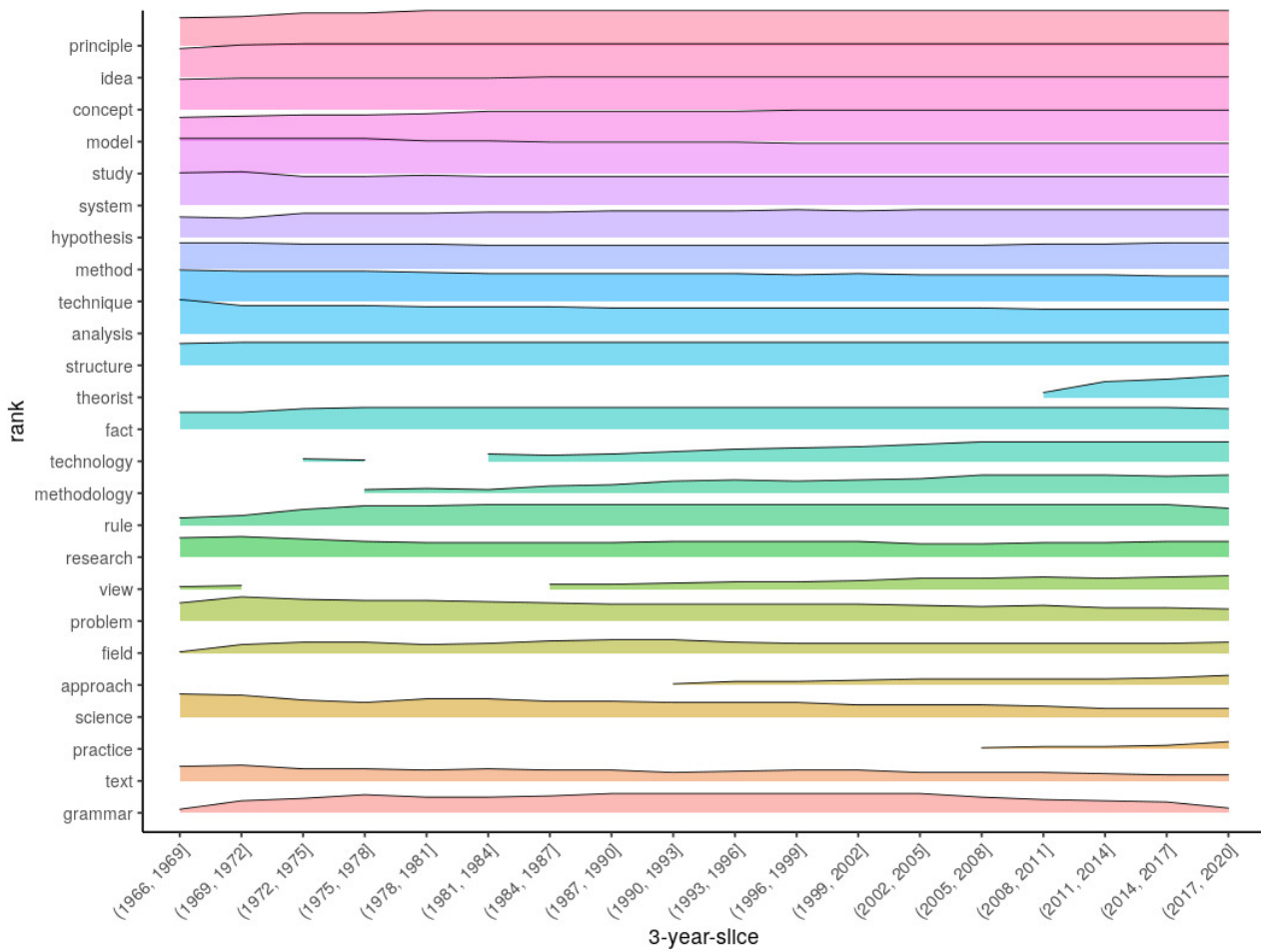
Figure 5. Overview of the top-25 global best-ranked terms and their accumulated cosine similarity to the theory vector through time.

We used a threshold of the top-25 ranks, that means anything below that threshold is not visualized in the plot. The cosine similarity is based on max-pooling of the contextualized embeddings of each term within a 3-year slice.

described for [Figure 4](#). Interestingly, technology and methodology only appear in the top-20 most-similar terms in the early 1990s. It could be argued that technology's appearance in the top 20 is related to the dissemination of large and searchable (textual) databases, and more generally to the advent of the World Wide Web and its manifold technological inventions and implications. This seems to be somehow related to the need for a methodology, in other words a more abstract reflection, systematization, and theorization of methods.

While methodology seems to co-evolve with technology, *method* on the other hand remains rather stable and rather close to *theory* throughout time. Surprisingly, Underwood characterizes the 1990s as the period where theory building was missed (70). At the same time, the 1990s seem to play a crucial role in DH's epistemic culture, as the periodization approach introduced by Todd Presner and Jeffrey Schnapp, who suggest two main waves of DH, also starts precisely here. The first wave, from the mid-1990s to the early 2000s, focuses "on large-scale digitization projects and the establishment of technological infrastructure." The second wave, which entails the period from the early 2000s until 2010, is "deeply generative, creating the environments and

tools for producing, curating, and interacting with knowledge that is 'born digital' and lives in various digital contexts." While this periodization has been criticized by some, we can definitely add to this debate with our observation that significant semantic shifts can be observed during this period (Berry 4). As technology and methodology gain new meanings, they also become more similar to the contextual embeddings of *theory.*

CONCLUSIONS

Our second case study presents an onomasiological investigation of *theory* within DH research. A key insight from the comparison of *theory* and related epistemic concepts shows that most acquired their full contextual meaning early, in the first five years of our corpus. Afterwards, the concepts remained rather stable, in terms of their self-similarity on the one hand, and their similarity to the concept of *theory* on the other. A further investigation of the epistemological role of tools in DH is very much needed.

Another interesting insight comes from the comparison of the two contextual embeddings of *theory* and *model*. It became apparent that *model* comes closer to contexts of representation. Moreover, the relation of (opposite) terms to *theory* such as *experiment* and *tool* was found to be rather stable and expected. To put it simply, our contextual embeddings stories are more predictable and less eventful than the often polarized (research) discourse might convey. Looking at the time component of theory references, it is striking that the mid-1990s was a point when new (conceptual) stories began to contextually interfere with each other or were forgotten (Underwood). Further research approaches could therefore focus more on periodization issues.

4. Final remarks

In this paper we investigated narratives of *theory* through a computational conceptual history approach. The fundamental assumption was that a conceptual study of *theory* can shed light on research discourses and knowledge structures in DH's epistemic cultures. Our investigation was founded on the premise that our conceptual history approach can be regarded as a first foray into the field of a DH-specific history of science. Therefore, our semasiological and onomasiological studies aimed to provide a comprehensive picture of *theory* within DH research. We understand *theory* as one central concept of DH research, which is semantically ambiguous but also highly indispensable. In our article, we have therefore presented two possible storylines of *theory* that we encountered in our conceptual forays. Our first storyline (case study 1) was about the range, frequencies, and functions of the concept of *theory* in DH research articles. Our second storyline (case study 2) was about the contextual embeddings of *theory* and related concepts that are central to an epistemology of DH.

Our conceptual stories of *theory* emerged within the context of our specific experimental setting, which is obviously limited and biased by our specific corpus of DH journals as well as by our specific selection of dictionaries and methods. We want to highlight that we do not claim to have identified and sufficiently discussed central theoretical narratives in DH. Rather, we have presented a methodology that is inspired by current approaches to computational conceptual history, allowing us to make various forays into the development of the concept *theory*. This article is to be understood as an invitation to follow our approach and to contribute further storylines in order to draw a bigger, more complete picture of the role, function, and development of theory in DH over time.

Data Repository: https://doi.org/10.7910/DVN/DKPKGD

REFERENCES

Alvarado, Rafael C. "Digital Humanities and the Great Project: Why We Should Operationalize Everything and Study Those Who Are Doing so Now." *Debates in the Digital Humanities 2019*, edited by Matthew K. Gold et al., University of Minnesota Press, 2019, pp. 75–82, doi:10.5749/j.ctvg251hk.9.

Anderson, Chris. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete." *Wired*, 23 June 2008, http://www.wired.com/2008/06/pb-theory/.

Arnold, Eckhart. "Positivistischer Methodenfetischismus als Anathema der digitalen Geisteswissenschaften." *DHd 2018 Kritik der digitalen Vernunft. Book of Abstracts*, Feb. 2018, doi:10.5281/ZENODO.4622295.

Baroni, Raphaël. "Tellability." *The Living Handbook of Narratology*, 2019, https://www-archiv.fdm.uni-hamburg.de/lhn/node/30.html.

Bauer, Jean. "Who You Calling Untheoretical?" *Journal of Digital Humanities*, vol. 1, no. 1, 2011.

Berry, David M. "Introduction: Understanding the Digital Humanities." *Understanding Digital Humanities*, edited by David M. Berry, Palgrave Macmillan, 2012, pp. 1–20, doi:10.1057/9780230371934_1.

Berry, David M., and Anders Fagerjord. *Digital Humanities: Knowledge and Critique in a Digital Age*. Polity Press, 2017.

Betti, Arianna, and Hein van den Berg. "Modelling the History of Ideas." *British Journal for the History of Philosophy*, vol. 22, no. 4, July 2014, pp. 812–35, doi:10.1080/09608788.2014.949217.

Bode, Katherine. "The Equivalence of 'Close' and 'Distant' Reading; Or, Toward a New Object for Data-Rich Literary History." *Modern Language Quarterly*, vol. 78, no. 1, Mar. 2017, pp. 77–106, doi:10.1215/00267929-3699787.

Bohman, James. "Critical Theory." *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, 2021, http://plato.stanford.edu/archives/spr2021/entries/critical-theory.

Brandt, Christina. "Wissenschaft." *Erzählen: Ein interdisziplinäres Handbuch*, edited by Matías Martínez, Metzler, 2017, pp. 211–18.

Burckhardt, Daniel, et al. "Distant Reading in der Zeitgeschichte. Möglichkeiten und Grenzen einer computergestützten Historischen Semantik am Beispiel der DDR-Presse." *Zeithistorische Forschungen/Studies in Contemporary History*, vol. 16, no. 1, 2019, pp. 177–96, doi:10.14765/ZZF.DOK-1345.

Burdick, Anne, et al., editor. *Digital Humanities*. Paperback, MIT Press, 2012.

Cameron, Deborah, and Don Kulick. *Language and Sexuality*. Cambridge University Press, 2003, doi:10.1017/cbo9780511791178.

Castle, Gregory. "The Blackwell Guide to Literary Theory." *Blackwell Guides to Literature*, Blackwell, 2007.

Cecire, Natalia. "Introduction: Theory and the Virtues of Digital Humanities." *Journal of Digital Humanities*, vol. 1, no. 1, 2011.

Chatman, Seymour. *Coming to Terms. The Rhetoric of Narrative in Fiction and Film*. Cornell University Press, 1990.

Clarin-D. *DiaCollo Tool*. 2016, https://www.clarin-d.net/de/kollokationsanalyse-in-diachroner-perspektive.

De Bolla, Peter, et al. "Distributional Concept Analysis." *Contributions to the History of Concepts*, vol. 14, no. 1, June 2019, pp. 66–92, doi:10.3167/choc.2019.140104.

D'Ignazio, Catherine, and Lauren F. Klein. *Data Feminism*. The MIT Press, 2020, doi:10.7551/mitpress/11805.001.0001.

Drucker, Johanna. "Humanistic Theory and Digital Scholarship." *Debates in the Digital Humanities*, edited by Matthew K. Gold, University of Minnesota Press, 2012, pp. 85–95, doi:10.5749/minnesota/9780816677948.003.0011.

El Khatib, Randa, et al. "Prototyping Across the Disciplines." *Digital Studies/Le champ numérique*, vol. 8, no. 1, 2019, pp. 1–10, doi:10.16995/dscn.282.

Elliott, Jane, and Derek Attridge. "Introduction: Theory's Nine Lives." *Theory after Theory*, edited by Jane Elliott and Derek Attridge, Routledge, 2011, pp. 1–17.

Endres, Bill. "A Literacy of Building: Making in the Digital Humanities." *Making Things and Drawing Boundaries*, edited by Jentery Sayers, University of Minnesota Press, 2018, pp. 44–54, doi:10.5749/j.ctt1pwt6wq.7.

Fazi, M. Beatrice. "The Ends of Media Theory." *Media Theory*, vol. 1, no. 1, 2017, pp. 107–21.

Finlayson, Mark A. "ProppLearner: Deeply Annotating a Corpus of Russian Folktales to Enable the Machine Learning of a Russian Formalist Theory." *Digital Scholarship in the Humanities*, vol. 32, no. 2, 2017, pp. 284–300.

Fisseni, Bernhard, et al. "Annotating with Propp's *Morphology of the Folktale*: Reproducibility and Trainability." *Literary and Linguistic Computing*, vol. 29, no. 4, Oct. 2014, pp. 488–510, doi:10.1093/llc/fqu050.

Flanders, Julia, and Fotis Jannidis. *Knowledge Organization and Data Modeling in the Humanities*. 1 Jan. 2015.

---, editors. *The Shape of Data in the Digital Humanities: Modeling Texts and Text-Based Resources*. Routledge, 2019.

Friedrich, Alexander, and Chris Biemann. "Digitale Begriffsgeschichte? Methodologische Überlegungen Und Exemplarische Versuche Am Beispiel Moderner Netzsemantik." *Forum Interdisziplinäre Begriffsgeschichte*, vol. 5, no. 2, 2017, pp. 78–96.

Giulianelli, Mario, et al. "Analysing Lexical Semantic Change with Contextualised Word Representations." *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, edited by Dan Jurafsky et al., Association for Computational Linguistics, pp. 3960–73.

Gold, Matthew K. "The Digital Humanities Moment." *Debates in the Digital Humanities*, edited by Matthew K. Gold, University of Minnesota Press, 2012.

Gutiérrez de la Torre, Silvia E., et al. "The Many Faces of Theories in DH Journals: Toward a Dictionary of Theoreticians Mentioned in DH." *Proceedings Digital Humanities Conference Tokyo*, 2022, pp. 215–17, https://dh2022.dhii.asia/dh2022bookofabsts.pdf.

Haase, Christian, et al. "SCoT: Sense Clustering over Time: A Tool for the Analysis of Lexical Change." *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*, EACL, 2021, pp. 198–204.

Hacking, Ian. *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*. Cambridge University Press, 1983, doi:10.1017/cbo9780511814563.

Hall, Gary. "Has Critical Theory Run Out of Time for Data-Driven Scholarship?" *Debates in the Digital Humanities*, edited by Matthew K. Gold, University of Minnesota Press, 2012, pp. 127–32, doi:10.5749/minnesota/9780816677948.003.0015.

Hamilton, William L., et al. "Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change." *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, vol. 1, 2016, pp. 1489–501, doi:10.18653/v1/p16-1141.

Harpold, Terry. "The Underside of the Digital Field." *Digital Humanities Quarterly*, vol. 6, no. 2, 2012, http://www.digitalhumanities.org/dhq/vol/6/2/000141/000141.html.

Harré, Rom. "Some Narrative Conventions of Scientific Discourse." *Narrative in Culture: The Uses of Storytelling in the Sciences, Philosophy, and Literature*, edited by Christopher Nash, Routledge, 1990, pp. 81–101.

Harris, Zellig S. "Distributional Structure." *Word*, vol. 10, no. 2–3, Aug. 1954, pp. 146–62, doi:10.1080/00437956.1954.11659520.

Hess, Carla W., et al. "The Type-Token Ratio and Vocabulary Performance." *Psychological Reports*, vol. 55, no. 1, Aug. 1984, pp. 51–57, doi:10.2466/pr0.1984.55.1.51.

Jannidis, Fotis, and Julia Flanders. "A Gentle Introduction to Data Modeling." *The Shape of Data in the Digital Humanities: Modeling Texts and Text-Based Resources*, edited by Julia Flanders and Fotis Jannidis, Routledge, 2019, pp. 26–95.

Jatowt, Adam, and Kevin Duh. "A Framework for Analyzing Semantic Change of Words Across Time." *Proceedings of the 14th ACMIEEE-CS Joint Conference on Digital Libraries, 9/8/2014 - 9/12/2014, London, United Kingdom, 2014 IEEE/ACM Joint Conference on Digital Libraries (JCDL)*, edited by George Buchanan, IEEE Press, 2014, pp. 229–38, doi:10.1109/jcdl.2014.6970173.

Jockers, Matthew L. *Macroanalysis: Digital Methods and Literary History*. University of Illinois Press, 2013, doi:10.5406/illinois/9780252037528.001.0001.

Kahmann, Christian, et al. "Detecting and Assessing Contextual Change in Diachronic Text Documents Using Context Volatility." *Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 2017, doi:10.5220/0006574001350143.

Kleymann, Rabea. "Softwareprototyp: Zur Theoriemündigkeit Experimenteller Software in Den Digital Humanities." *Wovon sprechen wir, wenn wir von Digitalisierung sprechen? Gehalte und Revisionen zentraler Begriffe des Digitalen*, edited by Martin Huber et al., 1st ed., CompaRe, 2020, pp. 119–38.

Knorr Cetina, Karin. *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard University Press, 1999, doi:10.4159/9780674039681.

Kollmeier, Kathrin. "Begriffsgeschichte und Historische Semantik." *Docupedia-Zeitgeschichte*, 2012, doi:10.14765/ZZF.DOK.2.257.V2.

Koselleck, Reinhart. *The Practice of Conceptual History: Timing History, Spacing Concepts*. Translated by Todd Samuel Presner, Stanford University Press, 2002, doi:10.1515/9781503619104.

Koselleck, Reinhart, and Michaela Richter. "Basic Concepts in History: A Historical Dictionary of Political and Social Language in Germany." *Contributions to the History of Concepts*, vol. 6, no. 1, Jan. 2011, pp. 1–37, doi:10.3167/choc.2011.060102.

Landow, George P. *Hypertext: The Convergence of Contemporary Critical Theory and Technology*. 2. print, Johns Hopkins Univ. Press, 1992.

Latour, Bruno, and Steve Woolgar. *Laboratory Life: The Construction of Scientific Facts*. Princeton University Press, 2013, doi:10.2307/j.ctt32bbxc.

Liu, Alan. "Where Is Cultural Criticism in the Digital Humanities?" *Debates in the Digital Humanities*, edited by Matthew K. Gold, University of Minnesota Press, 2012, pp. 490–510, doi:10.5749/minnesota/9780816677948.003.0049.

Luhmann, Jan, and Manuel Burghardt. "Digital Humanities—A Discipline in Its Own Right? An Analysis of the Role and Position of Digital Humanities in the Academic Landscape." *Journal of the Association for Information Science and Technology*, vol. 73, no. 2, June 2021, pp. 148–71, doi:10.1002/asi.24533.

---. "Same Same, but Different? On the Relation of Information Science and the Digital Humanities A Scientometric Comparison of Academic Journals Using LDA and Hierarchical Clustering." *Proceedings of the 16th International Symposium of Information Science (ISI2021): "Information between Data and Knowledge – Information Science and Its Neighbors from Data Science to Digital Humanities*, edited by Christian Wolff and Thomas Schmidt, Regensburg, 2021, pp. 173–99, doi:10.5283/EPUB.44944.

Lyotard, Jean-François. *The Postmodern Condition: A Report on Knowledge*. Repr. Manchester Univ. Press, 1986.

Malazita, James W., et al. "Digital Humanities as Epistemic Cultures: How DH Labs Make Knowledge, Objects, and Subjects." *Digital Humanities Quarterly*, vol. 14, no. 3, 2020, http://www.digitalhumanities.org/dhq/vol/14/3/000465/000465.html.

Martinc, Matej, et al. "Leveraging Contextual Embeddings for Detecting Diachronic Semantic Shift." *arXiv*, Mar. 2020, http://arxiv.org/abs/1912.01072.

McCarty, Willard. "Humanities Computing." *Encyclopedia of Library and Information Science*, 2003, pp. 1124–35.

Michel, Jean-Baptiste, et al. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science*, vol. 331, no. 6014, Jan. 2011, pp. 176–82, doi:10.1126/science.1199644.

Mikolov, Tomas, et al. "Distributed Representations of Words and Phrases and Their Compositionality." *Advances in Neural Information Processing Systems*, vol. 2013, pp. 3111–19, http://arxiv.org/pdf/1310.4546.

Moretti, Franco. *Distant Reading*. Verso, 2013.

Müller, Ernst, and Falko Schmieder. *Begriffsgeschichte und Historische Semantik: Ein kritisches Kompendium*. Suhrkamp, 2016.

---. "Begriffsgeschichte und Wissenschaftsgeschichte." *Geschichte und Gesellschaft*, vol. 44, no. 1, Mar. 2018, pp. 79–106, doi:10.13109/gege.2018.44.1.79.

---. *Begriffsgeschichte: Zur Einführung*. Junius, 2020.

Pawlicka-Deger, Urszula. "The Laboratory Turn: Exploring Discourses, Landscapes, and Models of Humanities Labs." *Digital Humanities Quarterly*, vol. 14, no. 3, 2020, http://www.digitalhumanities.org/dhq/vol/14/3/000466/000466.html.

Plotnitsky, Arkady. "Science and Narrative." *Routledge Encyclopedia of Narrative Theory*, edited by David Herman et al., Routledge, 2005, pp. 514–18.

Porsdam, Helle. "Digital Humanities: On Finding the Proper Balance Between Qualitative and Quantitative Ways of Doing Research in the Humanities." *Digital Humanities Quarterly*, vol. 7, no. 3, 2013, http://www.digitalhumanities.org/dhq/vol/7/3/000167/000167.html.

Presner, Todd. *Digital Humanities 2.0: A Report on Knowledge*. 19 Apr. 2010, https://cnx.org/contents/iqMKXpSE@1.4:J0K7N3xH@6/Digital-Humanities-2-0-A-Report-on-Knowledge.

Presner, Todd, and Jeffrey Schnapp. *Digital Humanities Manifesto 2.0*. 2009, http://humanitiesblast.com/publications/.

Ramsay, Stephen, and Geoffrey Rockwell. "Developing Things: Notes Toward an Epistemology of Building in the Digital Humanities." *Debates in the Digital Humanities*, edited by Matthew K. Gold, University of Minnesota Press, 2012, pp. 75–84, doi:10.5749/minnesota/9780816677948.003.0010.

Ransom, John Crowe. "Criticism, Inc." *The Virginia Quarterly Review*, vol. 13, no. 4, 1937, pp. 586–602.

Risam, Roopika. *New Digital Worlds: Postcolonial Digital Humanities in Theory, Praxis, and Pedagogy*. Northwestern University Press, 2019, doi:10.2307/j.ctv7tq4hg.

Rivkin, Julie, and Michael Ryan, editors. *Literary Theory: An Anthology*, 2nd ed., Blackwell Publication, 2007.

Rockwell, Geoffrey, et al. "The Trace of Theory: Extracting Subsets from Large Collections." *Digital Humanities 2016: Conference Abstracts*, Jagiellonian University & Pedagogical University, 2016, pp. 85–92, https://dh2016.adho.org/abstracts/139.

Ryan, Marie-Laure, et al. "Narrative." *Routledge Encyclopedia of Narrative Theory*, edited by David Herman, Routledge, 2005, pp. 344–48.

Schreibman, Susan. "Computer-Mediated Texts and Textuality: Theory and Practice." *Computers and the Humanities*, vol. 36, no. 3, 2002, pp. 283–93, doi:10.1023/a:1016178200469.

Schwandt, Silke. "Digitale Methoden für die Historische Semantik: Auf den Spuren von Begriffen in digitalen Korpora." *Geschichte und Gesellschaft*, vol. 44, no. 1, Mar. 2018, pp. 107–34, doi:10.13109/gege.2018.44.1.107.

Selden, Raman, et al. *A Reader's Guide to Contemporary Literary Theory*. 5th ed., Pearson Longman, 2006.

Shen, Dinghan, et al. *Baseline Needs More Love: On Simple Word-Embedding-Based Models and Associated Pooling Mechanisms*. arXiv, 2018, https://arxiv.org/abs/1805.09843.

Sinclair, Stéfan, and Geoffrey Rockwell. Voyant Tools, 2016, https://voyant-tools.org/.

Sula, Chris Alen, and Heather V. Hill. "The Early History of Digital Humanities: An Analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004)." *Digital Scholarship in the Humanities*, vol. 34, no. 1, Nov. 2019, pp. 190–206, doi:10.1093/llc/fqz072.

Terras, Melissa, et al, editor. *Defining Digital Humanities: A Reader*. Taylor & Francis, 2013.

Thelwall, Mike, et al. "Academic LGBTQ+ Terminology 1900-2021: Increasing Variety, Increasing Inclusivity?" *Journal of Homosexuality*, 2022, pp. 1–25, doi:10.1080/00918369.2022.2070446.

Underwood, Ted. "Theorizing Research Practices We Forgot to Theorize Twenty Years Ago." *Representations*, vol. 127, no. 1, 2014, pp. 64–72, doi:10.1525/rep.2014.127.1.64.

van Pelt, Tamise. "The Question Concerning Theory: Humanism, Subjectivity, and Computing." *Computers and the Humanities*, vol. 36, no. 3, 2002, pp. 307–18, doi:10.1023/a:1016160114582.

van Vree, Frank, et al., editor. *History of Concepts: Comparative Perspectives*, Amsterdam University Press, 2010.

van Zundert, Joris J. "Screwmeneutics and Hermenumericals: The Computationality of Hermeneutics." *A New Companion to Digital Humanities*, edited by John Unsworth et al., John Wiley & Sons Ltd, 2016, pp. 331–47.

Vogelsang, Kai. "Conceptual History: A Short Introduction." *Oriens Extremus*, vol. 51, 2012, pp. 9–24.

Warwick, Claire. "Building Theories or Theories of Building? A Tension at the Heart of Digital Humanities." *A New Companion to Digital Humanities*, edited by John Unsworth et al., John Wiley & Sons Ltd, 2016, pp. 538–52.

Weingart, Scott, et al. *The Index of Digital Humanities Conferences*. Carnegie Mellon University, 2020, doi:10.34666/k1de-j489.

Wevers, Melvin, and Marijn Koolen. "Digital Begriffsgeschichte: Tracing Semantic Change Using Word Embeddings." *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, vol. 53, no. 4, May 2020, pp. 226–43, doi:10.1080/01615440.2020.1760157.

Zhelezniak, Vitalii, et al. *Don't Settle for Average, Go for the Max: Fuzzy Sets and Max-Pooled Word Vectors*. arXiv, 2019, https://arxiv.org/abs/1904.13264.