



Rethinking History

The Journal of Theory and Practice

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rrhi20

Textuality as amplification: reconsidering close reading and distant reading in cultural history

Hannu Salmi

To cite this article: Hannu Salmi (31 May 2024): Textuality as amplification: reconsidering close reading and distant reading in cultural history, Rethinking History, DOI: [10.1080/13642529.2024.2360318](https://doi.org/10.1080/13642529.2024.2360318)

To link to this article: <https://doi.org/10.1080/13642529.2024.2360318>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 31 May 2024.



Submit your article to this journal [↗](#)



Article views: 32



View related articles [↗](#)



View Crossmark data [↗](#)

Textuality as amplification: reconsidering close reading and distant reading in cultural history

Hannu Salmi 

Department of Cultural History, University of Turku, Turku, Finland

ABSTRACT

This article discusses the idea of distant reading and explores the ways in which it can be conducted in research. It focuses especially on how distant reading can contribute to the study of cultural history, which is often regarded as a domain of close reading. The article argues that distant reading methods can successfully be applied in the analysis of cultural transmission in the past, where it is often essential to combine the study of textual signification with the idea of textuality as material flow. The article draws on an example from press history and especially discusses text reuse detection as a strategy for identifying textual amplification.

ARTICLE HISTORY Received 6 June 2023; Accepted 20 May 2024

KEYWORDS Digital history; textuality; close reading; distant reading; cultural history; text reuse

Introduction

Since the 1990s, the process of digitalisation has transformed not only research practices but also conditions of possibility for knowing, discovering phenomena, and establishing new knowledge. There are millions of images easily available in digital collections, and millions of pages of digitised newspapers in the repositories of national libraries worldwide. This plenitude of source material is received and interpreted by humans, but it is also increasingly read by machines. Texts are not only something we have gradually learned to interpret, but they are also processed through optical character recognition (OCR) and handwritten text recognition (HTR), identified as texts, and their semantic structures analysed through computational methods. It can even be argued that in the present situation, we are not only readers, but we are also objects of reading in

CONTACT Hannu Salmi  hansalmi@utu.fi  Department of Cultural History, University of Turku, Turku, FI 20014, Finland

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

new and imaginative ways. The algorithmic ‘reading’ of our words and deeds is an urgent issue. This means that the cultural-historical context of reading has profoundly changed over the last few decades.

These changes create the need to reconsider the idea of reading as a research practice. If the nature, quality, and extent of source material now available for research have changed, it is natural that the toolbox of research must also be renewed accordingly. The notion of close reading, cultivated by cultural historians and other researchers in the humanities, has been challenged by the idea of ‘distant reading’ in the 2000s, as already highlighted in the introduction of this special issue. The term was advocated by the literary scholar Franco Moretti (2000, 57) in his article ‘Conjectures on World Literature’, published in the *New Left Review*. Moretti argued that distant reading enabled the reader ‘to focus on units that are much smaller or much larger than the text: devices, themes, tropes – or genres and systems’. Hence, the reader had to have distance to fully grasp the larger overarching elements that would not draw the attention of the reader if s/he was too close to the text. Furthermore, smaller, subtle nuances might gain weight in the research material only when they accumulate *en masse*. An individual case would appear singular, but in a larger framework, it might have wider ramifications that only reveal themselves within a bigger picture. In his book *Graphs, Maps, Trees: Abstract Models for a Literary History*, Moretti (2005, 1) continued by maintaining that this distance ‘is not an obstacle, but a specific form of knowledge: fewer elements, hence a sharper sense of their overall interconnection. Shapes, relations, structures. Forms. Models’. It has to be added that Moretti was not the first to make a distinction between close and distant reading. It is obvious that the idea of ‘closeness’ already entailed its counterpart. When closeness was emphasised, it aroused the question of distance. For example, the founder of liberation theology, Gustavo Gutiérrez underlined the role of close and distant reading of the Bible. For him, close reading meant drawing on the present-day context and the meaning of scripture today, while distant reading referred to the reading of texts in their historical context (Gibbs 1996, 282). It was Moretti, however, who made distant reading a popular concept among humanities scholars.

In his original provocation in the *New Left Review*, Moretti (2000, 57) presented close and distant reading as opposites. He even claimed that distance is ‘a condition of knowledge’. In the 2000s, after Moretti, there have been many variations and interpretations of the idea of ‘closeness’, and its (supposed) opposite ‘distance’. There are many publications that

try to tackle this dichotomy, arguing that it is a constructed opposition in the first place, and goes beyond the dualistic setting (Jänicke et al. 2015; Nicholson 2020). This article continues this effort and discusses the challenges in combining practices of close reading with the aggregating and analysing of massive amounts of data. The article draws on examples of the study of text reuse in the nineteenth-century press and the extraction of reused text passages with computational methods. Within the press, particular themes or items became amplified through the nonlinear process of reprinting texts from paper to paper in an almost infectious manner.

This article revisits the idea of distant reading and explores the ways in which it can be conducted in research. It focuses especially on how distant reading could contribute to the study of cultural history, which is often regarded as a domain of close reading. The article argues that distant reading methods can be successfully applied in the analysis of cultural transmission in the past, where it is often essential to combine the study of textual signification with the idea of textuality as material flow, which, again, participates in producing cultural phenomena through amplification. Through the discussion on the reprinting of texts, and a concluding case study on the newspaper coverage on Johann Strauss the Elder and his Sophie Waltz, the article draws attention to the problem of repetitiousness in textual sources and how the historical significance of this reiteration could be acknowledged. By these means, by emphasising the importance of textual circulation and texts as material flow, the article presents a research strategy for going beyond the division into close and distant reading. This strategy draws on the methods of distance reading to perceive the materiality of texts, which further provides an essential perspective on the insights offered by close reading.

Close reading distance

Today, the idea of distant reading is often mentioned and debated in the study of the digital humanities. It has also meant the return of statistical analysis and quantitative methods to the humanities after an era of more interpretive approaches in cultural studies.¹ Distant reading is often presented rather straightforwardly as a way of analysing ‘hundreds or thousands of books’, or as a research strategy that is probably due to its statistical or quantitative approach, more ‘objective’ than close reading (Mone 2016). In close reading, the reader is the ‘analytical tool’, while in distant reading, the researcher draws on other kinds of tools, such as

algorithms and software. There are however several aspects that should be considered when interpreting distant reading. Moretti's thinking arose from literary studies, where there has always been debate on different methods of reading. The idea of close reading was one of these, emphasised by both American New Criticism and Russian Formalism. As a practice, close reading was originally against biographical interpretation of literary works and stressed the importance of exploring the inner structures of texts on the level of words and syntax. Today, close reading has gained many interpretations, as this special issue indicates, and, in historical studies, it often refers to contextual reading that is sensitive to unfolding structures of meaning. In literary studies, many other catchwords for reading have been emphasised. The idea of 'symptomatic reading' is based on the argument that the real meaning of a text 'lies in what it does not say' and therefore the reader tries to go beyond the surface and track down hidden layers of meanings. As a reaction to 'symptomatic reading', 'surface reading' was coined to take the face value of the text seriously and analyse 'the complexity of literary surfaces' that had been 'rendered invisible by symptomatic reading' (Best and Marcus 2009, 1–21).²

In his work, Moretti emphasised the simultaneous study of many texts through digital methods. From the perspective of historical studies, it is essential to be aware of the fact that his formulations of distant reading were, at least in its first instance in the *New Left Review*, meant to be against the literary canon. His starting point was to challenge the conventional and canonical understanding of world literature. Moretti (2000, 55) maintains:

'What does it mean, studying world literature? How do we do it? I work on West European narrative between 1790 and 1930, and already feel like a charlatan outside of Britain and France ... "I work on West European narrative, etc ..." Not really, I work on its canonical fraction, which is not even one percent of published literature. And again, some people have read more, but the point is that there are thirty thousand nineteenth-century British novels out there, forty, fifty, sixty thousand—no one really knows, no one has read them, no one ever will'.

Moretti (2000, 57) continues by arguing that close reading 'necessarily depends on an extremely small canon', since it is not possible to closely analyse as many texts as the understanding of world literature would require. This is why new methods are necessary. In fact, the researcher should 'learn how *not* to read'.

It seems that Moretti's provocation was addressed to literary scholars who analysed one single author or work and thus supported the existing canon. The choices of the researcher always carry values and judgements and participate in the construction of the notion of literature. This is crucial to point out when drawing on Moretti's ideas on distant reading in other fields of humanities. Furthermore, it must be noted that Moretti's representation of literary studies does not necessarily do justice to the plenitude of different approaches in his own field. Anyway, in historical studies, the relationship of the researcher to texts and textuality is clearly different from Moretti's image, since historians often aim at incorporating as many sources as possible into their investigations, instead of concentrating only on a few texts, let alone a single one.

Moretti made an interesting reference to historians too. He mentioned Marc Bloch's famous phrase that one needs 'years of analysis for a day of synthesis' (56–7). He continued by discussing the synthetical approaches of Fernand Braudel and Immanuel Wallerstein. In the end, he remarked that, for example, Wallerstein's works are very condensed. The author has crystallised his 'years of analysis' into a 'one-third of a page'. This shows Moretti's aim: he highlights the efforts to find new ways for creating abstractions based on the massive amounts of texts. For him, visualisations serve as a means to synthesise those observations that have been made through computational methods. This is what he further developed in his book *Graphs, Maps, Trees: Abstract Models for a Literary History* (Moretti 2005; see also Jänicke et al. 2015, 2–3).

The idea of distance is clearly central to distant reading. The researcher should not be too close to the text: '[T]he more ambitious the project, the greater must the distance be' (Moretti 2000, 57). Regarding historical studies, this view can, without a doubt, be also contested. It is important to draw on big data, which is available today for research, and apply novel ways of exploring these sources from a distance and return *from* the distance back to the close encounter of individual texts, and consider how the bigger picture changes our view on the details. As the next step, this article will discuss the question of big data and continue by suggesting a research strategy to combine close and distant reading.

Organising big data

As a research strategy, distant reading is often considered in situations where the researcher has to extract interpretations from a massive number of texts. The plain quantity of individual texts might be so high that

careful close reading would not simply be possible. The repository of American historical newspapers, *Chronicling America*, maintained by the Library of Congress, today includes 21.4 million pages of digitised content (Chronicling America 2024). Project Gutenberg in turn offers 70,000 digitised books as open access (Project Gutenberg 2024). It is easy to imagine historically relevant research questions that could draw on these vast collections. Without a doubt, these collections are not complete. There were many more newspaper pages printed in the United States from the eighteenth to the twentieth century, and to be sure 70,000 books are only a tiny fraction of all printed books. But these collections offer the possibility for asking those kinds of research questions that require a wide timeframe. By drawing on digitised newspapers, it would be possible to concentrate – for example – on how human–animal relations were debated and how this discourse changed over time. One could develop this idea further by asking how modernisation appeared as a gradual shift from horse power towards more complex transport technologies, or how it manifests itself in the explosion of domestic animal production and the increasing consumption of meat. There are other possible primary sources for answering these questions, but digitised newspapers offer an illuminating perspective on the larger framework of these historical changes. The choice of these kinds of questions does not mean that the researcher would only rely on distant reading techniques, since it would still be important to return *from* the distance and look at individual references to human–animal relations in detail.

Regarding digitised newspapers, the source material consists not of original papers, but of scanned images that have been further processed by recognising the texts with OCR software. The big data comprises newspaper pages as text or xml files that are thus in machine-readable format. This is necessary for the use of different distant reading techniques. This big data is often too diffuse or unstructured for the researcher and must be processed further. For example, the previous idea of human–animal relations in the newspaper corpus means that the researcher probably has to trace those kinds of articles or texts that deal with the issue and then extract them. This induces a new subset of data that could then be taken under scrutiny. In practice, although based on machine-readability, distant reading is not reading at all, at least if reading is understood as a conscious exercise, where the reader continuously compares what is read with what has been previously adopted, speculates how what is read will develop in the future, and draws conclusions on how the text makes meaning (Salmi, 2000, 32–3).

In practice, distant reading comprises tools that help organise big data. ‘Reading’ might start with statistical tools that help to assess the data at hand. It is important to measure the proportions of the dataset so that it can be illustrated and evaluated. This also helps in analysing the results of distant reading procedures. If the researcher creates his/her own subset of human–animal relations, it will help determine the characters, words, lines, pages, or issues. Word frequencies can also be illuminating: how many times, for example, ‘horses’ occur in the dataset. Furthermore, the data can be processed by tracking down word collocations, which means the automated analysis of sequences of words that tend to co-occur in the material (Graham, Milligan, and Weingart 2016, 159–194). In this case, it could inform, for example, the words that appear in connection to a ‘horse’, which again can be used in the analysis of meaning making around horse power. The role of horses can, of course, be studied through close reading and its essential densifications located in large datasets, but there are aspects that only reveal themselves through distant reading techniques in their full scale. Such aspects include, for example, the ways in which particular discourses, expressions, or word collocations become amplified in the material. In Finland, for example, horse power was significant until the 1950s, but gradually, the motorisation of the country changed the situation dramatically. Motorisation had already begun in the 1920s and 1930s, but the Second World War slowed down the development. During the 1960s, horse rides had already become a theme of nostalgia and an epitome of rural life. In the study of these kinds of long-term changes, quantifiable evidence can help analyse the intensity and magnitude of the research topic of the past.

The proportions of the dataset assist researchers in grasping the size and quality of their material. Moretti emphasised the role of visualisation, and certainly, these propositions can be presented in visual form, but they can be communicated in a textual and numerical form as well. There are easy-to-use tools for counting these properties, such as AntConc, which is a freeware toolkit for text corpus analysis. For visualisation, there are also open-sourced programmes that can be easily tested. One of the most popular is Voyant Tools, a web-based application for text analysis and visualisation.

These suggestions for organising big data are only the beginning. There are many more advanced methods that can be employed to take ‘distance’ from the texts. The array of distant reading techniques can be further illustrated by looking at the website *The Programming Historian*, which offers ‘novice-friendly, peer-reviewed tutorials that help humanists

learn a wide range of digital tools, techniques, and workflows to facilitate research and teaching'. Many of the tutorials do not require previous expertise or advanced technological skills but offer step-by-step guidelines for all interested. The website contains peer-reviewed tutorials in English, French, Portuguese and Spanish. In May 2024, the English part comprised 108 lessons on many methods and tools, such as the use of Mallet, but it also included more advanced guidelines, such as using R-language for data wrangling and management or Python for stylometrics. Under the heading 'Distant Reading', the website has 15 tutorials, and these themes show what methods are on a preliminary level associated with distant reading techniques. In addition to data wrangling, stylometrics, and topic modelling, these lessons include, for example, the study of text similarity with Python, corpus analysis with AntConc, basic text mining options with Python, and an introduction to a MySQL database to store and filter data. The lessons also include sentiment analysis, which is a method for measuring the degree of positivity or negativity within a group of texts.³ These distant reading strategies could be accompanied by many more, and the ways of assessing big data are undergoing continuous change and development (Guldi 2023). There are easy-to-use tools but also very complicated and sophisticated methods that require either fluent programming skills or engagement in interdisciplinary collaboration. Franco Moretti (2013, 214–22) was especially interested in networks and visualising, for example, character networks of fictitious texts.⁴ Shawn Graham, Ian Milligan, and Scott Weingart (2016, 195–234) have written a concise introduction to network analysis and many other distant reading techniques in their book *Exploring Big Data: The Historian's Macroscope*.

Moretti advocated the idea that distant reading would help the researcher analyse and understand overarching issues that otherwise would remain hidden. While this is certainly an aspect worthy of attention, it is furthermore important to stress that, more than actual reading, distant reading can be seen as an array of strategies in organising data and thus also preparing it for other reading methods, such as close reading. Many of the aforementioned techniques try to capture the dimensions of texts as a flow. In these examples, what is at stake is not a single meaning-making text but a flow of texts, propositions of which can be, and should be measured to understand the textual context of a single manifestation. The use of OCR'd scanned documents often stresses the text only as a series of characters, words, sentences, and paragraphs and makes the material aspect of textuality invisible. In the historical world, a newspaper

column was part of the technological process aimed at producing printed pages on a daily basis. The next subchapter focuses on a particular distant reading technique, text reuse detection, which is useful for assessing newspaper publishing as a material flow and helps to identify how texts relate to each other. Through its overarching possibilities, it supports the close reading of a single text by showing how it was used and reused over the course of time and how it was given cultural gravitation through repetition.

Text reuse and amplification

In the 2000s, computer-assisted methods proved to be effective in the study of how texts were copied and reused in the past. These practices include various forms of reuse. They can cover consciously republished copies of earlier content, from short quotations to long articles, and textual templates that were used in publishing regular information. They can also include official announcements or circulars that were sent to several newspapers at the same time, and even viral news that have been forwarded without consciously considering them as reprints.⁵

The making of copies has always been a method of reproducing more texts. Before the advent of the printing press in the fifteenth century, manuscripts were copied by hand, either by following the text line-by-line or by writing down the text from a dictation. Slaves copied texts in the Roman empire, and monks produced copies of classic texts in medieval monasteries. The availability of texts and material for reading has been at the heart of literary cultures. It may be argued that the cultural significance of a text derives not only from its internal structures of meaning making but, simultaneously, from its material existence and from its concrete manifestations. In the study of classical and medieval texts, it has been crucial to follow the routes of different manuscript editions to compare their variations and assess the *Wirkungsgeschichte* of the text. The interpretation of these editions must take the material dimensions of the text into consideration and conclude how they influence the analysis.

For historians, it is often essential to determine how the texts of the past relate to each other. This may happen through close reading and careful interpretation of sources. Texts refer to each other but also to non-textual sources. They make references, quotes, and paraphrases. In the end, the historian has found a network of interrelated texts and sources. These kinds of connections can unfold not only through close

reading, but also via distant reading techniques. Text reuse detection is one of these, although it has its limitations, since many of the detection methods are based on the recognition of similarities between sequences of characters or words and might therefore not detect paraphrases or other modifications of a text. Still, text reuse detection can effectively locate instances where a particular passage has been reproduced word by word.

Text reuse has been studied with computational methods by many projects in the 2000s by detecting morphological, syntactic, or semantic similarities between texts (Clough et al., 2002; Cordell and Smith 2017; Franzini, Franzini, and Büchler 2016; Gaizauskas et al. 2001; Lundell et al. 2023; Salmi et al. 2021). One of the most well-known solutions is Passim software, developed by David A. Smith, for the Viral Texts project. It has been successfully used in the study of nineteenth-century American newspapers (Smith, Cordell, and Maddock Dillon 2013; Smith, Cordell, and Mullen 2015). The software draws on word n-grams to recognise duplicate texts, along with cases where texts are different either because of OCR errors or editorial work (Smith, Cordell, and Maddock Dillon 2013).

In another project lead by the author of this article and based on Finnish newspapers, a character-based solution was developed, drawing on previous work in bioinformatics, especially on the software BLAST (Basic Local Alignment Search Tool), which was originally tailored to align similarities in biomedical sequences, but it is also applicable to text, strings of characters. In the analysis that involved five million pages from the advent of newspaper publishing in Finland in 1771 until 1920, the project could find 13.8 million clusters of reused texts or text passages (Salmi et al. 2021, 14–28). These clusters included many kinds of materials, not only journalistic content but also advertisements, announcements, fictitious short stories, poems, and humorous anecdotes. The abundance of the latter forms of content makes it challenging to explore, and there is obviously much to be done in the history of media publicity in all its richness.

The fact that newspapers copied each other's contents is no news in itself. It was the habit of newspaper editors to look for suitable content from other publications, home and abroad. While this has been acknowledged by media historians, the repetitive, amplifying nature of publishing has still been difficult to assess on its full scale before the availability of digitised sources. The benefit of the Finnish case was that in Finland, all published issues have been digitised, which gives more gravity to the

distant reading results. The 13.8 million clusters that have been found represent the repetitive character of newspaper publishing from the dawn of the press in Finland up to the early twentieth century. Simultaneously, these clusters offer the possibility to analyse the role of viral publishing and the reprinting of contents effectively within a short time. They also offer the possibility to see how certain texts floated in time and how they were printed within an extensive time span, either continuously or with shorter and longer intervals. The results of the project showed that 85% of all clusters occurred within one year, while the rest of the clusters were longer textual chains. There were almost 290 clusters, the span of which was over 140 years. This means that, in these cases, the first instance of the text was from the late eighteenth century, and the last one was reprinted in the twentieth century.

The study of text reuse has many applications. It can be used to support close reading of individual texts, or groups of texts, for example, when the researcher would like to analyse *en masse* how, for example, earlier sources have been quoted. Lincoln Mullen (2016) has made an illuminating exploration of how the Bible was quoted in the nineteenth-century American press. This may be helpful in the analysis of a single text when the researcher can identify cases in which the same quote has been employed as well. This may contribute to locating differences in meaning making. Another, perhaps even more fruitful, aspect is that it can shed light on the quotation practices of newspaper publishing in general. If the historian is interested, for example, in the role of the Bible in the nineteenth-century imagination, it is not only the Bible as a text collection that counts but the fact that the Bible was dispersed and fragmented, entangled with other texts, also in ways where the origin of the quote has been hidden or remained unmentioned. It may further be argued that the culture of texts is, and was, in the nineteenth century, a rhizomatic amalgamation of textual fragments and memes that had different origins and had an impact on contemporary thinking and worldview in any given moment of the past.

The study of reuse also gives ground for another idea on textuality. Historians have always been interested in the impact, and changing interpretations of influential texts from the Bible to the Communist Manifesto. Texts have been quoted and paraphrased, shortened, and widened, or reprinted in both original or edited formats. For this work, distant reading techniques can offer new avenues. Bibliographic metadata has been successfully employed for the study of book history, along with full-text analysis through various computer-assisted methods (Lahti et al.

2019). The study of text reuse in particular can be a fruitful method to assess the extent to which a particular text, and the cultural significance it holds, was, and became fostered in the past. In this way, textuality can be viewed from the perspective of cultural amplification. Repetition, or text reuse in itself, is not enough to assess the ramifications of the flow of texts. It is important to pay attention to how reused texts aggregate and get prominence in the past.

An empirical example can be used to illustrate this. The following case is based on a newspaper text about a waltz composed by Johann Strauss the Elder. It shows how different prints and reprints formed a network of material instances in the past, and this textual cluster can shed light on the interpretation of a single text. On 4 January 1843, the Finnish newspaper *Åbo Underrättelser* published a short story entitled ‘Strauss and the Sophie Waltz’ (‘Strauss och Sophia-valsen’). The paper does not indicate the nature of the story, but it starts by referring to Johann Strauss the Elder and his Sophie Waltz that had become very popular. It also tells us that the text is quoted from a ‘German newspaper’. What follows is a poetical, romantic narrative from the life of Johann Strauss, telling of his supposed love affair with ‘Sophie, daughter of a count’, whose eyes were ‘bluer than the clear sky of Italy’. The love affair proves to be an impossibility; however, since Sophie has already been promised to a nobleman. Time passes, and after a while, this nobleman comes to Strauss without knowing about the love affair and asks the composer to prepare a waltz for their wedding. Strauss agrees, devotes passionately to his work, and channels all his painful memories into the piece. Finally, the waltz is premiered, and it immediately enchants everyone, especially Sophie, who ‘waltzes, waltzes, and waltzes without stopping’. Finally, exhausted by dancing and the turmoil of emotions, she drops dead in the arms of her husband. (Anonymous, ‘Strauss och Sophia-valsen’. *Åbo Underrättelser*, January 4, 1843).

It is obvious to the reader that this text is a fictitious story, a novelette, although it refers to other newspapers and presents itself as a translation from the German press. It does not reveal the exact source, nor does it tell the author of the story. The fictitious nature of the text is expressed by the utmost poetical language, its strongly emotional undertone, and its narrative structure, since the story resembles *contes fantastiques*, inspired by E. T. A. Hoffmann, which were popular at the time. The story also reminds the reader of those fictional texts in the press that surrounded celebrities of the early nineteenth century, such as Niccolò Paganini and Franz Liszt (Salmi

2016, 135–53). The press became a platform for the imagination of popular figures and echoed their fame. Certainly, Johann Strauss the Elder was one of the most well-known public figures of the time, and his contribution as a waltz composer was acknowledged everywhere in Europe. As a single text, the story published by *Åbo Underättelser* participates in the construction of celebrity culture, but it may be argued that its true nature and capacity in this process cannot be illuminated only through one textual instance. It should be set into a larger context of similar texts. It is possible to explore how celebrity culture, in this case Johann Strauss the Elder's fame, was amplified through a network of texts, reprinting similar texts again and again. This material flow of texts amplified the presence of the composer in public life.

The database of text reuse in the Finnish press shows that the text had been published earlier in Finland by the magazine *Wanadis* on 18 February 1840 (Anonymous 1840). Both versions of the text were published in Swedish, which was the dominant language of the press in Finland at the time. Furthermore, a consultation with the newspaper database of the Royal Library of Sweden reveals that the publications by the Finnish press were not the only ones since the same text also appeared in *Dagligt Allehanda* in Stockholm on 27 December 1842. It becomes also evident that, out of all Strauss waltzes, the Sophie Waltz, originally titled *Sophie-Tänze*, was particularly popular, and newspapers included hundreds of advertisements of it for domestic music making. These included, for example, F. Weller's piano arrangement for two hands, with a short description: 'after a Romantic story on Strauss's life'.⁶ The practices of text reuse thus participated in the construction of fame on multiple levels, through a novelette but also through advertisements.

The cultural veil of the story on the origins of *Sophie-Tänze* lived through the century. In 1890, the Austrian monthly magazine *Oesterreichische Lesehalle* published a short story, 'Nur ein Musikant. Eine Liebesgeschichte aus dem Leben des Walzerkönigs Johann Strauss', which is a longer novelette but includes a similar kind of narrative ending with the death of Sophie (Mskzt 1890). The flow of texts had persistently continued to maintain the memory of the story.

In the study of past textualities, I argue that it is essential to pay attention to how a particular text was amplified through repetition. The story on the origins of Sophie Waltz was echoed by the press and domestic music making, and these entangled processes participated

in constructing the cultural representation of Strauss, along with the contemporary understanding of the waltz and its bodily ramifications.

Conclusion: towards the study of textuality as a material flow

This article has discussed the notion of distant reading and the challenge of big data for cultural historians. The issue is relevant if only because the toolbox of the cultural historian, and the historian more broadly, has traditionally emphasised the importance of close reading. This article has aimed at bridging the ideas of ‘closeness’ and ‘distance’ and argued that by applying distant reading as a methodological strategy, texts can be interpreted as material flow, which again enriches the interpretation of texts. The article has used the abundance of newspaper publishing in the modern era as an example of this: one text has significance in itself, but the text is continuously positioned and re-positioned as part of this flow. Furthermore, this flow does not only comprise characters, sentences, and paragraphs, it also entails natural and synthetical ink and printing paper that is technologically processed and, of course, printing machines that, in the early nineteenth century, started to produce tens of thousands of pages per day to be distributed geographically via new forms of transportation. These processes became quicker and more efficient decade by decade, which helps explain how the press finally expanded without any conscious plan rhizomatically. This enabled the simultaneous publication of various stories, sometimes opposing ones.

The challenge for future research will be that if textuality is understood as material flow, how could an individual researcher acknowledge the material dimensions of this mobile culture of texts and concurrently be sensitive to the detail. The analysis of this material flow is influenced by the time period under scrutiny, the time span being considered and the type of data available to the researcher. This article has explored the concept of distant reading, the tools available for organising data – which are, of course, in a state of constant flux – and suggested ways of combining distant and close reading, arguing, through examples, that they are not opposites but complementary practices that can be used simultaneously and in parallel to enrich the analysis at hand. It is precisely this combination that holds great potential for future researchers. How this unfolds in the end depends on the research question, which adjusts the balance between reading strategies.

Notes

1. This part of the article is based on my previous discussion on distant reading in (Salmi 2020, 33–8).
2. Marcus has also used the expression ‘just reading’, see (Marcus 2007, 75).
3. For further details on various methods of distant reading, see *The Programming Historian* 2024.
4. *The Programming Historian* offers several lessons on networks for self-learning. For further details on network analysis, see *The Programming Historian* 2024.
5. On previous research on text reuse, see Salmi et al., ‘The reuse of texts in Finnish newspapers and journals, 1771–1920: A digital humanities perspective’, 14–19; Paju et al., (2023) ‘Towards an Ontology and Epistemology of Text Reuse: Cycles of Information Flows in Finnish Newspapers and Journals, 1771–1920’, 253–273.
6. See, for example, *Svenska Biet*, June 8, 1843, *Linköpingsbladet*, November 22, 1843; *Dagligt Allehanda*, February 9, 1844; *Stockholms Dagblad*, March 5, 8 and 9, 1844.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Hannu Salmi is a Professor of Cultural History at the University of Turku, Finland, and was nominated Academy Professor for the years 2017–2021. He is an historian of the nineteenth and twentieth centuries and the author of volumes, such as *Nineteenth-Century Europe* (2008) and *What is Digital History?* (2020). He is also the editor of *The Routledge Companion to Cultural History in the Western World* (2020) together with Alessandro Arcangeli and Jörg Rogge.

ORCID

Hannu Salmi  <http://orcid.org/0000-0001-8607-6126>

References

- Anonymous. 1840. Johan Strauss och Sophia-valsén. *Wanadis*, February 18, 1840.
- Anonymous. 1843. Strauss och Sophia-valsén. *Åbo Underrättelser*. January 4, 1843.

- AntConc. 2024. Accessed May 14, 2024. <https://www.laurenceanthony.net/software/antconc/>.
- Best, Stephen, and Sharon Marcus. 2009. "Surface Reading: An Introduction." *Representations* 108 (1): 1–21. <https://doi.org/10.1525/rep.2009.108.1.1>.
- Chronicling America: Historic American Newspapers. 2024. Accessed May 14, 2024. <https://chroniclingamerica.loc.gov/>.
- Clough, P. D., R. Gaizauskas, S. L. Piao, and Y. Wilks. 2002. "Measuring Text Reuse." *Proceedings of Association for Computational Linguistics (ACL2002)*, 152–59. Philadelphia, PA: ACL.
- Cordell, Ryan, and David Smith. 2017. "Viral Texts: Mapping Networks of Reprinting in 19th-Century Newspapers and Magazines." Accessed May 14, 2024. <http://viraltexts.org>.
- Franzini, G., E. Franzini, and M. Büchler. 2016. *Historical Text Reuse: What Is It?* Accessed May 14, 2024. <http://www.etrapp.eu/historical-text-re-use/>.
- Gaizauskas, R., J. Foster, Y. Wilks, J. Arundel, P. Clough, and S. L. Piao. 2001. "The METER Corpus: A Corpus for Analysing Journalistic Text Reuse." *Proceedings of Corpus Linguistics* 2001, 214–23. Lancaster: Lancaster University. <https://eprints.lancs.ac.uk/id/eprint/52137/>.
- Gibbs, Philip. 1996. *The Word in the Third World: Divine Revelation in the Theology of Jean-Marc Éla, Aloysius Pieris and Gustavo Gutiérrez*. Roma: Editrice Pontificia Università Gregoriana
- Graham, Shawn, Ian Milligan, and Scott Weingart. 2016. *Exploring Big Data: The Historian's Macroscope*. London: Imperial College Press.
- Guldi, Jo. 2023. *The Dangerous Art of Text Mining: A Methodology for Digital History*. New York: Cambridge University Press. <https://doi.org/10.1017/9781009263016>.
- Jänicke, Stefan, Greta Franzini, Muhammad Faisal Cheema, and Gerik Scheuermann. 2015. "On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges." *Eurographics Conference on Visualization (EuroVis)*, edited by R. Borgo; F. Ganovelli, and I. Viola. Accessed May 14, 2024. <https://www.etrapp.eu/wp-content/uploads/2015/07/paper.pdf>.
- Lahti, Leo, Jani Marjanen, Hege Roivainen, and Mikko Tolonen. 2019. "Bibliographic Data Science and the History of the Book (C. 1500–1800)." *Cataloging and Classification Quarterly* 57 (1): 5–23. <https://doi.org/10.1080/01639374.2018.1543747>.
- Lundell, Patrik, Hannu Salmi, Erik Edoff, Jani Marjanen, and Heli Rantala, eds. 2023. *Information Flows Across the Baltic Sea: Towards a Computational Approach to Media History*. Mediehistoriskt arkiv 56. Lund: Föreningen Mediehistorik arkiv. <https://doi.org/10.54292/s6au8axqht>
- Marcus, Sharon. 2007. *Between Women: Friendship, Desire, and Marriage in Victorian England*. Princeton: Princeton University Press.
- Mone, Gregory. 2016. "What's Next For Digital Humanities." *Communications of the ACM*. Accessed May 14, 2024. <https://cacm.acm.org/news/whats-next-for-digital-humanities/>.
- Moretti, Franco. 2000. "Conjectures on World Literature." *New Left Review* 1 (January–February): 54–68.

- Moretti, Franco. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*. London: Verso.
- Moretti, Franco. 2013. *Distant Reading*. London: Verso.
- Mskzt, N. 1890. "Nur ein Musikant. Eine Liebesgeschichte aus dem Leben des Walzerkönigs Johann Strauss." *Oesterreichische Lesehalle: Monatsschrift für Belehrung und Unterhaltung* 119 (November): 321–30.
- Mullen, Lincoln. 2016 *America's Public Bible: Biblical Quotations in U.S. Newspapers*. Accessed May 14, 2024. <http://americaspublicbible.org/>.
- Nicholson, Bob. 2020. "Counting Culture; Or, How to Read Victorian Newspapers from a Distance." *Journal of Victorian Studies* 17 (2): 238–46. <https://doi.org/10.1080/13555502.2012.683331>.
- Paju, Petri, Heli Rantala, and Hannu Salmi. 2023. "Towards an Ontology and Epistemology of Text Reuse: Cycles of Information Flows in Finnish Newspapers and Journals, 1771–1920." In *Digitised Newspapers – a New Eldorado for Historians?*, edited by Bunout , and Estelle, 253–273. Oldenbourg: De Guyter. <https://doi.org/10.1515/9783110729214-001>.
- The Programming Historian. 2024. Accessed May 14, 2024. <https://programminghistorian.org/>.
- Project Gutenberg. 2024. Accessed May 14, 2024. <https://www.gutenberg.org/>.
- Salmi, Hannu. 2016. "Viral Virtuosity and the Itineraries of Celebrity Culture." In *Travelling Notions of Culture in Early Nineteenth-Century Europe*, edited by Asko Nivala; Hannu Salmi, and Jukka Sarjala, 135–53. New York: Routledge.
- Salmi, Hannu. 2020. *What Is Digital History?* Polity: Cambridge.
- Salmi, Hannu, Petri Paju, Heli Rantala, Asko Nivala, Alekski Vesanto, and Filip Ginter. 2021. "The Reuse of Texts in Finnish Newspapers and Journals, 1771–1920: A Digital Humanities Perspective." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 54 (1): 14–28. <https://doi.org/10.1080/01615440.2020.1803166>.
- Smith, David A., Ryan Cordell, and Elizabeth Maddock Dillon. 2013. "Infectious Texts: Modelling Text Reuse in Nineteenth-Century Newspapers." *Proceedings of the Workshop on Big Humanities*, 86–94. Washington, DC: IEEE Computer Society Press.
- Smith, David A., Ryan Cordell, and Abby Mullen. 2015. "Computational Methods for Uncovering Reprinted Texts in Antebellum Newspapers." *American Literary History* 27 (3): E1–E15. <https://doi.org/10.1093/alh/ajv029>.
- Voyant Tools. 2024. Accessed May 14, 2024. <https://voyant-tools.org/>.