

Digital Humanities and historical newspaper research

©Huub Wijffes

This text is part of the research project Pillarisation and Depillarisation tested in digitized Media Historical Sources. The project is made possible thanks to the generous support of Dutch Royal Library, the Dutch Institute of Advanced Studies NIAS and Netherlands eScience Centre

Using digital newspaper collections in historical research is quite new, but the problems and possibilities in this kind of research actually can be quite old. In this exploratory article the focus is firstly on the rise of digital society, digital sources and digital humanities, especially history. Zooming in on the more specific potentials and problems, a closer look is dedicated to digital newspaper research in historical context.

Prospects for Digital Humanities and History

In a 2013 special issue of *BMGN - Low Countries Historical Review* historians debated possibilities, problems and pitfalls of Digital History without coming to some sort of agreement. That seems logical. Although working with a computer was practised by some historians since the nineteen sixties, Digital History is still at the beginning of development. Basic questions about the availability and controllability of sources and about the new methods required for digital research still need an answer. A functional and open accessible infrastructure for digital humanities research and research presentation is not operational in most countries. Still, despite all technical and methodological problems and obstacles, digital humanities bears great possibilities for new research that in nature is 'global, trans-historical and trans-media'. These impressive claims divide the historical profession between 'stalwart believers and underwhelmed agnostics'.¹

Although the first handbook on Digital Humanities was published in 2004, it builds on traditions in using computers in historical research that goes back to the rise of computer aided research in the late nine forties.² Digital Humanities nowadays is still an experimental field of academic research and education, making connections of traditional humanities methodologies (for example historical hermeneutics) with tools researchers can use to curate or access online collections and to analyze big data sets.

Research of this kind triggers mixed feelings. Sceptics tend to say that the digital revolution till now didn't create a real paradigmatic or methodological revolution, but in the heart is a 'practical revolution' making relatively simple keyword searches in singular online sources far more easy.³ Optimists, like Rens Bod in his 2012 inaugural lecture at the University of Amsterdam, say that we are going to revolutionize humanities to an all-encompassing version 3.0. He stated that after the establishment of hermeneutical and critical traditions of humanities 1.0 in the nineteenth and twentieth century, we are now involved in finding historical patterns in digital big data in humanities 2.0. That is roughly the same as media historian Bob Nicholson calls 'the digital turn in cultural

¹ Van Eijnatten et.al. (2013), 57

² Clavert and Noiret (2009), 15-26; Berry ed., (2012); Schreibman (2004).

³ Nicholson (2013), 59-73.

history 2.0'. A modern media historian is looking for patterns and developments that cannot be traced with traditional methods of media research.⁴ For the future Bod sees the big challenge in finding a combination of 1.0 and 2.0 in humanities 3.0: a stage where hermeneutical and critical traditions are applied to digitally found 'patterns'.⁵

This idea of phases in humanities or historical sciences that are determined by the nature and availability of sources (analogue or digital) and the goal of historical research (interpreting unique events in narrative forms or reconstructing and analyzing 'patterns') seems to create the division in humanities between believers and sceptics. It looks like the general division between people who see a total new society arise because of digital media technology and people who stress the power of culture to adjust to these challenges.⁶ In the historical debate the cultural historians see the pleas for a totally new humanities research field around 'the digital paradigm' as a threatening takeover by quantitative scientist with an unlimited belief in rationality. The 'mechanisation' of the heuristic process can repress critique.⁷

You can also see it as a big challenge for the renewal of humanities research, maybe especially as a methodological problem. According to Rieder and Röhle for example a challenges in developing and using digital methods is the question if statistics and algorithms reach a higher level of objectivity then human interpreting. A second question is about the visual output that is so strong in digital humanities research because of the large amounts of 'infographics' and animations. Are these results of higher importance then other output? It is of course tempting, because visualisation gives us an (sometimes animated) image, and for some people visual material (often called 'evidence') is more powerful then evidence in words, that is often called 'argumentative'.⁸

Looking back at the practical results of Digital History research the last ten years one should say that expectations about a revolution shouldn't be too high. Most historians see the digital world just as a convenient place for fast and efficient browsing in the rich information sources available there and not as a vital environment for historical analysis. Historians who seriously explore the exiting new ways of interactive presentation of historical knowledge are still scarce. So, instead of preaching or predicting a revolution or paradigmatic change, we can better critically reconsider and revise vested historical methods and techniques in a growing digital environment.

Since the increasing importance of digital communication and digitized historical sources in the nineteen nineties of the 20th century, interest in what this means for historical sciences is growing.⁹ This interest doesn't seem to grow out of an urgent need for different views on history, but because of the awareness that digital data and software are increasingly guiding our world and can therefore be decisive for historical knowledge and understanding. Or as Lev Manovich wrote, speaking about 'softwarised culture': 'software plays a central role in shaping both the material elements and many of the immaterial structures which together make up *culture*.'¹⁰ The big questions here

⁴ Nicholson (2013), 63.

⁵ Bod (2012); 'Forum: the end of humanities 1.0' (2013)

⁶ Macnamara (2014); Bardoel and Wijfjes (2015), 11-29

⁷ Fickers (2013)

⁸ Rieder and Röhle (2012), 67-84.

⁹ About the history of history and computing see: Zaagsma (2013), 3-29

¹⁰ Manovich (2008), 15.

are: to what extent should we understand software in order to understand culture and are we digitally literate enough for this task?

The first question that calls for an answer is if we can understand history just by looking at and analysing digital sources. For our dominantly digital contemporary culture one cannot deny the indispensable relevance of digitally born sources of course. But what about historical culture that is not only created in analogue forms but can only be accessed through analogue ways until the moment comes that all sources will be digital? As we shall see in the review of digital newspaper research this lack of digital historical sources can be a real problem, that should be tackled on basis of classical source critique: the need to evaluate the value and restrictions that relevant sources (or the lack of them) offer for answering specific historical research questions.

The question about the necessity of digital literacy is pressing, but harder to answer. Of course, every specific research requires deep understanding of the methods used for delivering answers, but understanding of digital methods is difficult to master for humanities scholars because they require specialized knowledge about mathematics and data mining. This knowledge is restricted to insiders; for most historians a computer is a black box that presents amazing solutions for difficult problems.¹¹ Because of their ignorance of the algorithmic logics driving this black box they make themselves increasingly dependent of machine logic (in stead of the cultural logic of historical debate) and people and companies that control these new forms of softwarised analytics.¹²

Historians should be more aware that there is a big difference between statistically or algorithmically significance computers and software engineers show and the culturally or historically significance that is created by historians able to contextualize history from other angles than algorithms. Close cooperation of specialists in both fields is the obvious solution, but fruitful solutions can only grow out of historical questions and expertise and not out of the bare fact that there are digital sources.

Interdisciplinary research effort of this kind still is at the very start of development, but it shows that we miss something vital when close and contextualised reading in historical sources and in depth analysis of certain articles are skipped. Browsing and reading historical newspapers for example gives opportunities to see historical context of newspaper content more clearly. So the suggestion that digital humanities research can best be performed in a closed digital environment with big data, is a misunderstanding of the value of analogue research forms like browsing and in depth analysis of singular sources.¹³

Undoubtedly, new text mining methods bear a promise because they can overcome some manual browsing limitations. In principle all texts are available for fast computer aided analysis, no longer dependent on indexing or coding and with possibilities for unlimited combinations of associative searches.¹⁴ Euphoria is so big that some say 'manual browsing and sampling in various forms (...) are no longer

¹¹ Rieder and Röhle (2012), 67-84

¹² Pasquale (2015)

¹³ Broersma (2011)

¹⁴ Van Eijnatten et.al. (2013), 73; Bingham (2012); Hart and Lim (2015).

necessary'.¹⁵ But there is doubt. Even in the same article where this rather optimistic quote was placed, it is concluded that 'text mining techniques will displace but not replace traditional hermeneutic methods.'¹⁶

This is underlined by the argument that 'the way in which computers work is not automatically compatible with the way historians work'. Historical research is not only a question of processing big data sets on an algorithmic way, how sophisticated the methods are or will be.¹⁷ History is also building arguments by critically interpreting hybrid information from multiple and varies sources. Supposing that this can be done far better in computer-based methods, shows a technological deterministic ideology that is antagonistic to cultural influences. The clash between this technological determinism or 'solutionism' and cultural criticism is a general characteristic of contemporary academic discussion about the nature and future of humanities.¹⁸

This debate seems to be a repetition of earlier debates in the historical sciences about the ideological question if history should hermeneutically focus on unique events or on analyzing structures, patterns and quantifiable units and parts. The question relates directly to the more practical problem if historians should use 'documents' or 'data'.¹⁹ In this respect the debate about the digital turn in humanities research has some acquaintance to the lively debate in the nineteen seventies about the need to integrate sociological and economic theory and methodology in historical research.²⁰

Cultural historians tend to think that we cannot interpret complex historical or cultural processes without a notion about what constitutes and drives culture. Sole use of quantitative data, the quest for 'patterns' and turning history into a social science therefore are to limited, or even misleading. In the classic words of cultural historian Robert Darnton: 'the social scientists live in a world beyond the reach of ordinary mortals, a world perfectly organized in perfect patterns of behaviour, peopled by ideal types, and governed by correlation coefficient that exclude everything but the most standard of deviations'. Such a world can never be joined with the messiness of history.'²¹ This critique is familiar to the critique on 'algorithmic culture' that is formulated in digital society. Critics say that this relying on code computer languages and algorithmic reasoning are incompatible with the critical-argumentative thinking that is the basis of humanities research.²²

This kind of critical scepticism creates an artificial antagonism between quantitative and qualitative methods; in the practice of historical or digital humanities research both clusters of methods are used next to each other in a complementary way.²³ If the historical debate about the fast development of social and economic history since the nineteen sixties for example shows anything, it is that the structural approaches stimulated by social and economic history didn't replace but strengthened cultural, political, biographical and other non-quantitative historical approaches. In

¹⁵ Van Eijnatten et.al. (2013), 73

¹⁶ Van Eijnatten et.al. (2013), 75

¹⁷ Piersma and Ribbens (2013), 57

¹⁸ Liu (2012), 490-509; Morozov (2013); Fickers (2012); Fickers (2013).

¹⁹ Piersma and Ribbens (2013), 82-85

²⁰ Bertels (1973)

²¹ Darnton (1990), 60. Compare: Fickers (2013)

²² Fish (2012); Van Dijck (2014).

²³ Van Dijck (2014).

media historical research for example a 'cultural turn' was seen in the nineteen nineties. From focusing on institutional production and societal processes, attention shifted to the meaning of media contents in the specific historical context of media reception by publics, each with a different cultural background.²⁴

This indicates that we now can make 'the digital turn' without losing the strengths of cultural approaches. Progress can be made if we understand what digital cultural data are, what digital tools exactly do in finding what sort of results, and how these results can be fitted and contextualized in a broader ensemble of historical sources. 'Computationally supported thinking doesn't have to be dehumanising (...) but can give us greater powers of thinking and larger reach for our imaginations...'.²⁵ Of course you must acknowledge that there is a difference between the traditional close reading a limited amount of texts and detached reading of large amounts of data. Historians however should not become what they aren't: computer scientists. They should use new methods to expand their horizon and possibilities to answer questions.

This creates dilemmas because of the sometimes-alienating concentration in digital historical research on technological possibilities and the shrewdness of digital tools as such.²⁶ This triggers the debate on the end of humanities as we know it, creating a new dominant paradigm about history to be understood not as a set of unique social and cultural phenomena largely determined by distinction, deviance and coincidence but as a cohesive culture that can be understood just by using shrewd algorithms and spectacularly visualize the results.²⁷ Even data analysts acknowledge that 'there is a risk that we look more carefully at the technical components of the datasets than the historical context of the information that they represent.'²⁸

A more or less logical meaning was found with the traditions in qualitative research in social and economic history, but a separate infrastructure arose in bringing digital historical data together. In later phases digital tools to analyse these data were added to these growing archives. Research projects were set up in years to follow, generally bringing together young historians with more experienced researchers coming from computer sciences. Experiments were therefore more based on technological challenges than in tackling historical questions.

Nonetheless, the use of digital and more traditional historical sources and combining that with the abundance of historical information on the Internet has become a more or less normal part of the professional historical field. The big challenges not only lie in the analysis of digital sources, but in developing a way of life as a historian in the digital world.²⁹ In this it is of utmost importance to acknowledge that most archival sources are not digitized yet en shall not be digitized in the coming decades because of the enormous costs. Solely relying on digital analysis is therefore to limited in scope and even dangerous because it feeds the popular idea, cultivated in new media society, that only information that is instantly online available is relevant. In this respect the old cornerstone of historical research to look critically at the surviving

²⁴ Hilmes (2010); Hartley (2012), 27-58

²⁵ Berry, ed., (2012): 'Introduction', 12.

²⁶ Weller ed., (2013): 'Introducton'.

²⁷ See for this debate the contributions made in 'Forum: the end of humanities 1.0'.

²⁸ Prescott, 'The Deceptions of Data', as cited in: Zaagsma (2013), 24. Also: Prescott (2012).

²⁹ Zaagsma (2013), 17-18.

sources, acknowledging that they give a limited and very specific picture of historical reality, increases in importance.

We can perhaps bridge the rather artificial, undesirable and not very fruitful collision of the digital and cultural perspective on history by focusing on the numerous practical and methodological problems that need to be solved. Starting with the practical problems: the digitisation of sources and the preservation of original (analogue) sources come with considerable costs. Making digital versions of analogue sources can be costly and reaching a state of completeness therefore takes a lot of time. Maintaining expertise about the context of the original sources and the handling of digital bearers also costs a lot of money. It is unclear who will pay for this all, you can only say that creating facilities for scientific research in principle is publicly funded. How vital that is, is shown ultimately in the many fights about copyrights on digital sources. This copyright problem is decisive for the lack of completeness in historical sources like newspapers, magazines, films and broadcasting material. Black hole after 1945?

Furthermore it is still unclear how stable the technology of digital bearers is. Very obvious are problems connected to the use of digital copies, like the failure-rate in OCR-data, obstacles limiting free access and use (especially problems with copyright and ownership), the lack of uniformity in data and developing the required software expertise in searching digital databases and presenting the results.

Looking at the possibilities of using digitized sources in historical research, it must be stressed that it is not only a case of easy consultation of single sources and combining hybrid archival material in an interactive setting. Digital archives create digital data that can be analyzed as such if we have data analyzing tools for that. You can ask yourself too if a programmer of intelligent software is a full-grown scholar in humanities or just is an attendant of a machine. Both are not valid, but we should think more about the value we give to certain parts of humanities research.

These should however be used critically because the assumptions of digital toolmakers and analysts aren't neutral. 'Theory is already at work on the most basic level when it comes to defining units of analysis, algorithms, and visualisation procedures.'³⁰

Valid here are also the traditional historical guidelines to look carefully and critically at the unique materiality and historical context of sources and not to rely on just one source or method. In the combination of different analyzing methods lies the greatest contribution of professional historians to debates about the essence and relevance of history. Looking at one type of digital historical source – newspapers - can clarify further what problems and prospects historians are faced with.

Digital research and historic newspapers

Analyzing historical newspapers is getting a different dimension when we see this as analyzing big data. Manually browsing in newspapers (on paper or microfilms) automatically gave some historical context of the content of articles, the position in relation to other content, the cultural forms and media genres to be found in these sources. When analyzing digital newspaper data however a researcher should be aware that he is doing decontextualized research. He should also get used to the idea that in

³⁰ Rieder and Röhle (2012), 70.

principle scarcity of sources no longer exists; the big problem in digital history is how to cope with abundance.³¹

The digitization of historical newspapers undoubtedly has stimulated research, but eagerness to use the sources came together with an awareness that new problems arose, especially problems connected to the storage and retrieval of and the access to the data.³² Storage and free access are of course classical problems. From the perspective of historical research free availability of complete and uniform sources has always been vital. The historical infrastructure that was built in the 19th and 20th centuries is the results of this endeavour: public accessible archives, concise and extensively annotated source publications, heritage institutions guarding complete and contextualised collections and long term research projects.

Digital storage seemed to offer new solutions. But in a digital environment completeness and uniformity cannot be guaranteed either. Although millions have been invested in digitisation projects, still only a fraction of historical newspapers is available for research purposes. Lack of money, but also the scattering of collections and problems related to copyright protection can still be decisive for the success of research efforts.³³ So, a researcher wanting to work with complete newspaper data, needs to be able to organize, improvise and negotiate. He also needs funding, that can be substantial.

A central storage or database of all digital newspapers on national level doesn't exist, even in countries with a powerful national library infrastructure, like most Western European countries. In the first place there are big collections like British Newspaper Archive (subscription), Library of Congress (free), ProQuest Historical Newspapers and Newspaper Archive Library Edition (subscription), or the free collections of the Royal Dutch Library (Delpher), the Staatsbibliothek in Berlin (Zefys), Bibliothèque Nationale de France (Gallica) and National Library of Australia (Trove). Next to these newspaper archives all kinds of specialized – regional, local, thematic - collections pop up in the digital world. Each collection has own interfaces, standards and/or tariffs for accessibility and use. Connections between these separate collections are hard to establish, making really new digital search methods like text mining and network analysis complicated. Other important collections like the commercial Lexis-Nexis are based on text only and therefore totally ignore the visual dimension of news.³⁴

But even where complete editions of certain newspapers are kept, copyright problems create limitations in use. Especially twentieth century newspapers cannot be used free of all charge. Retrieval and consultation in a shielded environment can be possible, but publication of results of this consultation is problematic. Sources older than the twentieth century are less problematic, but here special problems of quality arise. The incompleteness of editions and the inconsistency of typography and layout create a growing amount of failures that spoils the results of digital searches. 'When we digitise a newspaper, it is fundamentally changed (...) sources are remediated and not just reproduced', Bob Nicholson rightly remarked.³⁵

³¹ Broersma (2012), 35-37.

³² Deacon (2007); Broersma (2011)

³³ Vgl. Berkhout (2011).

³⁴ Deacon (2007); Broersma (2011); Maurantonio (2014).

³⁵ Nicholson (2013), 61 and 64.

Obviously digital approaches seem to offer more possibilities than 'old' methods like browsing through newspapers, reading some selected and relevant content and interpreting that in relation to other sources for historical knowledge. The relevance of this kind of research analysis increased when media historical research made 'a cultural turn' in the nineteen eighties. The focus shifted from the history of institutional and political background of media institutions to the cultural meaning of media content for publics.³⁶ In this respect availability of content sources and methods to analyse them were vital.

Tailor-made methods were developed in the context of every specific research. Media historian Frank van Vree for example analysed the content of four major Dutch newspapers regarding their attitude towards Nazi Germany between 1933 and 1939. The sections on the historical context of the press in this period are just as long as the actual content research that can be characterized as a historical discourse analysis strongly focusing on opinion articles and background stories in the four newspapers. Because of the labour intensive work of this sort of analysis not the entire content of the newspapers could be included. Nor could vital sections of the Dutch press in this period be included, like the national neutral or regional press, so questions can be raised about the representativeness of this research for the interpretation of 'public opinion'.³⁷ In a later study into the cultural transformation of the leading national newspaper *de Volkskrant* in the nineteen sixties and seventies, the focus was also restricted to certain carefully selected sections of the newspaper. In comparable studies of similar developments in newspapers, the same restrictions were characteristic.³⁸

More recently methods have been developed to look more systematically at the long-term development of journalistic practices or genres. Media historian Marcel Broersma kicked off this research by making a long-term analysis of the content of one newspaper during 250 years. Style and genre analysis were integrated in thorough contextualised research of the institutional and political development of this newspaper.³⁹ Following the same lines, but with more emphasis on a single genre within several (international) newspapers was the research of Frank Harbers. He analysed the development of the reportage in newspapers in three countries from 1880 till contemporary society. Rutger de Graaf reconstructed the connections between the content of pamphlets and newspapers in nineteenth century Dutch society.⁴⁰

Although the scope of these studies was not primarily to analyse digital data but to solve a historical problem, the data itself were in majority gathered using specific coding schemes guaranteeing objectivity with intercoding correction. The advantage of these methods is that the coding is tailor made for answering very specific historical questions and that automated or technical failures were excluded. The disadvantage was of course the limited content that could be researched. Generally speaking only samples were taken coming from selected periods like one month every five or ten year. As long as there is no safe and sound method of automating the search for a specific and

³⁶ Wijfjes (1999).

³⁷ Van Vree (1989).

³⁸ Van Vree (1996); Mulder and Koedijk (1996); Wolf (2009).

³⁹ Broersma (2002).

⁴⁰ De Graaf (2010); Harbers (2014).

complex historical entity like ‘reportage’ or ‘comment article’, sample research of this kind will be necessary.

A typical example of the necessity of contextualising digital research questions is shown in an exploratory study of the theoretical concept of ‘pillarisation’ in Dutch history. A project called *Verrijkt Koninkrijk* aimed to analyze the digital texts of historian Loe de Jong in relation to this ‘pillarisation’, a long term process of societal and political segmentation characteristic of Dutch culture roughly between 1900 and the nineteen sixties. It showed that De Jong in his 14-volume book about the Netherlands during the Second World War didn’t write about concepts like ‘zuilen’ (pillars) and ‘verzuiling’ (pillarisation), but in related concepts like ‘volksdelen’ (sections in national community). Researchers also found that these words weren’t used with the same and uniform connotation. So alternative queries had to be developed, taking into account that pillar is a broad concept with different meanings on different levels. To get a grip on that, contextualised research is necessary. A researcher should also look at the sentiment in which the more detailed concepts were used. All this requires a lot of research time and sufficient expertise.⁴¹

Developing complex and tailor-made digital search methods that can tackle specific problems forms one of the big challenges of Digital Media History. This also relates to the problem how to retrieve and analyse visual or iconic elements within newspapers, like photographs, cartoons, maps and graphics. The search for the proliferation of iconic photographs in public debates for example has just begun.⁴²

Some consideration has to be made about the nature of digital data. In what way and to what depth are these data constructed, assembled or stored? An important question for example is what metadata are connected to the data. In principle an enormous potential of metadata can be added to each document, but in practice only a limited amount is added because of restrictions in costs and data handling. Crowd sourcing can be a solution, but despite the success of crowd sourced knowledge databases like Wikipedia, there is still doubt about the value and reliability for scientific purposes.

The quality of Optical Character Recognition (OCR) in scanned documents can seriously influence the amount of failures in the digital search possibilities, especially in documents that require specialized knowledge to read or interpret.⁴³ OCR-failures are a special problem in almost all texts produced before 1850, because of the inconsistency in typographic form and lay out in older periods.⁴⁴ You can see that for example in the digitized collection of historical newspapers in the Dutch Royal Library, where OCR-failures increase considerably in time: the older the original bearer the more failures are detected. It is estimated that this runs up to more than 90% for some seventeenth and eighteenth century newspapers. But a relatively new word like ‘verzuiling’ of which historical research has shown that it was developed in the nineteen fifties of the twentieth century, shows up two times in 18th century Dutch newspapers available in the search engine Delpher of Royal Library. In the nineteenth century 33 results show up as ‘verzuiling’ while in the original newspapers are mentioned: verzameling,

⁴¹ Piersma and Ribbens (2013), 91-95

⁴² Kleppe (2012).

⁴³ Jeurgens (2013). 34

⁴⁴ Smits (2014), 141

vervulling, verzetten, verzoeking, verzoening, verzorging, vergoeding and verzanding. In the twentieth century period before the first proper use of 'verzuiling' in 1952, more than 35 OCR-failures pop up.

There are several methods for OCR-failure correction, but none has developed into a definite solution. British Newspaper Archive and National Library of Australia for example allow users to correct OCR-errors and add tags they think are relevant for the article in question.⁴⁵ Dutch Royal Library works with a large group of volunteers to re-type the articles in the digital collection of seventeenth century newspapers on basis of the OCR.

Strange and others also point at OCR-errors and other technical obstacles, like the lack of expertise metadata on document level. But their conclusion on basis of a very clearly outlined selection of 19th century newspaper research is that correction of OCR-failures (in their data set: around 20%) is 'desirable but not essential' in this kind of topical research. It is different with failure-rates run up to more than 80% in older newspapers with problematic lay out and typography. Ideal of course is reducing failures, preferably by double manual correction, but the time and costs connected with that in big data sets are considerable. But adding metadata can enhance the quality of results considerably.

Of course, the conclusion shall be different when we don't talk about small and targeted data sets, but have to handle big and diffuse data sets. For that maybe crowd sourcing can be a solution, but this raises questions about reliability and the role of expert knowledge. On top of that comes the problem that it doesn't give solutions for research with deadlines and demands about scientific verification and reliability. A big problem is that in order to efficiently excavate in big data you need tools that only highly skilled data-engineers can use or develop in close cooperation with language specialist and/or historians.⁴⁶

Some experience has been build up with open source mining technology in research of historical newspapers, for example with historical 'sentiment mining' programs WAHSP and BILAND. In these researches word clouds are created based on relative frequencies in the retrieved selection of documents in the corpus. It highlights negative or positive connotation, but this still needs further historical contextualisation because connotation constantly changes in time.⁴⁷ A program like Texcavator – developed by university of Utrecht to trace patterns in public discourse – is also coping with this problem.⁴⁸

Almost all historians working with historical media sources agree that the greatest added value of digital sources lies in deconstructing long term connections between contents that till now couldn't be connected. New software techniques for historical data mining facilitate historians who are looking for patterns in large amounts of texts like newspapers. Interpreting these patterns should be performed on basis of historical context information. An example offers the research project Transatlantis, reconstructing debates crossing regional, cultural and national borders trying to contribute to the debate about the supposed Americanization of European culture in the

⁴⁵ Nicholson (2013), 64.

⁴⁶ Strange et.al. (2014).

⁴⁷ Van Eijnatten et.al. (2013), 61

⁴⁸ Van Eijnatten et.al. (2014)

twentieth century. The theoretical concept used is 'reference culture', defined as 'spatially and temporally identifiable cultures that offer a model to other cultures and have exerted a profound influence in history'. This concept is researched in a set of digital historical sources, creating a network of references about the United States in the Netherlands between 1890 and 1990.⁴⁹

Tracing 'patterns' is indeed a goal of Digital Humanities research in general. But most historical researchers stress that these patterns only get real meaning if they are combined with contextualised research, for example qualitative interpretation of specific texts, words or visuals. With digital newspaper research we can trace the development and intensity of influential events and persons, but for the interpretation of how these construction were made in different periods we need to take a closer look at the content in its media and cultural context. To make it more concrete: with digital media sources we maybe can trace the complete media coverage of the Dreyfus-affair in French society in the twentieth century, but to say something about how this event was constantly redefined in different contexts we need to look at single media itself in connection to a broad cultural and political context of its time. For this we need digital research too, because it can facilitate us to zoom in on content that in a traditional way could only be found by time consuming browsing of newspapers or viewing many hours of broadcasting material.

Mediatization of politics in history

Overlooking the problems and possibilities of Digital Humanities research, maybe its best to ask what the digital turn means for a historical topic that has been thoroughly researched in other then digital ways. For example, what new light van digital research shed on a concrete historical problem like the relation of media and politics in long-term development? Sceptics might say: not a lot because you cannot reduce politics to the study of language in newspaper sources related to politics; a practice that seems unavoidable in digital data analysis. And historical language studies show that connotation constantly changes and doesn't form universal patterns in the long periods data analysis has a preference for.⁵⁰ Of course the believers oppose that it is best just to begin with formulating questions and developing methods to answer them. It will be a question of time, trail and error to solve the problems. And look at the enormous possibilities: data analysis facilitates the analysis of the whole of media content for example. And connections can be made to political sources in order to shed some light on the dynamic relationship between politics and media in historical context.

Before starting such a quest it is important to evaluate the theoretical concepts driving this kind of research and also make evaluations of the research that for decades political communication scientists and historians have tried to perform in labour-intensive historical source analysis and content analysis of media.

⁴⁹ Van Eijnatten et.al. (2013), 69

⁵⁰ Beyen (2013), 164-170

Literature

- Bardoel, Jo and Wijfjes, Huub (2015), 'Journalistieke Cultuur in Nederland: een professie tussen traditie en toekomst'. In: Bardoel, Jo and Wijfjes, Huub, eds. (2015). *Journalistieke Cultuur in Nederland*. Amsterdam University Press
- Berkhout, Karel (2011). 'Het Digitale Drama'. *NRC Handelsblad*, 10 september 2011
- Berry, David M., ed., (2012). *Understanding Digital Humanities*. Houndmills: Palgrave Macmillan.
- Bertels, Kees (1973). *Geschiedenis tussen structuur en evenement*, Amsterdam: Wetenschappelijke Uitgeverij
- Beyen, Marnix (2013). 'A Higher Form of Hermeneutics? The Digital Humanities in Political Historiography', *BMGN - Low Countries Historical Review* 128-4, 164-170
- Bingham, Adrian (2012). 'Reading Newspapers: Cultural Histories of the Popular Press in Modern Britain'. *History Compass* 10-2, 140-150
- Bod, Rens (2012), *Het einde van de geesteswetenschappen 1.0*. Inaugural lecture University of Amsterdam, 14 december 2012
- Broersma, Marcel (2002). *Beschaafde vooruitgang. De wereld van de Leeuwarder Courant 1752-2002*. Leeuwarden: Friese Pers
- Broersma, Marcel (2011). 'Nooit meer bladeren? Digitale krantenarchieven als bron.' *Tijdschrift voor Mediageschiedenis* 14:2, 29-55
- Clavert, Frédéric and Serge Noiret (2009). 'Digital Humanities and History. A new Field for Historians in the Digital Age'. In: ibidem (eds.). *Contemporary History in the Digital Age*. Brussels: Peter Lang, 15-26
- Darnton, Robert (1990). *The Kiss of Lamourette. Reflections in Cultural History*. New York: W.W. Norton
- Deacon, David (2007). 'Yesterday's Papers and Today's Technology. Digital Newspaper Archives and Push Button Content Analysis.' *European Journal of Communication* 22:1, 5-25
- De Graaf, Rutger (2010). *Journalistiek in beweging. Veranderende berichtgeving in kranten en pamfletten (Groningen en 's-Hertogenbosch 1813-1899)*. Amsterdam: Bert Bakker
- Fickers, Andreas (2012). 'Towards a New Digital Historicism? Doing History in the Age of Abundance'. *View. Journal of European Television History and Culture* vol.1 2012-1
- Fickers, Andreas (2013). 'Veins filled with the Diluted Sap of Rationality. A Critical Reply to Rens Bod', *BMGN - Low Countries Historical Review* 128-4, 155-163
- Fish, Stanley (2012). 'The Digital Humanities and the Transcending of Mortality'. *The New York Times*, January 9, <http://opinionator.blogs.nytimes.com/2012/01/09/the-digital-humanities-and-the-transcending-of-mortality/>
- 'Forum: the end of humanities 1.0'. *BMGN - Low Countries Historical Review* 128-4, 145-180
- Harbers, Frank (2014). *Between Personal Experience and Detached Information. The development of reporting and the reportage in Great Britain, the Netherlands and France, 1880-2005*. PhD University of Groningen
- Hart, Roderick P. and Lim, Elvin T. (2015). 'Tracking the Language of Space and Time, 1948-2008'. *Journal of Contemporary History* 46:3, 591-609

- Hartley, John (2012). *Digital Future for Cultural and Media Studies*. Chichester: Wiley-Blackwell
- Hilmes, Michele (2010). *Only Connect. A Cultural History of Broadcasting in the United States*. Belmont
- Jeurgens, Charles (2013). 'The Scent of the Digital Archive', *BMGN - Low Countries Historical Review* 128-4, 30-54
- Kleppe, Martijn (2012). 'Wat is het onderwerp op een foto? Kansen en problemen bij het opzetten van een eigen fotodatabase.' *Tijdschrift voor Mediageschiedenis* 14:2, 73-107
- Liu, Alan (2012). 'Where is Cultural Criticism in the Digital Humanities?', in: Matthew K. Gold (ed.), *Debates in the Digital Humanities* (Minneapolis 2012), 490-509
- Macnamara, Jim (2014). *The 21st Century Media (R)evolution: emergent Communication Practices*. New York: Lang
- Manovich, Lev (2008), *Software takes command*, 15:
http://softwarestudies.com/softbook/manovich_softbook_11_20_2008.pdf
- Maurantonio, N. (2014). 'Archiving the Visual. The Promises and Pitfalls of Digital Newspapers'. *Media History* 20:1, 88-102
- Morozov, Evgeny (2013). *To save everything, click here: the folly of technological solutionism*, (New York 2013)
- Mulder, Gerard and Paul Koedijk (1996). *Léés die krant! Geschiedenis van het naoorlogse Parool 1945-1970*. Amsterdam: Meulenhoff
- Nicholson, Bob (2013). 'The Digital Turn. Exploring the methodological possibilities of digital newspaper archives'. *Media History* 19:1, 59-73.
- Pasquale, Frank (2015). *The Black Box Society. The Secret Algorithms that Control Money and Information*. Cambridge (MA): Harvard University Press
- Piersma, Hinke and Ribbens, Kees (2013). 'Digital Historical Research. Context, Concepts and the Need for Reflection', *BMGN - Low Countries Historical Review* 128-4, 78-102
- Prescott, Andrew, 'An electric Current of the Imagination: what the Digital Humanities are and what they might become'. *Journal of Digital Humanities* 1:2 (2012)
- Rieder, Bernhard and Theo Röhle (2012). 'Digital Methods: Five Challenges', in: David M. Berry (ed.), *Understanding Digital Humanities* (Houndmills 2012)
- Schreibman, Susan (2004). *A Companion to Digital Humanities*. Malden: Blackwell
- Smits, Thomas (2014), 'Problems and possibilities of digital newspaper and periodical archives', *Tijdschrift voor Tijdschriftstudies* 36, 139-146
- Strange, Carolyn and Wodak, Josh and Wood, Ian (2014). 'Mining for the Meanings of a Murder. The Impact of OCR Quality on the Use of Digitized Historical Newspapers'. *Digital Humanities Quarterly* 8:1.
- Van Dijck, José ((2014). 'Big Data, Grand Challenges. Over Digitalisering en het geesteswetenschappelijk onderzoek.' Ketelaar-lezing 12,
http://www.clariah.nl/files/publicaties/Ketelaarlezing_2014.pdf
- Van Eijnatten, Joris, Toine Pieters and Jaap Verheul (2013). 'Big Data for Global History. The Transformative Promise of Digital Humanities', *BMGN - Low Countries Historical Review* 128-4, 55-77
- Van Eijnatten, Joris and Pieters, Toine and Verheul, Jaap (2014). 'Using Texcavator to map Public Discourse', *Tijdschrift voor Tijdschriftstudies* 35, 59-65

- Van Vree, Frank (1989). *De Nederlandse pers en Duitsland. Een studie over de vorming van de publieke opinie*. Groningen: Historische Uitgeverij
- Van Vree, Frank (1996). *De metamorfose van een dagblad. Een journalistieke geschiedenis van de Volkskrant*. Amsterdam: Meulenhoff
- Weller, Toni ed., (2013). *History in the Digital Age*. London: Routledge.
- Wijfjes, Huub (1999). 'Perspectief in persgeschiedenis'. *Bijdragen en Mededelingen betreffende de Geschiedenis der Nederlanden* 114:2, 223-235
- Wolf, Mariëtte (2009). *Het geheim van De Telegraaf*. Amsterdam: Boom
- Zaagsma, Gerben (2013), 'On Digital History'. *BMGN - Low Countries Historical Review* 128-4, 3-29