Digital History and the Study of Modernity

Pim Huijnen

Abstract

This introductory article to the special issue on digital history claims that digital history bears promise for the study of modernity. The German historian Reinhart Koselleck argued that the rise of modernity involved fundamental discursive shifts. These can be found in texts, of which many are rapidly becoming available in digitized form, while the steady rise of new text mining techniques provides unprecedented opportunities to explore these texts in innovative ways. However, digital historians need to use these techniques critically since most are not geared towards the study of history. In particular, historians have to account for comparability, chronology and language. The articles in this special issue offer showcases of how digital historians have shed new light on various aspects of modernity.

Keywords: conceptual history, digital collections, digital history, historical semantics, Koselleck, text mining

Introduction

Digital history is the term usually adopted to describe the historical profession's occupation with recent technological developments: the democratic access to digitization of historical material in massive quantities, the introduction of computational techniques for the analysis of historical data, the pervasive opportunities afforded by the internet and social media to both publish the results of analysis and communicate with colleagues and readers. A sturdy debate on the nature

of the changes the historical profession is going through accompanies these developments. How are we to gauge these changes? Do conventional habits and practices persist in new shapes and forms, such as the discussions between scholars on Twitter that have been replacing the telephone calls, or the online, 'enhanced' publications that are now superseding the traditional critical editions? Or are technological developments transforming the nature of the profession itself? Do methodological innovations bring about epistemological changes? While these debates have been going on for more than a decade,[1] they still continue to exercise many minds.[2]

If the epistemology of historical scholarship changes it will be due to the new methods and techniques.[3] The reason for this is that most of these techniques are not native to the scholarly field of history. They have been, and are being, developed mostly by disciplinary fields that are firmly embedded in computer science: computational linguistics, statistics, information retrieval, artificial intelligence. Quantitative methods are traditionally far stronger in these fields, as is the use of computational techniques to put them into practice. Moreover, the focus of most scholars in these disciplines is on finding idiosyncrasies in large quantities of data rather than on the historian's goal of piecing fragmentary evidence of the past together into a single coherent narrative. The techniques developed in these fields do not, therefore, necessarily apply to problems of historical scholarship. They are less suited to study broader contexts than to examine isolated problems. They are less focused on understanding remote situations and more on optimizing tasks.

However, this does not mean that historians should stay away from using methods that originally were not intended to be used by historians. On the contrary, there are at least two reasons why scholars interested in history should engage with these new methods. First, they offer excellent opportunities for historians to approach their source material in new ways, not unlike the previous adoption by historians of methods from fields like psychology, anthropology and economics. As such, they provide historians with new tools to extract information from the historical record. In this sense, the 'digital turn'[4] might be compared to the turn towards 'history from below' historians witnessed in the 1970s. This historiographic turn too involved new focal points ('ordinary' men and women, instead of predominantly male elites), different source material (testimonies, memories, diaries), as well as fresh perspectives on time

(microhistory instead of structural history, *Gesellschaftsgeschichte*, *histoire totale*).

Although the digital turn in itself has not initiated new areas of study, the material that is typically being digitized (books, public media, parliamentary minutes) massively impacts the type of studies that are done. The adoption of computational techniques alters the historians' perspective. Most importantly, it enables them to combine the longue durée with the microanalysis. Many techniques allow for great sweeps. They let us look at linguistic trends over decades or even centuries, like the changing classification of the way people walked in London.[5] They aid us in finding thematic patterns in or between entire oeuvres, catalogues and archives, such as the position of Darwin in nineteenth-century science or the impact of Darwinism on popular culture.[6] At the same time, they are much better at finding needles in haystacks than the human eye.[7]

The new methods and techniques introduce thus new types of questions into historical scholarship. These are not always revolutionary. The frequency over time of the word 'I' in public media in the nineteenth and twentieth centuries, for example, as a proxy for the process of individualization in this period,[8] is technically a trivial exercise. It is a question that would have been possible to answer without making use of the new techniques (given the availability of the sources), but no-one would have considered doing the research because of the tremendous amount of time and energy involved (as well as the inability of the human eye to adequately appreciate large quantities). The task can now be done on a household pc in a matter of minutes, if not seconds. The same goes for questions that investigate the relation between words. Attitudes of the twentieth-century Dutch public press towards the United States, for example, could be quantified by studying the frequency and variety of words connected to the adjective 'American'.[9]

The types of 'new' questions that computational techniques enable historians to ask of their material easily go beyond these basic examples. Machine learning techniques, in particular, have the potential to fundamentally change the historian's perspective by enabling the search for latent structures and patterns in large sets of textual or visual material. The trade-off of these new abilities is that it becomes increasingly difficult, if not downright impossible, for non-computer experts to explain how state-of-the-art algorithms found these structures: in

other words, what these structures mean.[10] The shift from causality to correlation, from knowing why to knowing what, celebrated elsewhere,[11] is something historians struggle with. The discussion about the extent to which the potential of these techniques weighs up against the level of control historians might have to give up, however, is still in its infancy.[12]

The more basic quantitative questions, such as the examples mentioned above, might not have the same potential to revise the way historians think about method, but they do have a very definite added value to conventional scholarship. Approaching questions that are similar to the ones historians have been used to study by applying the new methods and techniques might even be more interesting. The historical profession could learn a great deal from them, not in the least about itself. Adopting quantitative techniques to answer questions that are conventionally studied by qualitative methods motivates reflection. What is it exactly that historians do when they study the archives? Should 'gaining an understanding of the archives' not also take quantitative aspects into account, like fluctuations in the reference to particular themes or in the use of significant words? If so, and I believe this to be the case, these techniques could, at the very least, be used to provide more a systematic underpinning of such intuitions.

We could call this the 'because we can' argument for the use of computational techniques in historical scholarship. To it we can add the 'because we need to' argument. While the amount of historical information that is being digitized steadily increases, historians to an ever larger extent make use of born digital data. Information that has never been non-digital will increasingly become the central source of information for every historian who studies the recent past. This observation bears a massive promise but also a cause for concern. The information that is typically available in digital form since the democratization of the internet and the advent of social media is well suited to studying the history of mentalities. It represents exactly those aspects of private and social life that historians have desperately sought wherever living eyewitnesses or their written legacies are absent. Social media make abundantly clear what 'ordinary people' think of politics, culture and even each other. To single out clear melodies in this cacophony of voices is, however, as large a challenge as it is to make this information permanently available to future generations of historians. Much of

today's digital history can, therefore, be seen as paving the methodological way for our future colleagues.

One of the goals of this special issue is to illustrate how highly rewarding it can be for present-day historians to start adopting computational methods in the study of history. Why they can be particularly worthwhile for an examination of long-term developments like modernization will be argued below.

Context, Historicity, Language

The arguments presented in the previous paragraphs make it clear that the usage of computational techniques in historical scholarship is not self-evident. As the contributions to this issue also make clear, historians have make an effort to be able to apply these techniques to the idiosyncrasies of their profession. In practice this means that they have actively to account for three methodological aspects, which are not coincidentally at the core of history as a scholarly endeavour: context, chronology and language. Computational techniques rarely automatically provide for the first two aspects, at least not in a manner acceptable to historians. Scholars of history using these techniques will have to find ways to do this themselves. The third aspect, language, plays a quite different role. History is largely based on language, since historians do not have direct access to the past, but have to make do with mediated traces and testimonies of it. These come, to a large extent, in written form. Computational techniques, however, demand not so much a sensitivity towards language as general linguistic skills.

Let us take an example to illustrate this point. A powerful way to highlight the particularities of a given selection of documents within a larger corpus is the so-called tf-idf statistic. The weight or frequency of a term (tf) is inversely proportional to the number of corpus documents it is present in (inverse document frequency, idf). In a massive corpus of historical newspapers it can be used to get a sense of the way in which a given subset of documents stands out from overall newspaper language. The statistic will bring to the fore which words are used more often than one would have expected, given the average distribution of words within a set of articles containing a specific word. Words like 'the' might be frequent within the subset, but they are present in many

articles that do not form part of the set as well. Therefore, these words score low on the tf-idf scale (as they should, because words that are used in many different kinds of articles tend to be the less meaningful ones). Because tf-idf is more than a basic frequency count of words within the set, it will show what articles that contain a particular word 'are about'. In the digitized Dutch newspaper archive mentioned previously, for example, about 2000 articles (from a total of almost 50 million) published between 1850 and 1940 that contain the specific word 'eugenics' score high for words like 'race', 'marriage', 'sterilization', 'traits' and 'hereditary', but also for words like 'society', 'congress', 'science' and 'problem'. The method accurately shows that eugenics was discussed in Dutch newspapers mostly in the form of reports from meetings of medical or scholarly societies in the Netherlands. A closer look at the articles confirms the idea that eugenics was not a serious topic of debate in the Netherlands before World War II.[13]

Although the computational technique required is quite basic, the example illustrates the need to come to terms with chronology, context and language. The first is perhaps the most characteristic feature of historical scholarship. Historians are interested in change over time. Yet the example on 'eugenics' hardly accounts for this. The question whether Dutch newspapers debated the (dis)advantages of eugenics in the period 1850–1940 is a relevant question. Analyzing all documents within a time frame spanning nine decades as a single 'bag of words', implies a consistency between these articles that (apart from the fact that they all contain the word 'eugenics') might not actually be present. Historians in any case want to take into account changes within the period, to see whether the discourse changed, and if it did, why.

Given the lack of an explicit public debate about eugenics in the Netherlands in the second half of the nineteenth century and the first half of the twentieth, the tf-idf statistic does not provide many clues for changes in the way articles used the word 'eugenics'. It is possible, however, that the Dutch public debate included different forms of eugenic thinking. One way to investigate this is by studying articles containing words that imply eugenic thinking, like 'regulation', 'health' and 'race'. Analyzing articles that contain such combinations of words provides an interesting approach to studying the evolving discourse of eugenic thinking in Dutch newspapers in this period. This is also where the real strength of the tf-idf statistic comes in to play. Rather than just

provoking general statements about longer periods of time it allows for the highlighting of discursive differences between successive periods of time. In 1942, for example, reports about the 'German empire', 'community', 'blood' and 'race' were dominant. In 1910, on the other hand, articles containing the same combination of words spoke significantly often about the 'Indies', 'Europeans', 'natives', and used qualifications like 'Jewish' and 'anti-Semitic'. Still further back in time, in 1882 articles predominantly referred to 'quarantine', 'epidemic' and 'cholera', while the tf-idf statistic for newspaper discourse from 1867 highlighted 'animals', 'livestock', 'breeder' and 'contagious'. Do these words from different years represent a shifting discourse around eugenic thinking? Maybe not, but they do testify to evolving public preoccupations around a central theme and in doing so provide an abundance of pointers to aid further research.

Rather than focus on the period 1850–1940 it would be more interesting to look at the eugenics discourse over an even longer period of time. Questions worth investigating are whether eugenics theory and practices were discussed before the movement spread across the western world at the end of the nineteenth century, or whether eugenic ideas persisted after the Nazi atrocities permanently contaminated the idea (or at least the word). These are relevant historical questions, and ones for which digital techniques can make a difference. By analyzing larger periods of time than historians have traditionally done, current periodizations can be tested and, where applicable, overturned. There is a growing body of literature on the implementation of time scales into computational tools, especially topic modelling.[14] But in most cases historians still have to find ways themselves to meaningfully structure their data chronologically to account for change over time, as the above example illustrates.

The same can be said for contextualization. Few computational techniques support hermeneutic interpretations of a broader historical context. On the contrary, they are suited for narrower tasks. Instead of studying notions of eugenic thinking in the Dutch public sphere, to stick to our example, they enable us to calculate the frequencies of words, relative to their overall average frequency, in a subset of documents we already know to contain the word 'eugenics'. It is up to the user to embed such analytical steps into his or her research question, in other words: to operationalize contextualization. A practical way to do so is

to set up a comparative element in the analysis. As in our example, the results of computational techniques (in this case, words particular to the 'eugenics' subset) only gain meaning in comparison to something else.

The tf-idf statistic is in itself comparative by nature, relating the frequencies of a word in a subset to the number of articles containing that word. More insightful, however, would be to compare the language in a 'eugenics' subset within a particular time frame to that in another set of documents or in another time frame, to relate a 'eugenics' subset that represents a particular community to that of another community, or to compare the Dutch subset to one from another country. Such research designs would make the results from the tf-idf statistic more relevant. And the same is true for any other computational technique, whether topic models or word embeddings. As in conventional history, the application of a comparative approach is restricted only by the resourcefulness of the researcher, to find units that can be meaningfully contrasted.

The third aspect of computational text analysis in historical scholarship is language. The things that historians are commonly interested in cannot always be made explicit (for instance, the Enlightenment attitude, or generational conflicts). Nor are they easily abstracted into single keywords (such as 'modernity' or 'secularization'). Most computational tools, however, are keyword-based. Consequently, working with these tools inevitably involves, on a quite literal level, putting the object of study into words. This procedure, which in traditional historical scholarship forms an implicit part of the heuristic process, necessarily contains a degree of contingency. The 'eugenics' example is highly illustrative in this respect. Several synonyms of the word were current in Dutch language before World War II. Besides the most common Dutch word *eugenetica*, newspapers also used the variants *eugenetiek*, *eugeniek* and *eugenese* (all were taken into account in the analysis mentioned above). But the notion of eugenics could also be implied in a completely different word like *rassenhygiëne* (racial hygiene). In fact, the term *Rassenhygiene* was far more common to denote the eugenics ideology in German than either *Eugenetik* or *Eugenik*. Accounting for such word usage is, of course, crucial when comparing different languages computationally.[15] Breaking down a concept into lists of words that cover its meaning, as the 'eugenics' example illustrates in a very basic way, can be a fruitful way of overcoming the problem of selecting the most relevant spelling conventions, synonyms or word variations.

When studying the changing meanings of concepts, it might become relevant to adopt different definitions (i.e. different word lists) for different periods. This is where the linguistic challenge meets the challenge of chronology.

The three aspects of digital scholarship discussed in this section cannot be considered independently from one another. The comparative approach can be implemented, for instance, by relating multiple time periods to one another, or different expressions (for instance to understand how the discourse around 'eugenics' differed from that around 'racial hygiene'). The chronological and the linguistic elements are inseparable; changes in spelling over time make that clear. Together, these three aspects also make clear that historians can consign to computers only so much. Consequently, the researcher has to be in control of everything he assigns to computational analysis. He has to design his methodological workflow solidly from start to end and at the same time adjust or alter it flexibly whenever that may prove necessary.

Modernity

The contributions to this special issue are showcases of how historians deal with context, chronology and language. This is no coincidence. The study of modernity makes for an excellent experimental playing ground. The impact of the developments for which the term modernity is normally adopted can be studied over time and across different domains. In other words, they involve a clear sense of chronology and comparison. Moreover, besides new practices, institutions and experiences, the movement towards modernity as 'a discourse and practice'[16] involves linguistic and conceptual changes as well as changes in language use.

Crucial is the realization that the rise of modernity left its marks on discourse. The process usually called modernization was, amongst many other things, about changing hierarchies and gender roles, about individualisation, about shifting mentalities; it embraced new attitudes (that of the *flaneur*, for example), new reference points (such as the United States) and new meanings (for instance 'progress'). The rise of modernity involved, in other words, discursive shifts, which can be found in texts. Precisely these texts are rapidly becoming available

in digitized form. At the same time, new techniques can be helpful to explore these texts in new ways.

To help focus on the shifts that are most meaningful to study, Reinhart Koselleck is an excellent intellectual guide. Particularly, Koselleck's suggestion to think of the period between about 1750 and 1850 as a *Sattelzeit* turns our attention, both historically and analytically, to context, chronology and language. Koselleck maintained that the mental and cultural changes that took place in Europe during this period resulted in what we call 'modernity'. Thinking in terms of closed epochs (such as the 'Middle Ages') in combination with a fundamentally different conception of time are among the defining characteristics of Koselleck's *Sattelzeit*. Although the custom of defining time periods as closed was not new, Koselleck argues that it was typical for modernity to define itself as an (unfinished) epoch. He identified this as one of the signs of a perceived acceleration of time. Another new perception of how time was conceived was modernity's idea of an open future. Instead of relying on the repeatability of known things, modernity anticipated a changeable future. This changing mentality became manifest in new words like 'progress' or 'development'.[17]

Focusing on the manner in which contemporaries conceptualized the world around them, Koselleck was very sensitive to the ways in which historians construct (rather than merely reconstruct) the past. He warned for overgeneralizations by stressing the fleeting and, at times, variable ways in which historical actors gave meaning to the world around them. Historians should realize, he argued, that 'no historical movement can be adequately evaluated in terms of the self-same counterconcepts used by the participants of such a movement as a means of experiencing or comprehending it.'[18] This applied as much to characterizations of in-groups and their oppositional counterparts as it did to generalizations based on time. Influenced by Hegel and Heidegger, and borrowing from Ernst Bloch the notion of *Ungleichzeitigkeit*, Koselleck throughout his work stressed 'the non-simultaneity of the simultaneous' (*die Ungleichzeitigkeit des Gleichzeitigen*). A concept as grand as modernity, he maintained, is unable to account for the political, economic, technological and cultural developments that took place in the western world at distinctive moments and at different velocities. To talk about modernity, therefore, means to do justice to all 'decelerations and accelerations' as well as 'overlappings and temporal shifts'.[19]

This sensitivity represents Koselleck's own research agenda, brought into practice, for example, in the *Geschichtliche Grundbegriffe* project. For digital historians, both Koselleck's elaboration on concepts and counter-concepts and his ideas on the non-simultaneity of the simultaneous provide ample opportunities to account for comparability and chronology. Moreover, Koselleck has shown what the focus on linguistic aspects adds to our understanding of history. For example, he identified in the rise of modernity a process of the 'singularization' of concepts that he pinned down to the use of the definite article 'the'. Ideas about nations, parties and churches were not new to the *Sattelzeit*, but the idea that one could speak of 'the Nation', 'the Party', or 'the Church' certainly was.[20] Utilizing the linguistic structure of key concepts, this process of singularization is a practical way to find instances of what Koselleck regarded as the 'politicization' (*Politisierung*) of concepts during the *Sattelzeit*. It was one of the features of modernity he identified, next to *Ideologisierbarkeit* (which Richter translates as 'the ease with which concepts could be incorporated into ideologies', and which is characterized by a similar process of the singularization of words into 'isms'), *Demokratisierung* (democratization, which denotes the changes in the relations between authors and their massively expanded readership) and *Verzeitlichung* (the putting into words of the new perceptions and experiences of time).[21]

By focusing on these new characteristics of concepts, Koselleck aims to make explicit how historical modes of thought changed in the process of modernization. His classifications, at the same time, offer methodological tools for the conceptual historian to study these changes. It is for this reason that Koselleck himself was highly sensitive to the need to keep apart the perspectives of historical agents from those of the historian. Koselleck insisted that the way in which historians conceptualize the past resembles only to a small degree the way in which the people in that past made sense of their world: 'In general, language and sociopolitical content coincide in a manner different from that available or comprehensible to the speaking agents themselves'.[22]

Throughout his work, he elaborated on a number of methodological instructions to account for this fact. Firstly, Koselleck demanded that the (discursive, but also non-linguistic) socio-political context be brought into the study of historical semantics wherever possible.[23] Secondly, regarding the linguistic study of concepts, he distinguished between

several layers of meaning: the *Pragmatik* or practical use of concepts in their immediate context, the *Semantik* or common meaning of concepts, and the *Syntax* and *Grammar* or the syntactic and grammatical structure of concepts. These all have to be taken into account when investigating concepts.[24] Thirdly, in expounding the *Semantik* of concepts, he underlined the need to take into account both their semasiological and the onomasiological aspects. The study of concepts, in other words, involves not only the different uses of a concept but also the different words that denote the same concept.[25]

Using traditional methods, to take all these different investigatory aspects of concepts into account is an enormous task. The *Geschichtliche Grundbegriffe* project, for instance, which Koselleck undertook with Werner Conze, Otto Brunner and others, spanned no less than twenty-five years. Digital tools, however, have proven to be very helpful in studying some aspects of conceptual history more systematically. Established digital techniques like generating concordances or determining co-occurrences of a given keyword are effective ways of mapping the linguistic and discursive context in which words are used, as well as gaining a sense of their semasiological properties. More experimental techniques like word embeddings are promising tools to investigate semantically related words and, thus, cover the onomasiological qualities of concepts.

As a result of mass digitization, digital conceptual history is able to apply these techniques to larger quantities and more types of texts, covering larger periods of time. For this reason, the 'digital turn' has brought about a revitalization of conceptual history in general, and a renewed interest into the larger cultural patterns involved in the rise of modernity in particular.[26] Investigating how London's Criminal Court gradually adopted a distinction between violent and non-violent crimes between the late eighteenth and the early twentieth century, as Klingenstein, Hitchcock and DeDeo have done, would have been virtually impossible without digital tools: they utilized the full Old Bailey's Online dataset of 112,000 trial records between 1760 and 1913. The authors were able to map the 'civilizing process' of modernity during the long nineteenth century. Other scholars have used parliamentary records, sources of popular (folk) culture or literature to study similar long-term processes. Although 'distant reading' will not deliver the final assessment of modernity, digital techniques are making more tangible

the way modernity became present and was debated in Europe over a long period of time.

This Special Issue

The articles collected in this special issue demonstrate how sound digital historical scholarship can add to the study of modernity. They explicitly implement the aspects of contextualization (or comparability), chronology and language. The contribution by Peter De Bolla et al examines the evolution of political ideas in eighteenth-century Britain by zooming in on the words and the conceptual forms that 'provide the building blocks for the long extension of "modern political thought"'. Their axes of comparison, therefore, are mostly along chronological lines. By doing so, they do not only aim to expand traditions of intellectual history into modern political thought in terms of the periods covered, but also in terms of the underlying textual data, in this case the huge collection of books from the Eighteenth Century Collections Online (ECCO).

Jo Guldi likewise scrutinizes the British case. She evaluates different types of digital tools to measure the 'shifting concerns and self-understanding' of the British Parliament during the nineteenth century. She shows how the financial, political and global issues that we normally associate with modernity can be traced in parliamentary speech. While doing so, her measures of 'innovation and irrelevance' of particular speech makes manifest the extent to which particular factions within parliament were willing or reluctant to adopt new issues, but also new concepts and ideas. Guldi demonstrates how the explicit focus on language can add to the study of modernity, while enabling the comparison between different political factions.

Next to book publications and parliamentary records, newspapers make up a third type of textual data that mass digitization has enabled to be studied in large quantities. These form the material that Joris van Eijnatten and Ruben Ros have based their investigation on, which revolves around the question to what extent the conceptualization of 'Europe' in public media during the nineteenth and twentieth centuries was 'entangled' with that of 'civilization' and 'modernity'. While evidently linguistic in nature, the article by Van Eijnatten and Ros takes an

explicit comparative approach, making comparisons between successive periods as well as newspapers (and their respective communities of readers).

Notes

1 'Interchange: The Promise of Digital History,' *The Journal of American History* 95:2 (2008) 452–91.

2 Matthew K. Gold and Lauren F. Klein (eds), *Debates in the Digital Humanities 2019* (Minneapolis, 2019).

3 'Digital History and Argument,' White Paper (Roy Rosenzweig Center for History and New Media, November 13, 2017), https://rrchnm.org/argument-white-paper/, accessed 24 June 2019.

4 Bob Nicholson, 'The Digital Turn,' *Media History* 19:1 (2013) 59–73.

5 Joanna Guldi, 'The History of Walking and the Digital Turn: Stride and Lounge in London, 1808–1851', *The Journal of Modern History* 84:1 (2012) 116–44 (https://doi.org/10.1086/663350).

6 Timothy R. Tangherlini and Peter Leonard, 'Trawling in the Sea of the Great Unread: Sub-Corpus Topic Modeling and Humanities Research', *Poetics* 41:6 (2013) 725–49 (https://doi.org/10.1016/j.poetic.2013.08.002); Jaimie Murdock, Colin Allen, and Simon DeDeo, 'Exploration and Exploitation of Victorian Science in Darwin's Reading Notebooks', *ArXiv Preprint* (2015): http://arxiv.org/abs/1509.07175.

7 Martha Van den Hoven, Antal Van den Bosch, and Kalliopi Zervanou, 'Beyond Reported History: Strikes That Never Happened', in *Proceedings of the First International AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts* (Vienna, 2010) 20–28.

8 Jean M. Twenge, W. Keith Campbell, and Brittany Gentile, 'Changes in Pronoun Use in American Books and the Rise of Individualism, 1960–2008', *Journal of Cross-Cultural Psychology* 44:3 (2013) 406–15.

9 Melvin Wevers, 'Consuming America. A Data-Driven Analysis of the United States as a Reference Culture in Dutch Public Discourse on Consumer Goods, 1890–1990', unpublished PhD dissertation, Utrecht University, 2017; Pim Huijnen and Melvin Wevers, 'Digital Deconstruction. The Digital Turn and the Use of Media as Sources for Doing History', *Officina Della Storia* 13 (2015): https://www.officinadellastoria.eu/it/2015/09/14/digital-

deconstruction-the-digital-turn-and-the-use-of-news-media-as-sources-for-historical-research/; Melvin Wevers and Pim Huijnen, 'Mapping America in Public Discourse', in Michal Peprník and Matthew Sweney (eds), *America in Foreign Media* (Olomouc, 2014) 109–25.

10 Just as the techniques themselves, the explanation is mostly left to scholars from other fields. See, for example: Yoav Goldberg and Omer Levy, 'Word2vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method', *ArXiv Preprint* (2014): http://arxiv.org/abs/1402.3722; Jonathan Chang et al, 'Reading Tea Leaves: How Humans Interpret Topic Models', in *Advances in Neural Information Processing Systems* (Cambridge, 2009) 288–296, http://machinelearning.wustl.edu/mlpapers/paper_files/NIPS2009_0125.pdf, accessed 24 June 2019; D. Sculley and Bradley M. Pasanek, 'Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities', *Literary and Linguistic Computing* 23:4 (2008) 409–424.

11 Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Boston, 2013) 13.

12 A similar discussion has recently shaken up the field of literary criticsm after the publication of Nan Z. Da, 'The Computational Case against Computational Literary Studies', *Critical Inquiry* 45:3 (2019) 601–39 (https://doi.org/10.1086/702594); See also: Nan Z. Da, 'The Digital Humanities Debacle. Computational Methods Repeatedly Come up Short', *Chronicle of Higher Education* March 27 (2019): https://www.chronicle.com/article/The-Digital-Humanities-Debacle/245986; Ted Underwood, 'Dear Humanists- Fear Not the Digital Revolution', *Chronicle of Higher Education* March 27 (2019): https://www.chronicle.com/article/Dear-Humanists-Fear-Not-the/245987/.

13 Jan Noordman, *Om de kwaliteit van het nageslacht: eugenetica in Nederland 1900–1950* (Nijmegen, 1989).

14 See, for example: Xuerui Wang and Andrew McCallum, 'Topics over Time: A Non-Markov Continuous-Time Model of Topical Trends', in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, 2006) 424–433 (http://dl.acm.org/citation.cfm?id=1150450).

15 Pim Huijnen et al, 'A Digital Humanities Approach to the History of Science. Eugenics Revisited in Hidden Debates by Means of Semantic Text Mining', in Akiyo Nadamoto et al (eds), *Proceedings of Social Informatics 2013 and HISTOINFORMATICS* (Berlin, 2014) 71–85.

16 Joris van Eijnatten et al, 'Shaping the Discourse on Modernity', *International Journal for History, Culture and Modernity* 1:1 (2013) 5.

17 Reinhart Koselleck, *The Practice of Conceptual History: Timing History, Spacing Concepts* (Stanford, 2002) 165–66.

18 Reinhart Koselleck, *Futures Past on the Semantics of Historical Time* (New York, 2004) 158.

19 Koselleck, *The Practice of Conceptual History*, 160.

20 Koselleck, *Futures Past*, 156.

21 Melvin Richter, 'Reconstructing the History of Political Languages: Pocock, Skinner, and the Geschichtliche Grundbegriffe', *History and Theory* 29:1 (1990) 46–47.

22 Koselleck, *Futures Past*, 157.

23 Reinhart Koselleck, 'A Response to Comments on the Geschichtliche Grundbegriffe', in Hartmut Lehmann and Melvin Richter (eds), *The Meaning of Historical Terms and Concepts: New Studies on Begriffsgeschichte*, (Washington, 1996) 64; Reinhart Koselleck, *Begriffsgeschichten: Studien zur Semantik und Pragmatik der politischen und sozialen Sprache* (Frankfurt am Main, 2006) 9–31.

24 Koselleck, *Begriffsgeschichten*, 99–102.

25 Ibid., 62; Koselleck, *Futures Past*, 88.

26 Peter De Bolla, *The Architecture of Concepts: The Historical Formation of Human Rights* (New York, 2013); Jennifer A. London, 'Re-Imagining the Cambridge School in the Age of Digital Humanities', *Annual Review of Political Science* 19:1 (2016) 351–73 (https://doi.org/10.1146/annurev-polisci-061513-115924).

About the Author

Pim Huijnen is an assistant professor of digital cultural history at Utrecht University. His research focuses on the digital exploration of conceptual history and the history of mentalities. Previously he worked as a postdoctoral researcher in the project *Digital Humanities Approaches to Reference Cultures: The Emergence of the United States in Public Discourse in the Netherlands, 1890–1990*, and on a digital humanities project comparing Dutch and German public debates on genetics and eugenics. Huijnen has published 'Digital Deconstruction. The Digital Turn and the Use of News Media as Sources for Historical Research',

Officina della Storia 13 (2015), with Melvin Wevers; and 'A Digital Humanities Approach to the History of Science. Eugenics Revisited in Hidden Debates by Means of Semantic Text Mining', A. Nadamoto et al (eds), *SocInfo2013 Workshop Proceedings* (2014), with Fons Laan, Maarten de Rijke and Toine Pieters. E-mail: p.huijnen@uu.nl