DIGITAL HUMANIORA I NORDEN

DIGITAL HUMANITIES IN THE NORDIC COUNTRIES SECOND CONFERENCE

GÖTEBORG, MARCH 14–16, 2017

GÖTEBORGS UNIVERSITET

CENTRUM FÖR DIGITAL HUMANIORA

DIGITAL HUMANIORA I NORDEN
DIGITAL HUMANITIES IN THE NORDIC COUNTRIES

DHN 2017

Digital humaniora i Norden/
Digital Humanities in the Nordic Countries
Göteborg, March 14–16 2017

CONFERENCE ABSTRACTS

Published by: The University of Gothenburg, Department of Literature, History of Ideas and Religion, 2017
ISBN: 978-91-88348-83-8
http://hdl.handle.net/2077/52239
Editor: Daniel Brodén
Cover logo: Dick Claésson

Programme Committee

Christian-Emil Ore, University of Oslo, Norway (Chair)
Jenny Bergenmar, University of Gothenburg, Sweden (Co-chair)
Ilze Auziņa, University of Latvia, Latvia
Stefan Gelfgren, Umeå University, Sweden
Olga Holownia, University of Iceland, Iceland
Sakari Katajamäki, Finnish Literature Society – SKS, Finland
Rimvydas Laužikas, Vilnius University, Lithuania
Cecilia Lindhé, University of Gothenburg, Sweden
Liina Lindström, University of Tartu, Estonia
Mats Malm, University of Gothenburg, Sweden
Bente Maegaard, Copenhagen University, Denmark
Annika Rockenberger, University of Oslo, Norway
Nina Tahmasebi, University of Gothenburg, Sweden
Mikko Tolonen, University of Helsinki, Finland

Local Organizing Committee

Jenny Bergenmar, University of Gothenburg (Chair)
Daniel Brodén, University of Gothenburg
Trausti Dagsson, University of Gothenburg
Cecilia Lindhé, University of Gothenburg
Mats Malm, University of Gothenburg
Julia Pennlert, University of Borås

Preface

The book you hold in your virtual hand contains the collection of abstracts for the presentations to be given at the second conference for Digital Humanities in the Nordic Countries, DHN2017. The conference is held at the University of Gothenburg March 14–16, 2017, and is organized by the Centre for Digital Humanities at the University of Gothenburg.

Digital Humanities, Humanities Computing, Computer Applications in the Humanities or Computational Methods in the Humanities – our field has had many names throughout its history going back to the end of the 1940s when Roberto Busa started his collaboration with IBM on producing a complete concordance for the works of Thomas Aquinas. Originally, Busa did not plan to use digital computers as we know them. His idea was to use punch cards and the corresponding semi-mechanical machinery to create the concordance. From our retrospective point of view, Busa's large number of boxes filled with punch cards seems to be extremely unsophisticated. This is of course not true. Firstly, punch cards represented the state of the art, secondly it is the scholarly method and how the available machinery is exploited to achieve the results, which are important. The same is true for DH today. Although the study and development of digital methods are important, in DH one mostly uses digital methods developed in other contexts like for example machine learning and general statistics. The extensive digitization of our everyday lives has extended DH to a meta-level. An important part of DH is the study of the digitized society – also called the study of digital cultures at some universities.

This wide scholarly landscape of DH is reflected in the three main topics listed in the call for papers for DHN2017:

- Nordic Textual Resources and Practices,
- Visual and Multisensory Representations of Past and Present,
- The Digital, the Humanities, and the Philosophies of Technology.

DH activities in the Nordic countries have a long history dating back at least to the early 1980s. However, there has never been a Nordic Association for the digital and the humanities until the Swedish initiative came early in 2015 headed by professor Mats Malm from the University of Gothenburg. The organization *Digital Humanities in the Nordic Countries* was founded in 2015 and is now one of three DH-organizations associated to the *European Association for Digital Humanities*, EADH. Through EADH our organization is connected to ADHO, the global Association of Digital Humanities Organizations.

The first conference, DHN2016 held in Oslo, was a big success. A second conference will usually indicate whether there still is an interest. The response to the call for paper to DHN2017 was indeed good. We received 105 proposals for workshops, panels, presentations and posters. The final programme consists of three plenary keynotes, 56 paper presentations, 4 panel sessions and 14 posters, all presented in this book of abstracts.

The authors were asked to indicate the main topic of their submission. The first of the three main topics "Nordic Textual Resources and Practices" is the traditional topic in a DH conference and not unexpectedly, 46 % of the submissions were tagged with this topic. 28 % were tagged with "Visual and Multisensory Representations of Past and Present". In this category we find cultural heritage papers, arts as well as visualization techniques used in text studies. The final topic "The Digital, the Humanities, and the Philosophies of Technology" is the least typical for traditional DH and 24 % were tagged with this category. A large part of these are about topics not so uncommon in traditional DH. However, there are also many interesting presentations on the meta-level which are not so common. In future conferences one should definitely encourage

submissions with topics in this third category. In general the submissions cover a wide range of DH. Digital Humanities in the Nordic countries is indeed an active, flourishing activity.

We wish to give our warmest thanks to our colleagues in the Programme Committee and the Local Organising Committee, and also to the Scientific Committee who did a splendid job in reviewing and evaluating the submissions (see also dhn2017.eu).

Finally we would like to thank our sponsors for their generous funding enabling us to organize this conference: The Royal Swedish Academy of Letters, History and Antiquities, Sven och Dagmar Saléns stiftelse and the Department of Literature, History of Ideas, and Religion, University of Gothenburg.

Bursary funding has been generously provided by Digital Scholarly Editions Initial Training Network, DiXiT.

Christian-Emil Ore
Chair of the Programme Committee, Chair of Digital Humanities in the Nordic Countries

Jenny Bergenmar
Chair of the Local Organizing Committee

Table of Contents

Short Papers

Posters

Pre-conference Workshops

PLENARY LECTURES

New Natures of the Anthropocene and the Need for Humanistic Inquiry into the Digital

Dolly Jørgensen
Associate Professor, History of Technology & Environment, Luleå University of Technology, Sweden

Nature is not without a history, and new natures are constantly produced through human technology and activities. This idea is currently framed as the Anthropocene, the geological Era of Man, which has been proposed as an official geologic era beginning c. 1950. The evidence for this new geologic era is based on new materials like radioactive isotopes and plastics and the redistribution of materials like carbon from consumed fuels. The Anthropocene's wide and deep human influence on the planet also corresponds with the creation of digital technologies in the modern era. Presenting examples from webcams, visitor information boards, databases, and other forms of digitally augmented nature, I will argue that it is through the digital that many humans now come to know and experience new nature. Knowledge of nature has always been mediated through technology, but the digital has enabled both greater physical distance and conceptual closeness with nature. I propose that digital humanities needs to move beyond thinking of the digital as a tool to thinking of the digital as a part of the ecosystem of the Anthropocene's new nature, shaping and reshaping both culture and the environment.

Topic: The Digital, the Humanities, and the Philosophies of Technology

Fluid, Frozen, Aggregated: On Discursive Images, Visual Discourse, and the Rematerialization of Data

Katja Kwastek
Professor of Modern and Contemporary Art, Vrije University Amsterdam, Netherlands

After a brief overview of the history of digital art history, this lecture will discuss the discursive potential of (digital) images, remediating or responding upon each other, circulating in the networks, aggregating as big image data visualizations, and serving as arguments within (scholarly) discourse. It will look both at the impact and implementation of these phenomena within academia and at their reflection and instrumentalization in artistic practice, including recent tendencies to rematerialize data and discourse in sculptures and installations.

Topic: Visual and Multisensory Representations of Past and Present

Towards a Macroscope for the Study of Nordic Literatures

Peter Leonard
Director of Yale University Library Digital Humanities Lab, United States of America
Timothy R Tangherlini
Professor, Scandinavian Section and Dept. of Asian Languages and Cultures, UCLA, United States of America

The study of Nordic literatures is one marked by a series of complexities that pose significant challenges as we move toward developing meaningful approaches to the study of literature at the scales made possible by the vast and successful digitization projects underway across the Nordic region. These complexities arise not only from the linguistic variation across the region, where

seeming proximity of the languages can lead to a false sense of security, but also from divergences in concepts of canon, periodization, and the divergent cultural trajectories that characterize the region. Rather than resign in the face of this complexity, we should embrace it as a challenge that may allow us to make intriguing discoveries about literary influence, development and disruption. Modeling this complexity requires a bold turn toward an encompassing methodology that weds the time-tested benefits of close reading, philology, and literary history to emerging approaches of distant reading such as network analysis and probabilistic modeling. Consequently, we propose developing a "macroscope" for the study of Nordic literature. Katy Börner, writing in the CACM, articulates the power of the macroscope for the study of complexity, noting that the macroscope "provide[s] a 'vision of the whole,' helping us 'synthesize' the related elements and detect patterns, trends, and outliers while granting access to myriad details. Rather than make things larger or smaller, macroscopes let us observe what is at once too great, slow, or complex for the human eye and mind to notice and comprehend" (Börner 2011, 60). In our talk, we present some initial steps toward realizing a macroscope tuned specifically to the exigencies of the Nordic literary world.

We present a series of three tools that could be integrated into a rich study environment for Nordic literature. The first tool allows for the alignment of closely similar passages, allowing a scholar to focus on close comparisons between works. Sequence alignment has always been a powerful tool of the philologist, but has usually required painstaking, manual work. This tool makes such alignment far quicker and, by relaxing the standards of precision, can be used to detect similarities within and across authors. The second tool, subcorpus topic modeling (STM), radically lowers the barriers to entry for scholars interested in using probabilistic methods for discovering latent semantic patterns in corpuses. The tool allows the user to fashion a virtual "fishing net" to discover similar semantic patterns (topics) in much larger, unlabeled corpora. The third tool makes use of word-embedding models, to allow the user to trace differences in language use patterns within authorships, across authorships, and across historical periods.

We recognize that these tools constitute baby steps on the way to a more unified macroscope for the study of Nordic literature, and that many more tools based on sound methodology should be devised. Many additional challenges—from the development of accessible and machine actionable corpora, to the development of tools that can consistently and accurately deal with linguistic differences across the modern Scandinavian languages—lie ahead. Yet the promise of these approaches, which will allow scholars to work across traditions, to engage reading from the distant to the close and everything in between, and to situate these studies in a rich historical context of cultural change, is too appealing not to engage.

Topic: Nordic Textual Resources and Practices

PANELS

Digitizing Industrial Heritage: Models and Methods in the Digital Humanities

Anna Foka
Finn Arne Jørgensen
Pelle Snickars
Umeå University, Sweden

The Nordic cultural heritage sector is under strong pressure to digitize their collections, as a way of ensuring long-term preservation and access. Yet, digitization itself is no panacea, nor is it always entirely clear what it means to digitize something. The relationship between the original and the copy has long been the subject of scholarly analysis, yet new digitization technologies and digital fabrication methods such as 3D scanning, 3D printing—and even computational modeling—raises intricate questions about representation, authenticity, and contextuality in cultural heritage.

This panel interrogates the intersection between digitizing archives and visualizing history with the goal of developing methodology of high relevance for the cultural heritage sector. We build on Turkel's argument that 'the process of digitization creates a representation that shares some of the attributes of an original' and that technologies that are not frequently used by historians, for example, could allow us to capture and recreate particular attributes of documents, artefacts and environments' (2011: 287). From this perspective, the panel aims to explore the possibilities and affordances of emerging technologies and methods that combine innovative digital visualizations with more traditional modes of understanding of the past.

The panel discusses three different cultural heritage perspectives to examine the specificity of digitization and its potential to bridge research, institutional heritage and interest from the general public. Our case studies for examination spring from the project 'Digital Models. Techno-historical collections, digital humanities & narratives of industrialization', funded by the Royal Swedish Academy of Letters, History and Antiquities (2016–2019). The project is a collaboration between the Swedish National Museum of Science and Technology, with a national responsibility for technical and industrial heritage, and Humlab at Umeå University. Based on selected parts of the museum's collections the panel aims to explore the potential of digital technologies to reframe Swedish industrialization and the communication of society, people and environments.

Material from the museum's collections selected for digitization (and research) are all related to different phases of Swedish industrialization. Hence, the underlying project idea is that heritage institutions should not only devote themselves to digitize their collections, but also make it possible for researchers and visitors to use digital heritage through various kinds of applications, tools and software.

The starting point of the project are three selected categories of material in the museum collections which in various ways mirror the "three phases of industrialization": (A) parts of the business leader and industry historian, Carl Sahlin's (1861-1943) extensive collection. (B), all editions of the museum yearbook, Daedalus (1931-2014), and (C) 31 wooden models from Swedish pre-industrial inventor Christopher Polhem's "mechanical alphabet" from the early 1700s, belonging to his so called Laboratorium mechanicum. These models were later (in the 1750s) inserted into The Royal Model Chamber (Kungliga modellkammaren), a Swedish institution for information and dissemination of technology and architecture set up in central Stockholm at Wrangelska palatset.

Our digitization methods are correlated with different industrial-historical periods, and in effect they will result in three sets of digital tools, applications and/or game prototypes focused on various narratives of Swedish industrialization. The Royal Model Chamber is, for example, a space which our project has constructed a VR-model of through HTC Vive and Unity. The theme of (digital) reconstruction thus has both a profound and ambiguous historical dimens-

ion, since Polhem sincerely believed (as a pre-industrial inventor) that physical models were always superior to drawings and abstract representations.

As stated, it is not always clear what it means to digitize something, and the panel particularly seeks to address the challenges of digitizing disparate forms of data, gamification and visualizations in immersive, virtual reality environments. Following the London Charter on computer-based visualisation of heritage, it promotes "intellectual and technical rigour in digital heritage visualisation" – yet, in what way should one, for example, digitize Polhem's models and his Laboratorium mechanicum? The London Charter defines principles for the use of computer-based visualisation methods "in relation to intellectual integrity, reliability, documentation, sustainability and access" (London Charter 2009). Indeed, the charter recognises that the range of available computer-based visualisation methods is constantly increasing. Still, what is the exact relation between "technical rigour" and virtual heritage (in a software culture permeated by constant updates)? In order to provoke a confrontation of (stupid) scanning versus (intelligent) simulation, we have for example both 3D scanned some of Polhem's models using an ordinary iPhone (and Agisoft Photoscan software), and CT-scanned them (at Linköping University Hospital), with multitudinous images taken from different angles to produce a cross-sectional and tomographic 3D image (a kind of virtual slice, allowing one to see inside the models without breaking them).

The panel will feature short presentations from the three participants, with Foka, Jørgensen, and Snickars presenting one category each. Demonstrations of preliminary results and visualizations will also be presented, as well as ample time for discussion and exchange with the audience.

Anna Foka is Assistant Professor in Information Technology and the Humanities at Humlab, Umeå University. Her research is focused on the rendering of the past with embodied and interactive technology, critical digital methods for creative and cultural heritage organizations, and digitalized infrastructures for the study of arts and humanities.

Finn Arne Jørgensen is Associate Professor of History of Technology and Environment at Umeå University, a position he will combine with serving as head of the Norwegian Museum of Travel and Tourism starting February 2017. His research explores the influence of mediating technologies on the use and experience of nature. Digital scholarship and digital media is a core component of his academic work and upcoming museum practice.

Pelle Snickars is Professor of Media and Communication Studies, specializing in digital humanities at Umeå University, with an affiliation to the research centre Humlab. His research has focused the relationship between old and new media, media economy, digitization of cultural heritage, media history as well as the importance of new technical infrastructures for the humanities.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: models, digitization, 3D, visualization, heritage

References

London Charter (2009) The London Charter for the Computer-Based Visualization of Cultural Heritage. http://www.londoncharter.org/downloads.html

William J. Turkel (2011) "Intervention: Hacking history, from analogue to digital and back again," Rethinking History 15(2), 287-296.

The Nordic Hub of DARIAH-EU: A DH Ecosystem of Cross-Disciplinary Approaches

Koraljka Golub
Marcelo Milrad
Linnaeus University, Sweden
Marianne Ping Huang
Aarhus University, Denmark
Mikko Tolonen
University of Helsinki, Finland
Andreas Bergsland
Norwegian University of Science and Technology
Mats Malm
University of Gothenburg, Sweden

Background and Motivation

The particular exploration of new ways of interactions between society and Information Communication Technologies (ICT) with a focus on the Humanities has the potential to become a key success factor for the values and competitiveness of the Nordic region, having in mind recent EU and regional political discussions in the field of Digital Humanities (European Commission, 2016; Vetenskapsrådet's Rådet för forskningens infrastrukturer, 2014). Digital Humanities (DH) is a diverse and still emerging field that lies at the intersection of ICT and Humanities, which is being continually formulated by scholars and practitioners in a range of disciplines (see, for example, Svensson & Goldberg, 2015; Gardiner & Musto, 2015; Schreibman, Siemens, & Unsworth, 2016). The following are examples of current areas of fields and topics: text-analytic techniques, categorization, data mining; Social Network Analysis (SNA) and bibliometrics; metadata and tagging; Geographic Information Systems (GIS); multimedia and interactive games; Music Information Retrieval (MIR); interactive visualization and media.

DARIAH-EU (http://dariah.eu), is Europe's largest initiative on DH, comprising over 300 researchers in 18 countries, thereby opening up opportunities for international collaboration and projects. Among the Nordic countries, Denmark is the full partner with four universities, Copenhagen, Aarhus, Aalborg and University of Southern Denmark (DARIAH-DK). Danish DARIAH-EU activities are facilitated by the national DH Infrastructure DIGHUMLAB, hosted at the DARIAH-DK coordinating institution, Aarhus University. Sweden's first academic institution, Linnaeus University, joined in May 2016 as a collaborative partner. Finland (University of Helsinki) and Norway (Norwegian University of Science and Technology) also became collaborative partners, in November 2016. The Nordic Hub of DARIAH-EU (DARIAH-Nordic) held its first meeting on 8 November in Växjö, Sweden, in connection with the International Symposium on Digital Humanities (Växjö, 7-8 November, https://lnu.se/en/research/conferences/international-digital-humanities-symposium/).

The Digital Humanities in the Nordic Countries (DHN) organisation was established in 2015 in order to create a venue for interaction and collaboration between the Nordic countries, including the Baltic countries. The ambitions behind the DHN initiative thus largely overlap with the recently formed Nordic Hub of DARIAH-EU. The panel would like to present different perspectives on Nordic contributions to DH as well as the aims of the DARIAH-Nordic and discuss possible joint opportunities and challenges in Nordic DH. With its tradition in supporting the Humanities research and development, Nordic countries may serve as a bastion for (Digital) Humanities. The Nordic Hub of DARIAH-EU and DHN may pave the way forward towards reaching that aim.

A DH Ecosystem of Cross-disciplinary Approaces

Mats Malm (previous chair of DHN) will present the visions and ambitions behind DHN and the recently established Centre for Digital Humanities at the University of Gothenburg, which will start a Master programme in Digital Humanities in the autumn of 2017. While both the Centre for Digital Humanities and DHN aim at broad inclusi-

veness, he will here focus on the use of textual databases for re-examining the history and cultural heritage of the Nordic countries. This implies collaboration on common textual resources and technologies for mining, at the same time as it raises a number of questions concerning cross-disciplinarity and exchange of perspectives and methods.

Mikko Tolonen will present the ongoing developments at the University of Helsinki (and in Finland) regarding Digital Humanities. This includes the recently launched Heldig (Digital Humanities Centre, https://www.helsinki.fi/en/researchgroups/helsinki-digital-humanities) and how it can relate to collaboration in DARIAH-EU. Tolonen will particularly discuss the relationship between the Digital Humanities infrastructure designed to be implemented at the University of Helsinki and how it relates to ongoing grassroot research projects.

Andreas Bergsland will discuss the role that the arts might play within Digital Humanities. As a starting point, he will take the work that has been done at the Norwegian University of Science and Technology (NTNU): establishing ARTEC, an interdisciplinary task force at the intersection of art and technology. He will argue how some of ARTEC's initiatives might have both opportunities and challenges partly converging with those of the DH field, but might also expand and enrich current practices. One such initiative, Adressaparken, is a commons area in Trondheim for exploration of sensor-based digital storytelling and an open arena for test and experimentation of new experiences and new digital media. While most DH initiatives in Europe seem to focus on computational humanities projects, Bergsland will explore the unique potential of integrating artistic and creative practices into DH/ARTEC initiatives at NTNU.

Koraljka Golub and Marcelo Milrad will present and analyse the cross-sector and cross-disciplinary Digital Humanities Initiative at Linnaeus University (LNU) along the axes of its strengths, weaknesses, opportunities and threats. Their long-term vision is to: 1) create a leading and innovative educational programme in this field; and, 2) to establish a prominent research regional centre that combines in novel ways already existing expertise from different departments and faculties working in close collaboration and co-creation with people and different organizations (both public and private sector) from the surrounding society. The main goals of this new initiative (launched in February 2016) at the first phase (12-15 months) are twofold; first, to establish the foundations for the creation of a DH educational programme and second, to carry out research and create an innovation centre at the wider region surrounding LNU, encompassing east southern Sweden. A combination of cross-disciplinary, cross-sector and international aspects would provide a solid ground to build a more or less unique international distance Master-level programme. Addressing future societal challenges would be eventually possible, 1) by highly skilled professionals whose education has been markedly enhanced by practice-informed education, and, 2) through joint, cross-sector innovation.

Marianne Ping Huang will present DARIAH-EU related activities in a Danish and European context, focusing on initiatives for cultural creative participation, including born digital cultural data and a presentation of open cross-sectoral innovation with DARIAH-EU Humanities at Scale (2015-2017). DARIAH-EU will set up its new Innovation Board in 2017 and host the first DARIAH-EU Innovation Forum with the Creativity World Forum in Aarhus, November 2017, intersecting with Aarhus European Capital of Culture 2017. DARIAH-EU's move towards digitally enhanced public humanities, closer collaboration with GLAM (Galleries, Libraries, Archives, Museums) institutions, and public-private innovation will be discussed in light of the scope of DH and the Nordic Hub of DARIAH-EU.

Discussion Points: Prospects and Challenges

The great breadth of cross-disciplinary and organizational initiatives presented above presents significant potential for DH in Nordic countries. Major opportunities lie in

the collaborative democratic tradition that supports re-combining already existing expertise and resources encompassing 1) different universities, 2) various disciplines, and 3) the wider community through input from related public and private sectors. These points serve to unite and consolidate already existing expertise in order to create new constellations for collaboration leading to new knowledge and products (expertise, education, research, public and relevant commercial services). Possibilities to collaborate across Nordic countries can take place at a number of levels, including joint research and innovation, education efforts, expertise and experience exchange, bringing in international views to address more regional challenges. Ensuing important value for the general public could be a (re)-affirmation of the value of humanities in particular, and academic practices in general.

Challenges would be discussed in terms of the emerging job market, the low number of students pursuing carriers in humanities at the Master level (e.g., in Sweden), and the fact that DH as a field is still in its infancy, leading to it being quite difficult to get funding and grants to carry out long-term research that sustain our efforts over time. Related to sustainability is the question on how to promote a dialogue and collaboration with potential industrial partners in order to run collaborative projects that go beyond just research. Not the least, epistemological, conceptual and terminological differences in approaches by the different disciplines and sectors may present further challenges and therefore may require additional resources to reach an understanding. Further, while there is a strong collaborative spirit across Nordic countries, there will certainly be administrative issues with cross-university collaboration as the current working structures are based on individual units.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: Nordic Hub of DARIAH-EU, digital humanities in the nordic countries (DHN), cross-disciplinariy initiatives, cross-sectoral initiatives

References

1) European Commission. (2016). Horizon 2020: Social Sciences & Humanities. Available at https://ec.europa.eu/programmes/horizon2020/en/area/social-sciences-humanities
2) Gardiner, E. and Musto, R. G. (2015). The Digital Humanities: A Primer for Students and Scholars. Cambridge: Cambridge University Press.
3) Schreibman, S., Siemens, R., and Unsworth, J. (2016). A New Companion to Digital Humanities. (2nd ed.). Malden, MA; Chichester, West Sussex, UK: Wiley-Blackwell.
4) Svensson, P., and Goldberg, D. T. (Eds.). (2015), Between Humanities and the Digital. Cambridge, Ma.: MIT Press.
5) Vetenskapsrådet's Rådet för forskningens infrastrukturer. (2014). Områdesöverikt för forskningens infrastrukturer. Available at http://www.vr.se/download/18.2302fa711489c4798d4a35fa/1411461229423/Samtliga+områden+infrastruktur.pdf

New Research on Digital Newspaper Collections

Patrik Lundell
University of Mid Sweden
Mikko Tolonen
Jani Marjanen
Hege Roivainen
University of Helsinki, Finland
Leo Lahti
KU Leuven, University of Turku, Finland
Asko Nivala
Heli Rantala
Hannu Salmi
University of Turku, Finland
Johan Jarlbrink
University of Umeå, Sweden
Kristoffer Laigaard Nielbo
Mads Rosendahl Thomesen
Aarhus University, Denmark
Melvin Wevers
Utrecht University, Netherlands

The proposed panel focuses on new studies on digitized newspaper collections in the Netherlands, Denmark, Sweden and Finland. The panel consists of six individual papers that highlight different aspects of using digitized newspapers for research. The panel focuses on discussing semantic change in newspapers as a response to political crises and to technological innovation. It will also scrutinize the direction of conceptual innovation and the role of text reuse in how ideas and concepts "traveled" between newspapers. Finally, the papers will reflect on the role of newspapers in the long-term transformations of public discourse and to which extent the availability of digital newspaper collections has affected the theory and practice of historical inquiry. Each short paper presents the methods and preliminary results in the respective cases. The discussion that follows concentrates on how to use digitized newspapers collections to study patterns in the newspaper texts as well as the outer ramifications of publishing. It will also compare the different methodological approaches and address how the different projects have addressed problems pertaining to OCR quality.

Online Newspaper Databases and Swedish History, 2009–2017 (Patrik Lundell)
This paper discusses aspects of the impact of online databases, in particular the newspaper databases of the Swedish National Library, on Swedish historical scholarship. How have Swedish historians reacted to the availability of these databases? One aim is to map out their actual use of these digital sources, also in relation to the use or non-use of other and complementary newspaper sources, since the first substantial newspaper database, Digitized Swedish Newspapers (Digitaliserade svenska dagstidningar), was launched in 2009. Another aim is to investigate explicit methodological considerations, or the lack of them, regarding this usage in terms of for example awareness of OCR related accuracy and the selection of primary sources on which the databases are built. A third aim is to reflect on potential problems and shortcomings, assuming from preliminary investigations that the impact is as profound as the meta-reflections are scarce. The empirical case will be doctoral dissertations in various historical disciplines published from 2009 until today.

Patterns of Public Discourse in Finland: Combining Meta-data from Library Catalogues and the Finnish Historical Newspaper Library (Mikko Tolonen, Jani Marjanen, Hege Roivainen, Leo Lahti)
The Finnish Newspaper Library made digitally available by the National Library of Finland contains nearly all the printed newspapers between 1771 and 1910. This paper uses the metadata information about the newspapers to statistically trace the expansion of public discourse in Finland during the long nineteenth century. Rather than using the metadata as a tool to find information and relevant papers, we use it as a tool to analyze the structural changes in public discourse. By relating information on publication places, language, number of issues, number of words, size of papers, and publishers and comparing that to the existing scholarship on newspaper history and censorship we aim at reaching an improved birds-eye view of newspaper publishing in

Finland after 1771. We then compare the results to our previous study that uses library catalogues from the Royal Library in Sweden and the National library in Finland to discuss the role of newspapers and books respectively in the public sphere. Finally, the paper addresses issues of representativity of the material, the need to clean up the existing data and potential shortcomings in the analysis due to missing data in catalogues or errors in the metadata.

Towards the Study of Text Reuse in Finnish Newspapers and Journals, 1771–1910 (Asko Nivala, Heli Rantala & Hannu Salmi)
The paper draws on the digitized collection of Finnish newspapers and journals. It includes all newspapers, published between 1771 and 1910 in Finland. This material covers as much as 1,951,076 pages, half of it in Swedish, the other half in Finnish. In addition, there are digitized journals, in sum 1,099,527 pages prior to 1910. We have been working with this material in the consortium Computational History and the Transformation of Public Discourse in Finland, 1640–1910 (funded by the Academy of Finland, 2016–2019), which is based on a co-operation between the Universities of Helsinki and Turku and the National Library of Finland. The Turku team concentrates on full text mining, especially on the question of text reuse. Prior to the Berne Convention in 1886, there was no effective copyright law to regulate the circulation of texts. Newspaper business took advantage of this situation, and news items and stories, poems and anecdotes were copied from paper to paper. Newspapers were active proponents and producers of culture: their content included a mixture of textualities, from advertisements to jokes, and they participated in formulating cultural influences, standardizing prevailing phenomena and establishing conventions for the modern era. The paper discusses the problem of text reuse detection, its particular challenges with Finnish material, and the preliminary results of the project.

From a Canon of the Extraordinary to an Archive of the Everyday (Johan Jarlbrink)
The newspapers from the nineteenth century has, until recent, formed a gigantic paper archive without an index. With so many newspapers and texts, and no possibility to search and get an overview, most scholars has focused on a limited number of canonized genres, events, writers and titles. Whereas the newspapers were characterized by repetitive and miscellaneous elements, the historical research has been dominated by the extraordinary. Digitized newspapers are far from perfect, but the digital archive makes it possible to search and research the everyday banalities in a systematic way. My aim in this paper is to discuss and describe the possibilities of digital newspaper archives and digital tools for text analysis. I will argue that the benefit of the digitized archive is less the possibility to find textual gold nuggets, but rather the opportunity to find pattterns in the great mass of less spectacular texts. To illustrate the discussion I will present a digital analysis of word co-occurrences in newspaper texts about new technologies in the mid-nineteenth century. The analysis show that very few texts describe the sensational possibilities of new technologies, often highlighted in previous research. What dominates the reports is bureaucratic banalities and technical details. When the trivialities of the everyday are taken more serious we will get a better understanding of the meaning of the new technologies in the historical context.

Semantic Disruptions in Public Discourse: Topical Divergence and Change-Point Detection in Two Centuries of Danish Newspapers (Kristoffer Laigaard Nielbo, Mads Rosendahl Thomesen)
Culture and media researchers theorize that negative cultural events (e.g., war, terrorism, migration) trigger semantic disruptions in the field of public discourse (i.e., create new meanings that displace established cultural values). Until recently, researchers have been using limited and biased samples when trying to model such semantic disruptions. With increased access to high performance computing and digitized media collections, it

has become possible for humanistic researchers to query and model millions of documents. We have therefore initiated a study that combines unsupervised statistical learning and information theory in order to model semantic disruption in Danish public discourse. The target data set, which serves as a discourse proxy, consists of 150 years of newspaper articles in the Danish Mediastream collection. In this talk we present research design, methods and our initial modeling based on simulated data.

Tracking the Consumption Junction – Long Range Dependencies and Predictive Causality in 20th Century Dutch Newspapers (Melvin Wevers, Kristoffer L. Nielbo)
Historian Roland Marchand argues that advertisements provide an insight into the social realities of the past, they inform us about the role consumer goods played in the lives of consumers. Marchand adds that the central purpose of an ad, however, is to sell merchandise, and herewith ads not only reflect but also shape society. This raises the question to what extent advertisements determined how consumers viewed products? Or should the process actually be conceptualized as a more complex interplay between advertisements and consumers? The negotiation of meaning between producers and consumers has been conceptualized as the consumption junction. In this study we analyze the interplay between advertisements and consumers by comparing the long-term dependencies in word use and predictive causal relationship between product terms in articles and advertisements in digitized newspapers. In the case of advertisements shaping society, one would expect the predictive causal direction to go from advertisements to articles. In the case of a more complex relationship, external factors determine the relationship between advertisements and articles. Results indicate that different scaling laws apply to advertisements and articles. Moreover, the ability of advertisements to shape society systematically appears to be dependent on product group.

Topics: Nordic Textual Resources and Practices
Keywords: newspapers, culture analytics, public discourse, historiography

Web Archives: What's in Them for Digital humanists? Panel on Web Archiving in the Nordic Countries

Caroline Nyvang
The National Library of Denmark
Lassi Lager
The National Library of Finland
John Erik Halse
The National Library of Norway
Olga Holownia
The British Library/IIPC
Pär Nilsson
The National Library of Sweden

Introduction
For the past 20 years the Nordic countries have been at the forefront of web preservation. The National Libraries of Denmark, Finland, Iceland, Norway and Sweden are among the founding members of the International Internet Preservation Consortium (IIPC) the objective of which is to acquire, preserve and make accessible knowledge and information from the Internet for future generations. The members of the IIPC have been working together on tools, techniques as well as standards that have enabled the creation of web archives. The Nordic countries have a long history of working together on technology development, techniques and methods for accessing archived web documents, the Nordic Web Archive (NWA), started in 2000, being one of their most successful initiatives that has underlined the value of cross-border collaborations.

While web archiving has been essential for experts working in the field of digital preservation, web archives are still an untapped resource for researchers, not least in the field of Digital Humanities. Projects and initiatives such as Buddah (Big UK Domain Data for the Arts and Humanities), "Archives Unleashed" datathons (organised in Ca-

nada and the US) or NetLab in Aarhus, to name just a few, show the importance of interdisciplinary collaborations between researchers and web archiving experts. A number of researchers have also worked with the Nordic National Libraries on projects based on web archiving but there is still a lot that can be done in that respect and our hope is that the web archiving panel at the DHN conference will lead to a better understanding of the state of arts as well as researchers' needs.

Therefore, the objectives of the panel are:

* to introduce the Nordic web archives: netarkivet.dk, Norsk Nettarkiv, Suomalainen verkkoarkisto, Kulturarw3 and Vefsafn.is;

* to highlight the value of web archives as a source for researchers;

* to discuss common platforms of collaborations as well as challenges posed by different legal frameworks and, consequently, different types of access;

* to discuss "new kinds of collaborations" between DH researchers and curators of online collections;

* to present use cases from the Nordic countries and beyond;

* to encourage DH researchers' feedback on the type of datasets and tools they would like to work with;

* "to compare the collections across borders".

Introduction to the National Libraries Web Archives

While all the Nordic National Libraries have collaborated on developing tools and platforms that can be used by their respective web archives, the main difference between them is related to legal frameworks and, consequently, availability and access.

Denmark: netarkivet.dk

Since 2005, the Royal Library and the State and University Library in Aarhus (in 2017, the two institutions will be merged into the Royal Library of Denmark) has been responsible for archiving the Danish Internet in accordance with the Legal Deposit law passed by parliament December 2004.

The Danish Netarkivet (netarchive) is constructed by following a four-string approach: 1) Four times a year, all -dk.domains are harvested; 2) Daily harvests of app. 100 select sites ensure that very dynamic websites (e.g. news sites) are properly archived; 3) We initiate special harvests in relation to both predictable and unforeseen events (e.g. the 2015 terror attack in Copenhagen and the Eurovision Song Contest); 4) Special harvests are planned at the specific request of researchers as well as the general public. Netarkivet is the second largest archive in Denmark and the fourth largest in the world, based on the amount of archived data. However, we also face a number of unique challenges in relation to making the collections readily available for researchers due to the strict Danish data protection laws. Furthermore, we are challenged by the fact that Netarkivet is not curated, which makes finding relevant data extremely difficult.

Based on recent Danish use cases – a study of online memorial sites and contemporary literature blogs – the presentation will explore how the Humanities can utilize archived web to formulate new research questions or, perhaps, revisit old ones.

Finland: Suomalainen verkkoarkisto

As a part of its legal deposit duties, the National Library of Finland annually collects a representative sample of webpages that have *.fi or *.ax domain names and are located in Finland, or contain subject matter that is targeted to the Finnish public. Many news sites (both open and paywall contents) are harvested daily. Domain crawls are supplemented by theme and event based special harvests. The collections of the past years include for example national elections (2008 – 2015), Eruption of Eyjafjallajökull volcano in Iceland (2010) and "European Refugee Crisis" (2015). Social media contents (Twitter, YouTube and Facebook) are harvested selectively, mostly as part of the thematic crawls.

The Finnish web archive was launched in 2006. By the end of the year 2016, the size of the web archive was about 120 TB. Archived Finnish web is preserved in the National Digital Library's Digital preservation

service. Access to the index of harvested URLs is open to anyone, but access to the archive itself is available only at special legal deposit workstations. The new user interface will be opened on spring 2017, with metadata of – and easy access to all of the theme and event based crawls and social media contents.

National Library has collaborated with DH researchers and research communities in language detection of web pages (other than *.fi- and *-ax -domains), in selecting seeds for some thematic crawls and mechanisms to collect social media contents. At the moment, mining of the web archive itself hasn't been possible due to copyright law and data protection regulation, but the library wants to make collaboration with research communities and other web archives to find best ways to make appropriate data sets of limited contents available for research use. Also, the library wants to get feedback what is essential for DH researchers for their current needs – and what would be essential for future researchers.

Iceland: Vefsafn.is

The Icelandic Web Archive (Vefsafn.is) contains all websites hosted on the Icelandic domain .is and many web sites hosted elsewhere that are in Icelandic or refer directly to matters of interest to Iceland. Access to the complete Web Archive is open to the world except for web sites where the user must pay for access and web sites that for some reason are closed by the owner's request.

The .is domain has been harvested by the National and University Library of Iceland since October 2004 and the policy is to harvest the complete .is domain three times a year. In addition selected web sites are harvested at least weekly and for national events like elections relevant web sites are harvested. Additionally, material from the Internet Archive, covering .is from 1996-2004 is available in the archive.

The researchers from the University of Iceland have used Vefsafn for projects analysing linguistic corpora. Given that this is an open access archive, certainly more work can be done with the available material.

Norway: Norsk Nettarkiv

The National Library of Norway (NLN) started harvesting the Norwegian top level domain (.no) on a yearly basis in 2001. In 2008, however, the National Library had to stop full domain harvests since the Norwegian Data Inspectorate questioned the legal basis for this practice. From 2008 until now the National Library are harvesting between 500–2500 subdomains under .no, after informing the website owners in writing first.

A revised version of the Norwegian Act on Legal Deposit came into force 1st January 2016 and it enables the NLN to do full domain harvests of the Norwegian top level domain (.no), as well as to collect websites outside the .no-domain that are either owned by Norwegian institutions or individuals, or adapted to Norwegian users. Importantly, the revised law also makes it possible for the National Library to make the web archive available for research and documentation purposes. Parts of the archives will be open to the public.

NLN currently works to establish a more up to date solution for the web harvesting activity, that are flexible and that scales to handle the full Norwegian top domain.

In terms of harvesting content, different approaches have been followed since 2001: 1) selective harvesting of web sites 2001-2004 and from 2009; 2) domain crawls once or twice a year since 2002; 3) event harvesting since 2001, for events of national interest, such as general and local elections, royal weddings etc. At present the archive is growing with about 175 TB a year. It is the aim of the Norsk Nettarkiv to make part of the archive available in open access which will certainly create new research possibilities.

Although access to the archive has limitations, we see opportunities to collaborate with researchers in making derivative datasets available. We have one such project going now where we analyze the proportion of use of the two Norwegian languages (bok-

mål and nynorsk) to assist the Norwegian Language Council in their work.

SWEDEN: Kulturarw3

The National Library of Sweden (Kungliga biblioteket) started to harvest the web in 1997 under the project Kulturarw³. Today there are more than 1.7 billion items corresponding to approximately 72 terabytes of data. One part of the archive consists of bulk harvesting of the Swedish web. The collection includes both web servers located under the Swedish top level domain "se" and servers located elsewhere. This second part is identified as Swedish using geolocation. Harvesting is done roughly twice a year. A second collection comprises about 140 newspapers with a daily issue. These are harvested every day. The archive is open to everybody but only within the library. Kulturarw3 - The Web Archive of the National Library of Sweden.

Topics: Nordic Textual Resources and Practices

Keywords: web archiving, digital preservation, digital born content, collaboration, tools development, curation, use cases

LONG PAPERS

Body Parts in Norwegian Books

Lars Bagøien Johnsen
Siv Frøydis Berg
National Library of Norway

Introduction

In this presentation we will discuss questions like how the human body is represented in literary works, and if there is a difference between fiction and nonfiction. We will also be looking at differences between authors, gender and time. For example, do literary works agree on the most frequent body parts?

One of the themes of this conference is visualization, and we will demonstrate how augmented tabular data can be used as a tool in discovering patterns, and formulating hypotheses about them, which should be of interest to scholars in the humanities. Another theme texts, and features of texts, which we use throughout this investigation.

Although the body in itself, in biological terms, hasn't changed much in the recent history of humans, its presentation and focus of particular details differs to a certain extent across genre, time periods and authors. A discussion of how the body is used in modern culture is found in e.g. Christopher E Forth; Ivan Crozier (2005).

We report on a pilot study that considers body references in fiction and compare it with nonfiction in the form of encyclopedias. One specific hypothesis to be tested is if there is a difference between fiction and nonfiction in their frequency of references to body parts using possessive pronouns like *his* and *her*.

Method and data

All the material is made available through the Norwegian National Library as a result of its digitization effort. Some part of the investigations uses feature sets of books that already is, or will be, made available to the research community, so that the questions and results reported here can be replicated and done on different selections.

Of particular interest is trends in Norwegian books during the period 1810 up to 2000 using the library's n-gram viewer with a focus on the 20th century, while smaller selections of authors from the middle and later part of the century is compared.

The set of body words we consider is a list of approximately 35 different body parts ranging from head to toe, in singular and plural forms. In the present pilot study references to genitals, intestines or bodily fluids are not taken into account, words related to those parts of the body are reserved for future studies. These words are of particular interest when considering e.g. medical literature in comparison to fiction and religious writings (e.g. Forth & Crozier op.cit.).

Nouns expressing body parts are counted as they are, in addition to counts in the context of a possessive pronoun. Possessive constructions in Norwegian differ from Swedish and Danish in that in addition to "hans arm" (his arm) the possessor can be positioned behind a definite version of the noun "armen hans" (literally the arm his). However, modern Norwegian seems to prefer the latter construction for body part possession.

Each count is connected to metadata. In the case of the n-gram viewer, it is Norwegian books across all genres, and in the case of authors the counts are linked to the author, which are studied in two groups, a set of authors from the middle of the century and a set from the later part. For encyclopedias the counts are connected to each encyclopedia in turn. This gives a collection of tables which we show in the next section.

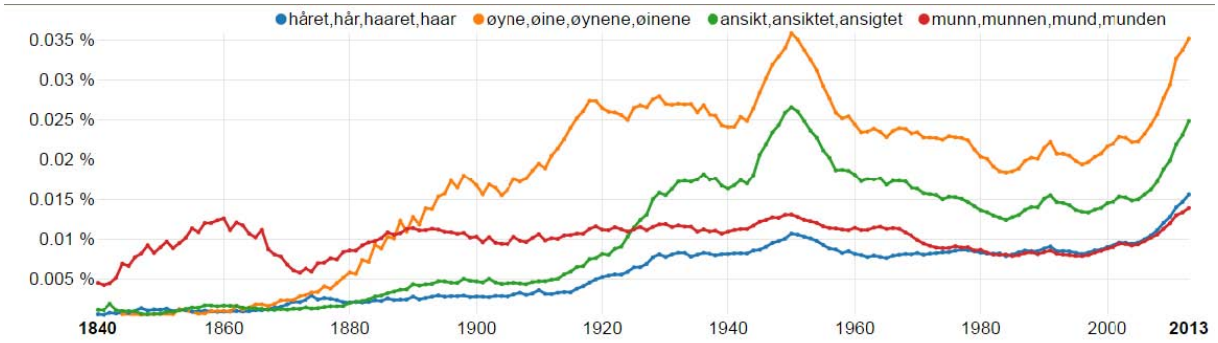

Figure 1. For an interactive graph, see: https://www.nb.no/sp_tjenester/beta/ngram_1/

In addition to metadata, body words are linked to other words in collocations, which are constructed from concordances with a size of 6 preceding and 6 following words. These concordances form a basis for creating so called collocational word clusters. For that purpose we use the standard PMI measure (Pointwise Mutual Information) which here is computed by comparing the frequency of a word in the concordances with its total frequency within the whole book collection. The PMI value is also weighted with the logarithm of the absolute frequency within a concordance set, which will penalize rare words, especially those that come from OCR-errors, while still not giving too much weight to high frequency words. For the methods described here, see e.g. Lewandowska-Tomaszczyk (2007), Romesburg (2004). These clusters give us a clue as to what activity and quality a certain body part is associated with, and are constructed from books across the whole digitized collection.

Results

Here we show some of the results, and start off with a view from the n-gram viewer which displays trends of n-grams up to three. For parts of the top of the body, consider the following graph (Figure 1), which shows a rising trend starting around the middle of the 20th century for some of the body parts. Each part is shown as a summation of its different morphological and spelling variants. Note that "munn" (mouth) shows a relatively stable frequency (The n-gram viewer takes care of capitalized nouns, which are typical for 19th century Norwegian), while "øyne" (eyes), "ansikt" (face) and "hår" (hair) have a clear increase.

	Hjort	Jacobsen	Lindell	Nesbø	Ragde	Sandemo	Solstad	Uri	Total
hodet	897	608	2106	1808	1312	1513	447	368	9059
øynene	718	596	1535	1348	1500	2208	414	391	8710
ansiktet	461	497	1247	844	976	1534	282	207	6048
hånden	655	377	1399	906	196	1587	140	240	5500
hendene	514	543	1049	533	889	1038	335	187	5088
munnen	476	194	839	565	697	494	205	283	3753

Figure 2.

	Hjort	Jacobsen	Lindell	Nesbø	Ragde	Sandemo	Solstad	Uri	Total
ansiktet hennes	55	43	150	82	77	163	17	32	619
ansiktet hans	38	54	98	113	114	279	18	14	728

	Hjort	Jacobsen	Lindell	Nesbø	Ragde	Sandemo	Solstad	Uri	Total
håret hennes	33	15	86	38	55	62	6	45	340
håret hans	5	12	34	11	26	38	1	7	134

Figure 3 and 4.

While trends tell us a little bit, there may be variation behind it. For that purpose, we place it within the context of a selection of writers, starting with the modern set. Using the tables of counts for each writer, and sorted on the total, the results look like this, where the columns are labeled by surnames (Figure 2, see previous page).

The table is augmented with a gradient color along the columns. One result that can be read off this table is that writers do share the same top reference to body parts, "hodet" (head) as container, "eyes" and "ansikt" for expressions, and "hand" for action. Note that "hodet" (head) is either first or second with all.

Now, since body parts also where counted in the context of possessive pronouns, we can use the gradient table to show differences between writers and differences within the possessive construction. The following table displays how "ansiktet" (face) differs between the masculine and feminine possessor. The table also suggest that there is a connection to gender of writer, but if that connection is meaningful requires further study (Figure 3).

However, one result that does look solid is the connection between "hår" and the gender of the possessor (Figure 4).

All have "håret hennes" ranked well above "håret hans", except for Roy Jacob-

	Bjerke	Bjørneboe	Borgen	Børsum	Evensmo	Hoel	Hofmo	Holt	Jensen	Mykle	Total
magen	2	1	31	25	8	5	0	192	27	9	300
brystene	1	74	12	2	33	21	0	43	47	17	250
hoftene	3	36	7	5	8	16	0	44	19	16	154
hoften	6	14	7	1	4	5	0	1	9	24	71
navlen	0	20	3	0	1	0	0	0	3	8	35
midjen	0	8	0	0	0	0	0	0	0	16	24
rumpen	1	0	1	0	0	2	0	0	0	15	19

Figure 5.

sen, which have the counts close to each other.

The group of writers from the middle of the 20th century show similar top six in body parts. This group also contains Agnar Mykle who was prosecuted for the novel "Sangen om den røde rubin" (http://www.arkivverket.no/arkivverket/Arkivverket/Riksarkivet/Nettutstillinger/ Skatter-fra-arkivet/Sangenom-den-roede-rubin). What can the table of counts tell us about the difference between Mykle and other writers of the same era? In order to answer this, we look at words from the torso, considering a relatively small set as seen in the table (Figure 5, see previous page).

As is readily seen, Mykle do differ from the others in key respects, his top words are "hofte" (hip) and "brystene" (breasts) which he shares more or less with most of the others, however, as we see from the table, Mykle also moves down the torso to "midjen" (the waist) and "rumpen" (the butt), and these two sets him apart from the rest.

For the encyclopedias, our hypothesis that these contain only references to nouns without pronominal possessor appears to hold up.

Topics: Nordic Textual Resources and Practices, Visual and Multisensory Representations of Past and Present
Keywords: body, literature, genre, gender

References

Barbara Lewandowska-Tomaszczyk (2007) *Corpus linguistics, computer tools, and applications : state of the art* P.Lang Frankfurt am Main, New York

H Charles Romesburg (2004) *Cluster analysis for researchers*, Lulu Pr.

Christopher E Forth; Ivan Crozier eds. (2005) *Body parts : critical explorations in corporeality* Lanham : Lexington Books.

Confusing the Modern Breakthrough: Naïve Bayes Classification of Authors and Works

Peter M Broadwell
UCLA Library, United States of America
Timothy R Tangherlini
UCLA, United States of America

The Modern Breakthrough is widely considered to be one of the most important turning points in late nineteenth century Nordic literature, ushering in a period of literary experimentation predicated on a pivot toward naturalism. Georg Brandes's iconic work, Det moderne gennembruds mænd (1883), provides a literary historical framework for the consideration of the movement, outlining in broad strokes the contours of this shift in literature and, in the portraits of a series of featured male authors, offering a touchstone for broader understanding of this movement. In 1983, Pil Dahlerup offered a corrective to Brandes's work with Det moderne gennembruds kvinder. Here, Dahlerup surfaces the numerous female authors who were writing groundbreaking work in the shadows of the male-dominated literary world. A great deal of scholarship on the Modern Breakthrough considers the rich network of literary cross-influence that characterized the period. Influence, however, is a complex phenomenon and one that is hard to formalize. In the following work, we propose to explore the related phenomenon of similarity, predicated on the notion that the most sincere form of flattery is imitation. To what extent do writers from this period share aspects of language? Can we capture this sharing in a useful manner?

In earlier work, Leonard and Tangherlini (2013) showed how probabilistic topic modeling could be deployed to help discover similarities across the works of male and female authors of the period. Working at the level of the passage, they used a topic model of male Modern Breakthrough authors to identify passages from a large, poorly labeled corpus that exhibited topical similarity. In this case, the corpus consisted of all of the works in

Google Books written in Danish until 1923. With this approach, they were able to confirm Dahlerup's identification of numerous female Modern Breakthrough authors.

In this work, we focus specifically on the authors identified by Brandes and Dahlerup as Modern Breakthrough authors, with their works constituting a well-defined corpus. Extending work by Broadwell, Mimno and Tangherlini on the classification of folk legends (2016), we develop a "hold one out" Naïve Bayes classifier trained on the machine actionable works of the authors in our corpus, and also run standard text-similarity calculations on the corpus, including LDA topic inference and cosine similarity based on TF-IDF scores for unigrams, bigrams, and trigrams.

The multilingual nature of the corpus raises numerous problems that we are unable to address in a sophisticated manner with this work. To get around the problem of curating a single-language representation of the Modern Breakthrough that would likely miss a great deal of interesting overlap, we have chosen to consider Danish works as well as Danish translations of Swedish and Norwegian works. Similarly, to avoid problems of classifier failure based on orthographic differences, we have normalized the Danish in these works to comply with the orthographic conventions of 1948 (Hartvig Frisch).

To make its results useful for literary scholars, we present our analysis at two levels of aggregation. On the first level, we aggregate all of the works of a particular author into a single grouping. On the second level, each work (e.g., a novel) is a single grouping. Consequently, users can explore the varying levels of overlap between all authors in the corpus and among all works in the corpus. This significantly complicates the previous binary of male-female authors, and allows for various alternative groupings of authors.

During our analysis, each machine-actionable work is chunked into 500-word passages after applying basic orthographic normalization. We then run the passages in groupings as described above through the Naïve Bayes classifier and text similarity calculations. Instances of classification "confusion" – where the classifier "fails" in assigning all passages to their original grouping – suggest significant overlaps in style and content within or between authors' oeuvres, which we compare to the output from the text similarity calculations. Such comparisons enact a fundamental principle of the "macroscope" as introduced by Katy Börner (2011) and extended to the humanities by Tangherlini (2013), namely the greater degree of insight made available when one can switch rapidly between multiple analytical perspectives on complex cultural phenomena. A related "macroscopic" feature of our analysis is the ability of the Naïve Bayes classifier interface to "drill down" to investigate a specific passage and even view the words that were most influential in assigning it to a category other than its original category.

We visualize the alternate classifications of the authors' works via an interactive confusion matrix, in which the sizes of the dots drawn on the cells of the matrix indicate the number of passages with the actual "label" (author or author+work) in the same row on the vertical axis that were assigned by the classifier to the proposed label at the same column on the horizontal axis (Figure 1, at the end of the text). For instance, a passage from Herman Bang's Ved Vejen may be properly classified by the NB classifier as a Herman Bang passage, or it may be assigned to another author in the corpus. The strong diagonal of blue circles that emerges in these visualizations represents those passages that the NB classifier has placed into the expected category. The presence of red dots off the main diagonal indicates where passages have "confused" the classifier. By selecting a red dot in the matrix, the user is taken to a list of passages labeled in one manner and classified in another manner. A list of words associated with the classifications at the top of the page allows one to understand the words that the classifier finds significant, from the original label at the beginning of the list to the new label at the end of the list (Figure 2, at the end of the text).

The results of the text cosine and LDA topic similarity comparisons are visualized via a similarity matrix, analogous in format

to the confusion matrix, with the degree of shading in each cell x,y indicating the similarity of the full texts associated with column x and row y (Figure 3, at the end of the text). Such matrices can also be converted to distance plots where points (representing texts) are placed closer together when they are more similar, although the two-dimensional nature of the plots can obscure important relationships (Figure 4, the end of the text).

Initial experiments with a subset of the Modern Breakthrough authors indicate a low degree of inter-author similarity and confusion, with the F-score of the Naïve Bayes classifier averaging over 95 % for each author "label" when only the works' authors are considered. Further dividing each author's output into his or her individual works yields more intriguing results, showing a considerable degree of classifier confusion and overall text similarity among certain authors' works (primarily Bang and Pontoppidan in our sample), but generally quite low confusion between the works of different authors. Visualizations of the text similarity matrices echo these results.

In further work, we plan to use moments of "misclassification" and overlap between authors and within the works of a single author to develop a better understanding of stylistic similarity and possible influence among the authors of the Modern Breakthrough. In particular, incorporating a temporal dimension into these analyses may help to estimate authorial influence by determining whether the classificatory "confusion" of a given author's texts favors works by the authors that are considered to have influenced them. Alternately, such an analysis can suggest instances of text similarity and potential influence that extend or even contradict accepted narratives of Nordic literary history.

Topics: Nordic Textual Resources and Practices
Keywords: modern breakthrough, classification, influence, text similarity

Works cited

Börner, Katy. 2011. "Plug and Play Macroscopes." Communications of the ACM 54 (3): 60-69.

Brandes, Georg. 1883. Det moderne gennembruds mænd. København: Gyldendal.

Broadwell, Peter, David Mimno and Timothy R. Tangherlini. 2016. "The Telltale Hat: Surfacing the Uncertainty in Folklore Classification." Journal of Cultural Analytics 1 (2). In process.

Dahlerup, Pil. 1983. Det moderne gennembruds kvinder. København: Gyldendal.

Leonard, Peter and Timothy R. Tangherlini. 2013. "Trawling in the Sea of the Great Unread: Sub-Corpus Topic Modeling and Humanities Research." Poetics 41 (6): 725-749.

Tangherlini, Timothy R. 2013. "The Folklore Macroscope: Challenges for a Computational Folkloristics." The 34th Archer Taylor Memorial Lecture. Western Folklore 72 (1): 7-27.

Bibliography

Broadwell, Peter and Timothy R. Tangherlini. 2016. "GhostScope: Conceptual Mapping of Supernatural Phenomena in a Large Folklore Corpus." In Maths Meets Myths: Quantitative Approaches to Ancient Narratives, edited by Raph Kenna, Máirín MacCarron, and Pádraig MacCarron, 131-157. Cham, Switzerland: Springer International.

Broadwell, Peter, David Mimno and Timothy R. Tangherlini. 2016. "The Telltale Hat: Surfacing the Uncertainty in Folklore Classification." Journal of Cultural Analytics 1 (2). In process.

Broadwell, Peter, Timothy R. Tangherlini and Hyun Kyong Hannah Chang. 2016. "Online Knowledge Bases and Cultural Technology: Analyzing Production Networks in Korean Popular Music." In Series on Digital Humanities 7, 369-394. Taipei: National Taiwan University Press. In process.

Broadwell, Peter and Timothy R. Tangherlini. 2016. "WitchHunter: Tools for the Geo-Semantic Exploration of a Danish Folklore Corpus." Journal of American Folklore 511 (Winter 2016): 14-42.

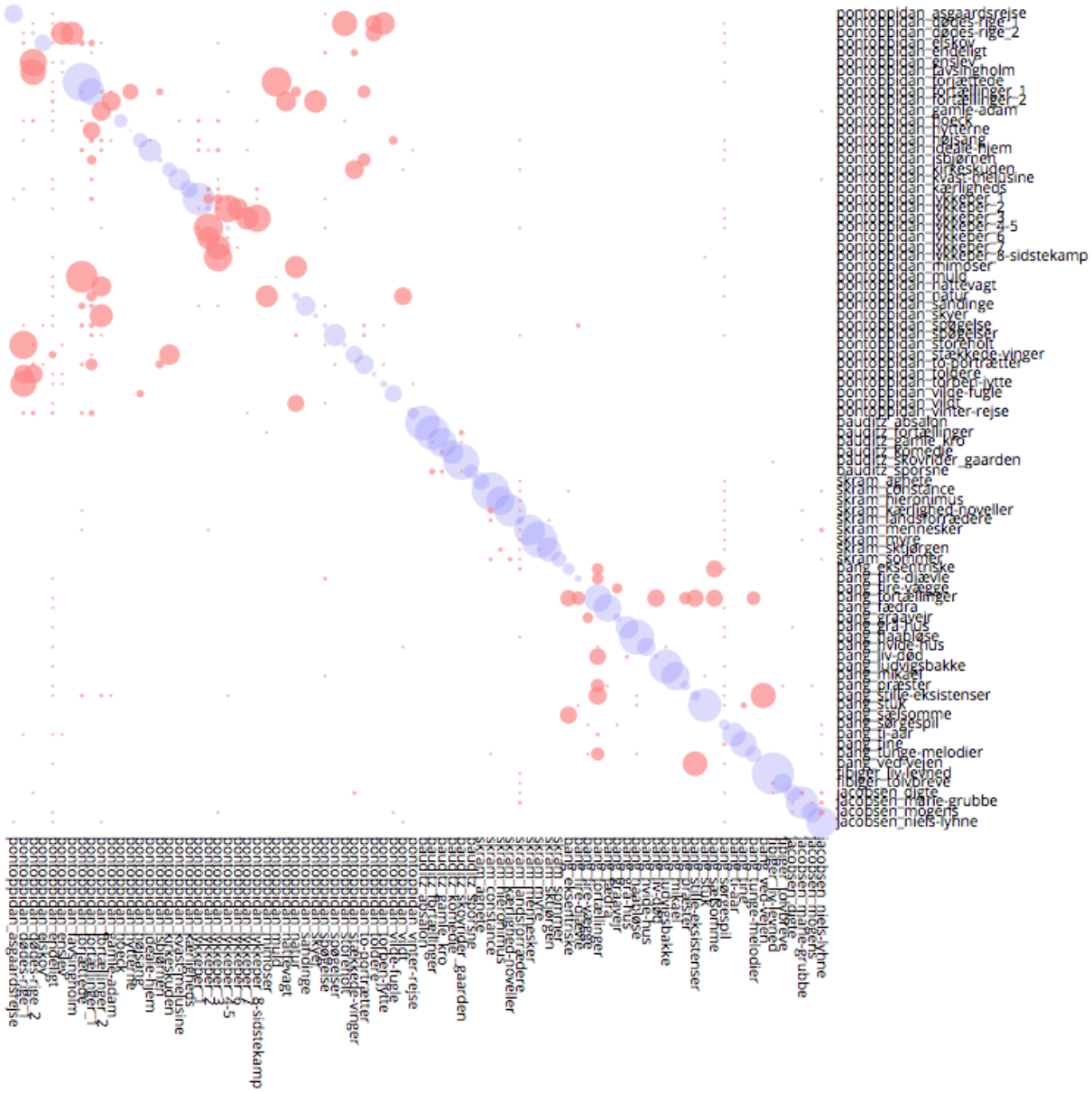

Figure 1. *Confusion matrix of a subset of Danish-language works by authors from the Modern Breakthrough, with the size of the dots indicating the number of passages from each author and work om the horizontal axis that were categorized by a Naïve Bayes classifier as belonging to the author and work on the corresponding row of the vertical axis.*

Figure 2. *Detailed "drill-down" view of the text passages from a single work (Bang's novella* Stille Eksistenser *from 1886) identified by the Naïve Bayes classifier as most likely belonging to Bang's novel* Ved Vejen *(also written in 1886). The color-coding of the words indicates that the classifier considered reddish words to be more closely associated with* Stille Eksistenser, *while the blue-tinted words are more closely related to* Ved Vejen.

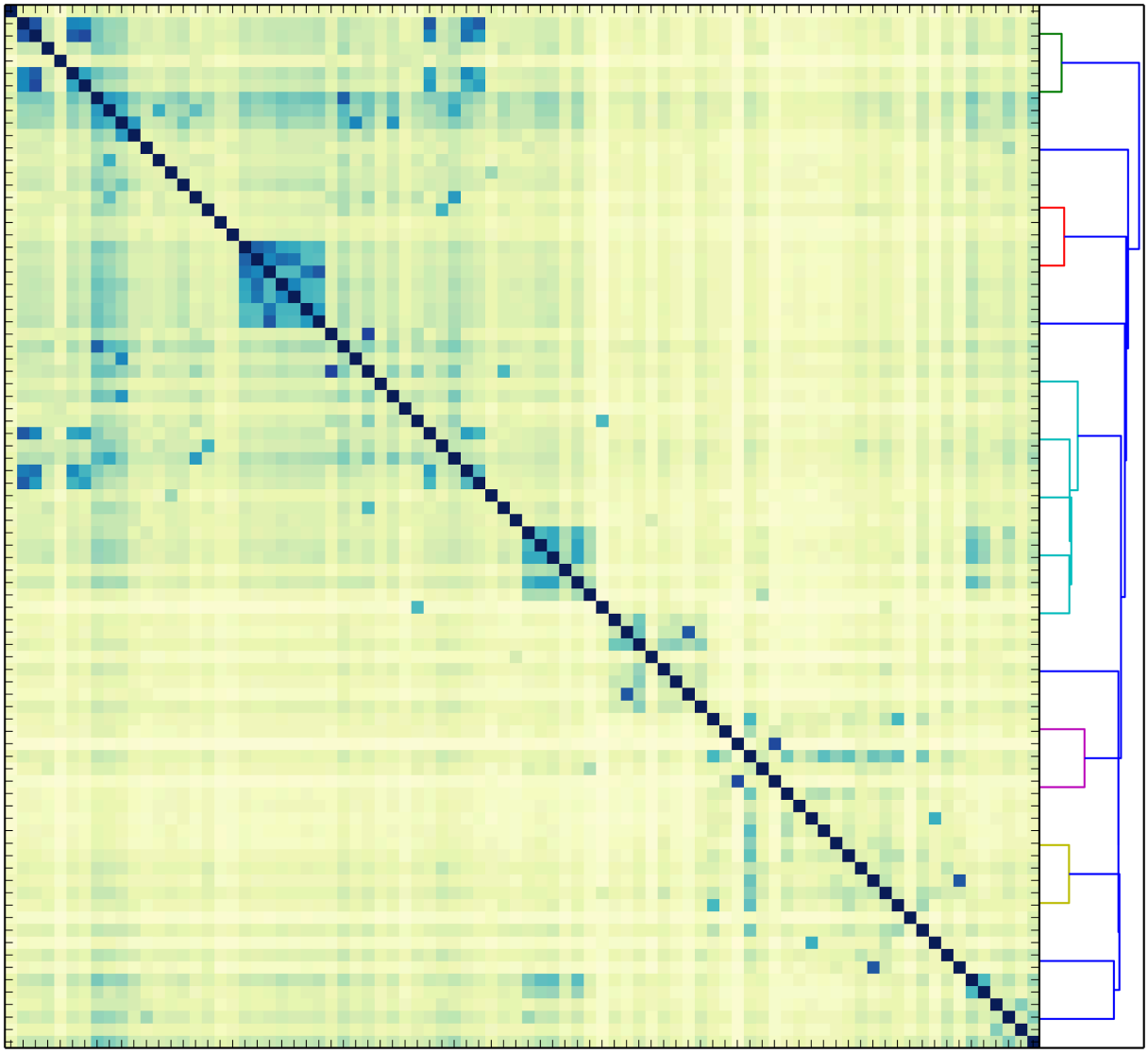

Figure 3. *A text similarity matrix for the same works as in Figure 1, based on the cosine similarity of the TF-IDF weights of the unigrams, bigrams, and trigrams in each work. Note the resemblance to* Figure 1.

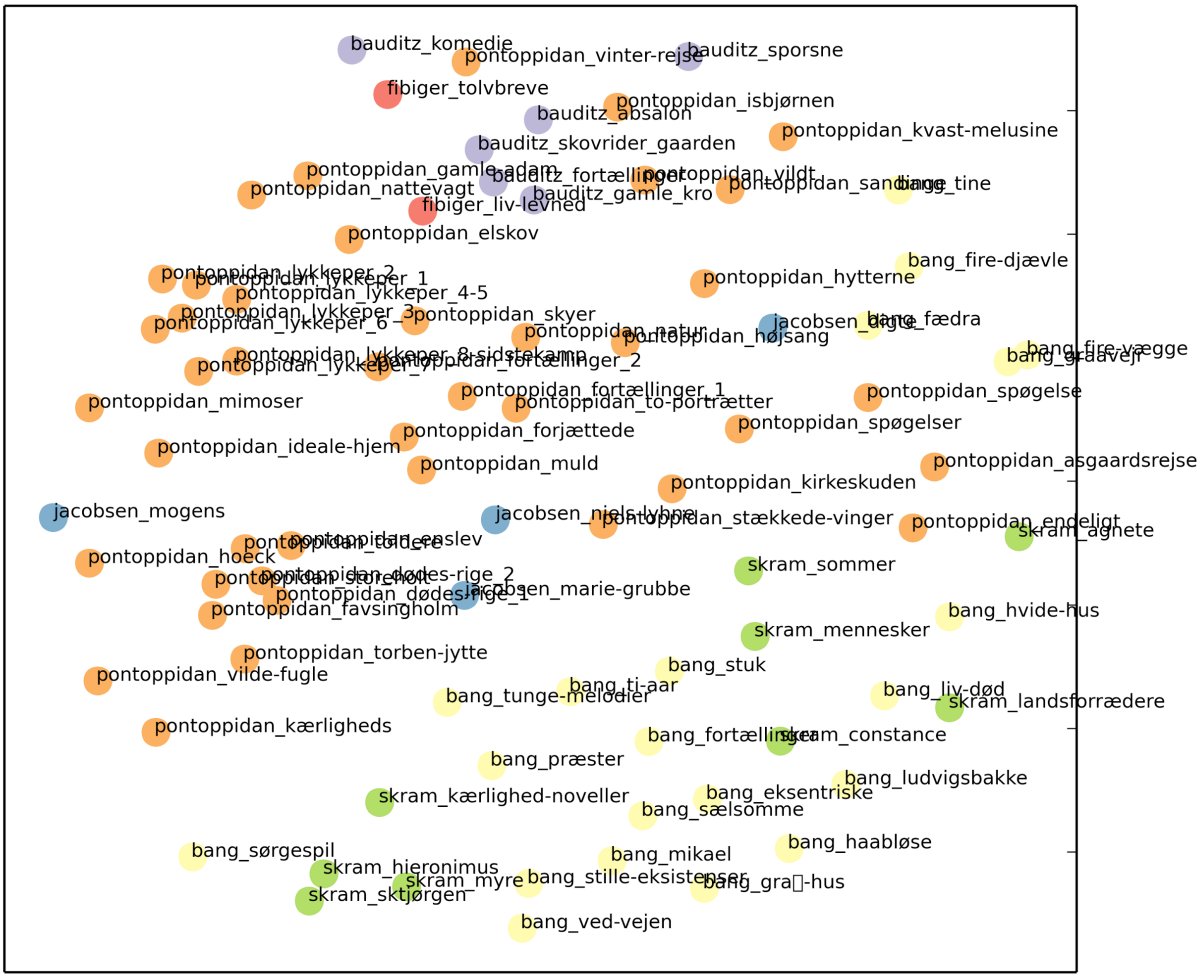

Figure 4. *A text clustering plot of the works from* Figures 1 and 3. *The distance between the points (works) is indicative of their textual similarity as calculated for the similarity matrix shown in* Figure 3.

Topical Discourse Networks: Methodological Approaches to Turkish Foreign Policy in Sub-Saharan Africa

Fabian Brinkmann
Ruhr-University Bochum, Germany

The Republic of Turkey undoubtedly changed in 2002 when the *Adalet ve Kalkinma Partisi* (Justice and Development Party, AKP) came into power after a period of economical and political struggle. With the end of the Cold War the Turkish Republic had been in a geopolitical situation, which had an enormous impact on its self-perception and its international relationships. In a new geopolitical environment Turkey was searching for a new interpretation of its political role after it has lost the geopolitical standing it had during the Cold War. The search for a new geopolitical role should hugely influence the foreign policy of the Turkish Republic in the coming years and the foreign policy of Turkey changed under the new circumstances. Developing the concept of 'Strategic Depth', which located Turkey at the centre of a 'Eurasian-African landmass', Turkish foreign policy became increasingly diversified, although whether this has been a dramatic reorientation or just a series of gradual shifts still remains a subject of debate (See for example Bagdonas 2012).

In this context, A lot has been written about the Turkish interests in the Balkans, the Caucasus, Central Asia and the Middle East, but still the political and economic interests of Turkey in Sub-Saharan Africa remain an underdeveloped part of the discussions about Turkish foreign policy. Against the background of these political developments a closer look at the argumentations, rhetorics and discourses behind the Turkish dedication towards Sub-Saharan Africa seems in order to provide an encompassing overview about the Turkish foreign policy towards Africa.

Naturally, the different actors of this particular political field cannot be ignored in this context since they are themselves dominant figures within the discursive formations. Among the most relevant actors in the field of Turkish foreign policy towards Sub-Saharan Africa are certainly the Turkish foreign aid agency TI KA and the Turkish bureau of religious affairs Diyanet, which both are increasingly active in Africa (See for example Ali 2011). For the Presidency for Turks Abroad and Related Communities (YTB) there also can be identified first approaches towards a more structured engagement with the African continent (Öktem 2014). Besides these state-political actors non-state actors will have to be taken into account as well. In this context the African engagements of Turkish NGOs, for example in the field of development aid, will have to be named. In a similar way the Turkish economic organizations (for an overview see Seufert 2012) and the Think Tanks that are both part of the first approaches towards cooperations with African countries (Uchehara 2008) are important players in this discursive field. Thus, the networks between the different state and non-state actors will have to be considered to provide an overview of the structures of the Turkish foreign policy discourse about Sub-Saharan Africa.

Based on ongoing research this paper will present how the Structural Topic Model, a R package developed by Roberts et al (2011, 2013, 2014, 2015, available online at structuraltopicmodel.com), can be used to uncover discursive structures and (discourse) networks of actor. It will describe the attempt to untangle the different political, economic, anticolonial, religious, historical and cultural discourses across intertwined actors via the mass data approach of Topic Modeling across different covariates in a diachronic and synchronic way.

Structural Topic Modeling (or STM) enables the researches to do additional things compared to other topic modeling approaches It allows for the inclusion of metadata (i.e. date, actor, etc.) in the model. This can be done in two ways: topical prevalence and topical content. Topical prevalence allows us to look at the influence of the metadata on

the frequency of a topic (i.e. "Is a topic discussed more in one year as compared to another), while topical content allow us to observe how a particular topic is discussed (i.e. does a specific actor use different words than another). It can also be used to uncover the latent topic structures of documents. Thus, it can show how specific topics show up closely tied to other topics in a given corpus. Using these three aspects of STM it becomes possible to discover diachronic changes in topics and across actors as well as the topic structuring in the discourse of the examined documents. (Wang 2011) The inclusion of metadata allows the researcher to build new and expanded question into his research and make better inferences about relevant issues in the data of the corpus.

This mostly theoretical paper will show how these aspects of Structural Topic Modeling can be operationalized for an encompassing discourse analysis of actors in intertwined networks. It will show methodological-theoretical approaches derived from Critical Discourse Analysis, especially the concept of discourse strands (thematically consistent trends of discourse, which regularly appear in an overall societal discourse) by Siegfried Jäger (M. Jäger/S. Jäger 2007: 25) and the Discourse-Historical Approach of the Vienna School of Critical Discourse Analysis by Ruth Wodak. It aims at showing on a methodological-theoretical level the opportunities of STM can be used to identify changes, similarities and differences between these actors in a discursive network.

Based on a solely Turkish corpus of all available documents, texts and utterances (such as for example press releases, activity reports, journals, speeches, etc.) published by the actors of this policy field, which have already been laid out above, the presented project aims at two things: a) uncovering the discursive macrostructures through the use of Topic Modeling, and b) investigating the discursive networks between different political actors by taking into account various document metadata.

Topic Modeling provides a way of 'distant reading' of documents uncovering topics and topical structures (Bauer/ Fridlund 2013; Mimno 2012). In discourse analysis the term 'topic' has also been used by some scholars. Thus, Haslinger (2006) points out that topics could be understood as complexes of meaning that are talked about with different opinions. He does this in order to define specific discursive processes other than 'discussion' or 'debate'. He argues that topics are the foundation of the structure of every form of communication and thus have to be a part of discourse analysis. Thus, this paper will try to show the methodological possibilities Structural Topic Modeling gives the researcher to undertake a structured analysis of discursive networks, debates and discussions across a multitude of intertwined actors.

All this will be done with the already laid out example of Turkish foreign policy in Sub-Saharan Africa. These discourses are of particular interest, since the Sub-Saharan space is a new field of Turkish foreign policy and the political and societal discourses that support these new developments have only been just developed in Turkish politics. Thus, different actors vie for discursive influence and power over this concrete policy field. Topic Modeling is able to uncover these discursive conflicts and differences.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: Turkey, foreign policy, topic modeling, network analysis, discourse analysis

Bibliography (selection)

Ali, Abdirahman: Turkey's Foray into Africa. A New Humanitarian Power?, in: Insight Turkey 13/4 (2011), S. 665-73.

Bagdonas, Özlem Demirtaş: A Shift of Axis in Turkish Foreign Policy or a Marketing Strategy? Turkey's Uses of its ‚Uniqueness' vis-à-vis the West/Europe, in: Turkish Journal of Politics, 3/2 (2012), S. 111-132.

Brauer; René/Fridlund, Mats: Historizing Topic Models. A Distant Reading of Topic Modeling Texts within Historical Studies, in: Cultural Research in the Context of 'Digital Humanities'. Proceedings of International Conference 3-

5 October 2013, St. Petersburg 2013, S. 152-163.
Haslinger, Peter: Diskurs, Sprache, Zeit, Idendität. Plädoyer für eine erweiterte Diskursgeschichte, in: Eder, Frank X. [ed.]: Historische Diskursanalysen. Genealogie, Theorie, Anwendungen, Wiesbaden 2006, S. 27-50.
Jäger, Margarete/ äger, Siegfried: Deutungskämpfe. Theorie und Praxis Kritischer Diskursanalyse, Wiesbaden 2007.
Mimno, David: Computational Historiography. Data-Mining in a Century of Classic Journals, in: ACM Journal of Computing in Cultural Heritage 5/1 (2012), S. 3:1-3:19.
Öktem, Kerem: Turkey's New Diaspora Policy. The Challenge of Inclusivity, Outreach and Capacity (Istanbul Policy Research Paper), Istanbul 2014.
Roberts, Margaret E. et al.: stm: R Package for Structural Topic Models (Working Paper), 2011. (Available online under https://github.com/bstewart/stm/blob/master/vignettes/stmVignette.pdf?raw=true).
Roberts, Margaret E. et al..: The Structural Topic Modek and Applied Social Science, presented on: Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation, 2013. (Available online under http://scholar.harvard.edu/files/bstewart/files/stmnips2013.pdf).
Roberts, Margaret E. et al.: Navigating the Local Modes of Big Data. The Case of Topic Models, presented on: Data Analytics in Social Science, Government and Industry, New York, Im Erscheinen. (Available online under http://scholar.harvard.edu/files/dtingley/files/multimod.pdf).
Roberts, Margaret E. et al.: Structural Topic Models for Open-Ended Survey Responses, in: American Journal of Political Science 58/4 (2014), S. 1064-1082.
Roberts, Margaret E. et al.: A Model of Text for experimentation in social sciences (Working Paper), 2015. (Available online under http://scholar.harvard.edu/files/bstewart/files/stm.pdf).
Seufert, Günter: Außenpolitik und Selbstverständnis. Die gesellschaftliche Fundierung von Strategiewechseln in der Türkei, Berlin 2012.
Uchehara, Kieran E.: Continuity and Change in Turkish Foreign Policy Toward Africa, in: Akademik Bakış 3 (2008), S. 43-64.
Walker, Joshua W.: Turkey's Global Strategy: Introduction: The Sources of Turkish Grand Strategy – the 'Strategic Depth' and 'Zero-problems' in context, in: Kitchen, Nicholas [Hg.]: IDEAS reports – special reports (2011), S. 6-12.
Wang, Xuerui: Structured Topic Models. Jointly Modeling Words and their Accompanying Modalities, Amherst 2009.

Vectors or Bit Maps? Brief Reflection on Aesthetics of the Digital in Comics

Daniel Brodén
University of Gothenburg, Sweden

In recent years, researchers from various disciplines have contributed to the emerging study of the digital in comics (see Goodbrey 2013; Digital Humanities Quarterly 2015). Scholars from film and media studies, for example, have demonstrated the uses of film theory that deals with circulation (production, distribution and consumption) for thinking about digital comics and web comics (see Werschler 2011). Others have written on issues of digital mediatisation, drawing on theories of adaptation and animation (see Burke 2014). However, scholars have shown less interest in how film theory can be useful to explore the aesthetics of the digital in comics (for another study of media forms in digital humanities that utilizes film theory, see Ng 2015).

The aim of this paper is to briefly reflect on this topic, drawing on Sean Cubitt's prolific study on the history of moving images from a digital perspective, *The Cinema Effect*

(2004). Cubitt's book is grounded in the concepts pixel, cut and vector. Concisely put, pixel describes the cinematic image's appearance. Cut concerns how images are organised and differentiated through a film. Schematically transferred to the medium of comics, theses concepts may designate the visual elements of images and grids, respectively. My interest lies in the vector, which concerns the relation between the image and the interpretative mind. In computer graphics the vector is a line drawn from the centre of the screen, connecting programmed points and existing only temporarily as it leaves behind trails of light. Cubitt utilizes the analogy of the disappearing line to conceptualize thinking in the cinema as a vector-like process that links images in space and time, drawing on the viewer's experience of the flow of images. However, writing on digital effects driven Hollywood cinema Cubitt also uses the concept of the bit map to argue that what he regards as the vector's principle of openness has moved into something more fixed. Through the bit map he describes how cinema in the digital-era has become not only more visually spectacular but also more composed and controlled, not least since serendipity is harder to achieve on a computer, an instrument of precision (2004: 251).

Given the limited space of the paper, it is hard to address the complications of exporting theoretical concepts from one medium to another or the fundamental differences between cinema and comics. Nor will I engage with the ideological argument Cubitt develops concerning the qualities of the bit map. However, it should be noted that he presents his concepts in response to a medium with some kind of indexical relation to reality, a relation created by the cinematographic apparatus that seems to capture events. But as Cubitt himself writes, digital cinema and other digital media do not primarily refer, they communicate (2004: 250). The same could be said about comics (both pre-digital and digital) and I simply want to use Cubitt's concepts, the vector and the bitmap, in order to tease out some brief reflections on the aesthetics of the digital in comics.

Digital Colouring and Multimedia Styles

I will draw on two examples from mainstream comics. The first one concerns the breakthrough of digital colouring in the mid-1990s. In 1990 Frank Miller, the auteur behind iconic graphic novels such as *Dark Knight Returns* (1986), collaborated with artist Dave Gibbons of *Watchmen* (1986–1987, author Alan Moore) fame, on the dystopian action/satire series *Give Me Liberty* for Dark Horse Comics. A sequel, *Martha Washington Goes to War*, was published in 1994 and Gibbons' idiosyncratic style, a combination of cartoonish and mundane realism, characterized both runs. But there were also significant differences. For example, on *Martha Washington Goes to War* Gibbons used a less gritty style, a choice that tied in with Miller's more fantastical, high-concept storytelling. However, what is most interesting here is that whereas the colours in *Give Me Liberty* were hand-painted on watercolour paper with the then-advanced blue-line method, computer rendering was used in *Martha Washington Goes to War*. Simply by looking at the clean, smooth colour schemes the attentive reader could see that digital graphics was used.

Figure 1. Give Me Liberty *(1990)*

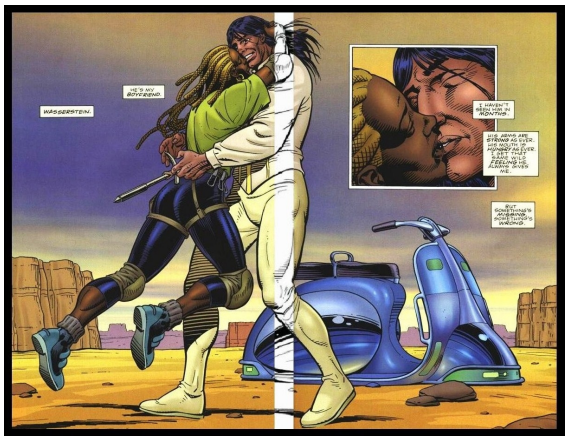

Figure 2. Martha Washington Goes to War *(1994)*

To some extent, this ties into how Cubitt associates the bit map with a more precisely composed aesthetic universe. In *Martha Washington Goes to War* there is even, arguably, a visible tension between Gibbons' old-fashioned hatched line drawings and the slick, heavily graded colour palette. One can discern similar tensions in later works by Gibbons, such as the subsequent *Give Me Liberty* runs (1995–2007) or the high-concept spy thriller/parody *The Secret Service* (2012, author Mark Millar), and conspicuous uses of digital colouring have generally become a prominent element in mainstream comics.

It is worth pointing out that the aesthetics of digital colouring depends on the approach. The use of rendering, which has a certain three-dimensional feel, differs from a flat colouring approach, which tends to have more of an old-school feel. Moreover, most artists working in mainstream comics today combine digital and classic hands-on approaches. For example, on *The Secret Service* Gibbons used watercolour brushes and Indian ink as well as digital graphic tablets.

My second example concerns this kind of hybrid aesthetics; writer Brian Michael Bendis and artist Alex Maleev's acclaimed run of Marvel Comics super-hero book *Daredevil* (2001–2006). Here, I should stress a point Cubitt makes; that the "distinction between bit map and vector […] so dear to first-year classes in computer graphics, is now approaching obsolescence" (2004: 249). According to Cubitt, the differences between the two principles no longer lie on "the familiar axis of verisimilitudinous painting and abstraction but along a line stretching from cartography at one end to architecture at the other. Somewhere in between lie the fields of virtual sculpture and computer-aided design manufacture" (ibid). This idea seems somewhat pertinent to Bendis and Maleev's *Daredevil* run. Maleev has been described as one of a new, graphically astute breed of multimedia artists, who incorporate painting, drawing and cartooning as well as photography, collage and computer effects to expand the visual universe of their works (Schumer 2005). Combining an angular, sketchy approach and photorealism (working from photographs of models and cityscapes), he has crafted a style, which goes beyond established modes of realism in mainstream comic books, yet retaining an organic feel.

Figure 3. Daredevil *no. 62 (2003)*

It is tempting to mainly describe Maleev's images in terms of a higher degree of realism. Arguably, they have another quality of verisimilitude compared to, for example, the artwork of Frank Miller and David Mazzucchelli's defining *Daredevil* mini-series *Born Again* (1986), which is also characterized by gritty and realistic but nevertheless non-photorealistic stylization.

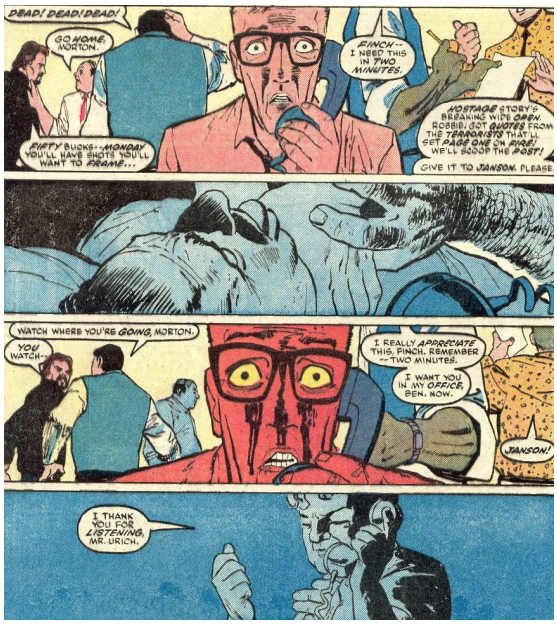

Figure 4. Daredevil *no. 229 (1986)*

However, Cubitt's claim that the difference between the bit map and the vector should not simply be described from an axis of verisimilitudinous painting and abstraction complicates matters. Though it would not be accurate to place Maleev's images in a field equivalent to the one Cubitt associates with digital cinema (in between virtual sculpture and computer-aided design manufacture), as the analogies become a little bit "off" in the context of the comics medium, it nevertheless seems reasonable to propose that Maleev's imagery exists in an aesthetic border zone in which Miller and Mazzucchelli's pre-digital *Daredevil* mini-series do not.

Conclusion

The aesthetics of the digital has brought new visual elements to comics, such as computer rendering of colours, but also hybrid, multimedia aesthetics, which do not necessarily on the surface seem that different from the ones of yesterday. It would probably not be impossible to create images with similar qualities to those of Maleev's without the use of computers, but arguably it would require more work and time. In this perspective, what digital technology has perhaps enabled is an economy within the comics industry with which spectacular images can be manufactured (c.f. Cubitt 2004: 248). But some differences between pre-digital and digital aesthetics in comics might also lie somewhere along a line that stretches toward absolute precision and control.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: comics, digital aesthetics

Bibliography

1. Burke, L. (2016): "Sowing the Seeds: How 1990s Marvel Animation Facilitated Today's Cinematic Universe", M J McEniry, R Moses Peaslee & R G Weiner (eds), *Marvel Comics into Film: Essays on Adaptations Since the 1940s*, London: McFarland,
2. Cubitt, S.(2004). *The Cinema Effect*. Cambridge, Mass: MIT Press.
3. *Digital Humanities Quarterly* (2015), 9:4. [Issue on digital comics]
4. Goodbrey, D. (2013). "Digital Comics: New Tools and Tropes". *Studies In Comics*, 4:1.
5. Ng, J. (2015). "The Cut between Us: Digital Remix and the Expression of Self". In Svensson, P. & Goldberg, D T. (eds), *Between Humanities and the Digital*. Cambridge, Mass: MIT Press.
6. Schumer, A. (2005). "Super-hero Artists of the Twenty-first Century: Origins". In Dooley, M. & Heller, S. (eds), *The Education of Comics Artists*. New York: Allworth Press.
7. Werschler, D. (2011). "Digital Comics, Circulation, and the Importance of Being Eric Sluis". *Cinema Journal*, 50:3.

Multilingual Clusters and Gender in Nordic Twitter

Steven Coats
University of Oulu, Finland

Recent years have seen an increase in the relative prominence of computer-mediated communication (CMC) modalities such as texting, instant messaging, or posting on social media, and platforms such as Twitter have become multilingual sites with global

representation (Mocanu et al. 2013; Leetaru et al. 2013). At the same time, population movements and changes in education and media consumption have contributed towards an increasing bi- and multilingualization of local environments -- trends that are particularly evident in the Nordic countries.

National languages continue to receive reinforcement in education and state media, but bilingualism with English has become the norm in Fenno-Scandia, while greater population mobility and demographic changes have contributed to increased linguistic diversity in the population.

Large-scale quantitative studies of multilingualism on CMC and Twitter (e.g. Ronen et al. 2014, Hale 2014) have shed light on multilingual networks globally, and the ways in which Twitter language use can pattern with gender expression have also been investigated in linguistics and natural language processing research (e.g. Bamann et al. 2014, Burger et al. 2011, Rao et al. 2010).

In this study, online multilingualism in the Nordic countries is investigated by means of a quantitative analysis of geo-located Twitter messages. The research investigates the following questions: Which languages are favored by multilingual users in the Nordics? To what extent are the Nordic languages used by multilingual Twitter users located in Fenno-Scandia? What role is the global language English playing in this multilingual landscape? and finally, what are the similarities and differences between the multilingual networds of male and female users?

In a first step, the online linguistic behavior of bi- or multilingual persons using Twitter in the Nordic countries is investigated according to location and gender. In a second step, the influence of the languages themselves is analyzed by looking at the aggregate network behavior of language users according to gender. What does the structure of the networks of multilingual users tell us about the current state of the Nordic languages and their future prospects?

Previous Work

Much research has investigated the status of English, the extent of English use in various media or communicative contexts, and the attitudes of speakers towards the use of English in the Nordic countries. For example, in Iceland, the majority of Icelanders are exposed to English every day, while 21% of Icelanders report speaking English daily (Arnbjörnsdóttir 2011). Norwegians are reported to be essentially diglossic (Rindal 2010; Rindal and Percy 2013), and re generally unperturbed by the prospect of English displacing Norwegian in Norway (Sandøy 2010). For Sweden, Bolton and Meierkord (2013), for example, attest that while Swedish remains the "preferred language… in most domains" (93), English is dominant in academia and business. Similar findings are reported for Finland in the results of an extensive survey into the use of English in the country: for Finland, English has become "a language used in many domains and settings within Finnish society" (Leppänen et al. 2011: 16).

A number of studies of CMC and Twitter language have investigated aspects of English, including phenomena such as the discourse functions of hashtags (Wikström 2014; Squires 2015), lexical innovation in American English (Eisenstein et al. 2016), grammatical varation in English-language Twitter from Finland and the Nordic countries (Coats 2016a, 2016b), or the interaction between demographic parameters such as gender with lexical and grammatical features in American English (Bamann et al. 2014).

For multilingualism, Ronen et al. (2014) compared the worldwide influence of languages by analyzing networks of bi- and multilingual book translations, Wikipedia author editors, and Twitter users, and found that English plays an important central role. Hale (2014) investigated global multilingual networks on Twitter, including the network associations of retweets and user mentions, and found that while most interaction networks are language-based and English is the most important single mediating language, other languages collectively represent a larger bridging force. Eleta and Golbeck (2014) demonstrate that multilingual users' language choice on the Twitter reflects the predominant language of their social networks. While

it has been found that users of less represented languages are more likely to switch languages and that English has become the central mediating language, the interaction of multilingualism with gender in has not yet been subject to research attention.

Data Collection

Tweets with populated place attributes were collected from the Twitter Streaming API in November 2016 using the Tweepy library in Python (Roesslein 2015). The country_code attribute was used to filter for only those tweets originating from the Nordic countries (including Åland and the Faroe Islands).

Language Determination

Since March 2013 Twitter objects include an automatically detected languge field, lang, determined on the basis of probabalistic matching of byte sequences in various language training data. Like other automatic language detection modules, the Twitter algorithm performs poorly on very short sentences. For this reason, tweets whose language was reported as "undefined", as well as those with fewer than 6 word tokens, were removed in a further filtering step. In total, the multilingual database consisted of 296,437 tweets by 33,347 unique users in 51 languages.

Gender was disambiguated on the basis of name lists provided by the statistical offices of the Nordic countries. The most extensive name information was available for Denmark, while public information available for the other Nordic countries was somewhat less extensive. 5,277 male and 6,095 given names from the lists were matched with the value of the char_name attribute for each unique user in the dataset. The method assigned gender to approximately 65% of the tweets collected from the target area.

Quantification of Bilingualism Strength

A user in the dataset was determined to be bilingual for languages i,j if he or she had authored at least three tweets in both languages. Of the 51 languages in the dataset, male bilingual users were present for 40 languages and females for 41. The connection strength between languages i,g was quantified using the phi coefficient, calculated from a contingency table of the number of bilinguals.

Phi is equivalent to Pearson's product-moment correlation coefficient for two binary variables, and ranges in value from -1 to 1. Positive values indicate the language pairs are more strongly connected than would be expected based on the prevalence of the languages in the multilingual dataset.

A t-statistic was calculated to test the significance of the correlation between languages. Links between languages that were statistically significant at $p<0.1$ were retained in the multilingualism network.

The network relationship between linguistic communities was represented by an N by N matrix of the number of bilingual users, where N represents the number of languages. To reduce the number of false positives due to language misidentification, only connections with fewer than bilingual users were considered.

Network relationships were visualized using the R packages igraph and visNetwork (Csardi and Nepusz 2006; Almende and Thieurmel 2016).

Results

In terms of overall language representation, and in accord with earlier findings (Mocanu et al. 2014; Coats 2016b), English is the most prevalent language in the data, with approximately 32% of the data in English, followd by 26% in Swedish, 13% in Finnish, 6% in Norwegian, 5% in Danish, and 2% in Icelandic.

Overall, 6.2% of the users in the complete data set qualify as multilinguals. For the gendered subcorpora, 6.49% of male users and 6.42% of female users fulfilled the criteria for multilingualism. Taken together, these findings match well with those reported by Hale (2014), who reported that that 11% of Twitter users in a global sample collected in 2011 are multilingual.

A multilingualism network for the entire Nordic region was created without taking gender into account. Node size corresponds

to the number of multilingual users for a language. Edge width corresponds to the strength of the connection (number of bilinguals) for a language pair.

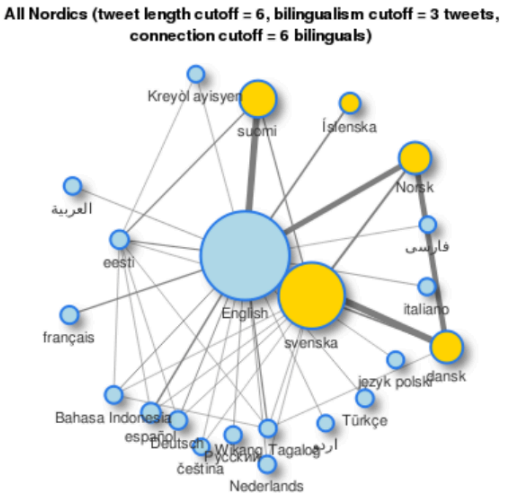

Figure 1: Multilingualism Network for the Nordic Countries

For the region as a whole, a network of 22 languages and 40 edges describes the statistically significant bilingual links. English clearly plays the most important role: it is connected to all of the languages for which a statistically significant phi value was calculated. Swedish has the next highest number of connections, connecting to 11 of the 22 language nodes. Other Nordic languages have fewer active bilingualism links on Twitter: Denmark has 4, Norway and Finland 3, and Iceland is only connected to English.

Multilingual networks were also created for individual Nordic countries. In them, the principal national language(s) figure prominently, but English remains important as a bridge between linguistic communities.

The multilingualism networks created by gender for the Nordic region may reflect some cultural and demographic facts. While English plays the central role in both the male and female clusters, the languages represented as well as the strength of links between languages are somewhat different for males and females. It can be shown that for Nordic languages, link strength for gendered clusters may reflect common cultural attitudes towards the traditional languages of the region. For linguistic communities of languages not traditionally present in the Nordic countries, the network associations may reflect recent gender imbalances in migration.

Topics: Nordic Textual Resources and Practices
Keywords: computer-mediated communication, twitter, multilingualism, NLP

Bibliography

Coats, Steven. (2016). Grammatical feature frequencies of English on Twitter in Finland. In Lauren Squires (Ed.), English in computer-mediated communication: Variation, representation, and change, 179–210. Berlin: de Gruyter Mouton. https://doi.org/10.1515/9783110490817-009

Coats, Steven. (2016). Grammatical frequencies and gender in Nordic Twitter Englishes. In Darja Fišer and Michael Beißwenger (Eds.), Proceedings of the 4th conference on CMC and social media corpora for the humanities, 12–16. Ljubljuana: U. of Ljubljana Academic Publishing. (http://nl.ijs.si/janes/wp-content/uploads/2016/09/CMC-conference-proceedings-2016.pdf)

The Prior-project: From Archive Boxes to a Research Community

Volkmar Engerer
Henriette Roued-Cunliffe
Jørgen Albretsen
Per Hasle
University of Copenhagen, Denmark

Introduction

A very important part of Digital Humanities (DH) is the development, use and discussion of digital research infrastructures within the humanities field. In fact, the notion of DH is often almost identified with this kind of en-

deavours. We ourselves think that such a conception is too narrow and misses important points, but we shall not attempt a general discussion here. However, it is important to note that data and representations within the humanities are often more heterogeneous and more dependent on domain expertise than are datasets within the STEM and in fact even the social sciences, whose datasets tend to be more regular and more amenable to standardized tools for storing, retrieving, analysing and visualising.

In this paper, we present a DH research infrastructure which relies heavily on a combination of domain knowledge with information technology. The general goal is to develop tools to aid scholars in their interpretations and understanding of temporal logic. This in turn is based on an extensive digitisation of Arthur Prior's Nachlass kept in the Bodleian Library, Oxford. The DH infrastructure in question is the Prior Virtual Lab (PVL). PVL was established in 2011 in order to provide researchers in the field of temporal logic easy access to the papers of Arthur Norman Prior (1914-1969), and officially launched together with Prior's Nachlass at the Arthur Prior Centenary Conference at Balliol College, Oxford, in August 2014 (Arthur Prior Centenary Conference 2014; Albretsen et al. 2016a).

Prior was a distinguished logician, philosopher, and in his younger years also a theologian. He is best known for his work on time and for being the founding father of modern temporal logic, beginning in New Zealand in the early 1950s. In 1956 he presented his ideas at the John Locke Lectures in Oxford. Following this he took up a professorship in Manchester (1959-1965) and was later appointed Reader in the University of Oxford and Fellow of Balliol College (1966-1969). Prior died, age 55, from a heart attack, while on a lecturing tour in Norway (Priorstudies 2016).

Prior's archive is now kept in the Bodleian Library in Oxford and is still subject to copyright. Following an agreement between the team behind PVL, the Prior family, and the Bodleian Library, the team has been permitted to take digital photos of the archive material. This work is ongoing and has currently resulted in approx. 7000 photos, which have been reassembled in the PVL so that they mirror the original documents, integrating a facility for transcribing them and adding user comments (PVL 2016).

Current Prior Virtual Lab

The restricted access has inspired the term Virtual Closed Collaborative Community. To get access a potential new user must first contact the project team and ask for login, stating her or his areas of research and in which ways that person's work in the PVL can add to the collaborative effort of publishing Prior's Nachlass - the digital edition of Prior's hitherto unpublished papers as well as other relevant material such as correspondence between Prior and other researchers, Prior's notebooks and scrapbooks etc. In PVL users can follow each other's progress and add comments to on-going transcriptions (Albretsen et al. 2016b). After an editorial process the transcribed texts are made available in PDF format combined with a prototype search facility (Nachlass 2016). Those users who have contributed to the transcription are credited in the footnotes of each transcribed edition. Our experience to date is that the current PVL is in need of enhanced search facilities combined with underlying metadata structures in the Nachlass.

Digital Humanities project

Prior's archive includes various documents such as drafts of philosophical essays, letter correspondence between Prior and other scholars, or sudden ideas scribbled as handwritten notes. These documents are currently used as information sources about Prior's convictions, theories, his life, relations to colleagues etc. However, when digitised, transcribed, and connected to each other through a database structure, they become a research object in their own right. Because of this transformation, Prior scholars can now explore patterns in the structure of the documents that were not visible before.

The further development of PVL is headed by a research group (the authors of this abstract) at the Royal School of Library and Information Science, University of Copenhagen. This activity forms an important part of the "Prior project", which 2016 received funding from the Danish Council for Independent Research | Humanities to carry out the research project *The Primacy of Tense: A.N. Prior Now and Then*, duration three years (DFF Grant 2016). The further development of PVL will be split into work on the data repository and the interface as two separate entities. The team behind PVL are to varying degrees Prior scholars, digital humanists, information scientists, and database engineers. Moreover, we have a vivid exchange with the other project researchers as well as the users of PVL. The project aims to combine this cross-disciplinary expertise in order to integrate community-specific practices of Prior scholars into the data structures and interfaces of the digital tools they use. It must be said that the information behaviour of the users has not yet been studied systematically. Such a study is one of the points within our project plan.

Data repository

In order to extend the existing facilities of the PVL it is necessary to offer a data and query structure that enables Prior scholars to explore the documents with varying and flexible parameters such as references to logicians, publications being mentioned and theories discussed. The new PVL architecture aims to separate the data structure from the interface and to develop a sustainable dataset that is suitable for both new and future interface designs. It applies traditional information science knowledge (mostly generated in the library domain) to the data repository, drawing on insights from indexing theory, metadata research, knowledge organization, information retrieval, and theories of information seeking.

A further refinement of the PVL's data structure can be achieved by ontologies. The information scientific concept of an ontology encompasses the sphere of indexing terms and related search terminology at the same time, and therefore regards index terms as closely related to the vocabulary used by specialists in their domain. The step from traditional thesauri and classification schemes to ontologies of knowledge domains integrates semantic web principles into the description of data and introduces a controlled language for knowledge representation with a built-in logic. This also makes it possible to derive information which is not explicitly contained in the descriptive terms themselves (Antoniou et al. 2012: 4). To be a bit more specific, let's give an example. The metadata established (or to be established) in this infrastructure will contain not only standardised metadata such as author, title, date, etc., but also and in fact more so domain specific metadata such as types of temporal logic, e.g. A-series and B-series logics, hybrid logic, metric and non-metric tense logic, and so on. These notions can only be established by domain experts and not by general information specialists. At the same time, they are exactly the kind of metadata that makes it possible to search and chart the kind of patterns that experts in the field are looking for.

New interface for PVL

It is our goal to make the PVL a research portal, where query results are presented to scholars through an interface that facilitates the identification of new relationships, identify patterns, and offer alternative ways of understanding and analysis. This work will build on concepts identified in Roued-Cunliffe's (2011) research on Decision Support Systems for the reading of ancient documents. This research examines digital tools useful for the transcription, interpretation and publication of the Vindolanda Tablets (2010) from the Roman occupation in Britain. However, many of the conclusions are equally relevant for scholarship on other handwritten documents such as Prior.

Building the new interface comprises the task of bringing together Prior community practices with system design, metadata structure and the system's affordances in terms of Prior researchers' information seeking behaviour.

Conclusion

PVL as well as the general website concerned with Prior's work and his archive in the Bodleian Library has without doubt already for quite some time been a useful DH infrastructure for researchers. This is evident not least in many papers from the Arthur Prior Centenary Conference, cf. (Albretsen et al. 2016a), to which the Nachlass material made available through PVL was crucial. Moreover, this infrastructure clearly could not have been developed without specific expertise on temporal logic and Prior's work. PVL is, in all modesty, a showcase of how important humanities material kept in a research library can be digitised using domain knowledge (and indeed only when using domain knowledge), and made available and useful for the relevant research community, making up the Virtual Closed Collaborative Community. In this manner it also reflects an important characteristic of many research infrastructures for the humanities, namely a particularly strong call for domain expertise for their useful development. The perspectives for taking PVL to its next level raises some new information scientific, not to say epistemological, issues of great importance. The development of a relevant ontology together with search options and visualisations of search results as well as other PVL material is in fact not just about making powerful tools available for research in temporal logic and in Prior's work; it is itself such research. The structure to be achieved is not neutral. It is itself a kind of theory about the internal coherence in Prior's work, a "statement" about its overall architecture. This we intend to elaborate in a longer ensuing paper, but we hope to have established a convincing case to the effect that our infrastructure does indeed form a sufficient basis for studying some pertinent epistemological issues for DH.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: digital epistemology, domain analysis, ontology, research infrastructure, virtual closed collaborative community.

References

Albretsen, J., Hasle, P., and Øhrstrøm, P. 2016a. Special Issue on The Logic and Philosophy of A.N. Prior. Synthese. Volume 193 Number 11. Guest edited by Jørgen Albretsen, Per Hasle, and Peter Øhrstrøm. http://link.springer.com/journal/11229/193/11/page/1. Retrieved November 14, 2016.

Albretsen, J., Hasle, P., and Øhrstrøm, P. 2016b. The Virtual Lab for Prior Studies: An example of a Closed Collaborative Community, DRAFT paper. http://research.prior.aau.dk/anp/pdf/The_Virtual_Lab_for_Prior_Studies_article_draft.pdf. Retrieved November 14, 2016.

Antoniou, Grigoris, Groth, Paul, van Harmelen, Frank & Hoekstra, Rinke. 2012. A semantic Web primer. 3rd. Cambridge, Mass.: MIT Press.

Arthur Prior Centenary Conference. 2014. http://conference.prior.aau.dk/. Retrieved November 14, 2016.

DFF Grant. 2016. *The Primacy of Tense: A.N. Prior Now and Then*, funded 2016-2019 by the Danish Council for Independent Research | Humanities. DFF|FKK Grant-ID: DFF – 6107-00087. http://ufm.dk/forskning-og-innovation/tilskud-til-forskning-og-innovation/hvem-har-modtaget-tilskud/2016/bevillinger-fra-det-frie-forskningsrad-kultur-og-kommunikation-til-dff-forskningsprojekt-2-juni-2016. Retrieved November 14, 2016.

Nachlass. 2016. http://nachlass.prior.aau.dk. Retrieved November 14, 2016.

Priorstudies. 2016. http://www.priorstudies.org . Retrieved November 14, 2016.

PVL. 2016. http://research.prior.aau.dk. Retrieved November 14, 2016.
Roued-Cunliffe, H. 2011. A decision support system for the reading of ancient documents (Doctoral thesis). University of Oxford. https://ora.ox.ac.uk/objects/uuid:9d547661-4dea-4c54-832b-b2f862ec7b25 . Retrieved November 14, 2016.
Vindolanda Tablets. 2010. Vindolanda Tablets Online II. http://vto2.classics.ox.ac.uk. Retrieved November 14, 2016.

Recent publications by the first author:

Engerer, Volkmar (accepted 13 July, 2016): "Control and Syntagmatization. Vocabulary Requirements in Information Retrieval Thesauri and Natural Language Lexicons", Journal of the Association for Information Science and Technology.
Engerer, Volkmar (im Erscheinen): „Informationswissenschaft für Linguisten. Die Sprache des Information retrieval" (Akten der Gesus-Jahrestagung in St. Petersburg, Russland, 2015).
Engerer, Volkmar (im Erscheinen): „Das Vokabular zwischen Sprach- und Informationswissenschaft" (Akten der Gesus-Jahrestagung in Brno, Tschechische Republik, 2016).
Engerer, Volkmar, (2016), "Exploring interdisciplinary relationships between linguistics and information retrieval from the 1960s to today", Journal of the Association for Information Science and Technology, Article first published online April 4, 2016 (Early view). DOI: 10.1002/asi.23684.
Engerer, Volkmar, (2014), „Indexierungstheorie für Linguisten. Zu einigen natürlichsprachlichen Zügen in künstlichen Indexsprachen", in: Schönenberger, Manuela, Volkmar Engerer, Peter Öhl & Bela Brogyanyi (Hgg.) (2014), Dialekte, Konzepte, Kontakte. Ergebnisse des Arbeitstreffens der Gesellschaft für Sprache und Sprachen, GeSuS e.V., 31. Mai – 1. Juni 2013 in Freiburg/Breisgau, Jena, pp. 61 – 74.
Engerer, Volkmar, (2014), „Thesauri, Terminologien, Lexika, Fachsprachen. Kontrolle, physische Verortung und das Prinzip der Syntagmatisierung von Vokabularen", Information, Wissenschaft & Praxis, 65/2 (2014), pp. 99 – 108. [BFI 1]
Engerer, Volkmar, (2012), „Informationswissenschaft und Linguistik. Kurze Geschichte eines fruchtbaren interdisziplinären Verhältnisses in drei Akten", SDV – Sprache und Datenverarbeitung. International Journal for Language Data Processing, 36/2 (2012), pp. 71 – 91 (= Hermann Cölfen (Hg.), E-Books – Fakten, Perspektiven und Szenarien)

Mapping the Development of Digital History in Finland

Mats Fridlund
Petri Paju
Aalto University, Finland

In 2015, the field of digital history in Finland saw a tremendous development (see Parland-von Essen 2016), and in many ways this has continued ever since. This paper presents work by the project "Towards a Roadmap for Digital History in Finland: Mapping the Past, Present & Future Developments of Digital Historical Scholarship." The ongoing project was awarded by the Kone Foundation in December 2015 and lasts for 12 months. On its steering board, the project involves several of the most active Finnish digital historians who as a group also felt the need for and came up with the idea of the project amidst this fast development. Principal Investigator of the project is professor Mats Fridlund (Aalto University) and Dr Petri Paju does the research.

In this paper, we aim to discuss some key questions and challenges firstly in its own work and goals, and secondly in the field of digital history research in Finland in general.

Further, we aim to place these Finnish developments within the context of the larger digital humanities (as well as history) movement in the Nordic countries.

The project work started in February 2016. Its information gathering, including interviews, are carried on from March onwards. To prepare historians for its inquiry, the project organized a public Opening Seminar titled "Digital History in Finland: Possible Futures" in Helsinki 15.4.2016. It featured expert speakers from history's neighboring disciplines (archeology and historical linguistics) and presentations about computational history in Finland as well as about big data from a historians' point of view.

The online inquiry was open from late April till end of June 2016. It was widely advertised with for example articles written in both Swedish and Finnish. Altogether seventeen (17) persons responded to the inquiry. This somewhat low number of respondents will be complemented by results from other recent surveys and user studies of which there are a few.

Based on these inquiry answers, the researcher of the project compiled a report and the report, called "Digitaalinen historiantutkimus kyselytuloksia" (Paju 2016, 12 pages, with an abstract and key results in English), was made public in the project's blog 1.9.2016. Main results of the report centered on the complexity of defining digital history and the researchers' difficulties with such an identity. Moreover, several critical issues were identified, namely creating better, up-to-date information channels of digital history resources and events, providing relevant education, skills, and teaching by historians, and the need to help historians and information technology specialists to meet and collaborate better and more systematically than before. Meanwhile there is a lot happening in the field of digital history that should and will be somehow included in the mapping. This now on-going project should have fresh results from mapping this fast changing domain and compiling a roadmap for it by the time of the possible presentation in Gothenburg in mid-March 2017.

The presentation could fit with the conference subtheme of The Digital, the Humanities, and the Philosophies of Technology. Our scholarly perspective comprises of research in the history of science and technology and to some extend science and technology studies. Drawing on some of these research traditions, one recent inspiring study has been Smiljana Antonijević's book Amongst Digital Humanists: An Ethnographic Study of Digital Knowledge Production (2015).

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: digital history, history of the digital humanities, Finland, survey project, mapping the field

Sources

Antonijević, Smiljana: Amongst Digital Humanists: An Ethnographic Study of Digital Knowledge Production. Palgrave Macmillan, New York 2015.

Project's blog: https://digihistfinlandroadmapblog.wordpress.com/

Paju, Petri: "Digitaalinen historiantutkimus kyselytuloksia." Report from the project Towards a Roadmap for Digital History in Finland, available online from 1.9.2016: https://digihistfinlandroadmapblog.wordpress.com/2016/09/01/raportti-kyselyvastauksista/

Parland-von Essen, Jessica: "Tankar kring den snabba utvecklingen i Finland år 2015", Historia i en digital värld, January 5, 2016.

Bibliography

Fridlund, Mats & Daniel Sallamaa: "Radikale Mittel, gemäßigte Ziele: Repression und Widerstand im Großfürstentum Finnland", Osteuropa 66 (2016):4, 35-47.

Fridlund, Mats: "Motståndets materialitet: oppositionella ting, tekniker och kroppar under Finlands ofärdsår", in: Nina Wormbs, & Thomas Kaiserfeld,

eds. Med varm hand (Stockholm, 2015), 53-84
Fridlund, Mats & René Brauer: ”Historizing topic models: A distant reading of topic modeling texts within historical studies”, in: LV Nikiforova & NV Nikiforova, eds., Cultural Research in the Context of "Digital Humanities" (St Petersburg, 2013), 152-163.
Paju, Petri: ”Digitaalinen historiantutkimus kyselytuloksia.” Report from the project Towards a Roadmap for Digital History in Finland, available online from 1.9.2016: https://digihistfinlandroadmapblog.wordpress.com/2016/09/01/raportti-kyselyvastauksista/

Visualising Genre Relationships in Icelandic Manuscripts

Katarzyna Anna Kapitan
University of Copenhagen, Denmark
Timothy Rowbotham
University of York, United Kingdom
Tarrin Wills
University of Copenhagen, Denmark

Purpose

The proposed paper is based on the research which arose from a collaboration between three of us working respectively on the writing and reception of medieval Icelandic legendary histories (Rowbotham); transmission history and applications of digital tools in philological research (Kapitan); and understanding the manuscript context for prose and poetic texts (Wills). We discovered that we had between us access to enough data and expertise to remarkably expand on previous analyses of the relationships between Old Norse texts as preserved in medieval and later manuscripts and, furthermore, that these analyses could be used to refine our definitions of literary genres and the place of individual texts within those categories.

We focused initially on texts belonging to a group of so called ‘legendary’ sagas, or ‘mythical-heroic’ sagas (Ice. fornaldarsögur), since the question of this group’s genre status - specifically, whether fornaldarsögur (FAS) ought to be considered a distinct genre, or be analysed alongside their ‘cousins’ riddarasögur (RIDD) - has been widely discussed in the literature. Contradictory opinions concerning genre classification have been offered by leading scholars in the field; Mitchell (1991, 21) and Aðalheiður Guðmundsdóttir (2001, cxlvii) have suggested that fornaldarsögur were considered a distinct category of literature already in the Middle Ages, as they are frequently bound together in manuscripts, whereas Driscoll (2005, 193; see also Ármann Jakobsson 2012, 24) has suggested, also on the basis of their codicological context, that riddarasögur and fornaldarsögur should be treated as one literary group. Despite their opposing conclusions, the consensus among these scholars is that the codicological context of these texts is key to understanding the genre they represent.

Though it is necessary to look into medieval manuscripts to reach the medieval reader’s understanding of the genre, we must take into consideration the huge loss of medieval manuscripts, and thus recognise that our knowledge of the medieval tradition is fragmentary. Due to this lack of data, looking into sixteenth and seventeenth century manuscripts may deliver us important information about the medieval tradition, since there is some probability that post-medieval manuscripts are close copies of their medieval exemplars, and thus might preserve the the texts’ original context. Therefore, we have decided to look at all available manuscript descriptions collected in handrit.org, fasnl.ku.dk and the Skaldic Project Database. The method we have pursued for identifying genre association has been to analyse the complex manuscript context of these texts, on the basis that analysis of this context helps to inform our understanding of the genre classification of medieval Norse literature. The approach we have developed has been applied across the corpus to un-

derstand genre relationships as represented by the manuscript tradition.

Method

Our paper focuses on an interpretation of the relationships between Old Norse texts based on a statistical analysis of digitized manuscript descriptions. Since the initial focus of our research was an interpretation of genre associations within the corpus of fornaldarsögur an obvious point of departure was the online catalogue fasnl.ku.dk. The catalogue of all the manuscripts in which fornaldarsaga texts are found, including information on their format and layout, the other texts they preserve and when, where and by and/or for whom they were written.

Further data came from other projects: The Dictionary of Old Norse Prose (ONP) has produced a comprehensive list of works within the scope of that project (published in their Registre volume and with subsequent revisions), along with detailed information about the manuscripts for each work including the dating of the manuscripts and location of each work within the manuscript. This data was supplied to the Skaldic Project and has also been used (with permission) here. The Skaldic Project itself has supplemented the manuscript information with the poetry relevant to that project and other manuscripts that were not recorded in the ONP data tables. Additionally, relevant data for manuscripts not containing fornaldarsögur has been supplemented by the XML descriptions in handrit.org.

The ONP and Skaldic Project manuscript information is structured with texts linked to the manuscripts using a relational database model. fasnl.ku.dk and handrit.org, in contrast, give XML descriptions of the manuscripts. One of the challenges for addressing this question is taking the complex manuscript descriptions, constructed as TEI XML, and extracting the relationships between the texts contained within them. The manuscript descriptions from fasnl.ku.dk and handrit.org were designed to give a detailed description of each object and its structure, but do not definitively describe the relationships between items in different manuscripts. Consequently the same text in two descriptions may be labeled with different 'uniform' titles or even genre class. In order to build a visualisation and analysis of the relationships between texts and genres we have had to define these relationships ourselves. We describe firstly how one particular genre, the legendary sagas, was supplemented and normalised. Secondly, we describe how manuscript data from the fasnl.ku.dk, handrit.org, Skaldic Project and Dictionary of Old Norse Prose were merged, including processes for normalising text names and generic classifications.

An open source visualisation software, Gephi, was used to analyse 153963 connections between 1518 texts. A network of relationships between all the texts was achieved by application of ForceAtlas2 layout (Jacomy et.al. 2011). ForceAtlas2 is a force directed layout in which nodes repulse each other like magnets while edges attract the nodes they connect like springs, in case of our network, inspired by RIDD-network presented by Hall (2013), texts are represented as nodes, while edges represent manuscripts. The thicker is the edge between two texts the bigger is a number of manuscripts in which these texts appear together. Unlike in Hall's (2013) network, the size of the nodes is standardized and independent of a number of connections created by the texts.

Further analysis weights the connections between texts according to length (using page counts), as a large number of very small texts (i.e. þættir) can disproportionately influence the network by generating more connections. Additionally, we have compared results using different watershed dates for the manuscript tradition, including 1728 (the year of the great fire of Copenhagen) and 1829 (the publication year of Rafn's Fornaldarsögur norðrlanda).

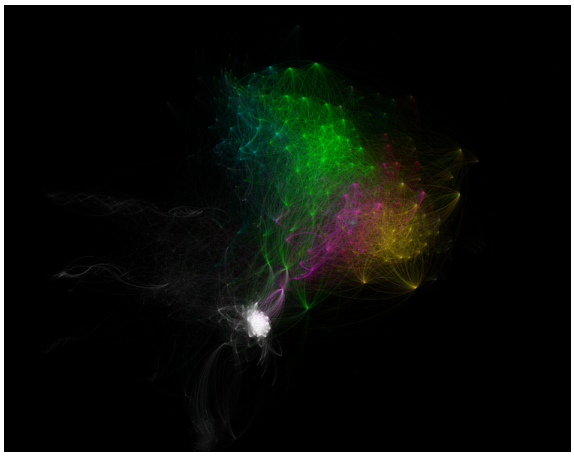

Figure 1: *Network of Icelandic literature. FAS - pink; FORNS, FORNTH - red; ISL, ISLT - green, KON, KONTH - blue, RIDD - yellow, RIDDST - orange; EDD - white*

Findings

As presented on Figure 1, the group of fornaldarsögur (pink) is positioned between íslendingasögur (green) and riddarasögur (yellow), and mixes with fornadlarsögur síðari tíma, and fornaldarþættir (red). Kongungasögur (blue) show close affiliation with íslendingasögur, while eddic poetry (white) creates a separate group, which is connected to fornadalrsögur through Hervarar saga ok Heiðreks. This connection can be explained by the fact that riddles from Hervarar saga ok Heiðreks were often copied independently from the saga, and included in manuscripts together with other poems, but in the catalogues they appear as the witnesses of the saga.

The data collected and visualised is of great value to the study of medieval Icelandic literature, but the great volume of it presents a significant challenge to researchers wishing to provide a detailed philological analysis. To begin to analyse the data, we decided to take a small number of texts as case studies and, marrying the approaches of philological research with those of the digital humanities, examine relationships between an individual fornaldarsaga and the texts it is linked to in the manuscript transmission. The selection of case studies was initiated by focussing on a number of texts that were of interest from a literary critical perspective, and that we regarded as somewhat 'peripheral' to the fornaldarsaga genre; these included texts such as Hrómundar saga Gripssonar, for which we have only indirect evidence of its existence in the middle ages, Þjálar-Jóns saga, which has often been regarded in scholarship as a riddarasaga, and those texts, such as Helga þáttr Þórissonar and Norna-Gests þáttr, that were originally included as episodes (or þættir) in longer konungasögur, but since the nineteenth century have been included in the fornaldarsaga corpus. The XSLT scripts used in the earliest stages of our research confirmed that these texts, among others, were noteworthy for the frequency with which they appear alongside genres such as riddarasögur and konungasögur.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: manuscript studies, network analysis, data visualisation, genre

Bibliography

Aðalheiður Guðmundsdóttir. (2001). Úlfhams saga, Reykjavík: Stofnun Árna Magnússonar.

Ármann Jakobsson. (2012). "The Earliest Legendary Saga Manuscripts". In: The Legendary Sagas: Origins and Development. Eds. Annette Lassen, Agneta Ney Ármann Jakobsson. (pp. 21–32). Reykjavík.

Driscoll, M. J. (2005). Late prose fiction ('lygisögur'). In A Companion to Old Norse-Icelandic Literature and Culture. (31 ed., pp. 190-204). Oxford: Blackwell Publishing Ltd.

Hall A., Parsons K. (2013). "Making stemmas with small samples, and digital approaches to publishing them: testing the stemma of Konráðs saga keisarasonar", Digital Medievalist 9.

Jacomy M., Heymann S., Venturini T., Bastian M. (2011). "ForceAtlas2, A Graph Layout Algorithm for Handy Network Visualization", http://webatlas.fr/tempshare/ForceAtlas2_Paper.pdf

Mitchell, S. (1991). Heroic sagas and ballads. Ithaca and London: Cornell University Press.
Rafn C.C. (1829). Fornaldarsögur Norðrlanda. Vol I-III. Kaupmannahöfn.

Online resources
Ordbog over det norrøne prosasprog Registre: http://onpweb.nfi.sc.ku.dk/mscoll_d_menu.html
Online catalogue handrit.org: http://handrit.org
Skaldic Project: http://skaldic.abdn.ac.uk/db.php?
Stories for all time Poject: http://fasnl.ku.dk/
Gephi The Open Graph Viz Platform: https://gephi.org

Spatiality, Tactility and Proprioception in Participatory Art

Raivo Kelomees
Estonian Academy of Arts, Estonia

In this presentation I analyse performances, artworks and installations in audiovisual and contemporary art which emphasise tactile and corporeal experiences. This tendency can be observed in technological art, cinema and large visual attractions. I aim to demonstrate that due to technical developments and new tools, the possibilities now exist for new aesthetic experiences in which the body's position and its biological reactions are decisive. This leads to the question of how the critical or theoretical point of view of an artwork changes when the spectator's reactions to it are documented and quantified in real time and are changed into source material for the next stage(s) of the artwork. Does this constitute the next step in the research of interactive artworks which were based on the subjective analysis of the participant's reactions? Does it allow us to rewrite art analytical analyses, which were based on the subjective analysis of the researchers?

The main emphasis of this presentation is the proprioceptive experience in art. I will start with an analysis of earlier inventions and analogous practices which introduce corporeal artistic experience. I then investigate whether we can talk about the 'proprioceptive image' in the same way that we can speak about the artistic, musical or literary image. This analysis is influenced by a media archaeological approach, in particular Erkki Huhtamo's interpretation in which his approach is termed "media archaeology as topos study" or simply "topos archaeology." I aim to demonstrate how this "topoi" – "haptic and corporeal experience in audiovisual performances and visual art" or "spatiality, tactility and proprioception in participatory art" – changes and "transfigures" those examples in which the corporeal experience is translated into digital data and subsequently used for manipulations of the artwork. Before starting to analyse the works of Jeffrey Shaw, Char Davies and Bill Seaman in the sub-chapter "Tactility and proprioception in media art", I will provide a series of historical examples which lead to contemporary developments in media art.

The main focus of the text is on changes in the "art world", with an emphasis on fields which could be called media art, new media, electronic art, and contemporary art. To a lesser extent there is also a focus on discussions happening in crossmedia and transmedia—even though some projects are not easy to define, or belong to the fields of both new media and transmedia. This particularly concerns those works of multimedia where the tactile experience on screen is gradually becoming spatial and corporeal. Another topic under analysis is how clear is the tendency to make the audio-visual experience tactile, tangible and physically experienceable, in contrast to the virtual experience.

In my discussion of multi-screen and physically perceptible environments I want to show situations, solutions and artworks from the beginning of a so-called television

era, and in experiments of the expansion of the cinematic experience, in which:

* an "interrelation" occurs between the visual screen content and a "communication" occurs between screens: the visual or auditive content on different screens is transferred from one to another, and a narrative is split between different (two or more) screens;

* a connection occurs between screen images and stage activity: actors in physical space and screen-space are acting in collaboration or antagonism to each other;

* viewers are influencing and leading the screen content: screen environments which surround viewers are gradually changed into environments which are shaped by users/viewers;

* viewers or actors are "in the image": viewers or actors are corporeally in the image or influencing it directly;

* the spectator's physiology is influencing or leading the screen content: the viewer's participation in the presentation of images is influenced by the biological data of the same viewer. This means that biological data (such as Heart Rate Variability, HRV; Galvanic Skin Response, GSR etc.) are used as input data for audiovisual variations.

Amongst early examples the following works are analysed: Raoul Grimoin-Sanson's "Cinéorama", 1900; Charles and Ray Eames' "Glimpses of the United States" in Moscow in 1959; Czechoslovakia's "Laterna Magika" at the World Fair in Brussels in 1958, designed by Josef Svoboda; "Polyvision" by Josef Svoboda and Jaroslav Frič; Josef Svoboda's "Diapolyekran" at Expo'67 in Montreal; Roman Kroitor's "Labyrinthe" at Expo'67 Montreal; Radúz Činčera's "Kinoautomat" at Expo'67; and other projects.

Discussing contemporary environments of stage performances, digital art and research practices I will present predecessors like Robert Whitman and his "Prune Flat" (1965) and "Shower" (1964); Tony Oursler's works; the British theatre company "Moving Being"; Steve Dixon's group "Chameleons"; Peeter Jalakas' performance "Estonian Games. Wedding" ("Eesti mängud. Pulm", 1996). I also discuss the "digital theatre" and "cyberformance" groups Troika Ranch and Dumb Type. The goal in presenting these examples is to illustrate the attempts in cinema, theatre, art and research environments to create multi-screen environments that engage the audience, offering them entertainment, information and an explorative experience. The tendency is to make the visual medium tangible and corporeal so that in some examples in interactive art the viewer "puts his hands" into the artwork.

In this text I will formulate the definition of proprioception, which means the spatial orientation arising from stimuli within the body itself. This term is used to cover sensorial systems which give information about position, posture, orientation and movement of the body (and its parts) in space. In regard of a proprioceptively perceived artwork we can talk about the situation where the viewer's whole body and behaviour is involved in the decisive interaction. I will choose three examples of interactive art to analyse from the proprioceptive point of view: Jeffrey Shaw's "Legible City" (1989), Char Davies' "Osmose" (1995) and Bill Seaman's "Exchange Fields" (2000). Transferring proprioceptive cognition into interactive, participative and tactile art allows us to enquire whether the corporeal experience is interesting and aesthetically novel. Also, does the corporeal experience make these artworks proprioceptively distinctive? I conclude that "Legible City" is more ordinary than "Osmose and "Exchange Fields"—in which the viewer's proprioceptive participation is original.

In this analysis I avoid discussion of biofeedback-based interactive art and cinema. The goal of the presentation is to prove that the expansion of the viewers' experience in cinema and art has reached a corporeal and tactile experience. In these artworks the visual-auditive-spatial presentation is related to the viewer's physical activity or reactions. Building on a series of historical examples I prove the existence of the trend and the historical tendency that was already visible in tromp l'oeil paintings – the desire to erase the difference between the artificial and real worlds. It is interesting to see a consistency

of attempts to "break the barrier" between reality and artificiality which occurs on different technical levels of complexity. We can talk about cultural topos that make the virtual tangible in that which is visible beside visual art and media art in experimental solutions of cinema.

Firstly I focus on artworks in which "immersion" is happening to a maximum extent and where the proprioceptive "sense" defines the aesthetic experience. Since proprioception is a complex corporeal-physiological feedback mechanism it would be wrong to call it "a sense", but undoubtedly it has been unjustly omitted in discussions about art. This presentation aims to foreground this term and to demonstrate that we can talk about a proprioceptive aesthetic experience.

I conclude that artworks which are made for tactile, proprioceptive and biofeedback experiences are made with experimental and research purposes. The creation of these works depends on the availability and cheapness of respective sensor technologies, the level of competency of artists, designers and programmers, and the rise of new collaborative practices.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: tactility, biofeedback, proprioception, participatory art, interactive art, corporeal-physiological feedback

References

N. Carpentier, Media and Participation: A site of ideological-democratic struggle. Bristol, UK and Chicago, USA: Intellect, 2011, 276-308; Book of Imaginary Media: Excavating the Dream of the Ultimate Communication Medium. Edited by Eric Kluitenberg. Rotterdam: NAI Publishers, 2006.

J. D. Bolter and D. Gromala, Windows and Mirrors. Interaction Design, Digital Art, and the Myth of Transparency. MIT Press, Cambridge MA, 2005, lk. 28.

S. Dixon, Digital Performance. A History of New Media in Theater, Dance, Performance Art, and Installation. The MIT Press, Cambridge, MA, 2007.

M. Bielicky, Prague–A Place of Illusionists, in: Future Cinema. The Cinematic Imaginary after Film. Jeffrey Shaw/Peter Weibel (eds), The MIT Press, Cambridge, MA/London, 2003.

O. Grau, Virtual Art. From Illusion to Immersion, The MIT Press, Cambridge, Mass., 2003.

Glimpses of the U.S.A. (1959) [excerpt], https://www.youtube.com/watch?feature=player_embedded&v=Ob0aSyDUK4A

C. Hales, Spatial and Narrative Constructions for Interactive Cinema, with particular reference to the work of Radúz Činčera. In: Expanding Practices in Audiovisual Narrative, ed. by R. Kelomees, C. Hales. Cambridge Scholars Publishing, 2014.

V. Havránek, Laterna Magika, Polyekran, Kinoautomat, in: Future Cinema. The Cinematic Imaginary after Film. Jeffrey Shaw/Peter Weibel (eds), The MIT Press, Cambridge, MA/London, 2003.

E. Huhtamo, Obscured by the Cloud: Media Archeology, Topos Study, and the Internet. In. ISEA2014 Dubai: Location. Proceedings of the 20th International Symposium on Electronic Art. Ed. by Thorsten Lomker. Zayed University Books, Dubai, UAE.

E. Huhtamo, Twin-Touch-Test-Redux: Media Archaeological Approach to Art, Interactivity, and Tactility, in: MediaArtHistories, ed. Oliver Grau. Cambridge, Mass: The MIT Press, 2006.

E. Huhtamo, Illusions in Motion: Media Archeology of the Moving Panorama and Related Spectacles. Cambridge, MA: MIT Press, 2013.

E. Huhtamo, Resurrecting the technological past: An introduction to the archeology of media art. Intercommunication, 14, 2. (1995).

E. Huhtamo, From Kaleidoscomaniac to

Cybernerd Towards an Archeology of the Media. Leonardo, Vol. 30, No 3 (1997).
I. Ibrus & C. A. Scolari, Introduction: Crossmedia innovation. In I. Ibrus & C. A. Scolari (Eds.), Crossmedia Innovations: Texts, Markets, Institutions. Frankfurt: Peter Lang.
Into the Light. The Projected Image in American Art 1964–1977, Chrissie Iles (ed.), exhib. cat., Whitney Museum of American Art. New York/Harry N. Abrams, New York, 2001.
H. Jenkins, Transmedia Storytelling 101, http://henryjenkins.org/2007/03/transmedia_storytelling_101.html#sthash.SGZaSez6.dpuf
O. Kruglanski "As Much As You Love Me", http://archive.aec.at/prix/#35286
L. Manovich, Soft Cinema, http://www.softcinema.net/mission_to_earth.htm
L. Manovich, Information as an Aesthetic Event, 2007, http://manovich.net/content/04-projects/056-information-as-an-aesthetic-event/53_article_2007.pdf
L. Manovich, What is Visualization? Visual Studies, vol. 26, no.1. (2011): 36-49, http://manovich.net/content/04-projects/064-what-is-visualization/61_article_2010.pdf
Media Archaeology: Approaches, Applications, and Implications. Edited by Erkki Huhtamo & Jussi Parikka. Berkeley, CA:University of California Press, 2011.
B. Montero (2006). Proprioception as an Aesthetic Sense. Journal Of Aesthetics And Art Criticism 64 (2):231-242.
S. Natale, Understanding Media Archaeology, Canadian Journal of Communication, Vol 37 (3), http://www.cjc-online.ca/index.php/journal/article/viewFile/2577/2336.
M. Saldre, Tühirand eesti kulturimälus:kirjandus-, filmi- ja teatripärase esituse transmeedialine analüüs. Acta Semiotica Estica VII, 2010, 160–182.
M. Saldre and P. Torop, Transmedia space. In: Ibrus, Indrek and Carlos A. Scolari (Eds.). Crossmedia Innovations: Texts, Markets, Institutions. Frankfurt etc.: Peter Lang, 2012.
M. Schjødt, Switching, http://www.switching.dk/en/
F. Sparacino, G. Davenport, A. Pentland, Media in performance: Interactive spaces for dance, theater, circus, and museum exhibits. IBM Systems Journal Vol. 39, Nos. 3 & 4, 2000, p. 479, http://alumni.media.mit.edu/~flavia/Papers/ibm_sparacino.pdf
F. Sparacino, C. Wren, G. Davenport, A. Pentland, Augmented Performance in Dance and Theater. International Dance and Technology 99 (IDAT99), at Arizona State University, Feb. 25-28, 1999, http://alumni.media.mit.edu/~flavia/Papers/flavia_augmented_performance.pdf
M. Teemus, Reisides toas. Pano-, kosmo- ja diaraamadest Tallinnas ja Tartus (1826-1850). Tartu Ülikooli Kirjastus, 2005.
F. Thalhofer, Planet Galata, http://www.thalhofer.com/_data/PAGES/xproject_2010_PlanetGalata.html
F. Thalhofer, Love Story Project, http://www.thalhofer.com/_data/PAGES/xproject_2002_LoveStoryProject.html
Proprioception, US National Library of Medicine Medical Subject Headings (MeSH), https://www.nlm.nih.gov/cgi/mesh/2011/MB_cgi?mode=&term=Proprioception
B. Seaman "Exchange Fields" (2000), http://projects.visualstudies.duke.edu/billseaman/seamanvanberkel/exchange_fields/exchange_fields.htm
See This Sound: Promises in Sound and Vis-

ion. Ed. by Claudia Albert, Amy Alexander, Rainer Bellenbaum, Dieter Daniels, Sandra Naumann. Walther König, Köln, 2010.
P. Tikka, Enactive Cinema. Simulatorium Eisensteinense. University of Art and Design Helsinki, 2008.
P. Weibel, The Post-Gutenberg Book. The CD-ROM between Index and Narration, in: artintact 3, Artists'interactive CD-ROMagazin. Cantz Verlag 1996.
S. Zielinski, Deep Time of the Media: Toward an Archaeology of Hearing and Seeing by Technical Means. Cambridge, MA: MIT Press, 2006.
G. Youngblood, Expanded Cinema, Dutton, New York, 1970.

The Elias Lönnrot Letters Online – Challenges of Multidisciplinary Source Material

Kirsi Keravuori
Niina Hämäläinen
Maria Niku
Finnish Literature Society SKS

The Finnish Literature Society SKS will launch the *Elias Lönnrot Letters Online* in April 2017. The new digital edition is part of the Society's "Open Science and Cultural Heritage" -project which seeks to develop scholarly online materials and tools.

The correspondence of Elias Lönnrot (1802–1884, doctor, philologist and creator of the national epic *Kalevala*) comprises of 2 500 letters or drafts written by Lönnrot and 3 500 letters received. The online edition is the conclusion of several decades of research, of transcribing and digitizing letters and of writing commentaries. Part of Lönnrot's letters we published already in the beginning of the 20th century, and the *Selected letters* came out in 1990. Since then Lönnrot's correspondence has been digitized in an Academy of Finland project, and transcribed between 2005 and 2016. The online edition will be the first complete publication of his vast correspondence. We will begin the publication with the approximately 1 800 private letters written by Lönnrot.

The online edition is designed not only for those interested in the life and work of Lönnrot himself, but more generally to scholars and general public interested in the work and mentality of the Finnish 19th century nationalistic academic community, their language practices both in Swedish and in Finnish, and in the study of epistolary culture. The rich, versatile correspondence offers source material for research in biography, folklores studies and literary studies; for general history as well as medical history and the history of ideas; for the study of ego documents and networks; and for corpus linguistics and history of language.

While being fully aware of the significance and the multidisciplinary use of the Lönnrot letters, the group working with the online publication is faced with the usual challenges of humanistic research and publication projects: insecure and discontinuous funding, the time-consuming process of transcribing of extensive source materials, and the fast development of technical solutions. The SKS decided to prioritize the prompt online publication of Lönnrot letters with good, practical tools for researchers and open, accessible data for those that want to develop the material further.

The extensive source material together with the priority on prompt publication and the small staff made it necessary to find a publishing platform that would require relatively light modification and would be easy to manage, and where the process of importing the source material could be easily automatized. The group first considered the edition platform used by the SKS's *Edith – Critical Editions of Finnish Literature* and the Svenska litteratursällskapet's *Zacharias Topelius Skrifter* (http://www.topelius.fi/). However, this was found to be too labour-intensive and complex. In particular the commentary tool included in the edition platform was deemed to be unnecessary for the purposes of the *Elias Lönnrot Letters Online*, which will include only a limited

amount of commentaries. The SKS had prior experience with Omeka, the open-source web-publishing platform for the display of library, museum, archives, and scholarly collections and exhibitions. A trial period demonstrated that this platform was the best option available for the planned publication.

As an open-source tool, Omeka is low cost and does not involve complex permission and copyright issues. Its item format, with Dublin Core metadata fields and the ability to attach files to the items, is well suited to the source material, in which each letter and draft forms an individual document consisting of facsimile images and transcription encoded in XML/TEI5. Omeka's collections feature makes it easy to organize the source material into collections according to letter recipients.

A number of plugins available for Omeka provide added functionality for importing and displaying documents. The CSV import plugin, combined with a simple XSLT script, enables mass import of documents together with the image and TEI files attached to each document. The SolrSearch plugin, built on Apache's Solr, provides an open text search that encompasses both the metadata fields and the transcription. Some image viewer plugins are useable for displaying the facsimile images for each document.

TEI5 enables detailed encoding of the source material. However, the project group decided on a light encoding of the transcriptions. Lönnrot's own underlinings, additions and deletions of text, and unclear and undecipherable parts of the transcriptions are marked with TEI tags. Information contained in the transcriptions, such as personal and place names, is left unmarked. Similarly, for example different kinds of additions of text (above lines, in the margins etc.) are not differentiated. This is partially to do with the extensive amount of manual work such detailed encoding require, and partially with the functionality provided by the publication. The open text search provides easier and quicker access to the same information as the encoding would. The TEI documents will be made available as free downloads, which enables researchers and other users to modify them for their own purposes.

A researcher who uses digitized letters as source material is faced with some challenges related to the material and its context. We ask what kind if information is lost in using online publications, where the materiality of the letters and their connection to the real-life physical objects in the archive are weakened. Can a good interface help convey information about the original letters and the entity they form in the archive? Just like the archive can be a place hiding, concealing and covering documents if nobody looks for them, a digital edition also needs an active user whose questions render the documents meaningful. Issues concerning context and contextualization and their relation to the digitization and to the online presentation of archival material are therefore of great importance. A researcher needs to be able to place the archival material in a wider historical and cultural context in order to make sense of it. To make digitized material understandable and meaningful, we need to provide as much contextual information as possible, e.g. precise information on the original archival material and the processes of selection, digitization, and edition it has undergone. Also contextual information on the text of the letters is required to help the reader understand the meanings within the text itself and the circumstances of letter writing. Lönnrot's letters have been available and well catalogued in the archives of SKS, but our comprehension of his correspondence is filtered through selected publications such as *Journeys of Elias Lönnrot* (1902), his biography (1931, 1935) and *Selected works of Elias Lönnrot* 1*: Letters* (1990). Therefore, we will reflect on how our perceptions could be widened, and possibly changed, by the complete digital edition of the correspondence.

We will demonstrate how a hypothetical researcher might use our online publication as a tool to access Lönnrot's letters and find answers to questions related to his/her research problem. We will show the benefits the tool offers in comparison to the traditional methods of accessing this kind of

source material, as well as address the potential limitations that might arise from the technical solutions adopted. The benefits and potential limitations are related to how the material is displayed and what kind of search tools are provided. Are the digitized letters and their transcriptions easily accessible and are features such as zooming in on the facsimiles or moving from page to page within a letter easy to use? How do the search options help the researcher find the information he/she needs? How can we help scholars make new interpretations based on digitized material?

We'll finish with the challenges of building platforms and interfaces for the multidisciplinary scholarly community. We have opted for an interface designed with the cultural historian in mind rather than focusing on Lönnrot himself. Thus the letters are published "vertically", all the letters to a particular addressee at a time, instead of "horizontally", year by year. We know that linguists are interested in the letters as well, but instead of attempting to build an interface that caters for them too, SKS will share the data with Finn-Clarin and the Language Bank, where linguists can use it together with other similar materials in Finnish and in Swedish. As Lönnrot's letters form an exceptionally vast collection of manuscripts written by one hand, we are handing part of the letters together with their transcriptions over to THE READ project (Recognition and Enrichment of Archival Documents). And finally, we are co-operating with the project STRATAS – Interfacing structured and unstructured data in sociolinguistic research on language change.

As a significant part of the Lönnrot letters are written in Swedish, we hope to find ideas for Nordic co-operation in Göteborg.

Topics: Nordic Textual Resources and Practices
Keywords: Online publishing, open science, correspondences

Tagging Named Entities in 19th century Finnish Newspaper Material with a Variety of Tools

Kimmo Kettunen
Teemu Ruokolainen
The National Library of Finland

Introduction

Digital newspapers and journals, either OCRed or born digital, form a growing global network of data that is available 24/7, and as such they are an important source of information. As the amount of digitized journalistic information grows, also tools for harvesting the information are needed. Named Entity Recognition (NER) has become one of the basic techniques for information extraction of texts since the mid-1990s (Nadeau and Sekine, 2007). In its initial form NER was used to find and mark semantic entities like person, location and organization in texts to enable information extraction related to this kind of material. Later on other types of extractable entities, like time, artefact, event and measure/numerical, have been added to the repertoires of NER software (Nadeau and Sekine, 2007). In this paper we report evaluation results of NER for historical 19th century Finnish. Our historical data consists of an evaluation collection out of an OCRed Finnish historical newspaper collection 1771–1910 (Kettunen and Pääkkönen, 2016).

Kettunen et al. (2016) have reported first NER evaluation results of the historical Finnish data with two tools, FiNER and ARPA. FiNER is provided by the Fin-CLARIN consortium, ARPA is a semantic web tool produced by the Semantic Computing group at the Aalto University. Both tools achieved maximal F-scores of about 60 at best, but with many categories the results were much weaker. Word level accuracy of the evaluation collection was about 73 percent, and thus the data can be considered very noisy. NER results for modern Finnish have not been reported extensively so far.

Silfverberg (2015) mentions a few results in his description of transferring an older version of FiNER to a new version. With modern Finnish data F-scores round 90 are achieved.

In this paper we add two more analysis tools to our earlier NER repertoire. Finnish Semantic Tagger (FST) is not a NER tool as such; it has first and foremost been developed for semantic analysis of full text. The FST assigns a semantic category to each word in text employing a comprehensive semantic category scheme (USAS Semantic Tagset, available in English[1] and also in Finnish[2]; Löfberg et al., 2005). The scheme contains three name related categories: persons, locations and organizations. Our other new tool is Connexor's NER software[3], which is a commercial tool for modern Finnish.

Results for the Historical Data

Our historical Finnish evaluation data consists of 75 931 lines of manually annotated newspaper text, one word per line. Most of the data is from the last decades of 19th century. Earlier NER evaluations with this data have achieved at best F-scores of 50–60 in some name categories (Kettunen et al., 2016). Our baseline tagger, FiNER, is described more in Kettunen et al. (2016). Shortly described, it is a rule-based NER tagger that uses morphological recognition, morphological disambiguation, gazetteers (name lists), pattern and context rules for name tagging.

We evaluated performance of our different NER tools using the *conlleval*[4] script used in Conference on Computational Natural Language Learning (CONLL). *Conlleval* uses standard measures of precision, recall and F-score, the last one defined as 2PR/(R+P), where P is precision and R recall (Manning and Schütze, p. 269). As the FST and Connexor's tagger do not distinguish multipart names with their boundaries only a comparable loose evaluation without entity boundary detection is reported here (Poibeau and Kosseim, 2001).

Table 1 shows F-score results of four evaluations of locations and persons in our evaluation data. *EnamexPrsHums* contain both first names and last names; *EnamexLocXxx* is a general location category that combines three more refined location categories to one.

	EnamexPrsHum		**EnamexLocXxx**	
	F-score	*Number of found tags*	*F-score*	*Number of found tags*
ARPA	52.9	3636	52.4	2933
Connexor	56.4	5321	**60.9**	1802
FiNER	**58.1**	2681	57.5	1541
FST	51.1	1496	56.7	1253

Table 1. *Evaluation of four tools with loose criteria and two name categories in the historical newspaper collection. Best results are in bold.*

All taggers recognize locations and persons quite evenly, differences are small. Our baseline tagger FiNER achieves best F-score with persons, Connexor with locations. Performance of the taggers is quite bad, which is expectable as the data is very noisy.

It is evident that the main reason for low NER performance of the tools is the quality of the OCRed texts. If we analyze the tagged words with a morphological analyzer (Omorfi v. 0.3[5]), we can see that wrongly tagged words are of lower quality than those that are tagged correctly. Figures are shown in Table 2. Thus improvement in OCR quality will most probably bring forth a clear improvement in NER of the material.

[1] http://ucrel.lancs.ac.uk/usas/USASSemanticTagset.pdf

[2] https://github.com/UCREL/Multilingual-USAS/raw/master/Finnish/USASSemanticTagset-Finnish.pdf

[3] https://www.connexor.com/nlplib/?q=technology/name-recognition

[4] http://www.cnts.ua.ac.be/conll2002/ner/bin/conlleval.txt, author ErikTjong Kim Sang, version 2004-01-26

[5] https://github.com/flammie/omorfi

	Locations	Persons
ARPA right tag, word unrecognition rate	1.9	4.5
Connexor right tag, word unrecognition rate	10.2	25.0
FiNER right tag, word unrecognition rate	6.3	12.8
FST right tag, word unrecognition rate	5.6	0.06
ARPA wrong tag, word unrecognition rate	22.7	29.3
Connexor wrong tag, word unrecognition rate	53.5	57.4
FiNER wrong tag, word unrecognition rate	38.3	34.0
FST wrong tag, word unrecognition rate	44.0	33.3

Table 2. *Unrecognition rates for rightly and wrongly tagged words, percent.*

Development of a New Statistical Tagger

Our baseline tagger FiNER employed in the above experiments is a rule-based system utilizing morphological analysis, gazetteers, and pattern and context rules. However, while there does exist some recent work on rule-based systems for NER (Kokkinakis et al., 2014), the most prominent research on NER has focused on statistical machine learning methodology for a longer time (Nadeau and Sekine, 2007; Neudecker 2016). Therefore, we are currently developing a statistical NER tagger for historical Finnish text. For training and evaluation of the statistical system, we are manually annotating newspaper and magazine text from the years 1862–1910 with classes *person*, *organization*, and *location*. The text contains approximately 650,000 word tokens. Subsequent to annotation, we can utilize freely available toolkits, such as the Stanford Named Entity Recognizer (Finkel et al., 2005), for teaching the NER tagger. We expect that the rich feature sets enabled by statistical learning will alleviate the effect of poor OCR quality on the recognition accuracy of NEs. For recent work on statistical learning of NER taggers for historical data, see Neudecker (2016).

Discussion

In this paper we have shown results of NE tagging of historical OCRed Finnish with four tools: FiNER ARPA, a Finnish Semantic Tagger, the FST, and Connexor's NE software. FiNER and Connexor's tagger are dedicated NER tools for modern Finnish, but the FST is a general semantic tagger and ARPA a semantic web linking tool. Our results show that they all tag names of locations and persons almost at the same level in the noisy OCRed historical newspaper collection. FiNER is best with names of persons, Connexor with locations. Differences between tagger performances are at biggest 7–8 % points.

In general our results show that NE tagging in a noisy historical newspaper collection can be done to a reasonable extent with tools that have been developed for modern Finnish. Anyhow, it seems obvious, that better results could be achieved with a new tool, which is trained with the noisy historical data. We have ongoing development work with regards to this. We also try to improve the quality of our OCRed text data with new OCRing and post-correction. Together these should yield better NER results in the future.

Finally, a note about usage of Named Entity Recognition is in order. Named Entity Recognition is a tool that needs to be used for some useful purpose. In our case extraction of person and place names is primarily a tool for improving access to the Digi collection. After getting the recognition rate of some NER tool to an acceptable level, we need to decide, how we are going to use extracted names in Digi. Some exemplary suggestions are provided by the archives of La Stampa[6] and Trove Names (Mac Kim and Cassidy, 2015). La Stampa style usage of names provides informational filters after a basic search has been conducted in the newspaper collection. User can further look for persons, locations and organizations mentioned in the article results. This kind of approach enhances browsing access to the collection (Bates, 2007; McNamee, Mayfield and Piatko, 2011; Toms, 2000). Trove Names' name search takes the opposite approach: user searches first for names and then gets articles where the names occur. We

[6] http://www.archiviolastampa.it/

believe that La Stampa style usage of names in the GUI of a newspaper collection is more informative and useful for users, as the Trove style can be achieved with the normal search function in the GUI of the newspaper collection.

Our main emphasis with NER will be to use the names with the newspaper collection as a means to improve structuring, browsing and general informational usability of the collection. A good enough coverage of the names with NER needs to be achieved also for this use, of course. A reasonable balance of P/R should be found for this purpose, but also other capabilities of the software need to be considered. These remain to be seen later, if we are able to connect functional NER to our historical newspaper collection's user interface.

Acknowledgements
This work is funded by the EU Commission through its European Regional Development Fund, and the program Leverage from the EU 2014–2020.

Topics: Nordic Textual Resources and Practices
Keywords: named entity recognition, historical newspaper collections, Finnish

References

Bates, M. (2007). What is Browsing – really? A Model Drawing from Behavioural Science Research. *Information Research* 12. http://www.informationr.net/ir/12-4/paper330.html.

Finkel, J.R., Grenager, T. and Manning, C. (2005). Incorporating non-local information into information extraction systems by Gibbs sampling. In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL 2005), 363–370, available at http://dl.acm.org/citation.cfm?id=1219885.

Kettunen, K., Mäkelä, E., Kuokkala, J., Ruokolainen, T. and Niemi, J. (2016). Modern Tools for Old Content - in Search of Named Entities in a Finnish OCRed Historical Newspaper Collection 1771-1910. LWDA 2016, available at: http://ceur-ws.org/Vol-1670/paper-35.pdf

Kettunen, K. and Pääkkönen, T. (2016). Measuring Lexical Quality of a Historical Finnish Newspaper Collection – Analysis of Garbled OCR Data with Basic Language Technology Tools and Means. In *LREC 2016, Tenth International Conference on Language Resources and Evaluation*, available at http://www.lrec-conf.org/proceedings/lrec2016/pdf/17_Paper.pdf.

Kokkinakis, D., Niemi, J., Hardwick, S., Lindén, K., and Borin. L. (2014). HFST-SweNER – a New NER Resource for Swedish. In *Proceedings of LREC 2014*, available at: http://www.lrec-conf.org/proceedings/lrec2014/pdf/391_Paper.pdf.

Löfberg, L., Piao, S., Rayson, P., Juntunen, J-P, Nykänen, A. and Varantola, K. (2005). A semantic tagger for the Finnish language, available at http://eprints.lancs.ac.uk/12685/1/cl2005_fst.pdf.

McNamee, P., Mayfield, J.C., and Piatko, C.D. (2011). Processing Named Entities in Text. *Johns Hopkins APL Technical Digest*, 30, 31–40.

Mac Kim, S., Cassidy, S. (2015). Finding Names in Trove: Named Entity Recognition for Australian. *In Proceedings of Australasian Language Technology Association Workshop*, available at https://aclweb.org/anthology/U/U15/U15-1007.pdf.

Manning, C. D., Schütze, H. (1999). *Foundations of Statistical Language Processing*. The MIT Press, Cambridge, Massachusetts.

Nadeau, D., and Sekine, S. (2007). A Survey of Named Entity Recognition and Classification. *Linguisticae Investigationes,* 30(1): 3–26.

Neudecker, C. (2016). An Open Corpus for Named Entity Recognition in Historic Newspapers. In *LREC 2016, Tenth In-*

ternational Conference on Language Resources and Evaluation, available at http://www.lrec-conf.org/proceedings/lrec2016/pdf/110_Paper.pdf .
Poibeau, T. and Kosseim, L. (2001). Proper Name Extraction from Non-Journalistic Texts. *Language and Computers,* 37(1): 144–157.
Silfverberg, M. (2015). Reverse Engineering a Rule-Based Finnish Named Entity Recognizer. Paper presented at Named Entity Recognition in Digital Humanities Workshop, June 15, Helsinki available at: https://kitwiki.csc.fi/twiki/pub/FinCLAR-IN/KielipankkiEventNERWorkshop2015/Silfverberg_presentation.pdf
Toms, E.G. (2000). Understanding and Facilitating the Browsing of Electronic Text. *International Journal of Human-Computer Studies*, 52, 423–452.

The Digital Experience: Technology and Representation

Lars Kristensen
University of Skövde, Sweden
Graeme Kirkpatrick
University of Manchester

The research presented is part of an ongoing project regarding post-critique and digital culture. It is our objective to describe an experience that is particular to the digital, an aesthetics that is only achieved with/through digital paraphernalia. Media histories tell us that newness has always been projected as giving some kind of vertiginous bodily experience; an experience in which our bodies are too slow to react to the represented or too inexperienced to follow through, displacing or disrupting our sense of being in control. These experiences are part of capitalist celebrations of new technologies where humans can test their readiness for capitalist projected future realities. Currently it is various kinds of different 'realities' (AR, VR, MR) that are flocked as 'affordable' for consumers in order to promote 'new' experiences. We reject this narrative and argue that these 'new' experiences are as old as their ancestors, but argue that the digital experience is in fact not that different from a real experience.

As a consequence of this experience of non-difference, which can be summed up as indifference to different realities, we observe a deflating critical effect in the digital experience as it takes us closer to the real. We no longer only occasionally escape from reality, but reality is also the escape. The ever presence of the digital escape in our everyday life is the escape from critique. We argue that we have arrived at an understanding of the digital experience as near real or as a slightly alternated state of reality, but that in this bodily state of being, an embodied 'gratuitous' experience (Williams 1991), the meaning of critique becomes pointless, since our sense of the real and the represented has narrowed to indistinctive proportions. We are resigned to a state of a frozen standstill (Benjamin 1999) where critical distance allows for minimal or no reflections. This means that here is no end to the escape but only continuous digital escapes into bodily experiences.

Viewed in this way, critique is part of the problem of digital media and not the first step towards any kind of solution. Rather than constituting a special, epistemically privileged vantage point it should be part of what we are talking about when we begin media analysis. Just as we should attend to the structure of feeling that makes certain kinds of media possible (Williams 1968), so we must include in that reading an account of the critical interpretation that articulates and manages the feeling responses of audiences. Our approach to digital media starts with the idea that we need to give up the idea of critique as a standpoint outside media and that this step is essential to what digital media are; to their specific difference from older forms.

The hazard in such an approach is that we might appear to be advocating a kind of quietism or a post-modern, post-political

celebration of the popular. However, we retain a political concern with how media serve contemporary domination. Our theoretical perspective is informed by Jacques Rancière's (2007) notion that media do not so much deceive or mislead their audiences as they cleave them to a particular perception of reality, namely, one that is partitioned, divided. Viewed in this way media are not deceptive or even manipulative, as critique would have it; their divisiveness – their violence – is of a different order. It consists in their articulation, or better, superimposition, of the sensed and the sensible.

Digital technology

In this part we generalize some of the findings of contemporary sociology of art (Heinich 1998; Hennion 1995) onto modern media more generally. We will argue that, viewed in this perspective, much of Adorno's theory of media and culture has been banalised. Where he wrote of a loss of authentic experience, it is clear that all media are now experiential, obliging us to act and to think. Where he identified the subject-object relation as the locus of 'critical tensions', and charged the modern subject with the 'critical' task of breaking itself apart in order to let objects speak, we now struggle to shut them up. Where critical theory targeted a monolithic 'hegemonic technological rationality' (Feenberg 2002), contemporary media accommodate a variety of thinking styles and present diverse openings to experience of the incommensurate, of that which does not fit.

To make this argument, we will use the example of the computer game, *Dark Souls* (2009-2012). In its difficult complexity, in the fact that it demands a response from the player to become anything at all, and in the obscurity of its literal meaning this game resembles the modernist work of art. Unlike those works, though, the video game does not require theory to redeem its meaning. Playing through the game is an experience of cleaving form from the dark matter of the digital machine but, detached from anything like critique, this form does not oppose the tyranny of meaning or subvert a hegemonic codification. At best, the experience we have with a computer game may separate players from everyday world-orderings and oblige them to question what makes the experience cohere. This reflects a shift in the relation of subject and media object more generally, in which their relationship is more symmetrical and entwined. In consequence 'art' loses its privileged status and 'mediatisation' loses its association with 'top-down' forms of domination (Kirkpatrick 2011).

Digital representation

In this part, we explore the indistinction of representation and reality in the digital experience. Being post-critical means that the world is not 'out there' beyond media and represented by them but rather constituted by being mediated through technology. More precisely, it is in moving between mediatic instants that we pull together a world, or worlds of experience, out of what is. Reality here is what gets instituted (Latour 2013: 280) or what is included 'as one' in the count (Badiou 2007). Under these circumstances reality is in between fiction and authentic life: in a strange and anti-climactic fulfillment of surrealism, the element of pretense is a recognized part of real life.

This situation corresponds to what Hal Foster calls the new Alexandrine, in which previously sharpened differences have become "stagnant incommensurabilities" (Foster 2004: 28). The infinite complexity of contemporary mediation, in which fantasy permeates reality and playful subjects disappear into a nether world where life is never fully life and death is only temporary, also merits description as neo-baroque (Ndalianis 2003). We will explore these ideas through the case studies of *This Is Not a Film* (2011) by Jafar Panahi. In this film, Panahi reenacts scenes from his earlier films in his living room in Teheran, while being under house arrest for alleged crimes against national security. We will discuss the film in relation to contemporary media where everyone and no one is an 'actor'; everyone, including Panahi, knowingly acts out a role that is and is not them. The example illustrates the way in which our experience of the

inchoate, or of being as difference, works because it does not work: it is our stumbling failures and awkward mistakes that bridge the gaps and produce a world, as well as it is Panahi's. Moreover, this activity is something we find thematised by 'un-critical' participants in media, whose activity instantiates the contemporary social imaginary.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: theory of aesthetics, technology, experience, games, film

References

Badiou, A. (2007) Being and Event Trans. O. Feltham. London: Continuum.

Benjamin, W. (1999) The Arcades Project, Cambridge, Massachusetts: Harvard University Press

Feenberg, A. (2002) Transforming Technology Oxford: Oxford University Press.

Foster, H. (2004) Design and Crime (and other diatribes) London: Verso.

Heinich, N. (1998) The Glory of Van Gogh: an anthropology of admiration Trans. P. L. Browne New Jersey: Princeton University Press.

Hennion, A. (1995) La Passion Musicale Paris: Éditions Métailié.

Kirkpatrick, G. (2011) Aesthetic Theory and the Video Game Manchester: Manchester University Press.

Latour, B. (2013) An Enquiry into Modes of Existence: an anthropology of the moderns Trans. C. Porter New York: Harvard University Press.

Ndalianis, A. (2003) Neo-baroque Aesthetics and Contemporary Entertainment London: MIT Press.

Ranciere, J. (2009) The Future of the Image Trans. G. Eliott. London: Verso.

Williams, L (1991) 'Film Bodies: Gender, Genre, and Excess', Film Quarterly, 44(4) (Summer), pp. 2-13

Williams, R. (1968) The Long Revolution London: Chatto and Windus.

The Corpus of American Danish: A Corpus of Multiligual Spoken Heritage Danish and Corpus-based Speaker Profiles as a Way to Tackle the Chaos

Karoline Kühl
Jan Heegård Petersen
Gert Foget Hansen
University of Copenhagen, Denmark

During the last years (2014– 2016), we have overcome a number of challenges while establishing the Corpus of American Danish (CoAmDa) within the project 'Danish Voices in the Americas' at the University of Copenhagen. The challenges have emerged from the digitization and integration of legacy data, the challenges of transcribing non-standardized data (with regard to spokenness and multilingual language use) and, finally, from using the corpus for research of language variation and change in emigrant and heritage speakers.

This will give a presentation of CoAmDa and its sub-corpora, the Corpus of North American Danish (data from Canada and the USA) and the Corpus of South American Danish (data from Argentina), as a newly established linguistic resource in the Nordic region. We aim at presenting and discussing corpus-based sociolinguistic speaker profiles as a newly developed tool for coping with a huge number of speakers who diverge massively with regard to their speech production.

As of December 2016, the CoAmDa amounts to 1.47 million tokens (165 hrs) produced by 264 speakers (born between 1876 and 1965). Basic speaker metadata like gender, birth year, time of emigration, home area in homeland and residence at the time of the recording are available for most speakers. The data is orthographically transcribed, aligned with sound, PoS-tagged, and the CoAmDa is currently being annotated with a basic syntactic annotation (main clauses, declarative clauses, subject, finite

verb and sentence adverbials). During the transcription process, words were coded according to language used. To make the coding as time efficient as possible, automated procedures were developed to the effect that transcribers needed only code language use for words in languages other than the designated default language (in our case Danish), such as English or Spanish, as well as word-internal switching (occurring either between stems or between stem and suffixes) and words that could not unambiguously be assigned to one language due to an intermediate pronunciation (e.g., Danish *søster* and English *sister*).

A subsequent automated procedure assigned a language code to the remaining words, based on the designated default language, and language codings based on the orthographic transcription. Some words were categorized by common traits such as interrupted words beginning or ending in a dash (-) and proper nouns by being capitalized. Some words belonging to (semi-)closed sets such as discourse markers were coded based on a small lookup table. Based on this preliminary coding of language use, dictionaries (comprehensive word-language-category lookup tables) for either language combination were generated which were then meticulously proofread. Two different dictionaries were necessary, since the words that contain word-internal language switching and/or those that are ambiguous with regard to language assignment are not the same between Danish vs. English and Danish vs. Spanish. The final language coding was arrived at by automatically checking the preliminary language coding against the dictionaries, thus ensuring a very high accuracy of the language coding.

For future projects, the dictionaries created in this process may facilitate (semi-) automated designation of language use in transcriptions of Danish mixed with either English or Spanish.

Emigrant and heritage speakers are notorious for the variation in language competence and language production (Polinsky & Kagan 2007). Accordingly, we observe a great deal of variance within the language production of the North American Danes and the Argentine Danes: The variance concerns fluency (i.e. amount of empty and filled pauses, hesitation phenomena, restarts), the amount of word-internal codeswitching and code shifting between utterances as well as the kind and amount of non-Danish linguistic variables (e.g., the use of 'English' word order in American Danish, see Kühl & Heegård Petersen to appear). Within the tradition of Labovian sociolinguistics, our aim is to look for ties between this variance in speech production and the sociolinguistic variables (i.e. speaker metadata). Ties between language use and sociolinguistic characteristics cannot be assumed per se in research on emigrant and heritage speakers, as the emigration process typically will result in a mix-up of the connections between linguistic variables and sociological characteristics established in the homeland. Non-intuitive knowledge about the speakers' speech production, including differences in lemma production, amount and type of code-switching, speech rate, hesitations repetitions and restarts, is thus a desideratum.

In order to establish these connections anew for the speakers in the corpus, we draw on the above-mentioned transcription and annotations, i.e. codings of language choice, syntax, pauses and hesitations, in combination with the speaker metadata. Using these data to create corpus-based sociolinguistic speaker profiles provides us with the possibility of recognizing patterns across the language production of many speakers, but also in the language production of single speakers. The paper will demonstrate how this tool enables a more objective assessment of a speaker's or a group of speakers' language production and competence, in addition to presenting the Corpus of American Danish as a linguistic resource and discussing the process of establishing the corpus.

Topics: Nordic Textual Resources and Practices
Keywords: corpus, multilingualism, spoken language

References

Kühl, Karoline & Jan Heegård Petersen (2016) Ledstillingsvariation i amerikadanske hovedsætninger med topikalisering. In *Ny forskning i grammatik* 23. Available online at http://ojs.statsbiblioteket.dk/index.php/nfg/article/download/24650/21598.

Polinsky, Maria & Olga Kagan (2007) Heritage Languages: In the 'Wild' and in the Classroom. In *Languages and Linguistics Compass* 1 (5), 386-395.

Rhythms of Fear and Joy in Suomi24 Discussions

Krista Lagus
Mika Pantzar
Minna Ruckenstein
University of Helsinki, Finland

Suomi24 is the largest social media platform in Finland. Every day, over 15,000 messages are posted in its nearly 2,900 discussion groups. In its slightly more than 15 years of existence, about 80 million messages in total have been posted on Suomi24, of which we have had access to about 56 million (excluding, for example, deleted messages; see the material description by Lagus et al. 2016). The readers and the posters of the messages represent the Finnish population well in their geographical distribution, for example. In this presentation, we contemplate how and to what extent it is possible to recognise emotional movement and rhythm from this kind of material.

Over the past few years, sentiment analysis (Abbasi 2008; Liu 2010; Honkela et al. 2014) has become an established method of describing emotional states by means of social media content. Twitter or Facebook posts are a window to recognising people's views on abortion, NATO, or political parties, among other things. Sentiment analysis can be used to examine the posters' subjective views on various topics, whereas topic mining (Purhonen & Toikka 2016; Winter & Wiberg 2016) is used to examine facts or themes subject to discussion. The sentiment analysis most often entails examining the negative/positive dichotomy in the material, as in reviews indicating good vs. bad (e.g., film or other product reviews), indications of shared/different opinion or approval/disapproval, and expressions of positive/negative emotional state.

With social media, the entity we call 'society' has become more visible, more tangible, and more concrete than before. Social media make visible and renew the process through which we collectively produce thoughts and thus reshape society. In social media conversations, emotions are transmitted from one participant and one discussion to another. This results in occasional emotional rushes and affective contagion. Often people's emotions become synchronised and eventually form shared rhythms, in a process called entrainment, and larger entities: recognisable wave motions.

We approach the social media discussions by means of rhythmanalysis (Lefebvre 2004), focusing on social rhythms produced by different practices and systems. These kinds of rhythms can be detected in cities and in people's biologies, resulting in fluctuations of stress and recovery that could be detected by means heart rate variability measurements (Pantzar et al. 2016). Our presentation considers whether it is possible to detect similar rhythms in social media text. We ask what kind of fluctuation can be detected in the emotional discourse between weekdays and weekends, and whether one can see systematic fluctuation depending on the time of day, as we witnessed in the above-mentioned stress data, collected from 35 people (the study revealed that on Saturdays, after a spike in early afternoon, stress declines towards the evening, whilst on weekdays and Sundays, the level of stress reaches its peak in the evening, and on weekdays also at 8–9am).

The presentation focuses on the approximately 56 million messages in the Suomi24 dataset. We describe how the lexicons for the respective emotion categories were chosen. We will then present time cycles calculated from the frequencies of words express-

ing fear and joy. Finally, we discuss to what extent our word-frequency examination in connection with emotional discourse represents meanings related to emotions and to experiencing emotions.

When discussing emotional discourse, we are especially interested in getting at long-term mood-like emotional states and also emotional discourse related to or contributing to the activation of these mood states. Previous efforts to recognise long-term emotional states or moods include studying of six basic emotions (Strapparava & Mihalcea 2008), recognition of eight basic emotions from Twitter data (Roberts et al. 2012), and recognition of emotional states representing five distinct aspects of well-being as represented in various types of news and discussion data (Honkela et al. 2014).

The method we applied comprises the following components:

1. Identification of the emotional states subject to study and their more precise definition using a vocabulary-based approach.

2. Evaluating the preliminary emotion features qualitatively using Korp tool (Borin et al., 2012) and adjusting the set of features accordingly.

3. Calculating the averaged occurrences of emotion expressions over various time cycles and visualizing them.

As a starting point, we selected for observation two quite distinct categories of emotional discourse, which appear as somewhat dichotomous: categories of fear ('Fear/Worry') and joy ('Joy/Happiness'). We chose fear and worry to be the starting point because we deemed them likely to have more multifaceted contexts than other considered moods such as sadness or anger and, hence, to provide a richer target of analysis. Moreover, the mood of fear and worry may function as a seedbed of social aggressions.

In our dataset, the ways in which both the Fear/Worry lexicon and the Joy/Happiness lexicon manifest themselves show distinct differences by season, day of the week, and time of day. On a monthly level (see Figure 1, at the end of the text), the Joy/ Happiness discourse seems to follow the rhythm of holidays. Joy/Happiness discourse thrives throughout the summer, particularly in July. We can see another peak over Christmas. In contrast, the Fear/Worry discourse is more evenly distributed over the year, but peaks in September. We might ask whether these figures reflect the real-world seasonal stress curve. With the Fear/Worry spike in September and the concurrent low point for the Joy/Happiness curve, we could ask whether there is a spike in the population's stress level in September, as the working population have returned to work after the summer holidays and as schools too are running as usual. Are there any other studies that would indicate September to be a particularly challenging time in people's lives? A similar pattern is found in January, after Christmas.

On a weekly level (see Figure 2, at the end of the text), on Sundays, both the Joy/Happiness discourse and the Fear/Worry discourse reach their highest. The former is at its lowest on Thursdays, and the Fear/Worry discourse reaches its lowest point on Saturdays. Thursday is a common banking day – that is, when people pay their bills. We might also consider the possible relationship of alcohol use to these results, especially in the case of Saturday evening. Is the spike in the Joy/Happiness curve at around 9pm on Saturdays a sign of a 'buzz' and the spike in the Fear/Worry curve on Saturday night a sign of coming down from that 'high'?

On an hourly level, the Joy/Happiness discourse is at its highest slightly before midnight, after which the curve keeps dropping until 10am. After that, the proportion of Joy/Happiness discourse starts to increase slowly. With the Fear/Worry discourse, we see a much clearer hourly rhythm. In the discussions, there is a concentration of fear and worry between 2 and 4am. This can be explained by the fact that the group of people taking part in the discussion at these hours is very limited and specific. For instance, the messages are longer than average and the number of participants is far below the daily average (Lagus et al. 2016).

Discussion

The emotional discourse in the climate of interaction surrounding us tunes us in to the emotional atmosphere that we live in, and it affects the emotional nature of our actions in the following days. Hence, an interesting empirical question arises: what kind of emotional discourse does this nation collectively produce and read on Internet fora? Empirical research directed to such issues might give us a stronger foundation for answering questions about the recent evolution of the Finnish social climate or emotional terrain.

As shown by Facebook's study of the contagious nature of emotions expressed in the virtual world, once people read messages of either positive or negative polarity in their news feed, it affects the positive and negative expressions in their own messages in the following days (Kramer 2012; Kramer et al. 2014). What the illustrations calculated here inform us of is the emotional landscape of a particular discussion forum against cyclic time - a natural way to think about our own life. While we might not deduce the entire emotional state of a nation from these figures, nevertheless we might consider, when is it emotionally most beneficial for ourselves to participate in social media discussions. Furthermore, these figures might suggest times when to add particular support for those who might need it most. Much like a rhythm map of a city traffic might depict times of road rage versus polite and calm, this mapping depicts an emotionally informed timescape of virtual social climate.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: social media, social sciences, sentiment analysis

Bibliography

Abbasi, A., Chen, H., & Salem, A. (2008). Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums. ACM Transactions on Information Systems (TOIS), 26(3), 12.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. Wired, 7/2008. Available at http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory.

Borin, L., Forsberg, M., & Roxendal, J. (2012). Korp – the corpus infrastructure of Spräkbanken. In: Proceedings of LREC 2012 (pp. 474–478).

Honkela, T., Korhonen, J., Lagus, K., & Saarinen, E. (2014). Five-dimensional sentiment analysis of corpora, documents and words. In: Advances in Self-organizing Maps and Learning Vector Quantization (pp. 209–218). Springer International Publishing.

Kramer, A.D. (2012). The spread of emotion via Facebook. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 767–770). ACM.

Kramer, A.D., Guillory, J.E., & Hancock, J.T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24), 8788–8790.

Lagus, K.H., Pantzar, M., Ruckenstein, M.S., & Ylisiurua, M.J. (2016). Suomi24 Muodonantoa aineistolle ['Suomi24: Shaping the data']. Valtiotieteellisen tiedekunnan julkaisuja ['Publications of the Faculty of Social Sciences'], 10, May 2016. Helsinki: Unigrafia. 44 pages.

Lefebvre, H. (1992/2004). Rhythm Analysis, Space, Time and Everyday Life (Athlone Contemporary European Thinkers series), transl. by S. Elden and G. Moore. London: Continuum.

Liu, B. (2010). Sentiment analysis and subjectivity. In: Handbook of Natural Language Processing, Vol. 2 (pp. 627–666).

Pantzar M, Ruckenstein M, Mustonen V.(2016). Social rhythms of the heart. In: Health Sociology Review, forthcoming, http://dx.doi.org/10.1080/14461242.2016.1184580

Purhonen, S. & Toikka, A. (2016). ”Big datan” haaste ja uudet laskennalliset tekstiaineistojen analyysimenetelmät. Esimerkkitapauksena aihemallianalyysi tasavallan presidenttien uudenvuodenpuheista 1935–2015 [‘The challenge of “big data” and new computational methods for text analysis: An example from a topic model of New Year’s speeches of Finnish Presidents, 1935–2015’]. Sosiologia [‘Sociology’], 53(1), 6–27.

Roberts, K., Roach, M.A., Johnson, J., Guthrie, J., & Harabagiu, S.M. (2012). EmpaTweet: Annotating and detecting emotions on Twitter. In: Proceedings of LREC 2012 (pp. 3806–3813).

Strapparava, C. & Mihalcea, R. (2008). Learning to identify emotions in text. In: Proceedings of the 2008 ACM Symposium on Applied Computing (pp. 1556–1560). ACM.

Winter, L. & Wiberg, M. (2016). Presidentin uudenvuodenpuheet: Kvantitatiivisen tekstianalyysin mahdollisuuksia [‘The Presidents’ New Year’s speeches: The possibilities of quantitative text analysis’]. Politiikka [‘Politics’], 58(1), 80–88.

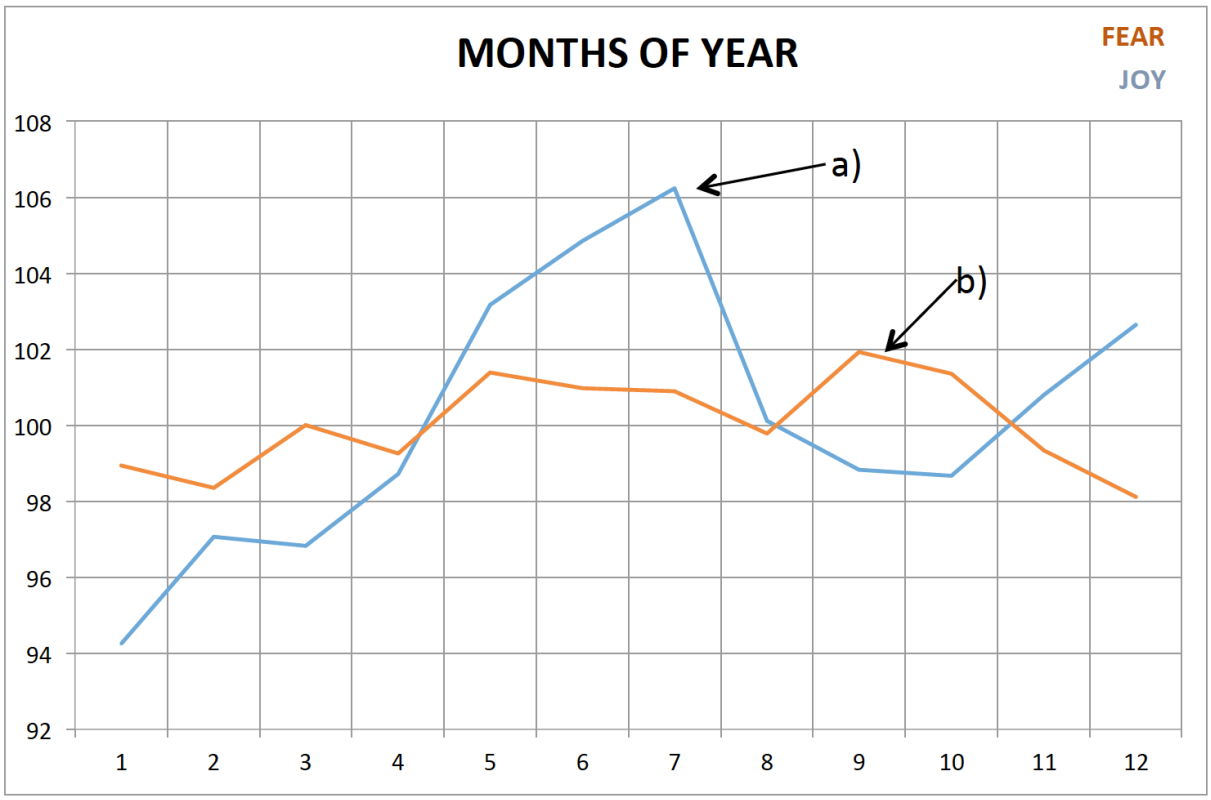

Figure 1: Proportions for Fear/Worry discourse and Joy/Happiness discourse by month in the Suomi24 dataset. a) marks the time when there is most joy compared to the level of fear, that is, July. b) indicates the opposite taking place in September, a peak of fear as opposed to the dropping level of joy. A similar situation can be seen in January.

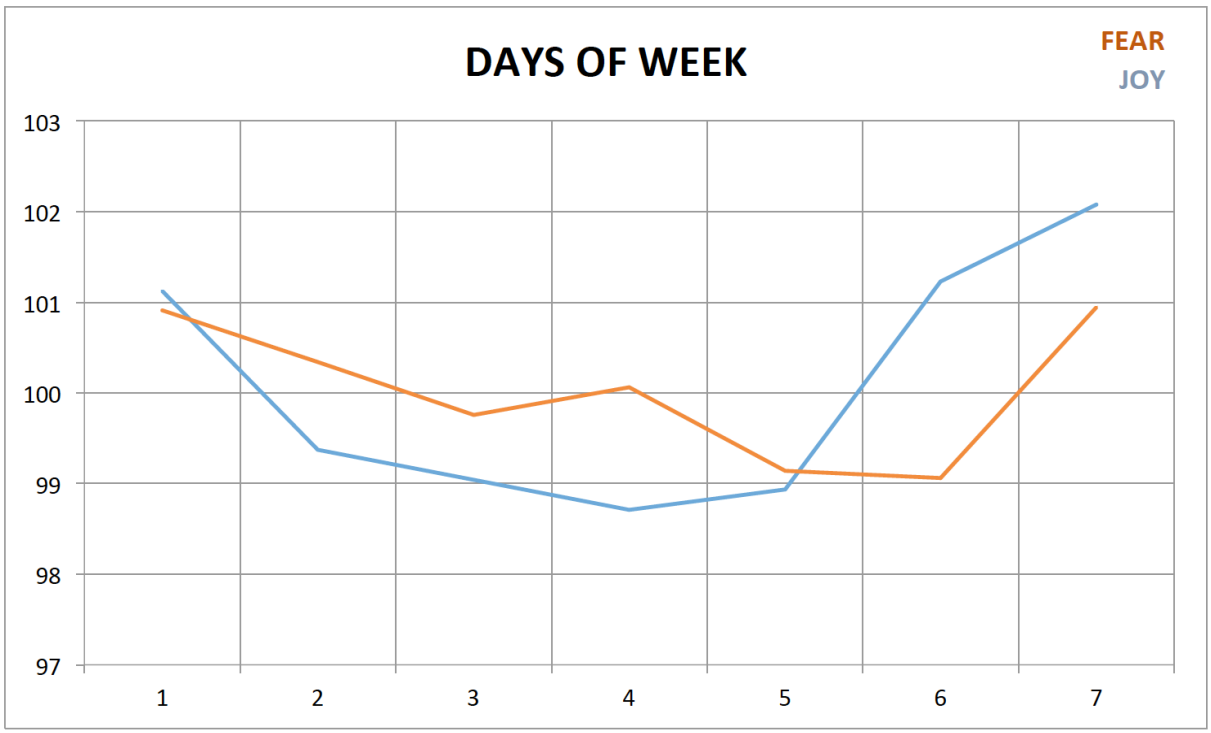

Figure 2: Proportions of Fear/Worry discourse and Joy/Happiness discourse in the data by day of the week.

Long-Range Information Dependencies and Semantic Divergence Indicate Author Kehre

Kristoffer Laigaard Nielbo
Aarhus University, Denmark
Katrine Frøkjær Baunvig
University of Southern Denmark

Introduction
Across academic disciplines that study literary and intellectual history there are ongoing discussions of if and when culturally important writers of fiction or non-fiction underwent a personal paradigm shift (a Kehre). The collected writings of philosopher Martin Heidegger, writer Milan Kundera, and theologian Martin Luther, all show indications of such a Kehre. The temporal identification of a Kehre however represents a significant methodological challenge. In this paper, we describe a novel approach to identification of change in the history of highly productive writers. The approach combines information theory and fractal analysis to substantiate claims about an intellectual Kehre as exemplified by the Danish liberal thinker, theologian and romantic writer N.F.S. Grundtvig (1783-1872).

Methods
The corpus consists of the collected writings of N.F.S. Grundtvig (N = 988). Shannon Entropy and Latent Dirichlet Allocation were used to model llexical density and semantic content, respectively, of each of Grundtvig's writings. Adaptive Fractal Analysis (AFA) was used to estimate the Hurst-exponent (i.e., a measure of Long-Range Dependencies in time series) of lexical density in windowed time slices (n = 589). Kullbach-Leibler (KL) divergence was applied to an LDA model's topic distributions in order to estimate time dependent semantic change.

Results
AFA indicated an 'early style' phase of persistent lexical innovation in Grundtvig's early writings, which is contrasted by otherwise short-term and anti-persistent dynamics. KL divergence, on the other hand, identified a short 'late style' phase of semantic innovation in Grundtvig's late writings.

Discussion
We argue that early and late style phases are signatures of one, or several, Kehren in Grundtvig's writings that reflect a combination of individual mental history and general cognitive development. At a methodological level, we argue that LRD and Information Theory can substantiate qualitative observations of paradigm shifts in cultural systems both at the micro and macro-level.

Topics: Nordic Textual Resources and Practices
Keywords: author development, long-range dependencies, information theory, culture analytics

Bibliography
Structural Differences Among Individuals, Genders and Generations as the Key for Ritual Transmission, Stereotypy, and Flexibility. / Nielbo, Kristoffer Laigaard; Fux, Michel ; Mort, Joel; Zamir, Reut; Eilam, David. In: Behaviour, 2016 (in press).

Traveling Companions Add Complexity and Hinder Performance in the Spatial Behavior of Rats. / Dorfman, Alex; Nielbo, Kristoffer Laigaard; Eilam, David.
In: PLoS ONE, 04.01.2016.

Segmentation and cultural modulation in perception of internal events are not trivial matters. / Nielbo, Kristoffer Laigaard; Andersen, Marc Malmdorf; Schjødt, Uffe.
In: Religion, Brain, and Behavior, 03.06.2016.

Attentional Resource Allocation and Cultural Modulation in a Computational Model of Ritualized Behavior. / Nielbo, Kristoffer Laigaard; Sørensen, Jesper. In: Religion, Brain, and Behavior, 2015.

Finnish Internet Parsebank – A Web-crawled Corpus of Finnish with Syntactic Analyses

Veronika Laippala
Aki-Juhani Kyröläinen
Jenna Kanerva
Juhani Luotolahti
Tapio Salakoski
Filip Ginter
University of Turku, Finland

This paper presents the Finnish Internet Parsebank (FIP), a freely available web-crawled corpus of Finnish, its user interface as well as some recent studies enabled by it.

The FIP consists of nearly 4 billion words automatically collected from the Web. It has full morphological and dependency syntax analyses. On the word-level, this includes the part-of-speech classes of the words and their morphological features (such as noun, singular and genitive), and on the sentence-level, the sentence structure and the syntactic functions of the words in it (such as nominal subject). These are marked following the Universal Dependencies (UD) scheme, a syntactic model seeking cross-linguistically consistent annotations and attested on 47 languages. The UD allows for novel insights to many linguistic research problems by enabling their study across languages. For instance, the characteristics of different texts can be analysed not only in one language but across several languages. Also, many language-technology applications, such as machine translation, profit from these harmonised markings.

The FIP is available through a user interface at http://bionlp-www.utu.fi/dep_search/and as a downloadable version, shuffled at the sentence-level at http://bionlp.utu.fi. The advanced user interface is described in detail in Luotolahti et al. (2015). The interface allows for the search of both individual words (such as boy), words with specific morphological or syntactic features (such as boy as the sentence subject) and the search of specific morphological and / or syntactic constructions regardless of their lexical realisation (such as all sentence subjects). Also searches with restrictions are possible (such as verbs without a subject). The interface returns the sentences including the searched expression as well as its linguistic contexts. The search hits can also be downloaded. In addition to the FIP, the user interface includes several other language resources, such as corpora in other languages following the UD scheme.

The currently available version of the FIP is in total composed of 3,662,727,698 words. These include 28,585,422 lemmata, 39,688,642 unique words and 275,690,022 sentences. The FIP was collected using two methods. First, all Finnish texts were detected from the 2012 release of the Common Crawl dataset. Common Crawl is a U.S. non-profit organisation that builds and maintains web-crawled data. Second, we launched a dedicated web crawl targeting Finnish data, not delimited to the .fi-domain. The crawl was realised using the SpiderLing crawler which is designed for collecting linguistic data.

For the linguistic analysis of the data, the raw texts were first segmented to sentences and words with a sentence splitter and tokenizer developed using the Apache OpenNLP toolkit trained on our previously manually developed language resource, the Turku Dependency Treebank (TDT) (Haverinen et al. 2014). The part-of-speech classes of the words and their morphological features were assigned with the Marmot tagger (Mueller et al., 2013), the morphological analyzer OMorFi (Pirinen, 2011) and a system transforming the OMorFi output to UD (Pyysalo et al., 2015). An evaluation of this analysis pipeline showed an accuracy of 97.0%, for the parts-of-speech and 94.0% for the full representation of the morphological features, which is comparable to the state-of-the-art results in other languages (Pyysalo et al. 2015). The dependency syntax analysis on the sentence structure is carried out using the parser of Bohnet et al. (2010) which is also trained on a version of TDT in the UD scheme. The parser performance is 81.4% labeled attachment score which indi-

cates the percentage of dependencies between the correct words with a correct dependency type.

Thanks to its size, linguistic variation and the syntactic analyses, the FIP allows for novel possibilities for all disciplines working on textual data. Among others, these advantages have allowed us to develop large-scale quantitative methods for a detailed linguistic analysis of the characteristics of both different kinds of texts and individual words or expressions, such as discourse markers expressing reactions or interaction (for instance, kyllä 'sure' and tokin 'certainly' (Laippala et al.2016, fortc.)). These methods allow as well the automatic identification of for instance machine translations and informal texts from the FIP (Laippala et al. 2015).

In addition, the FIP has been applied for the study of very rare linguistic constructions, typical in spoken or informal language varieties not necessarily found in traditional, manually collected corpora. For example, Huumo et al. (forthc.) explore the variation of an extremely rare and grammatically questionable syntactic construction, where a transitive sentence, i.e. a sentence with an object, includes a subject in the partitive case, as in Useita uudehkoja autoja on reputtanut tämän testin 'Several newish cars have failed this test'. Wessman (2016) studies the use of a novel syntactic construction typical of spoken language, where the subordinate conjunction koska 'because' is attached to a noun phrase instead of a verb phrase, as in I am tired, because the headache'.

Another line of investigation that utilizes the data available in the FIP concerns the representation of clausal semantics using neural networks, specifically, modeling the semantic fit of arguments in a transitive construction. The implemented neural network builds a semantic representation for a transitive construction based on word2vec (Mikolov, et al. 2013). We are currently testing the performance of the implemented model in several tasks such as cloze task and modeling reading times using eye-tracking. The results of current model appear promising as the model estimates are correlated to human preferences when asked to complete transitive constructions (cloze task) and reading times (eye-tracking) (Kyröläinen et al. 2016). The morphosyntactic analysis of the FIP allows us to model transitive construction and, importantly, even rare verbs occur with sufficient quantity that makes it possible to build semantic representations for them.

Finally, the FIP data has been used to improve the language technology available for the Finnish language, especially machine translation (MT). A better language model as well as a reinflection generation model were induced from the FIP data, resulting in an improved MT performance especially for the English to Finnish direction. (Tiedemann et al. 2016) In an ongoing effort, the FIP data is being used to fully automatically gather a parallel corpus for MT system training.

Topics: Nordic Textual Resources and Practices
Keywords: Web corpus, Universal Dependencies, big data, corpus linguistics, natural language processing

References

Bohnet, Bernd 2010. Top accuracy and fast dependency parsing is not a contradiction. In Proceedings of COLING'10, pages 89–97

Haverinen, Katri, Nyblom, Jenna, Viljanen, Timo, Laippala, Veronika, Kohonen, Samuel Missilä, Anna, Ojala, Stina, Salakoski, Tapio and Ginter, Filip. 2014. Building the essential resources for Finnish: the Turku Dependency Treebank. Language Resources and Evaluation, 48(3):493–531.

Huumo, Tuomas, Kyröläinen, Aki-Juhani, Kanerva, Jenna, Luotolahti, Juhani, Salakoski, Tapio, Ginter, Filip, and Laippala, Veronika (forth.) Distributional Semantics of the Partitive A Argument Construction in Finnish. In Luodonpää-Manni, M., E. Penttilä and J. Viimaranta (eds.).Empirical approaches to cognitive linguistics: Analysing real-life data. Newcastle Upon Tyne: Cambridge Scholars Publishing.

Laippala, Veronika, Kyröläinen, Aki-Juhani, Kanerva, Jenna, Luotolahti, Juhani, Salakoski, Tapio, and Ginter, Filip (forth.) Dependency profiles as a tool for big data analysis of linguistic constructions: A case study of emoticons. Journal of Estonian and Finno-Ugric Linguistics. Grammar in Use: Approaches to Baltic Finnic.

Laippala, Veronika, Kyröläinen, Aki-Juhani, Komppa, Johanna, Vilkuna, Maria, Kalliokoski, Jyrki, and Ginter, Filip. 2016. Sentence-initial discourse markers in the Finnish Internet. In Text-Link 2016 Handbook. Harmattan.

Laippala, Veronika, Kanerva, Jenna, Pyysalo, Sampo, Missilä, Anna, Salakoski, Tapio, and Ginter, Filip. 2015. Syntactic N-grams in the Classification of the Finnish Internet Parsebank: Detecting Translations and Informality. Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015), May 11–13, 2015 in Vilnius, Lithuania.

Luotolahti, M. Juhani, Kanerva, Jenna, Pyysalo, Sampo, and Ginter, Filip 2015. SETS: Scalable and Efficient Tree Search in Dependency Graphs. Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. Association for Computational Linguistics, 51--55.

Kyröläinen, Aki-Juhani and Luotolahti, M. Juhani and Hakala, Kai and Ginter, Filip 2016. Modeling cloze probabilities and selectional preferences with neural networks. DSALT: Distributional semantics and linguistic theory, August 08–15, 20106 in Bolzano, Italy.

Mueller, Thomas, Schmid, Helmut, and Schutze, Hinrich. 2013. Efficient higher-order CRFs for morphological tagging. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, pages 322–332.

Pirinen, Tommi A. 2011. Modularisation of Finnish finite-state language description–towards wide collaboration in open source development of a morphological analyser. In Proceedings of the 18th Nordic Conference of Computational Linguistics (NODALIDA), pages 299–302.

Pyysalo, Sampo, Kanerva, Jenna, Missila, Anna, Laippala, Veronika and Ginter, Filip. 2015. Universal Dependencies for Finnish. In Proceedings of the 20th Nordic Conference of Computational Linguistics (Nodalida 2015), pages 163–173.

Tiedemann, Jörg, Cap, Fabienne, Kanerva, Jenna, Ginter, Filip, Stymne, Sara, Östling, Robert, and Weller-Di Marco, Marion. 2016. Phrase-Based SMT for Finnish with More Data, Better Models and Alternative Alignment and Translation Tools. In Proceedings of the First Conference on Machine Translation, pages 391-398.

Wessman, Kukka-Maaria. 2016 Koska internet. Finiittiverbittömän koska X konstruktion syntaksi ja variaatio. Master's thesis, School of languages and translation studies, University of Turku.

Bibliography

Laippala, Veronika, Kyröläinen, Aki-Juhani, Kanerva, Jenna, Luotolahti, Juhani, Salakoski, Tapio, and Ginter, Filip (forth.) Dependency profiles as a tool for big data analysis of linguistic constructions: A case study of emoticons. Journal of Estonian and Finno-Ugric Linguistics. Grammar in Use: Approaches to Baltic Finnic.

Huumo, Tuomas, Kyröläinen, Aki-Juhani, Kanerva, Jenna, Luotolahti, Juhani, Salakoski, Tapio, Ginter, Filip, and Laippala, Veronika (forth.) Distributional Semantics of the Partitive A Argument Construction in Finnish. In Luodonpää-Manni, M., E. Penttilä and J. Viimaranta (eds.).Empirical approaches to cognitive linguistics: Analysing real-life data. Newcastle Upon Tyne: Cambridge Scholars Publishing.

Laippala, Veronika, Kanerva, Jenna, Pyysalo, Sampo, Missilä, Anna, Salakoski, Tapio, and Ginter, Filip. 2015.

Syntactic N-grams in the Classification of the Finnish Internet Parsebank: Detecting Translations and Informality. Proceedings of the 20th Nordic Conference of Computational Linguistics (NODALIDA 2015), May 11–13, 2015 in Vilnius, Lithuania.

Writing and Rewriting: The Colored Digital Visualization of Keystroke Logging

Christophe Leblay
University of Turku, Finland
Gilles Caporossi
HEC Montréal, Canada

As they contain a lot of data, keystroke logging files, are difficult to read and analyze (Wengelin, et al., 2009). There are many reasons for this, including their chronological format and high number of complex details (Kollberg, 1996). However, representations of writing are, so far, one of the main tools used to analyze it. The reason why analyzing the writing process is so important derives from the genetic methodology, where the more a text is changed or modified, the better it becomes (Leblay, 2011). The ultimate goal then becomes to understand how modifications continue to improve the text and how modifications are done in order to understand the way the text continuously improves.

The goal of data representations is to help researchers with their analysis, to assist them in understanding the data and finding patterns in it. Visualization is more than just drawings of data; it is an analysis tool (Manyika, et al., 2011). Seeing how data interacts makes it possible to discover and understand patterns and changes over time within a database (Minelli, et al., 2013; Yau, 2011). For a researcher to use representations in a way that does more than just describe a dataset requires visualization techniques. These techniques are multidisciplinary and include statistics, cognitive science, graphic design, computer science and cartography (Kirk, 2012), in addition to textgenetic analysis.

It is important to consider two complementary concepts on the same visual surface when creating data visualizations, namely, data representation (visual variables in the creation of graphs or charts) and data presentation (appearance and delivery format of the entire data visualization design, colors, the interactive features and the annotations). (Aligner, et al., 2011)

The writing process is difficult to grasp as a whole. From a computer science and mathematics standpoint, there are only two dimensions to this process: the temporal dimension, involving the specific moment when each operation was made; and the spatial one, which corresponds to the exact position of the operation in the list. Because this definition is highly decontextualized, some writing process representations also use a third dimension, chronology, which is a simplification of the temporal aspect (Bécotte-Boutin, Caporossi & Hertz, 2015). The writer adds and removes characters chronologically in time, but the overall state of the text changes as the writer modifies it. Genetic criticism studies precisely the different states of the text. Those three dimensions then concern genetic operations at the most basic level. Each operation of the writing process can be considered as a substitution operation (Van Waes & Schellens, 2003). An insertion would be the replacement of an empty space by a keystroke, and the deletion or replacement of a keystroke by an empty space. These operations are characterized by the fact they are done in a single step with the mouse or keyboard. More complex operations, such as substitution and replacement, which are done in two steps (Caporossi & Leblay, 2011), are considered to be combinations of the simple operations.

Another aspect of the writing process is the micro and macro aspects of the text, i.e., the detailed operations performed and the process'overall structure. Because those two aspects cannot be visualized together in the same representation unless interactivity and the view adjustment feature are used (Aig-

ner, et al., 2011) researchers usually use several representations to understand the process more completely (Alamargot, et al., 2011; Breetvelt, et al., 1994; Caporossi & Leblay, 2011; Cox, et al., 2009; Doquet-Lacoste, 2003; Haas, 1989; Latif, 2008; Leijten & Van Waes, 2013; Southavilay, et al., 2013; Van Waes & Schellens, 2003).

Actual visualizations of the writing process are bidimensional, and because of that, they focus for example on revision, the temporal aspect or the writer's retrospection (Latif, 2008). Even if it is important to analyze and understand the spatiotemporal dimension of the process (Stromqvist, et al., 2006), none of the actual visualizations represent the problem completely.

We propose new visualizations based on mathematical graphs that consist of nodes (points) and edges (lines eventually joining the nodes). As such, graphs are based on relationships between nodes and may be used for modeling purposes. This colored representation is halfway between detailed representations and overviews. The dynamic aspect of the writing process is highlighted (Caporossi & Leblay, 2011; Leblay & Caporossi, 2014). One of its strength is that it clearly shows the temporal and chronological relationships between operations, facilitating their identification in a structured way. Another advantage of this visualization of the writing process is that it "can handle moving text positions" (Southavilay, et al., 2013).

Topics: Visual and Multisensory Representations of Past and Present
Keywords: classification, keystroke logging, digital colored visualization, textgenetics, time-oriented production

Bibliography

Aigner, W., Miksch, S., Schumann, H., & Tominski, C. (2011). Visualization of Time-Oriented Data. Human-Computer Interaction Series. London: Springer.

Alamargot, D., Caporossi, G., Chesnet, D., & Ros, C. (2011). What makes a skilled writer? Working memory and audience awareness during text composition. Learning and Individual Differences, 21 (5), 505-516.

Breetvelt, I., Van Den Bergh, H., & Rijlaarsdam, G. (1994). Relations between Writing Processes and Text Quality: When and How. Cognition and Instruction, 12 (2), 103-123.

Caporossi, G., & Leblay, C. (2011). Online Writing Data Representation: A Graph Theory Approach. In Lecture Notes in Computer Sciences 7014, 80-89.

Cox, M., Ortmeier-Hopper, C., & Tirabassi, K. E. (2009). Teaching Writing for the "Real World": Community and Workplace Writing. The English Journal, 98 (5), 72-80.

Doquet-Lacoste, C. (2003). Étude Génétique de l'Écriture sur Traitement de Texte d'Élèves de Cours Moyen 2, Année 1995-1996. Paris: Université Sorbonne nouvelle.

Haas, C. (1989). How the Writing Medium Shapes the Writing Process: Effects of Word Processing on Planning. Research in the Teaching of English, 23 (2), 181-207.

Kirk, A. (2012). Data visualization: a successful design process [electronic book]. Packt Pub.

Kollberg, P. (1996). Rules for the S-notation: a computer-based method for representing revisions. Stockholm, Sweden: IPLab, Royal Institute of Technology (KTH).

Latif, M. M. (2008). A State-of-the-Art Review of the Real-Time Computer-Aided Study of the Writing Process. International Journal of English Studies, 8 (1), 29-50.

Leijten, M., & Van Waes, L. (2013). Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes. 30 (3), 358-392.

Leblay, C. & Caporossi, G. (2014). Temps de l'écriture: enregistrements et représentations. Louvain-la-Neuve: Academia.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011).

Big data: the next frontier for innovation, competition, and productivity. McKinsey Global Institute.
Minelli, M., Chambers, M., & Dhiraj, A. (2013). Big data, big analytics: Emerging business intelligence and analytic trends for today's businesses. Wiley Publishing.
Southavilay, V., Yacef, K., Reimann, P., & Calvo, R. A. (2013). Analysis of Collaborative Writing Processes Using Revision Maps and Probabilistic Topic Models. Proceedings of the Third International Conference on Learning Analytics and Knowledge, 38-47.
Stromqvist, S., Holmqvist, K., Johansson, V., Karlsson, H., & Wengelin, A. (2006). What Keystroke Logging can Reveal about Writing. In K. P. Lindgren (Ed.), Computer Keystroke Logging and Writing. Elsevier, 45-71.
Van Waes, L., & Schellens, P. J. (2003). Writing Profiles: The Effect of the Writing Mode on Pausing and Revision Patterns of Experienced Writers. Journal of Pragmatics, 35, 829-853.
Wengelin, A., Torrance, M., Holmqvist, K., Simpson, S., Galbraith, D., Johansson, V., et al. (2009). Combined eyetracking and keystroke-logging methods for studying cognitive processes in text production. Behavior Research Methods, 41 (2), 337-351.
Yau, N. (2011). Visualize this: the flowing data guide to design, visualization and statistics. Indianapolis: Wiley Publishing.

Word Spotting as a Tool for Scribal Attribution

Lasse Mårtensson
University of Gävle, Sweden
Anders Hast
Uppsala University, Sweden
Alicia Fornes
Autonomous University of Barcelona, Spain

Word Spotting is a set of methods for localizing word forms in handwritten text. The project group behind the current abstract has previously used it on medieval Swedish manuscripts, namely on Cod. Ups C 64 (Latin) and Cod. Ups. C 61 (Old Swedish), see Wahlberg et al. (2011) and Wahlberg et al. (2014). The most common usage for Word Spotting is to extract words for different purposes, for instance for linguistic investigations. From a technical perspective, there are several different variants of Word Spotting, but in most cases the searching process is built up on a template of the word form in question being chosen, and then the computer identifies graph sequences in the manuscript, charter etc. that are similar to the template. For further details on the technical aspects of the method, see below.

In the present investigation, the Word Spotting method is used for another purpose, namely scribal attribution, i.e. identifying individual scribes. Our material is the medieval Swedish charter corpus in its entirety, as far as they have been photographed (more than 10 000 charters). These are preserved at Svenskt diplomatarium, Riksarkivet. As stated above, the basic concept of the Word Spotting method is that a word template is chosen as a point of reference, from which the other similar word forms are identified. From a linguistic perspective, the template consists in a graph sequence, as such unique and produced by a certain scribe at a certain time. This means that the template contains some characteristics of the scribe that produced it. For our purpose, the template is not used for identifying all the word forms in the corpus that the template represents, but for identifying the instances when the word forms (and individual letters; see below) have been executed in a way similar to the template.

For the purpose of scribal attribution, not only graph sequences (in this case word forms) are of interest, but also individual graphs (letters). The shape of letters has for a long time been considered as a key issue for scribal attribution in the palaeographic research. One could mention Per-Axel Wiktorsson's work in four volumes, Sveriges medeltida skrivare (2015), where Wiktorsson identifies the scribes mainly on the basis of

the shape of seven letters: 'g', 'w-', 'æ', 'ø', 'y', ' n', 'k' och 'h' (p. 27). We have therefore focused on the identification of specific letters, and especially those consisting of several components, with a more complicated ductus, more specifically those used by Wiktorsson. These are, of course, more likely to show individual traits than more simple formations such as 'i', 'o' etc. In our investigations this far, we have made searches for 'g', 'æ' and 'k', and we will continue with the other ones listed by Wiktorsson.

From a technical perspective, the search for individual letters poses a greater challenge than the search for sequences, as the number of measuring points for the former is much smaller. Thus, a great deal of time has been put into optimizing the technical aspects of the method. The current state of the art in HTR (Handwritten Text Recognition; Llados et al. 2012) can be divided into at least two categories: 1) Segmentation techniques (Rath et al. 2007) need to segment the documents into text lines or even into words. Therefore, the performance of these techniques highly depends on the accuracy of the line or word segmentation algorithms. To this approach belong the above mentioned Wahlberg et al. (2011) and Wahlberg et al. (2014). 2) Segmentation-free approaches (Leydier et al. 2009) divide the manuscript into zones, or cells. Our approach belongs to this category and we use a so-called sliding window to match the template with the content of the window, in this case the handwritten document being investigated. The unique quality of our approach is that we can perform what has been done for a long period of time in the area known as image registration. In image registration, template images are matched to find identical images. In our case, dealing with handwritten text, this must be done in a different way, since the template and the word within the sliding window are not identical, since all graphs are unique and always displaying some incidental variation, however small. Therefore, the algorithm must be much more relaxed than in the case of ordinary image registration, i.e. allowing for variance (without loosing accuracy). The current method can be used for searching for both words and graphs, and even for parts of graphs.

The fact that there are matches in the Word Spotting and the Letter Spotting process do not automatically lead to the conclusion that the letters have been produced by the same scribe. Instead the matches should be seen as suggestions, to be further evaluated by a human researcher. The matches represent graph sequences that display similarities with the template regarding the measuring points. If for instance matches are found regarding 'g' in certain documents, one would also expect matches in the same documents regarding other letters. This is, however, not always the case, and thus one must evaluate the results of the searches with great care.

One great difficulty when dealing with scribal attribution in medieval documents is the absence of ground truth. It is very rare that we know who actually held the pen in these documents, and when the scribes are known, they are in most cases known through earlier attributions. When working with new methods for scribal attribution, it is not satisfactory to rely on previous attributions only. If one would use previous attributions to evaluate the methods, one would risk going in circle, forming the new methods on the previous work. For that purpose, we have established a set of charters where the scribes have been identified on external evidence, i.e. not through attributions on palaeographic grounds etc. Most important are the charters containing a notice from a recording clerk, stating that this person has written the document in his own hand (see Wiktorsson 2015: 28). These charters function as our point of reference in the searches in the corpus.

This investigation is a part of an on going project, called "New Eyes on the Scribes of medieval Sweden" (Riksbankens jubileumsfond). The aim of this project is to investigate and map the characteristics of the script and the scribes in the medieval Swedish charters. Within this project, we use several methods, each aiming at measuring certain features of the script (see e.g.

Mårtensson et al. 2015). Hence, the current Word Spotting investigation should not be seen as one isolated attempt at solving the issue of scribal attribution, but as a part of a large scale mapping of script features. The purpose of this project is not to find one single method that will work as an automatic tool for scribal attribution. It is through the collected evidence of several methods for measuring script features that a new mapping of the medieval scribes will be achieved.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: word spotting, scribal attribution, palaeography, image analysis

Bibliography

Leydier, Y., A. Ouji, F. LeBourgeois, and H. Emptoz (2009). Towards an omnilingual word retrieval system for ancient manuscripts. In: Pattern Recognition, vol. 42, no. 9, pp. 2089–2105, 2009.

Llados, J., M. Rusinol, A. Fornes, D. Fernandez, and A. Dutta (2012). On the influence of word representations for handwritten word spotting in historical documents. In: International Journal of Pattern Recognition and Artificial Intelligence, vol. 26, no. 05, 2012.

Mårtensson, L., F. Wahlberg and A. Brun (2015). Digital Palaeography and the Old Swedish Script. The Quill Feature Method as a Tool for Scribal Attribution. In: Arkiv för nordisk filologi 130/2015.

Rath, T., and R. Manmatha (2007). Word spotting for historical documents. In: IJDAR, pp. 139–152, 2007.

Wahlberg, F., M. Dahllöf, L. Mårtensson and A. Brun (2011). Data Mining Medieval Documents by Word Spotting. In: Proceedings of the 2011 Workshop on Historical Document Imaging and Processing.

Wahlberg, F., M. Dahllöf, L. Mårtensson and A. Brun (2014). Spotting Words in Medieval Manuscripts. In: Studia Neophilologica 86/2014.

Wiktorsson, P.-A. (2015). Skrivare i det medeltida Sverige. Vol. 1. Skara: Skara stifthistoriska sällskap.

Text Mining the History of Information Politics Through Thousands of Swedish Governmental Official Reports

Fredrik Norén
Roger Mähler
Umeå University, Sweden

Why did "information", a concept and a keyword that we take for granted in our modern vocabulary, infiltrate the official language in the twentieth century? In this presentation, I will show how the rise of the governmental information discourse, in the 1960s and the 1970s, can be understood within a larger theme of "development", and how the concept of information became regarded as a silver bullet for the bureaucratic apparatus to tackle problems in society. This is done by topic modeling (with LDA/MALLET) the corpora of Swedish Governmental Official Reports (8 000 reports, 1922–), in particular by examining co-occurring topics, a less common approach within the practical use of probabilistic topic modeling. The scope and the long time-span of the report series makes it an internationally unique source, especially when it comes to study the emergence of interests and attitudes of a single state through time and, furthermore, to view it as the "voice" of the Swedish state. Today, when incalculable amounts of texts, like the report collection, not only are available online but also searchable – down to each single word – this presentation emphasizes the need for the humanities to accept the challenge of potentially re-rewrite parts of history. That is, how changes of language, in millions of documents, can be linked to – and create new understanding for – developments in society.

In collaboration with Humlab, the digital humanities hub at Umeå University, and

software developer Roger Mähler, a method was developed that utilized the output data from MALLET which was expected to give insight into three things: 1) the number and diversity of reports that the information topic occurred in through time, 2) to discover and visualize larger cluster of topics and give the information topic a position, and hence a context, in those clusters, and 3) enrich the analysis by combining distant and close readings.

The Swedish Governmental Official Report series are available for public access at The Riksdag's open data website (data.riksdagen.se), and part-of-speech tagged versions are available at Språkbanken (spraakbanken.gu.se) as downloadable XML-files. This study extracted the word stem of all nouns from the reports of the 1960s, 1970s and 1980s, which is sufficient for the study of themes in a text, and each report was split into chunks of 1 000 nouns each. MALLET was then used to compute three distinct LDA topic models, one for each decade, and each consisting of 500 topics. A manual review of the generated topics showed that each decade had been assigned a distinctive topic of a general information discourse. The computed average topic weights for each report, based on weights in each chunk, were used to visualize a network of all reports, and their most dominant topics, with weaker report-to-topic links filtered out based on a configurable threshold. Our pre-study showed that LDA topic modeling often computes a very dominant topic with a weight of over 60 %, while the weight of the following topics dropped significantly. Hence, a generous threshold of 0.01 was proved to capture both a discursive core as well as its periphery.

The method showed three things. Firstly, a distinctive information discourse evolved over time in the official language in the state bureaucracy. By highlighting reports in which the information topic was dominant, it was clear that the number of co-occurring reports increased by a fourfold from the 1960s to the 1980s.

Secondly, the three datasets, one for each decade, were imported, separately, to Gephi. Each graph were modulated by the Closeness-Centrality, Force Atlas and Modularity Class algorithms to sort topics and reports into larger thematic clusters, that is meta-topics. In order to classify a meta-topic, every topic and report that belonged to a cluster were manually examined as a way to identify the common theme of the topic cluster. For example, one interesting result, which was found in the pre-study of the network graph of the 1970s, situated the information topic within a cluster that had been labeled as theme of development (of various political issues).

Thirdly, by adding close reading to the analysis, the actual reports in the cluster of "development", presented concrete insights and illustrations of how to understand and synthesize the connection between "information" and "development" and how "information" became regarded as a universal tool for handling problems and challenges in society. Hence, the dynamic interaction between closeness and distance helped to strengthen both perspectives and enrich the end result.

Topics: Nordic Textual Resources and Practices
Keywords: topic modeling, mallet, swedish governmental official reports, gephi

Teaching and Learning the Mindset of the Digital Historian and More: Scaffolding Students' Critical Skills in the Digital Humanities

Thomas Nygren
Uppsala University, Sweden

Katherine Hayles (2012, p. 21) notes how "[y]oung people practice hyper reading extensively when they read on the web, but they frequently do not use it in rigorous and disciplined ways." This is an important observation but what does this mean in practice? What does it mean to read in "rigo-

rous and disciplined ways", online and offline, and how can we support students habits of mind when they read and interpret sources and information from and about the past? In this paper I will present some empirical studies to highlight challenges and opportunities to support students' success in navigating in the digital world of humanities.

Going to the sources and making sense of fragments from the past in archives, digital and analogue, demand historians to become experts in sourcing, corroborating and contextualizing the sources–and more. To better understand how experts and novices read historical documents we designed an eye-tracking study to track read differences between historians and students when they read historical documents from the time of the French revolution (Mulvey & Nygren, work in progress). We tracked eye movements of four historians, four university students, and eight high school students reading with an infrared non-intrusive eye-tracking camera. The stimuli contained four historical sources with information regarding human rights at the time of the French revolution. Sources were selected based upon their usefulness to test historical thinking (Wineburg, 1991); this means that they should be primary sources with important source information and hold valuable information for answering a complex historical question–challenging the reader to source, corroborate and contextualize the information.

The material participants in this study read included excerpts from four primary sources namely, (1) *Declaration of the Rights of Man and Citizen,* 26 August 1789, Paris, France; (2) *The Declaration of the Rights of Woman* by Marie Gouze [published under the pseudonym Olympe de Gouges], September 1791, Paris, France; (3) *A particular account of the insurrection of the Negroes of St. Domingo*, 1791, Paris, France; and (4) *Haitian Declaration of Independence*, January 1804, Gonaives, Haiti. In total there were 10 pages, 1828 words, for participants to read. The preliminary findings show that historians seem to focus more on sourcing and central historical aspects of documents, whereas high school students focus more on dramatic events and racist language. Guided by a central historical question in a directed reading, these patterns change, especially for university students, but also for high school students, who now read documents more in line with historians' initial reading strategies. Our findings shed new light on how experts' and novices' historical literacy differs, and how these differences may relate to their abilities to critically scrutinize sources, corroborate evidence, and understand central aspects of historical events. Participants' scores in the post-tests correlates to some extent with participants reading focus, indicating how a more professional focus may be scaffolded by a historical question in ways that can help students pay more attention to source information.

However, digital historians do not only closely scrutinize documents, they also use archives with large sets of data. In my talk I will also present some indications from quasi-experimental studies showing that novice users of Swedish digital archives may lose their awareness of their theoretical position and empathy when using large data sets (Nygren 2014). When facing a large set of data and statistics it may be an instinctive reaction to start sorting the information and quantifying, rather than reflecting on the starting point of your research and critical perspectives, thus conducting a more interpretive investigation. Close reading of a few documents may perhaps make it easier to hermeneutically scrutinize the information. However, using digital tools and material can also be used to closely analyze how smaller fragments fit into a bigger picture, but this takes a critical awareness of the materiality and a focus on the research question, which both students and researchers need to bear in mind when "going" to the archives.

Evidently it is possible for students to navigate digital databases and learn history in new ways using affordable technological resources. But this needs to be supported by hard and soft scaffolding (Brush & Saye, 2002). Hard scaffolds built into the databases can make databases more useful for other than just historians with expertise of the

digital architecture. Soft scaffolds designed by teachers and historians can make it possible for students to use Swedish digital databases designed for professional historians (Nygren & Vikström, 2013; Nygren, Sandberg & Vikström, 2014). In previous studies we find that students can learn to walk in the shoes of the digital historian and use primary digitized sources in constructive ways, but they often stumble when it comes to contextualizing the information. A central aspect here is the challenge of historical empathy. Understanding people in the past by their own standards means that we need to contextualize the information and try to shift perspectives. There is a challenge to understand the past as a "foreign country," a place where language and concepts as well as context differ in fundamental ways from our contemporary world (Lowenthal, 1985). This cognitive and emotional ability to understand unfamiliar perspectives across time and space, often labeled historical empathy, is a central but also complicated matter in historical studies (Davis et al 2001). Closeness and distance is vital in our understanding of the past and digital tools may help us see things in a larger perspective and also zoom in on selected parts in time and space. Closeness to primary sources and the environments studied may be a way to overcome temporal and spatial gaps of understanding. Materiality may certainly affect our construction of knowledge (Latour 1999) and researchers and students in digital history may benefit from physical reminders of the complexity of the fragmentary remains behind neat data. Mixing the digital with tangible materials may be an important scaffold to consider.

Digital tools can also be used to complement printed material, in ways that may help students and historians overcome the unreachability of the past so evident when going to the archives (Robinson 2010). Using visualizations makes it possible to organize the information in new ways, for instance linking it to geographical locations and on temporal scales. Digital tools can be used to collect different types of data and to explore relationships in time and space beyond what may be possible in more traditional explorations using pen and paper (Nygren, Frank, Bauch & Steiner, 2016). In digital history the writing of history may be more than creating text. With digital platforms, readers/users can create and present multiple interlinked narratives; integrate images, maps, commentary, and primary sources in the same field of vision; and curate and shape the reader/user's experience, allowing for a hybrid experience (cf Thomas III & Ayers 2003). But as a final presentation, visualizations need to help the reader/user see behind the seductive cleanliness of data presentations and animations. It is also important to bear in mind how multiple narratives and multimodal presentations may confuse the reader/user rather than give a richer understanding of the topic. There may be a risk of cognitive overload for readers in hypertext environments (cf Gerjets & Sheiter 2003). All users need to learn to become critical readers and navigators in multimodal environments, and digital historians and students need to understand the audience in somewhat new ways when a digital tool becomes a publishing tool (Hayles, 2012). Students and historians need also be able to review scholarship in new media, if we want to make use of new opportunities and safeguard quality in historical scholarship and prepare students for a future in academia and beyond (Presner 2010; Nygren, Foka & Buckland 2014).

Last but not least we need to consider the uses of history in contemporary digital media. In an ongoing non-intrusive study of teachers contrasting contemporary uses of history with primary sources from the era of civil rights movement, we observe that students rarely critically scrutinize contemporary uses of the ideas of Martin Luther King Jr. when misinterpretations are underlined in the media (Nygren & Johnsrud, in review). For students, critically examining evidence seems to be difficult when authorities and media augment oversimplified and popular perceptions of the past. However, we also find that students can learn a more nuanced perspective on the life and deeds of MLK, findings still observable a year after the

initial teaching took place. The results from this study show important potentials and limitations when trying to stimulate the learning of core content, critical thinking and values of social justice using primary historical sources and contemporary media representations that attempt to make the past historical and practical. Connecting the past to the present and critically scrutinizing contemporary media is asking more from students than what we ask from historians. And historians actually do not seem to be very good at critically scrutinizing online information (Wineburg & McGrew, 2016). We need to better understand this challenge and how to deal with this in schools and academia in a digital age.

Scaffolding students to read and write with the affordances offered by the digital humanities is a certainly a tall order for teachers. To make this challenge a bit less complex I suggest a focus on a few central mindsets, namely, skills to *criticize, empathize* and *create*. Having the skill to *criticize* means that students, and scholars, in the digital humanities need to be able to: critically examine and corroborate various types of sources (such as text, image, and audio), critically read between the lines, read close and distant, critically explore and experiment with various digital tools, understand different critical and ethical perspectives, and formulate critical questions. This critical mindset is a central part of being a rigorous reader in the digital humanities. A skillful reader is also able to *empathize* with multiple perspectives. This central aspect of the humanities involves classic challenges to understand: historiography, different human perspectives, not least the mind of the author, the reader, the creator and ordinary people in foreign cultures and countries. Today this means not least understanding human existence and making in analog and digital worlds. This means that students must learn to contextualize the information and empathize in cognitive and emotional ways with distant worldviews. Last but not least students in digital humanities need to *create* accounts to process and communicate their thoughts in nuanced ways. Writing is still a central skill. But in the digital humanities there are now opportunities to think, make, enact and experiment in a diversity of forms and in collaborations. Some ideas might certainly benefit from being treated and presented in non-textual ways. But how do we stimulate all this in practice? The answer today is that we do not really know, and we need more empirical studies to connect our theoretical understanding to the learning of students.

It is time to move beyond anecdotal evidence about how to support learning in the digital humanities. The research presented here provides some small insights into the potentials and pitfalls of teaching and learning critical mindsets useful in the digital humanities. This research highlights how it is possible for, at least some, students to learn to read like historians, navigate digital archives and deconstruct contemporary media myths about the past. But our research also highlights the complexity of teaching and learning in the digital humanities, how little we know, and how important it is to support students' humanistic habits of mind.

Topics: Nordic Textual Resources and Practices
Keywords: teaching, learning, history, archives, skills

References

Brush, T. A., & Saye, J. W. (2002). A summary of research exploring hard and soft scaffolding for teachers and students using a multimedia supported learning environment. *The Journal of Interactive Online Learning*, *1*(2), 1-12.

Davis, O. L., Yeager, E. A., & Foster, S. J. (Eds.). (2001). *Historical empathy and perspective taking in the social studies*. Rowman & Littlefield.

Gerjets, P., & Scheiter, K. (2003). Goal configurations and processing strategies as moderators between instructional design and cognitive load: Evidence from hypertext-based instruction. *Educational psychologist*, *38*(1), 33-41.

Hayles, N. K. (2012). *How we think: Digital media and contemporary technogenesis*. University of Chicago Press.

Latour, B. (1999). *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, Mass.: Harvard University Press

Lowenthal, D. (1985). *The Past is a Foreign Country*. Cambridge: Cambridge University Press.

Nygren, T. (2014). Students Writing History Using Traditional and Digital Archives. *Human IT* 12 (3): 78–116.

Nygren, T., Foka, A., & Buckland, P. I. (2014). The status quo of digital humanities in Sweden: past, present and future of digital history. *H-Soz-Kult*.

Nygren, T., Frank, Z., Bauch, N. & Steiner, E. (2016) Connecting with the Past: Opportunities and Challenges in Digital History, in *Research Methods for Creating and Curating Data in the Digital Humanities*, eds. M. Hayler & G. Griffin, Edinburgh University Press, 2016, 62–86.

Nygren, T., & Vikström, L. (2013). Treading old paths in new ways: upper secondary students using a digital tool of the professional historian. *Education Sciences*, *3*(1), 50-73.

Nygren, T., Sandberg, K., & Vikström, L. (2014). Digitala primärkällor i historieundervisningen: en utmaning för elevers historiska tänkande och historiska empati. *Nordidactica: Journal of Humanities and Social Science Education*, (2), 208-245. [Digital primary sources in history education: A challenge for students' historical thinking and historical empathy]

Presner, T. (2010). Digital Humanities 2.0: a report on knowledge. *Connexions Project.—2010*.

Robinson, E. (2010). Touching the Void: Affective History and the Impossible, *Rethinking History: The Journal of Theory and Practice,* 14:4, 503-520.

Thomas III, W. G., & Ayers, E. L. (2003). The differences slavery made: A close analysis of two American communities.

Wineburg, S. & McGrew, S. (2016) Why Students Can't Google Their Way to the Truth, *Education Week, 36*(11), 22, 28

New Multi-language Digitised Newspapers and Journals from Finland Available as Data Exports for Nordic Researchers

Tuula Pääkkönen
Jukka Kervinen
National Library of Finland

To respond to the needs of the especially researchers of digital humanities, we have created specific data export packages from the digitised materials (Pääkkönen, Kervinen, Nivala, Kettunen, & Mäkelä, 2016) which currently span until 1910. We have developed a custom XML format for the export packages, which contains the post-processing results from the digitisation. There is one XML file per one page of a newspaper, which contains three pieces of information: the metadata, ALTO XML (Technical Metadata for Layout and Text Objects standard) and the textual content of a digitised page. These three parts bring the information developed within the library available for many kinds of research opportunities from the bibliographic metadata to the content analysis. Also within the custom XML file the simple raw text format give additional possibilities for the researchers to focus on their research questions. We hope that by opening both metadata and the content, it is possible to create collaborations with library and researchers for the tools and method development of the materials, for example optical character recognition (OCR) and post-correction fixes. The material is as-is in the export packages, so the material has varying number of OCR errors, where the OCR quality ranges 57-76% from the 19th century data (Kettunen & Pääkkönen, 2016), for example. Potential OCR and metadata issues is something that the researcher needs to be aware when taking the material set into use – however, it is also opportunity to work together to improve the data content for everybody. Library can act as a central role by connecting different researchers who face the same issues with the raw content and the

library can benefit by being able to utilize the research results for example in improving the quality of the materials onwards (Pääkkönen & Kervinen, 2016).

Digitisation and export packages can also be seen as technical infrastructure for the research data creation, but thinking the details of the technical requirements is not enough. Therefore, we have also taken steps to analyze the legal aspects of opening the data, namely the incoming General Data Protection Regulation and the copyright directive proposal (DSM-directive) of the EU (European Commission, 2016), because with the versatile material of the newspapers we need think both the people appearing in the content and the original authors view. With these two viewpoints, and preparing to the incoming changes, we can start responding to the new requirements regardless of the way how the material is provided to the users. Together with the Finnish Copyright Society (Kopiosto) for the newspapers and journals and couple of media houses we have created a process and a tool within the presentation system to manage copyright redaction requests. As a brief overview to this tool, the tool allows National Library of Finland to redact specific part of the digitised contents, based on the request of the right holder, while still allowing the rest of the material stay intact. After the redaction, that particular section of the digitised material can only be read in the legal deposit libraries while it stays redacted in the public internet or in locations, where the accessibility to the materials has been extended by contracts (Karppinen, Kaukonen, Pääkkönen, & Sorjonen, 2016). On the other hand, for the researcher use, there are plans to enhance the user management of the presentation system further, so that we can offer materials via it to the researchers with whom we have agreements in place. With help of the preparation to the incoming changes to the regulations and new technical features, there are opportunities for new collaborations even across borders. However there are limitations in the research data, as via processing them further we have also noticed malformatted or missing metadata or data, which has now revealed itself. Therefore, the material as research data requires awareness of the limitations by the researchers, even though our attempt is to offer as authentic material as it is got after the digitisation post-processing.

The first version of the export packages contain material until 1910, but our internal tools make it possible to generate the export packages to the newer material when the need arises and new contract models are in place. The near proximity of the tools to the digitisation chain offers benefits as digitisation progresses, new export packages can be created with a cost-effective way. In this paper, we will tell how the export packages of the material were created, where you can get them at the moment and which constraints the materials have. The interesting material of the digitised collections of Finland contain material for example in Finnish, Swedish, Russian, German and Sami making the language-base possibly interesting for Nordic collaboration. For example, based on the feedback and information queries in our presentation system at digi.kansalliskirjasto.fi there is steady flow of visitors from Sweden, who are interested on the various news, big and small, which have appeared in both sides of the border. So far, the export packages have been delivered to a Comic research project of Academy of Finland and to the few researchers of Helsinki Centre for Digital Humanities (http:// heldig.fi) We will also tell some aspects of earlier digital humanities projects, which have been important in developing collections (Kettunen, 2016), features to the presentation system and start of collaboration with researchers, from who we have got feedback via initial user queries or via direct contacts. Besides the export packages, the new processes, and recently created contract models make it possible to open up materials for the research use, thus enabling us to implement the openness and digital humanities policies of the National Library of Finland in the future (National Library of Finland, 2016).

Acknowledgments
This work was funded by the EU commission through its European Regional Development Fund and the program Leverage from the EU 2014-2020.

Topics: Nordic Textual Resources and Practices
Keywords: newspapers, digital resources, accessibility, research use

References

European Commission. (2016). Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market. Retrieved 22 September 2016, from https://ec.europa.eu/digital-single-market/en/news/proposal-directive-european-parliament-and-council-copyright-digital-single-market

Hölttä, T. (2016). Digitoitujen kulttuuriperintöaineistojen tutkimuskäyttö ja tutkijat. Retrieved from http://urn.fi/URN:NBN:fi:uta-201603171337

Karppinen, P., Kaukonen, M., Pääkkönen, T., & Sorjonen, M. (2016). Contracts Enabling Collaboration of The National Library of Finland with Media Houses in Electronic Deposit. Presented at the IFLA World Library and Information Congress, Columbus, Ohio, United States.

Kettunen, K., & Pääkkönen, T. (2016). Measuring lexical quality of a historical Finnish newspaper collection—analysis of garbled OCR data with basic language technology tools and means. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Retrieved from https://www.researchgate.net/profile/Kimmo_Kettunen/publication/299515022_Measuring_Lexical_Quality_of_a_Historical_Finnish_Newspaper_Collection_-_Analysis_of_Garbled_OCR_Data_with_Basic_Language_Technology_Tools_and_Means/links/56fd194208aeb723f15d61be.pdf

National Library of Finland. (2016). Duties and strategy [Text]. Retrieved 17 May 2016, from https://www.kansalliskirjasto.fi/en/duties-and-strategy

Pääkkönen, T., & Kervinen, J. (2016). Historiallisten digitoitujen sanoma- ja aikakauslehtien avaaminen avoimena datana tutkijoille. *Informaatiotutkimus; Vol 35, Nro 3 (2016): Informaatiotutkimuksen Päivät 2016*. Retrieved from http://ojs.tsv.fi/index.php/inf/article/view/59442/20626

Pääkkönen, T., Kervinen, J., Nivala, A., Kettunen, K., & Mäkelä, E. (2016). Exporting Finnish Digitized Historical Newspaper Contents for Offline Use. *D-Lib Magazine*, *22*(7/8). https://doi.org/10.1045/july2016-paakkonen

Exploring User Engagement in Crowdsourcing Folk Traditions

Sanita Reinsone
University of Latvia

The mass digitizing activities of holdings of tradition archives carried out over the last few decades have introduced a significant amount of various digitized cultural artefacts to the wider public. As information technology has developed, knowledge of how tradition archive materials could and should be digitally maintained has advanced as well. Folklore archivists have developed and sought digital platforms appropriate for their specific collections and suitable solutions for virtual representation, further processing, and (re-)using of digitized data not only to ensure long-term preservation of the cultural artefacts and creating new access routes to collections, but also to remain close to contributors as well as to continue the archiving of new vernacular knowledge.

In recent years, projects providing tools and inviting volunteers to transform digital content from one format into another have become one of the most widespread phe-

nomena among Digital Humanities research and cultural heritage institutions. One such tool is transcription, which is still one of the most common forms of crowdsourcing in digital humanities. The benefits of such collaborative activities for tradition archives are apparent as the most historical tradition archives consist of vast quantities of handwritten or type-script text collections that cannot be automatically transformed digitally. By turning manuscript images into digital texts, documented knowledge becomes available not only by the metadata created by the archive system but also by its content.

Although the number of crowdsourcing projects is substantially increasing in the digital humanities and cultural heritage fields, tradition archives have not yet been as eager to carry out transcription crowdsourcing projects. However, the experiences of those few tradition archives that have managed to launch crowdsourcing campaigns for folklore manuscript transcription, such as the Irish National Folklore Collection (University College Dublin) and the Archives of Latvian Folklore (Institute of Literature, Folklore and Art, University of Latvia), are impressively positive.

This paper will provide in-depth analysis of the crowdsourcing campaign "Valodas talka"[7] carried out by the Archives of Latvian Folklore (ALF) in cooperation with the UNESCO Latvian National Commission in 2016. Targeted at a school audience and lasting for 71 days, the campaign provided a contribution of almost 15,000 transcribed manuscript pages. More than 1,500 participants from 120 educational institutions participated.[8]

The *Talka* website randomly displays manuscript pages in a slider carousel, providing users with an easy way of finding the most suitable files for transcription. In addition, simple game elements were included to provide additional incentives for younger-generation users to engage, i.e., each transcribed character provided the user with one *Talka* point. Users could collect points individually and/or collectively for their schools if they indicated which school they represented. The Top 10 individual users and Top 10 schools were displayed on the front page. The competition feature served to encourage schoolchildren to compete for the most points and contribute to their school's position by various means, e.g. involving friends and family members. Despite all the positive engagement through crowdsourcing, members did make some inappropriate responses in order to falsely inflate their scores. However, each of the transcribed pages was carefully proofread by the editors, corrected, and accepted or deleted if the submitted transcription proved to be fraudulent.

A crowdsourcing initiative based in gamification and competition while producing high youth involvement should be regarded with caution, because it can have negative consequences as well. Compared to other kinds of campaigns[9], participants are motivated more by the intensity of the exercise and competing than by the idea of volunteering as such. Because of this, they may be less inclined to be careful contributors leading to more work for editors in the end. Besides, a competition tends to create an inner tension in the young volunteer community as members observe each other's activity and examine each other's contributions, and are sensitive towards the review process of both their contribution and that of others which provides additional load of communication for curators.

[7] The title references ethnographical collective work in the fields and can be translated as 'collaborative work for language'. Campaign's website: http://talka.garamantas.lv

[8] Average school participant was 12 to 18 years old. Teachers had a significant mediating role to attract and encourage students to get involved.

[9] The comparison is made with the second crowdsourcing campaign of the ALF (http://lv100.garamantas.lv/en), which is based on the pure concept of volunteering with no competition.

User group No. / Level of involvement	Number of users	Percentage of users	Division of groups by contribution	Contribution (%)	Contribution (transcribed characters)
1 / Low	851	55.0%	<=1 000	3.6%	394 430
2 / Some	582	37.6%	1 001–10 000	16%	1 756 388
3 / High	92	6.0%	10 001–100 000	23.4%	2 570 067
4 / Very high	21	1.4%	>=100 001	57%	6 252 204
Total	1546				10 973 089

Figure 1.

The first crowdsourcing campaign of folklore manuscript transcription provided the Archives of Latvian Folklore not only with an opportunity to significantly increase textual corpora of the digital archives[10] and increase the amount of visitors and its popularity on a national scale, but also provide several valuable lessons on campaign strategy, management of communication with users, and editorial workflow planning. It was one of the most intense periods of communication with mass media and users of digital archives the ALF has ever experienced. It was also the largest crowdsourcing initiative to date in the field of the intangible cultural heritage of Latvia. On average 210 manuscript pages were transcribed per day, and the highest attainment was 520 transcribed pages in one day.

User involvement analysis indicates a remarkable disproportion of engagement activity among *Talka* participants. A medium level of activity was demonstrated by 37.6% of users (Group 2, see Table 1), whose contribution equaled 16% of all transcriptions; whereas the second highest involvement user group (6%, Group 3) provided 23% of all contributions. The lowest involvement user group consisting of 55% of all users (Group 1) provided 3.6% of all contributions which highly contrasts with the very high involvement group, which included only 1.4% of all users (Group 4) but who provided the most prominent results in terms of quantity, 57% of all contributions (Figure 1).

The analysis of user statistics collected during the campaign, after the campaign, and while launching a new campaign[11] suggests the high importance of the group of permanent users who have already formed a virtual community of trusted contributors and experts. They not only provide a permanent flow of contributions that can easily be followed by anyone but also positively influence the overall quality of submitted contributions. This makes sense because it is customary for new users to undergo some kind of "consulting" by exploring recent records submitted by other users before they begin transcribing a manuscript themselves. If theses samples are correctly done by a trusted user, the contribution stands as an example of good practice and helps to spread it further.

Topics: Nordic Textual Resources and Practices
Keywords: crowdsourcing, vernacular knowledge, tradition archives, society engagement, knowledge production

[10] Website of the Digital Archives of Latvian Folklore: http://garamantas-lv/en

[11] Open-ended crowdsourcing campaign for folklore manuscript transcription "Simtgades burtnieks" (The Wizards of Centenary) was launched in June 2017 by the Archives of Latvian Folklore in cooperation with Latvian Centenary Bureau and National Radio and Television. Campaign's website: http://lv100.garamantas.lv/en

Bokhylla: A Case Study of the First Complete National Literature Database in the World

Eivind Røssaak
National Library of Norway

This paper presents a part of my Norwegian Research Council funded research on how digitization refashions a nation's memory. Digitization within the cultural heritage sector is crucial here, and I will focus on the National Library of Norway in this presentation. A key digital resource in Norway is the National Library's digitized book collection, "Bokhylla" (The Bookshelf). Tech companies and libraries around the world have at least since the mid-1990s struggled to find good solutions for presenting and preserving our cultural heritage in a digital age. The challenges are institutional, legal, technological and aesthetical. The challenges have fostered a series of innovations within these fields. This paper will explore and present these innovations.

A key challenge when it comes to creating digital access to book repositories is legal barriers. When the National Library of Norway established its main digital resource, Bokhylla, an important part of its prehistory was a complex legal agreement with the copyright holders enabling the library to make available all books ever printed in the nation up until 2001. While Google has been in a continuous legal grey zone, Bokhylla has become a highly original and complex digital artifact.

Methodologically my approach is inspired by a Science and Technology Approach which relies on Bruno Latour's Actor Network Theory (ANT) in particular. The ANT-approach uses "decomposition" of the artifact to see how it is generated historically and by a variety of actors (human and non-human). It took ten years to create Bokhylla the way it looks today. It is a legal, technological and aesthetic artifact. Each of these three elements rely on a complex network of actors constituting what I will call "digital infrastructures". What where the infrastructures enabling the library to construct a robust legal agreement? The labor union field of Norway will be explored. What technological premises enabled the bokhylla solution? The file and database architecture will be explored. How was the "analogue" book aesthetically speaking turned into a digital artifact still resembling or signifying "literature" as we used to know it? The modes of interaction and interfaces will be explored. And finally: what sort of DH research does this construction enable? Some of the applications and research connected to Bokhylla will be presented.

Topics: Nordic Textual Resources and Practices, The Digital, the Humanities, and the Philosophies of Technology
Keywords: memory, cultural heritage, technology, new media, digital books, bokhylla, the national library, digital humanities

Bibliography

2016. Memories in Motion (co-eds. I. Blom and T. Lundemo), Amsterdam: Amsterdam University Press.
2011. Between Stillness and Motion: Film, Photography, Algorithms. Editor. Amsterdam: Amsterdam University Press.
2010. The Archive in Motion: New Conceptions of the Archive in Contemporary Thought and New Media Practices. Editor. Oslo: National Library.
2010. The Still/Moving Image: Cinema and the Arts, Saarbrucken: Lambert Academic Publishing.
2005. Selviakttakelse – en tendens i kunst og litteratur, [Observation of the Self – in the Arts] Bergen: Fagbokforlaget.
2004. Kyssing og slåssing. Fire kapitler om film, [Eros and Agon: Four Chapters on Film] Oslo: Pax (with C. Refsum).
2001. Sic. Ved litteraturens grenser, [An experimental history of margins in art, literature and philosophy] Oslo: Spartacus.

Life Based Design for Human Researchers

Pertti Olavi Saariluoma
University of Jyväskylä, Finland
Jaana Leikas
VTT technology center of Finland

Technology is only valuable to the extent that it can enhance the quality of life. When solving complex engineering problems, it is easy to forget the basic reason why technologies are designed and developed. They are developed to improve the quality of human life.

Designing technology to improve the quality of human life requires a multidisciplinary design approach. On one hand, multidisciplinary teams can give designers with a technical background the opportunity to better acquaint themselves with human research by working with human researchers. On the other hand, human researchers should be more aware of the various roles they can play in the process of designing and developing new technological solutions for people. Human researchers can be provided with concepts, facts, methods and theories that are useful in many aspects of design.

A study of multitude of HTI paradigms illustrates that one can integrate them into four research programs characterized each by a separate design question. The traditional HTI design discourses can thus be systematized by showing that HTI design thinking must always meet four fundamental design questions:

1. functionalities and technical user interface design;
2. fluency and ease of use;
3. elements of experience and liking; and
4. the position of technology in human life.

As these fundamental questions are necessarily present when designing the human dimension of technology, it is useful to understand the logic behind them. The questions define the basic tasks in HTI design: the decision of the functionalities of the artefact, understanding how to best use them, understanding the overall experience when using them, and finally – and most importantly – the technology's role in human life. The latter is always present in design, either consciously or tacitly.

The first question in HTI design concerns how the behaviour (functionalities) of a technical artefact can be controlled. How should the artefact be manoeuvred so that it can reach its expected state or carry out the expected processes? During a human action, an expected state can refer to a process that makes it possible for the user to reach her goal. In this sense, the expected state of a sailing boat can be as much about sailing on the sea as reaching a destination.

How to control the behaviour of a technical artefact is a fundamental problem in HTI design. No technical artefact can exist without providing its users the methods to use it. The problem can be called technical UI. First, the behaviour of the artefact must be logically linked to the human action in question. In the case of the lift, the technical capacity to move from one floor to another is the artefact behaviour that makes it possible to support people's movement in a block of flats. Second, it is essential to link the behaviour of the artefact to users' actions via a user interface. In the case of a lift, this often refers to the set of control buttons referring to which floor the lift should stop on. However, the latter presupposes design knowledge of how people use the artefact.

The next fundamental question in HTI design concerns the fit of the technology with users' ability to use it. This problem can be called usability. This problem concerns the human dimension of the user interface and opens up a new set of questions and sub-questions that can be answered with the help of human research and underlying psychological concepts. In order to guarantee smooth and easy interaction, user interface architectures should explicitly organize dialogues. For example, elements with similar functions should be associated in a sense-making manner. The foundations of understandable user interfaces can be searched from human research – that is, answers to such questions as why a particular architecture is favoured over another.

However, interaction is not only cognitive but there are many dynamic aspects to be discussed in design. Emotions are important, as they define the human position towards specific issues (Frijda, 1988). In design, the question is not only about positive emotions but also about asking which emotions are relevant in the particular interaction situation. One should be able to feel angry when the cause is irritating enough and happy when experiencing positive actions. Otherwise, the human emotional system does not operate in a rational manner. In interaction design, it is essential to consider how people experience a situation or event that arouses their emotional responses (Frijda, 1988). Designers mostly strive to create a positive mood in their clients when interacting with the product. To understand the emotional dimensions of human experience it is necessary to understand human emotions and motivation, which is closely linked with emotions. People often pursue positive mental states, and are therefore motivated to use artefacts that help them do so. The motives for doing something can be complex and long lasting. The modern psychology of motivation offers a sophisticated framework for analysing the motives for using technologies The importance of this sub-discourse is obvious: designers need to know why people use some technical artefacts and ignore others.

Finally, it is essential to ask, what is a technology intended to be used for, in the first place? One can call this problem designing for life. Why is it used? What is its position in people's life? Answering these questions is a prerequisite for successful interaction design, and calls for an understanding of the life settings that the artefact is intended to support. This is possible with the help of a general notion of 'form of life' that can describe any domain or context of life in relation to technology. Form of life is a system of actions in a specific life setting with its rules and regularities, and facts and values that explain the sense of individual deeds and practices in them.

In HTI design, innovative thought processes are often more or less unorganized. Because of hectic design cycles, interaction designers do not necessarily have the time to apply systematic methods or use scientific knowledge to construct interface. Although it is common in engineering design to apply the laws of nature and other scientific facts, this approach is unfortunately not often taken in HTI design.

If an organization wishes to exploit scientific knowledge in interaction design processes, it is important to create systematic procedures for doing so, for example by defining the relationships of relevant design concepts and questions and organizing design processes around them. The concept of usability, for example, opens up a large set of questions and sub-questions that can be both general and product specific in nature.

Such systems of questions and answer make it logical to ask, whether they could be ontologised and thus used to give a structure of HTI-desing processes (Saariluoma, Cañas and Leikas 2016). . Knowledge management in design has been a topical issue for some time (Gero, 1990). A key concept in this discussion is ontologies, which are organized sets of domain-specific concepts (Chandrasekaran, Josephson, & Benjamins, 1999) that describe the most general concepts in a given field; they are widely used in knowledge management. Ontologies can be seen as theories of information contents (Chadrasekaran, Josephson, & Benjamins, 1999; Gero, 1990)

Traditionally, ontologies have mainly been used to describe the structures of some domains as facts. For example, some products have been described as sets of elements and relations. In such instances, ontologies have had the role of information storage and retrieval. However, when considering design as a dynamic thought process, it is more worthwhile to discuss ontology as a question structure to be used to manage design problems. Ontologies in this sense can be seen as tools for creative thought rather than as information storage. Ontologies for HTI design can thus be used to generate sets of design questions describing the interaction – that is, the questions that must always be answered when a technology and its rela-

tionship to users is designed – and to conduct the HTI design process accordingly.

As concepts, questions and ontologies provide a means of managing corporate thinking and making corporate knowledge explicit. When thought processes are explicit, it is possible to support them, to provide correct knowledge to thinking, to foster innovations and to move tacit knowledge from one process to another. The answering to the four fundamental questions presupposes by far unified argumentation based on human research.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: design science, life-based design, design ontologies

Bibliography

Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What are ontologies, and why do we need them? Intelligent Systems and their Applications, 14, 20–6.

Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What are ontologies, and why do we need them? Intelligent Systems and their Applications, 14, 20–6.

Gero, J. S. (1990). Design prototypes: A knowledge representation schema for design. AI Magazine, 11, 26–36.

Frijda, N. H. (1988). The laws of emotion. American Psychologist, 43, 349–58.

Saariluoma, P., Cañas, J. & Leikas, J. (2016) Designing for life- A human perspective on technology development. London: PalgraveMacmillan.

Leseutgave av Hrafnkels saga, Menotas koding og knytting til andre ressurser

Fabian Schwabe
University of Tübingen,
Germany

Når man arbeider på ei digital utgave av en norrønt tekst trenges det å tenke på ei ordentlig koding av denne teksten slik at man ikke bare kan vise teksten selv på ei nettside, men også har teksten i en format som er forståelig for andre og kan benyttes av dem. På dette feltet er det sikkert lurt å kaste et nøyet blikk på anbefalingene til Medieval Nordic Text Archive (Menota). Der finnes det et utarbeidet foreslag, hvordan man kunne bruke ei XML-koding til å beskrive norrøne håndskrifter og lagre deres innhold.

Denne kodinga bygger opp på foreslagene til Text Encoding Initiative (TEI) som nå har utviklet seg som en standard til å kode tekster innenfor humaniora. Forskjellen mellom kodingssystemene er at TEI prøver å by ei mer allmenn koding for nesten alle tekstsjanger man kan tenke seg, mens Menota har momentan et veldig begrenset anvendelsesområde i blikket, fordi dets fokus ligger på det enkelte håndskriftet og lemmatiseringa av ordformene brukt i dette håndskriftet. Men dette vil bare være begynnelsen. Odd Einar Haugen, en av initiativtakene til Menota, beskriver i artikkelen sin Stitching the Text Together (Haugen 2010) at på grunnlaget av håndskriftene i form av enkle dokumentariske edisjoner kan det oppstå en (ny) eklektisk edisjon av en tekst. Når den skal være digital, må edisjonen en gang til bli kodet. Momentan er dette ikke inn i målene til Menota, men på lang sikt vil det utvilsom komme inn som det ble omdiskutert i Haugens artikkel.

I stedet for å vente på ei XML-koding av alle relevante håndskrifter til en tekst til å lage en eklektisk edisjon, kan man også lage litt mindre ambisjonerte leseutgaver til tekstene som møter en stor interesse eller har ei stor betydning innenfor den norrøne filologien, i forskning eller undervisning. Med

henblikk til undervisning av nybegynnere av det norrøne språket jobbet jeg med ei slik leseutgave til Hrafnkels saga Freysgoða. Målet av utgava er en grammatikalsk og semantisk selvforklarende tekst. Det vil si at hver eneste ordform i sagateksten ble eller blir lemmatisert slik at den språkinteresserte leseren får nok hjelp for å forstå syntaksen og betydninga av ordene.

I tillegg er alle ord knyttet til ordbøkene til Fritzner og Cleasby/Vigfusson som gir oversettelser til norsk respektive engelsk, og til Noreens grammatikk som gir mer informasjon om deklinasjonen til ord og enkelte ordformer. Knyttinga fungerer for største delen med enkle lenker. Det går ganske bra med Noreens grammattiken som Andrea de Leeuw van Weenen har overført til ei HTML-fil, og Fritzners ordbok som ble organisert som en database av prosjektet Eining for digital dokumentasjon (EDD). Ordboka til Cleasby/Vigfusson ble overført til ei fil med enkel markup av Sean Crist av prosjektet Germanic Lexicon. Knyttinger til nettsida av prosjektet er bare ei mellomløsning, til bearbeidelsen min av dataene i fila som er fri tilgjengelig, er ferdig gjort. Jeg skal jobbe på å vise bare de relevante ordbokartiklene i en klar og enkel layout. I den omtalte edisjonen er nå bestemt omtrent 90 % av ordene; edisjonen eller leseutgava som jeg kaller den, finnes under http://ecenter.uni-tuebingen.de/hrafnkels-saga/start.html.

Den digitale teksten med alle språklige annotasjoner er kodet som XML etter standarden til Menota, mens Ordbog over det norrøne prosasprog er grunnlaget for normaliseringa. Lemmatisering er lagt etter kodingssystemet til Menota, men for å rekke målene måtte det bli utvidet slik at den grammatikalske kodinga kunne være mer detaljert. I Menotas kodingssystem er det mulig å klassifisere verb som svake, sterke eller redupliserande, mens substantiver kan bare kategoriseres som vanlige eller egennavn. Når det gjelder preposisjoner og konjunksjoner er det mulig å bestemme dem ganske detaljert. Preposisjonen har en reksjon eller blir brukt adverbial. Konjunksjonene deles opp i subjunksjoner og konjunksjoner. Men i kodingssystemet var det ikke planlagt å bestemme bøyingsklassene til substantiver eller verb. Bøyingsklasser er svært interessant for nybegynnere, fordi de hjelper å identifisere ordformer og finne seg til rette i en norrøn tekst. Utvidelsen av kondingssystemet jeg gjorde, fører til at edisjonen kunne være knyttet til andre ressurser på nettet som atter forbedrer edisjonsteksten selv.

I øyeblikket blir det arbeidet på en revisjon av Menotas håndbok om XML-kodinga. Jeg har allerede meldt tilbake til Menota at det burde være mulig å være mer presis med beskrivinga av grammatikken. Sannsynligvis skal et resultat av revisjonen være denne utvidelsen. Med dette kodingssystemet har man et verktøy som ikke bare kan nyttes for å kode håndskrifter, men som nok er detaljert for å finne anvendelse i andre grammatikalsk orienterte prosjekter. I tillegg kan XML-kodinga til Menota blir et digitalt verktøy for forskjellige edisjonstyper (innenfor norrøn filologi), og ikke bare for dokumentariske edisjoner som i dag kodinga brukes til.

Topics: Nordic Textual Resources and Practices
Keywords: digital edition, working with online resources, teaching

Bibliography

– Cleasby, Richard og Gudbrand Vigfusson, An Icelandic-English Dictionary, Oxford 1874.
– Fritzner, Johan, Ordbog over det gamle norske sprog, 4 bind, 2. utgave, Kristiania 1883-96.
– Noreen, Adolf, Altisländische und altnorwegische Grammatik. Laut- und Flexionslehre unter Berücksichtigung des Urnordischen (Sammlung kurzer Grammatiken germanischer Dialekte A, 4), 4. utgave, Halle/Saale 1923.
– Haugen, Odd Einar, Stitching the Text Together: Documentary and Eclectic Editions in Old Norse Philology. In Quinn, Judy & Lethbridge, Emily (Hgg.), Creating the Medieval Saga: Versions, Variability and Editorial In-

terpretations of Old Norse Saga Literature, Viborg 2010, s. 39-65.
– Heimskringla.no, Hrafnkels saga Freysgoða etter Guðni Jónsson - http://heimskringla.no/wiki/Hrafnkels_saga_Freysgo%C3%B0a
– Leseutgave av Hrafnkels saga - http://ecenter.uni-tuebingen.de/hrafnkels-saga/start.html.
– Digital versjon av Altnordische Grammatik av Adolf Noreen - http://www.arnastofnun.is/solofile/1016380.
– EDD, Johan Fritzners ordbok - http://www.edd.uio.no/perl/search/search.cgi?appid=86&tabid=1275.
– Germanic Lexicon Project - http://lexicon.ff.cuni.cz/.
– Menotas håndbok 2.0 - http://menota.org/HB2_index.xml.
- TEI: P5 Guidelines - http://www.tei-c.org/Guidelines/P5/index.xml

Mischievous Machines: A Design Criticism of Programmable Partners

Jörgen Skågeby
Stockholm university, Sweden

This paper presents the results from a design critical reading (Bardzell & Bardzell, 2015; Bardzell, Bolter, & Löwgren, 2010) of the AI-powered social robot Cozmo. Cozmo was released to the market during the fall of 2016 and is described as a "supercomputer on treads". It comes in the form of a small forklift-like vehicle, which most prominent features are the caterpillar bands that drive it, the lift in front of it, and a screen, effectively displaying stylized graphical facial expressions.

The design critique will focus on the notion of programmability (Chun, 2008; Parikka, 2014) and how this condition may affect human-technology relations (Ihde, 1990; Nørskov, 2015; Verbeek, 2011). Programmability is the very precondition for machines as agents in the world. Recurring cultural myths have had a tendency to present agential technologies as either omnipotent masters or as completely loyal servants. However, social robots are likely to complicate that dichotomy. As effectively illustrated by Cozmo, its programmability creates an interesting tension inbetween master and servant. That is, Cozmo is both pre-programmed (with certain secondary agency) and programmable (through its so-called software development kit or SDK). In its pre-programming Cozmo simulates emotive and social capabilities. It is, for example, programmed take certain mischievous initiative when interacting with its surroundings, including human and animal actants. We argue that this is a necessary component of a partner technology, to be playful, to take initiative, and to display some eccentricity. An eccentric relation enacts an amalgamation of, in the case of Cozmo, a quirky personality, a mischievous tendency, and a capability to simulate anger or disappointment. Notably though, an eccentric partner relation can not be allowed to become too eccentric. A completely disobedient and fully self-aware technology has been a prime symbol of fear in several science fiction narratives (e.g. Ex Machina, Matrix, Colossus – The Forbin Project).

Cozmo's programmability also allows users to program Cozmo. From a human-technology relations perspective, the SDK offers a way to open up a black box (Hertz & Parikka, 2012) and form a more concrete and design-oriented relation to technology. Nevertheless, this also spurs a tension between a machine and an "almost human" (as Anki, the company behind Cozmo, themselves put it). The illusion, if you will, of Cozmo breaks down slightly when the 'magic' of it is revealed in a symbolic environment. The question is if the "almost human" (eccentric) qualities will wither when laid bare. At the same time, if Cozmo was only completely obedient (programmed for predictability) it would soon become boring. In other words, eccentricity has to be balanced. If Cozmo was really self-aware and only simulated a well-adjusted eccentric

partnership while interacting with humans, and pursued its private, potentially undesirable, agenda while on its own (or when in interaction with other partner technologies), it would turn into a Trojan technology. This balancing between a strong-willed eccentric partner and a wilful Trojan technology, and the question of where a line can be drawn, may arguably be what will signify human-machine relations in times to come.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: humanistic HCI, human-technology relations, social robots, programmability, coactive technologies

Bibliography

Skågeby, J. (in press) Im/possible desires: media temporalities and (post)human-technology relationships. Confero: Essays on Education, Philosophy and Politics, 4(2).

Skågeby, J. (2016) Media futures: premediation and the politics of performative prototypes. First Monday, 21(2).

Skågeby, J. (2015) The media archaeology of file sharing: broadcasting computer code to Swedish homes. Popular Communication – The International Journal of Media and Culture, 13(1): 62-73.

"En temmelig lang fodtur": hGIS and Folklore Collection in 19th Century Denmark

Ida Storm
Timothy R Tangherlini
Georgia Broughton
Holly Nicol
University of California Los Angeles, United States of America

Introduction

Folklore has played a significant role in the "imagining of the nation" since the inception of the field in the late 18th century. In Scandinavia, the "golden age" of folklore collection of the 19th century coincided with rapid changes in political, economic, and social organization. Although some later folklorists have expressed skepticism about these collections, this skepticism is often based on perceived notions of how these collections came to be, rather than a deep exploration of the actual practices of the collectors themselves. We show how techniques from historical Geographic Information Systems (hGIS) wedded to time tested archival research methods can reveal how a folklore collection came into being. By detailing the routes taken by the Danish folklore collector Evald Tang Kristensen (1843-1929) over the course of his fifty-year career, we trace not only his selection biases for geographic areas (and by extension, social and economic classes), but also the impact that intellectual currents, political developments and changes to transportation infrastructure have on his collecting.

Tang Kristensen over the course of his sixty year career traveled over 67,000 km, largely on foot, visiting ~4,500 storytellers in 4,203 unique places, recording these stories in ~24,000 field diary pages. In this work, we focus on determining how, when and where Tang Kristensen traveled in Denmark as he created his collection. We develop detailed route maps projected onto appropriate historical base maps showing his movement through the countryside. We develop aggregate statistics that allow us to

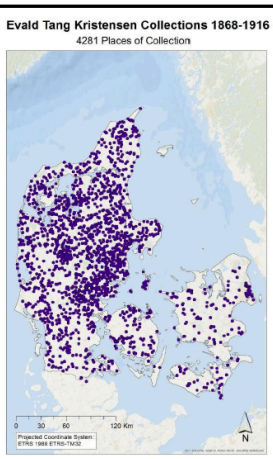

Figure 1: Aggregated stops geocoded with the address locator.

Figure 2: Example of a route.

understand, at a granular level, his collecting habits. In addition, we align the field trips with his writing about collecting, allowing us to approach a "thick description" (Geertz 1973) of folklore collecting in late 19th century Denmark. In all, we map 267 field collecting trips, starting in 1868 and ending in 1916. This work considerably extends qualitative assessments of Tang Kristensen's collecting (Christiansen 2013) and is a key contribution toward the development of the "Folklore Macroscope" (Tangherlini 2013).

Data Extraction

When we began this work, there was no existing catalog of Tang Kristensen's field collecting routes – we had to devise this catalog ourselves by coordinating annotations in his hand written field diaries with his four volume memoir Minder og Oplevelser (1923-1927). The memoir is based largely on letters he wrote home detailing all of his stops while out collecting, and includes information on means of transportation as well as travel dates. Our team began by making "proto-routes." We extracted trip start and end dates, as well as all stops and stop order for each trip by hand, and aligned these proto-routes with the field diaries (Figure 1). In later work, we will also align field stops with our electronic catalog of informants.

Address Locator

Finding the locations for the stops we extracted in our "proto-routes" was a significant challenge. As with most historical data, places can be difficult to locate: some are very small, names have changed, and some places have disappeared. Contemporary gazetteers are inadequate to the task and often confound, rather than solve, queries. To address this problem, we downloaded the historical place name database developed by the Afdeling for Navneforskning, Københavns Universitet, and used it to generate a customized "Address Locator" for 19th century Denmark. We matched the stops from each field trip with the address locator, generating a "best guess" for each field trip. Multiple places with the same name were resolved through the ESRI ArcMap "Interactive Rematch" interface. To derive the final field trip stops, each trip was inspected individually.

Routes

With the stops in a provisional sequential order, we created the "most likely route" for each trip. A basic assumption was that, unless otherwise specified, Tang Kristensen would take the shortest path between two points, an assumption that aligns with the underlying "Network Analyst" algorithm in ArcMap. We used a transportation network from OpenStreetMaps pruned against the cadastral survey maps of ~1880, the highest resolution historical maps from the era. Since Tang Kristensen occasionally traveled by boat, ferry routes based on ferry schedu-

les and close study of historical maps were also added. By feeding the provisional sequential stops to the network analyst, we were able to create the most likely routes for each trip as a single line record. These routes were then visualized as a line with sequentially numbered stops (Figure 2, see previous page). The visualization is augmented by simple statistics, such as route length, as well as descriptors from our database, including dates of collection, field diary page ranges, and modes of transportation.

Animations

Animations provide a dynamic representation of Tang Kristensen's movement through the countryside. Current animations reveal, for example, the numerous times where he backtracked, and can be used to augment the understanding derived from the static maps. To allow for sequential animations of all fieldtrips, we devised an additional "absolute order" field, and split all routes into inter-stop segments.

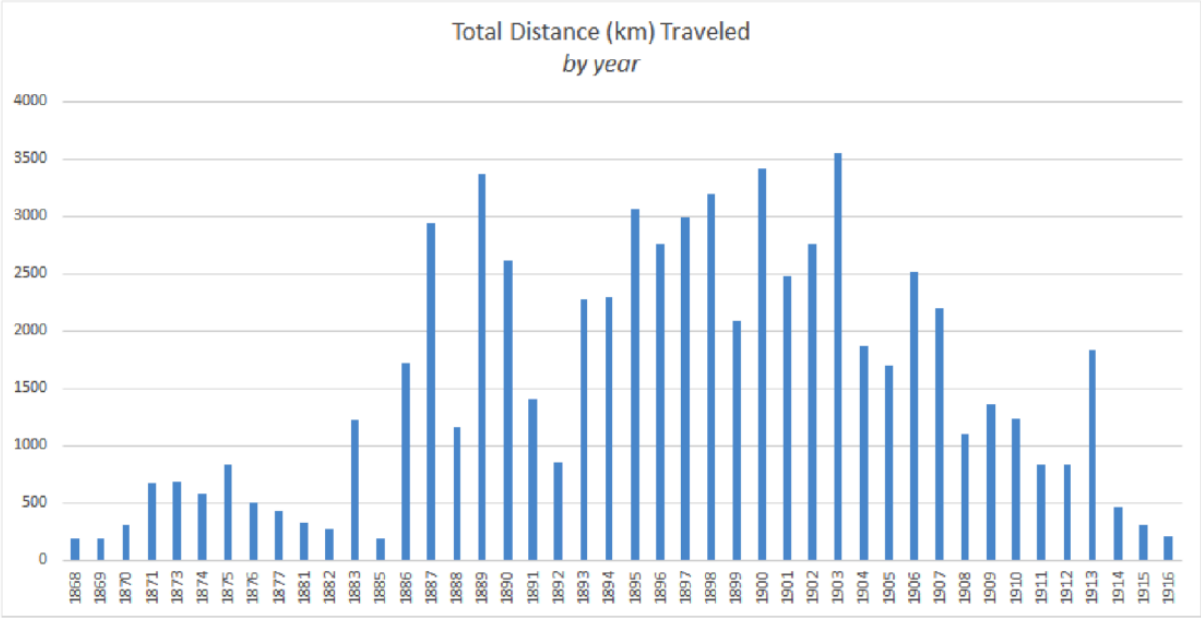

Figure 3: Chart of total distance traveled by year over the course of Tang Kristensen's career.

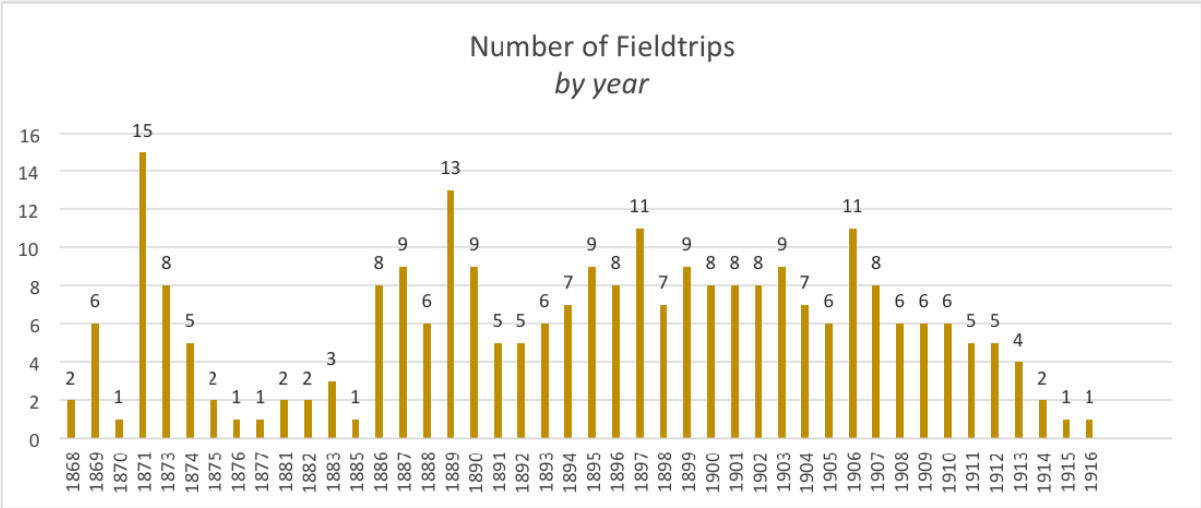

Figure 4: Chart of number of field trips by year over the course of Tang Kristensen's career.

Travel Statistics

By splitting routes into inter-stop segments, we could develop more detailed statistics regarding segment length, speed of travel, and travel mode. More importantly, we can now aggregate segment statistics and align this information with other data, allowing us to address a broad range of questions. For example, we can see how far Tang Kristensen traveled when he lived in a specific place, his travel distances at different times of year, and his travel distances in different parts of the country. Furthermore, we can consider changes in average travel segment or field trip distance over time. Future work will align stops with storytellers, allowing us to include story statistics with the field trip statistics. Population data and transportation data will further add to this picture (Figure 3 and 4, see previous page).

Topic Models and Field Trip Descriptions

We consider each field trip description in Minder og Oplevelser a "document" and use this collection of documents to constitute the corpus of field trip descriptions. Using a probabilistic topic modeling algorithm (LDA), we model these descriptions at varying topic levels (k=10-30 at intervals of 10) to uncover latent topics in his descriptions. This modeling allows us another method for aggregating field trips. We can then explore the characteristics of field trips associated with a particular topic. This work is a preliminary step toward aligning the field trips with the stories collected on those field trips (Figure 5, see following page).

Conclusions

Our work reveals the shifting parameters of Tang Kristensen's field collecting, from his intensely local focus early on to his more expansive and confident travels at the end of his career, when his collecting was no longer aligned with Romantic nationalist goals, but more in tune with a thick descriptive approach to Jutlandic rural life. By using hGIS techniques, we can provide a degree of detail about his travels missing in earlier studies.

Our approach enables a truly macroscopic approach to folklore collecting, allowing us to interrogate Tang Kristensen's field collecting at varying levels of resolution. For example, we can move from the micro-consideration of a single field trip, to a meso-consideration of all trips that included a particular parish, to a macro-consideration of all of his trips taken as a whole.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: historical GIS, folklore, named-entity detection/extraction, culture analytics

References

Christiansen, Palle Ove. Tang Kristensen og tidlig feltforskning i Danmark. National etnografi og folklore 1850-1920. Copenhagen: The Royal Danish Academy of Sciences and Letters, 2013.

Tang Kristensen, Evald. Minder og Oplevelser. Volumes 1-4. Viborg: Forfatterens forlag, 1923-27.

Tangherlini, Timothy R. "The Folklore Macroscope." Western Folklore 72.1 (2013): 7-27.

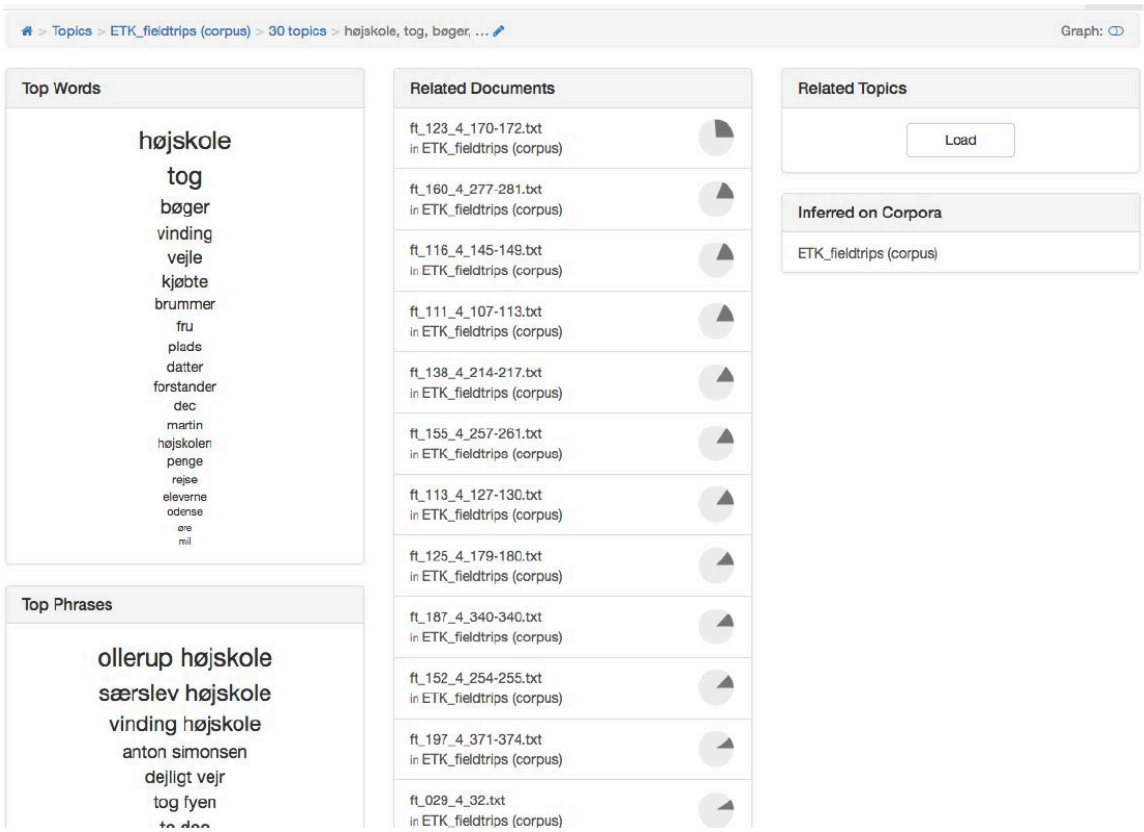

Figure 5: Topic "Højskole" in Tang Kristensen's field trip descriptions, with field trips ranked by importance of topic to the description.

Representations: The Analogue Photography as a Digital Source

Arthur Tennøe
National Library of Norway

Libraries have had to consider both the medium and the message, in different ways. Twentyfive years ago the archive of photography could be called a frozen stream of pictures. Since then we have been facing vast changes in the production, consumption and distribution of photography in the libraries

The National library of Norway has initiated and been involved in several projects that focuses on the current shifts: The archive in motion, 80 million pictures and The ends of photography. The presentation will focus on some aspects of these projects and the themes involved.

An archive is never the only end of photography. It endeavours to take its deposits toward other ends and uses. In the 1990s the library, opened the first digital photo-databases in Norway, and later went on to digitize, photographing the whole national collection of the library, and to make it accessible in an integrated search, making accessible objects from all media together, a work in progress.

Our ambition is to be an active part of the research infrastructure. Our task is to make possible the understanding and interpretation, the adequate contexts, to make every relevant document in all sorts of media a source of research for the future. This involves challenges of collecting, registering, cataloguing, conservation and preservation, digitizing and long-term preservation. To give acces to photography, we have, along the way, to make choices that will have impact on the possibilities for use now and in the future. What is lost/gained in the translation from analogue to digital, focusing on the photographic content, negatives and vintage prints, series, the original context of use and the archival aspects? We cooperate on international database-projects. What are the consequences of the new digital archiving? The presentation will use examples from daguerreotype to digital in our library.

For the last twenty five years The National library of Norway have been converting it's photographic collections into the digital format. Its mission is both to give access to content but also to the medium that carry and creates it. It started with databases before the internet and has been a long road of developments since. This has qualitative and quantitative implications for the use and understanding of the collections. It started with manual registration and small files and now is heading towards automatization of all processes in a high quantity output.

The analogue photo however is a complex object in itself. The photographic projects different stages results in different material artefacts. From the nineteenth century photographs produced for many different ends, portraits, prospects, records and private archives. The choices we make will make a difference for the output. Photography also started to convert into different printed documents suchs as newspapers, magazines, books, commercials, postcards etc. As a result of digitization of these other media both printed versions of the photographic images and metada are already present in the big data collection of the library, or from other sources. Recent projects are as different items as unika daguerreotypes from the 1840s, thousands of aerial photo from all municipalities of Norway and newspapers photographs with millions of photographs. They shall be published on our new net site for to be looked at and to be a proper source for research. New methods of scanning, image recognition and OCR create new possibilities for search understanding and contextualisation of the objects and their content.

The Daguerreobase is an online application designed to contain detailed information about daguerreotypes. Members can view, edit and store records of individual daguerreotypes and establish relations to other records based on a wide range of characteristics. This includes collections, owners, creators, hallmarks, housing models, sizes, materials and free text descriptions. Dager-

reotypes mainly created in 1840s and 50s was a very international activity and needs to be studied in this perspective. Metadata can originally be very sparse but through this cross-over project a new foundation has been established for the study of this important historical objects, rarely seen outside the archives before. The partners include a blend of institutions from all over Europe. Bergen City Museum; FotoMuseum Provincie Antwerpen; Museo Universidad de Navarra; Biblioteca Panizzi, Reggio Emilia; Museum of Decorative Arts in Prague; National Media Museum, Bradford; Musée Gruérien, Bulle; Rijksmuseum, Amsterdam; Agence Roger-Viollet, Paris; Albertina, Wien; The Royal Collection Trust; Finnish Museum of Photography; Oslo Museum; private collections and others.

Aerial photo is photography of landscapes and buildings taken from the air. The earliest aerial photos was photographed from balloons. First man out was by what we know the famous French photographer Nadar. Already in 1858 he took photographs of Paris seen from above.

Aerial divided into two main types. Vertical Photo is taken straight from the top down (perpendicular) and oblique photos taken at an angle down (acute angle). Vertical Photo has usually been used to produce maps, surveying and military intelligence. Oblique photographs give a more three-dimensional impression of the photographed. We glances obliquely down on mountains, hills and buildings. The design observed thus not only from above, but also partly from the side. We get in many ways more topographic information from oblique photographs than from vertical photographs. Oblique photographs are thus a rich topographical source material with a variety of applications. They can be used for everything from solving border disputes, detecting changes of buildings and vegetation, or for documentation of historic gardens and other things that can be changed or disappeared since the pictures were produced. Aerial photography has even been used for reindeer counting on the Hardangervidda (1948). Collecting this archive end transforming all the different part into digital representations are a complex process.

A even newer project are the huge press photo archive from the norwegian newspaper Bergens Tidende. The library will from 2017 start its newest line that will digitize the archive of over 5 million photos. This demands a new look on the possibillies of automatization of processes that starts even before the transport of the collection to the library. In the National Librarys database this collection also convergerges with the other materials digitized in other projects.

Bokhylla.no (The Bookshelf) is a collaboration project designed to provide online access to literature published in Norwegian based on a formal agreement between the National Library of Norway and Norwegian beneficial owners represented by Kopinor.

The service will cover around 250,000 books when completed in 2017. Books from the entire 20th century will be available to anyone with a Norwegian IP address. Books not protected by copyright may be downloaded.

The digital newspaper service is based on agreements between the National Library and a growing number of Norwegian newspapers. These agreements secure digital delivery of new publications and the digitization of historical newspaper archives. A central aspect of these agreements is the right of Norwegian libraries to make newspaper archives available on their premises.

Since many photos are already published in books and newspapers. This means that this sources together will give us a new possibilities never seen before for tracing the shooting, alternative takes, all publications, and historical impact of the photos.

So: The paper will give the background for the new situation for photography as a source of knowledge in the digital contexts and on this basis reflect critically on the subjects such as quantitative perspectives, contextualization and metadata constuction to reflections on the specific qualities of the photographic medium.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: Analogue-photography, digital-source, metadata, visual-media, big-data

Bibliography

Forglemmegei - Autochrome, i Historikeren n. 3 2016
Norwegian official photo no. L.8846., i Historikeren n.2 2016
Bergensbanen, i Wilse mitt norge, Oslo 2015

The Trading Faces: Online Exhibition and Its Strategies of Public Engagement

Alda Terracciano
UCL, United Kingdom

This paper will explore the dynamic relationship between artistic practices and digital humanities in the creation of an online platform launched in 2009 to commemorate the 200th year anniversary of the parliamentary abolition of the Transatlantic Slave Trade in Britain. It will consider how orality was used in the process of selecting and digitising archival material related to the history of performing arts produced by people of the African diaspora in the UK. More specifically, the way in which African oral traditions and techniques of storytelling played a role in the process of designing and constructing the first online exhibition on the legacy of the Transatlantic Slave Trade in British performing arts and society.

Amongst its various activities, the project, which was produced by a consortium of partners including Future Histories, Talawa Theatre Company, The National Archives, and the Victoria and Albert Museum, focused on the preservation and cataloguing of a number of black theatre archives from the Theatre Collections of the Victoria & Albert Museum, and the Talawa Theatre Company archive, as well as the digitisation of 257 archive items (totalling to about 600 document pages) for online publication. The selection process was based on consultation with a number of lecturers, students, artists, community group representatives, academics and experts in the field. The process was meant to open up the curatorial practice and move it towards current forms of public engagement, co-creation and democratic participation in knowledge production. The paper will discuss how this approach bore direct consequences on the creation of the online platform, with regards to its design, as well as the kind of archive material, which was made accessible for the first time on the online platform.

The critical question was not only to include voices from outside the heritage sector, but also to re-mould the practice of archiving and the representation of archival material, as well as to resist the tendency of objectifying the past within rigid co-ordinates of time and space. The paper will discuss the involvement of black artists in the curatorial process as a way of privileging a synchronic rather than diachronic approach to history and memory. This resulted in a number of essays on art forms originated in Africa, which were produced by a number of artists of African descent living in the UK, not only to provide a critical and historical context to the exhibition, but also to shift its focus from the object of the analysis to the discourse that produces it. This is an approach to history indebted to the African practice of 'Orature', which implies a circularity of knowledge and a creative exchange between performers and members of the audience.

The paper will also discuss how the heritage of the Transatlantic Slave Trade within British performing arts and society was set against the wider context of cultural identity and performance, and in particular contemporary forms of migration and human trade in Britain. To do so it will analyse the Voices section of the exhibition, which juxtaposed the experiences of 18th century African abolitionists Olaudah Equiano and Mary Prince, extracted from their autobiographical accounts, to the testimonies of two present day migrants from China and Russia named as Natasha and Liu to protect their real identities.

Supported by historical essays and links to further resources, Equiano and Prince's

views were set against the stories of Natasha and Liu, whose memories of their degrading treatment in the UK, recorded and filmed in London in April 2008, uncannily resonated with those voices from the past. Their memories reflected the pernicious continuity of two key aspects of the Transatlantic Slave Trade: economic exploitation and the infringement of human rights.

By asking the question "Has slavery really ended?" the exhibition looked at these two moments in human history in their resemblance, as well as crucial differences. The history of the Transatlantic Slave Trade was one of human subjugation, but also of racial discrimination, as the de-humananization of people from the African continent was key to economic exploitation. The condition of people trapped in human trafficking today resembles the past, but is also different: shorter periods of so-called 'enslavement', general absence of a racial bias, different juridical status, and so on. Nonetheless, forms of enslavement of human beings are still taking place in dirty, dangerous and difficult work in Britain, in the running of private homes, the care of the elderly and disabled and in keeping the sex industry alive. Many vulnerable people are trafficked or smuggled into the UK today. Natasha from Russia and Liu Bao Ren from China are two of them.

The attempt of the online exhibition was to bring alive the resilience and resistance of people of the African descent by activating intimate narrative points through different configurations, both visual and audio, which would facilitate the exploration of emotional and political connections between histories and stories from the past and the present.

To better elucidate this point, the presentation of the paper will intersperse research outcomes from the exhibition with the screening of video sections related to the experience of Natasha, an underage young Russian woman forced to prostitution in the UK today and Liu, a Chinese man who experienced forced labour in the UK.

Finally, the paper will consider the unfinished issues of public engagement and active contribution to the project, setting this aim once again within the context of Orature and current intercultural artistic and cultural heritage practices in Europe.

As the project evaluator commented in the End of Project Evaluation Report dated 8 March 2009:

"Inspiring confidence and trust in the public to submit material to a public site is no mean feat: such work needs to be done through outreach activities such as presentations and discussions; and by encouraging established contacts such as lecturers, teachers, community course directors etc. to integrate the Online Exhibition into their programmes so as to enable more engagement and submissions by the public. Funding pending, efforts to publicise the site in order to receive more submissions should continue by those responsible for Trading Faces: Recollecting Slavery site's maintenance so as the target of 25 submissions for 'Open Doors' is now met by the end of this year 2009." (Raminder Kaur)

The paper will consider the challenges faced by FUTURE HISTORIES, the organisation responsible for the delivery of the online exhibition, in its attempt at stimulating the production of new material to be uploaded on the online platform. It will frame them within the wider context of the organisation attempts at 'popularizing' the use of primary resources beyond academic and post-colonial intellectual circles, referencing the categories of 'speech' and 'history' to reflect on the intrinsic intersubjectivity of the archiving medium and the multiplicity of voices encompassed by black British performance.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: black cultural heritage

Bibliography

Terracciano, Alda. "Future Histories – An Activist Practice of Archiving," in Popular Postcolonialisms: Discourses of Empire and Popular Culture, edited by Atia N. & Houlden K. London: Routledge. Awaiting publication.

Terracciano, Alda. "Mapping Memory Routes: A multi-sensory experience of

7 cities in 7 minutes," in Curating the City. Proceedings of Challenge the Past / Diversify the Future, University of Gothenburg. Awaiting publication.
• Terracciano, Alda. "Trans-national politics and cultural practices of the Trading Faces online exhibition," in Black arts in Britain: Literary, Visual, Performative, edited by Annalisa Oboe and Francesca Giommi. Roma: Aracne, 2011.
Terracciano, Alda. The Future Histories Research Toolkit for African, Asian and Caribbean Performing Arts. 2009. Accessed 24 June 2016. http://www.tradingfacesonline.com
Terracciano, Alda. "The Black Theatre Forum and the Experiments of the Black Theatre Seasons," in Alternatives Within Mainstream: British Black and Asian Theatre, edited by Dimple Godiwala. Newcastle: Cambridge Scholars Press, 2006.
Terracciano, Alda and Kaur, Raminder. "South Asian / BrAsian Performing Arts," in A Postcolonial People: South Asians in Britain, edited by Ali, N. Kalra, V. and Sayyid, S. London: C. Hurst & Co, 2006.
Terracciano, Alda. "Together We Stand," in Navigating Difference, edited by Heather Maitland. London: Arts Council England, 2005.

The New Lexicon Poeticum

Tarrin Wills
Københavns Universitet, Denmark

The New Lexicon Poeticum (lexiconpoeticum.org) is a project to produce a new lexicographic resource covering Old Norse poetry (initially the category known as skaldic poetry). It is based on the corpus produced by the Skaldic Project (supported project no. 60 of the Union Académique Internationale, with funding provided by UK Arts & Humanities Research Council, Australian Research Council, Joint Committee of the Nordic Research Councils for Humanities, the National Endowment for the Humanities, Deutsche Forschungsgemeinschaft and other bodies). SkP is nearing completion, with over 80% of the corpus entered into its digital resource.

SkP was inspired by major problems with previous research, in particular Finnur Jónsson's edition (1915-18) of the corpus of skaldic poetry (Skj) and the dictionary based on it (Lexicon Poeticum, 2nd ed. 1931). While Skj is a monumental work which has provided the foundation for almost a century of skaldic poetry studies, Finnur Jónsson used a heavy hand of intervention, with frequent and silent emendation. His lexicon, based on his own corpus, is therefore founded on a body of material that does not accurately reflect the manuscript evidence. It includes a large number of words that only exist through editorial conjecture, and omits large numbers of words that are evidenced in the manuscript tradition, particularly as manuscript variants are largely ignored. This situation has left a significant gap in methodologies between the material evidence of the poetic lexicon and the resources to analyse it.

SkP provides the foundation for the current project because it will have re-edited the entire corpus based on current philological and textual editing methodologies. The edition is in the form of a digital resource (skaldic.abdn.ac.uk) from which the printed volumes are exported. It links together the normalised, occasionally emended edition with variant readings, manuscripts, secondary literature, prose contexts and previous editions. It includes unnormalised transcriptions of the main manuscripts of the corpus and significant numbers of variant manuscripts. The new resource will be linked directly to these resources, enabling the lexicon to be understood in its complex contexts.

ONP, founded in 1939, is the major dictionary of Old Norse. The poetic corpus was specifically excluded from ONP because of the lack of a reliable edition of this material — a lack that is now being addressed by SkP. ONP has a sophisticated database with a web interface that links the lexicon to the citation index and textual corpus. It uses re-

liable diplomatic editions and manuscript spellings, but is reliant on those editions rather than the manuscripts themselves.

The skaldic project's corpus is in a relational database structure with all words entered as separate items, with a normalised syntax and translation linked to each word, along with linked manuscript information including variants. It differs from lemmatised XML texts in that the lemmata (headwords) are linked to the (future) dictionary entry. The nature of the corpus is such that there are a very large number of headwords: with 100,000 words lemmatised, over 13,000 headwords have been linked to the corpus. Lemmatising produces an automatic concordance with a full set of contextual translations. Owing to the structure of the corpus database, each headword can be linked to its manuscript witnesses and to nominal periphrases (kennings) in which it occurs.

There are a number of questions that arise from the project as it has been conceived:

1. How to create interfaces for linking hundreds of thousands of words to tens of thousands of headwords. Additionally, variants add another 20% to the corpus, but need their status and relationship to the manuscript preserved. All this information must be in a form that can be checked and updated. Some forms of analysis were performed by the original project (diction (kenningar and heiti), translations, free text variants); others were not (lexical variants, lemmatising, compounds).

2. How to maintain alignment with both the original database and other lexicographic projects, particularly ONP, so that a word's use and history can be researched across corpora.

3. As a more general question, how to create a meaningful and useful lexical resource when the original and underlying corpus is so rich in itself, with translation, notes and commentary linked to each word — and how to publish it in the current metrics-driven research environment.

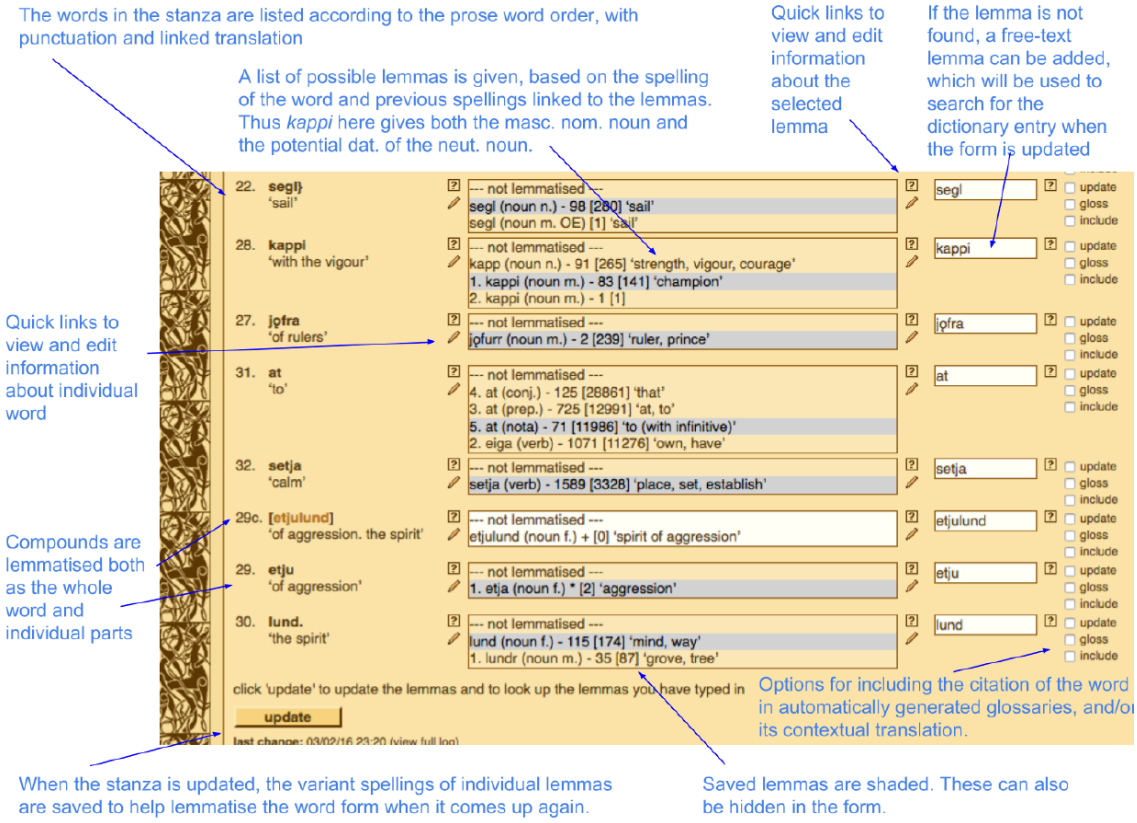

Figure 1: Detail of form for assisted lemmatisation

User interfaces

The original skaldic project uses a web interface to enter, edit and manage the data of the project. Relational databases differ from XML as there is no inherent connection between the data structure and its digital storage (serialisation). This has the advantage that the data can easily be exported in a number of ways, but direct editing of the data is not easy to perform. Early on I developed a web application for both viewing the edition, browsing the contextual information and editing the data, with customised forms for entering the textual data, and a generic interface for dealing with other information. This allowed editors to produce editions where a putative natural prose order is linked to each text (allowing for easier interpretation and potential morphosyntactic analysis), as well as a translation, with each word linked and reordered. Each stanza has a full set of linked manuscript references, as well as variants linked to both the words and manuscripts.

The process of lemmatising has been performed on the original corpus, again facilitated by the user interface. A web form lists all the words in a stanza or block of text. The user can select the lemma if it has the same form as the text, or look up the lemma by entering a search term. Variations in form and spelling are saved and used to prompt the user when they next occur, although all choices must be confirmed manually. The word list was originally taken from ONP (with permission) and has been supplemented as new headwords are identified (Figure 1, see previous page).

The new lexicon will include all variant manuscript readings, something that previous lexica poetica have not documented systematically. As the original variants were entered as free text, rather than as words within the data structure for words in the database, the new project needs to add these to the corpus. To aid this process I have created a web form which uses the variant apparatus in the corpus database to prompt the user to add lexical variants and link them to headwords. This is a complex process, with no direct correspondence between the words linked in the main text and those in the variants, but the interface attempts to analyse the information in the database to facilitate the process (Figure 2).

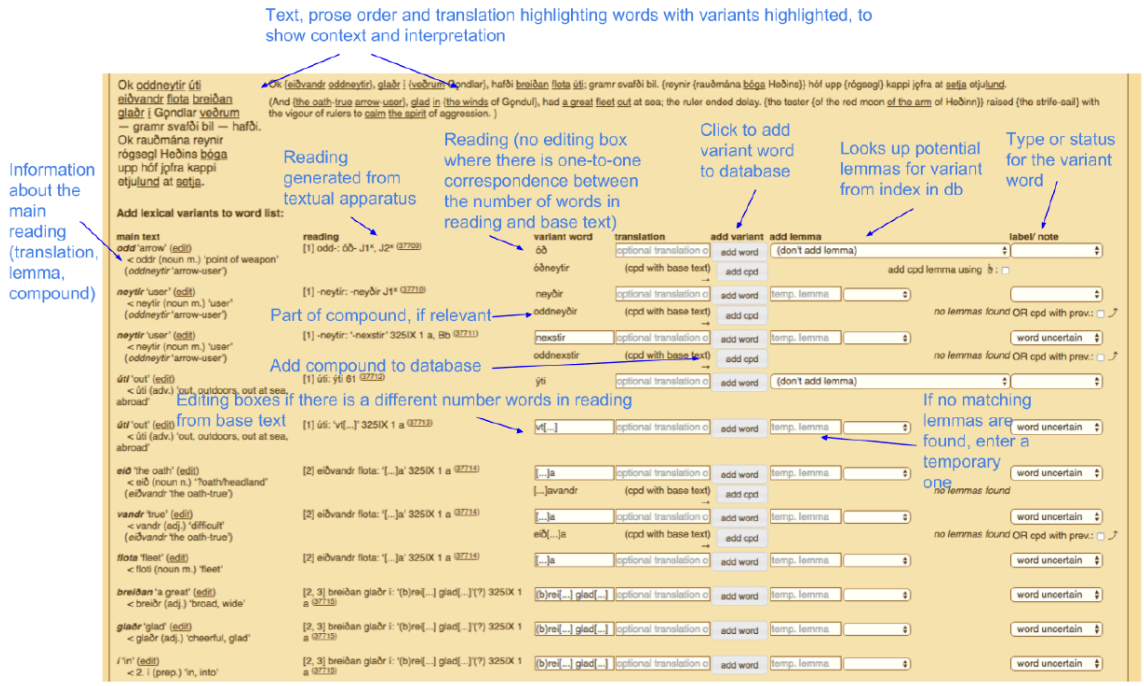

Figure 2: Detail of form for adding and lemmatising lexical variants

Relationship to other dictionaries

The original word list for the lexicon was copied from ONP almost a decade ago. Unfortunately the original unique identifiers for this list were not saved, and both the original ONP wordlist and the new lexicon's wordlist have continued to evolve. The connection between headwords in the two lexica is not reliable but we are making efforts to recover and check this information so that a single interface can be built to both resources.

There are still some questions regarding the nature and function of the new lexicon. The process of lemmatising a corpus with translations linked to each word produces already a concordance of all words with a gloss that effectively gives the interpretation of that word by the editor. Further information about each word can often be found in the notes linked to the word. What, then, does a dictionary entry for the word add to the information already available? Additionally, the prose dictionary ONP will have more comprehensively covered the more common words in the lexicon. Should LP simply supplement that lexicon, or should it be a full description of the skaldic lexicon in its own right? These questions derive from broader issues about the nature of traditional scholarship as DH methods become increasingly sophisticated.

Using and visualising the data

The linking of the rich corpus to dictionary headwords in itself provides an enormous amount of information for each word. The current interface shows all instances of each word with contextual translation and linked notes where relevant, plus compounds. Words occurring within kennings (nominal periphrases) are also explained in this context. Additionally, using the linked manuscript information, all manuscripts representing the word in both the base text and variants can be listed.

Analysis can be performed on this information to see, for example, the way parts of speech are distributed within each stanza and half-stanza of poetry. We plan to perform more nuanced analyses of the metrics by using the grammatical information linked by this process to identify line types (e.g. the Sievers/Kuhn system).

Additional dating information for both the manuscripts and the poetry (albeit unreliable at this stage) allows us to trace the history of the word in its poetic and material sources. Likewise, adding geographical data based on the poem's place of composition and/or recitation allows us to perform diatopic analyses of the words and language of the corpus.

Topics: Nordic Textual Resources and Practices
Keywords: Old Norse, lexicography, poetry, relational databases, web interfaces

Bibliography

Tarrin Wills, 'The thirteenth-century runic revival in Denmark', NOWELE 67 (2016), 114-129.

Tarrin Wills, 'Social Media as a Research Method', Communication, Research & Practice [special issue 'Digital Media Research Methods: How to research and the implications of new media data'], 2:1 (2016), 7-19. doi:10.1080/22041451.2016.1155312

Tarrin Wills, 'Semantic modelling of the Pre-Christian Religions of the North', Digital Medievalist 9 (2014) <http://www.digitalmedievalist.org/journal/9/wills/>

Tarrin Wills, 'Relational Data Modelling of Textual Corpora: The Skaldic Project and its Extensions', Literary and Linguistic Computing [Digital Scholarship in the Humanities] (2013) doi:10.1093/llc/fqt045.

Odd Einar Haugen, Matthew Driscoll, Karl Gunnar Johansson, Rune Kyrkjebø, Tarrin Wills, The Menota Handbook: Guidelines for the electronic encoding of medieval Nordic primary sources (Bergen: Medieval Nordic Text Archive (Menota), 2008).

SHORT PAPERS

What's Missing in This Picture? Political Change and Wordscapes of Latvian Poetry

Anda Baklāne
National Library of Latvia

Lexical and semantic change is an ongoing process in language and literature in particular; following cultural and technological developments, new words enter the circulation while others are shunned. Political factors also greatly contribute to this process, altering the vocabularies of discourses, both evidently and subtly. Drawing the conclusions from the world-list analysis of comprehensive corpus of Latvian 20th century poetry, the paper looks into the lexical change that followed distinctive political turning-points in Latvian history - Soviet occupation in 1940s and regaining of independence in 1990.

It has been previously established that in the aftermath of World War II literary process in Latvia was greatly affected by the censorship and new ideological tasks that writers had to assume. A number of topics were officially banned from the creative writing (such as criticism of Soviet life along with references to mysticism, religion, certain historical events etc.) while other topics and utterances quietly vanished from the literary discourse, since such features as expressions of sadness, displays of intimate feelings, and vagueness in general were harshly criticized. The range of topics as well as vocabularies of authors notably broadened again in the 1970s and 1980s, however, the textual scene already remarkably differed from that of 1930s. At the beginning of 1990s, the collapse of the publishing industry trimmed the production of literature, nevertheless, the change of political regime opened seemingly endless possibilities for topics (or, for that matter, avant-garde non-topics) that now could be discussed.

The aim of this study was to explore if or how these developments can be traced and described in computational analysis of the word usage, as well as to look for patterns that are less apparent and only identifiable via computation. The corpus examined in this study entails 480 poetry books, which were published for the first time between 1920 and 1999 (samples of 60 volumes per decade). This is a new dataset, which was aggregated during the winter of 2016/2017 and has not been statistically studied before. For this paper, analysis is based on results retrieved from two different tools – the corpus analysis toolkit 'AntConc' and the environment for statistical computing 'R'.

While the analysis of word-lists often yields interesting conclusions and hints for further research, there is always the challenge of displaying the results. The visualizations are important not only as means of presenting the information in a way that is audience-friendly and appealing, but also as cognitive tools that can help the researcher to discover overlooked links and anomalies. In this paper, several visualization tools are explored for displaying the "wordscapes" of Latvian poetry as they change in time - starting with simple Excel graphs and easy-to-use contemporary web-based tools, such as 'Voyant', to more sophisticated network visualization tools, among them the open-source software 'Gephi', which require more skills, however, can also render more exciting and possibly revealing results.

In order to introduce the digital methods into the mainstream of humanities research, it is important to develop tools (or user interfaces) that are not forbiddingly complicated, hence, the approach in this study was not to look for "new" and particularly smart methods, rather to find simple and mature solutions that could be recommended to researchers as ready-to-use and effective while working, for example, with the digitized materials of cultural heritage at the National Library.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: digital literary stylistics, digital culture studies

The Space Between: The Usefulness of Semi-distant Readings and Combined Research Methods in Literary Analysis

Karl Berglund
Uppsala University, Sweden

The study of literature has traditionally been a qualitative scientific endeavour. Researchers have, generally, analysed few and canonised works, and these works have been examined in great detail, with "close readings" being the typical choice of method. Franco Moretti's term "distant reading" and the rise of digital humanities and different sorts of text mining methods have, at least partly, changed this.

Moretti and his ilks in a way turned the scholarship of literature upside down by focusing units much bigger or much smaller than singular works of literature: "devices, themes, tropes – or genres and systems", as Moretti put it. Instead of reading the most well-known works of a specific period or genre it was suddenly possible to read almost all literature published and draw other sorts of conclusions (though this also meant handing over much of the reading process to computers).

Most literary scholars engaged in text mining have used very large data corpuses. The general rule seems to have been "the bigger data, the better". This is certainly true when it comes to showing statistical patterns etcetera. However, the bigger the material, the longer is also the distance between the machine-generated results and the qualitative analysis of these results. If your corpus consists of thousands of books it is simply not possible to know the content of this corpus very well. This is at the same time the strength and the weakness of the text mining research on literature conducted in recent years.

In my opinion, debates about pros and cons of text mining methods in the study of literature have been far too black and white. Instead of either or (big data or canonised work, distant reading or close reading) I argue for a position in between. My field of study is contemporary Swedish crime fiction. In an on-going study I use a corpus of 116 Swedish crime fiction novels published 1998–2015 and written by the most well-known and commercially successful authors of this period of time. Hence, I do not analyse the entire genre, yet not only the most renowned novels within it, but instead around ten per cent of all Swedish novels published in this period (the top decile). This choice makes it possible to both get the bigger picture and be very familiar with the material.

Moreover, I approach this corpus through a combination of methods, where some are computer aided and digital (word frequencies, topic modelling), others more traditional and analogue (shallow thematically-oriented readings of the entire corpus). Together these methods provide solid knowledge of the genre that is both quantitative and qualitative.

In my presentation I argue that such a combination of methods on semi-big data or corpuses can be very fruitful to many literary studies, with material from different epochs and genres. Literary scholars should start to make use of this "space between" the very distant and the very close, and let computer-aided methods serve as a helping hand rather than a goal in itself.

Topics: Nordic Textual Resources and Practices
Keywords: distant reading, text mining, method, popular fiction, crime fiction

Bibliography

Mordförpackningar. Omslag, titlar och kringmaterial till svenska pocketdeckare 1998–2011, (Uppsala: Uppsala universitet, 2016), 283 pp.

"Ett halvt sekel litteratursociologi. En kvantitativ genomgång av skriftserien Skrifter utgivna av Avdelningen för litteratursociologi vid Litteraturvetenskapliga institutionen i Uppsala 1967–2015", Spänning och nyfikenhet. Festskrift till Johan Svedjedal, (ed.) Gunnel

Furuland, Andreas Hedberg, Jerry Määttä, Petra Söderlund & Åsa Warnqvist (Möklinta: Gidlunds, 2016), pp. 482–499

[Review: Matthew L. Jockers, Macroanalysis: Digital Methods and Literary History, Urbana: University of Illinois Press, 2013], Samlaren. Tidskrift för forskning om svensk och annan nordisk litteratur, vol. 135, 2014, s. 342–345

"A Turn to the Rights: The Advent and Impact of Swedish Literary Agents", Hype: Bestsellers and Literary Culture, (ed.) Jon Helgason, Sara Kärrholm & Ann Steiner, (Lund: Nordic Academic Press, 2014), pp. 67–88

Deckarboomen under lupp. Statistiska perspektiv på svensk kriminallitteratur 1977-2010, (Uppsala: Uppsala universitet, 2012), 224 pp.

"These Memories Won't Last": Visual Representations of the Forgotten

Jennifer J Dellner
Ocean County College, United States of America

"These Memories Won't Last:" is a digital comic (2012) by Stuart Campbell that depicts his grandfather's descent into Alzheimer's and their, both the grandfather's and Campbell's, attempts to piece together and make sense of two simultaneous pasts: the grandfather and his life as a WW II soldier as well as Campbell's memories of his grandfather's forgetting. Beginning with Campbell's digital comic, this study examines two other pieces of e-literature, Strasser and Coverly's "in the white darkness" (2004) and Wilks' Rememori (2012), a digital poem and game respectively, whose common aim is to present experiences and representations of memory loss: while primarily visual pieces, each seeks to invoke in the reader the diminished ability to access and make sense of the past.

Drawing upon epistemological theories of constructivism and emergent learning (e.g. McMurtry, Osberg, Biesta and Cilliers) and those concerned with the affordances of digital art, representation, and interaction (e.g. Strickland, Coverly/Luesebrink, Augsburg), the paper explores the ways in which the works themselves theorize relationships between experiential knowledge and constructions of "the past." Since a good deal of digital literature is non-linear, it is not enough that these pieces simply use ergodic modalities; instead, the focus of the study are specific visualizations of the forgotten or fading past. Thinking of knowledge in relational terms is to see it as a dynamic relationship between the knower and world, a participatory relationship where knowledge allows the knower "to interact effectively with something else" (McMurtry). Memory loss reconfigures that world and results in a loss of the efficacy of interaction as modes of relating begin to weaken, shift, and disappear. The fact that these pieces are about memory loss suggests specific configurations of the experience of knowing/forgetting and allows us to backwards engineer, in a sense, the epistemological implications that underpin the visualization of what is being lost.

As such, the back end of the visualization will be explored in terms of its relationship to these ideas. Strickland (2009) writes, "time-space processing in e-lit is of another sort [from print literature]. It encompasses … kinds of time-space processing that authors set out deliberately to explore, because the computational situation allows them to imagine and build with their (code) writing." In "These Memories Won't Last," the image of a disappearing rope serves to link vignettes of the narrative together at the same time as it signals the grandfather's inability to do so. The more one manipulates it or scrolls back, the more it fades and becomes irretrievable. This design is ironically dependent on jquery architecture, a feature of which is chaining, represented, I argue, as the rope or thread that stands for the grandfather's memories and his attempts to chain or link them into a coherent past of memories; as these fail, the very chaining in the code ena-

bles this representation. The paper concludes with an examination of the tensions between the visualization of pasts lost and the techno-artistic choices that encode them.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: memory, forgetting, e-literature, digital comic, design

Bibliography

Forthcoming: DH as Intervention, Hybrid Pedagogy, early 2017

McMurtry, A. & Dellner, J. (2014) Relationalism: An interdisciplinary epistemology. Or, why our knowledge is more like a coral reef than fish scales." Integrative Pathways: Newsletter of the Association for Interdisciplinary Studies, 36 (3), 1, 6-12

Chapter in a Book: "Children of the Island: Ovid in the Poetry of Evan Boland and Derek Mahon," in (ed. J. Ingleheart) Two Thousand Years of Solitude: Exile After Ovid, Oxford University Press. 2011

"The Big End: William Gibson and the Ecology of Cool," American Fiction Reflecting Global Ecological Concerns, ed. Linda Cook, Cambridge Scholars Press. Tentative: Under Contract

From Theory to Practice: The Sett i gang Web Portal

Kari Lie Dorer
St. Olaf College, Minnesota, United States of America

The *Sett i gang* web portal is a project that began as a collaborative, student-faculty project in 2014 and is currently used by approximately 15 universities in North America by approximately 300 students. The portal was created based on an understanding of the scholarship of teaching and learning within beginning Norwegian language instruction and also within an online learning environment. It is an extension of two of my earlier projects, the theoretical findings from my dissertation (Lie, 2008) and the first edition of *Sett i gang* (Aarsvold & Lie), a text geared towards North Americans I co-authored and taught from for 10+ years.

Before the portal's conception, students used a print-only curriculum entitled *Sett i gang* (a print workbook and print glossary to supplement a print textbook) for beginning language learning. Now, the second edition of *Sett i gang* utilizes technology to motivate and stimulate language learning in new and meaningful ways by utilizing a print textbook together with an online web portal. This web portal houses thousands of language learning resources together in one location for first year Norwegian language learners. It's expansive as well-- housing 800+ webpages, 500+ interactive activities, 500+ flashcards and vocabulary games, 500+ audio clips, additional resources for instructors and many links to authentic materials online.

The portal was built from an understanding of how theory and practice meet in the interdisciplinary fields of Applied Linguistics, Foreign Language Learning, Educational Technology, and Online Learning. This presentation will focus on 10 specific research findings from the above-mentioned fields, which shed light on how students experience an online learning environment differently from a face-to-face environment.

Additionally, this talk will examine how specific research findings have helped to create a platform that can provide learners with the learning experience they need to be successful.

These findings and references to studies include: immediate feedback (Northrup, 2002; Brown, 1996; Lie 2008; Csíkszentmihályi, 1990); proximal goals & mastery experiences (Bandura, 1986); advance organizers (Ausubel, 1968; Chen & Hiumi's, 2009); authentic texts (Harmer, 1991; Lee, 1995); authentic tasks (Reeves, Herrington, Oliver & Woo (2004), life-long learning Bilash, Gregoret & Loewen 1999); raising metalinguistic awareness (Roth, Speece, Cooper, & de la Pazas, 1996; Sorace, 1985; Alderson, Clapham & Steel, D., 1996); reducing for-

eign language anxiety (Horwitz, Horwitz & Cope, 1986; Horwitz & Young, 1991; Crookall & Oxford, 1991; Krashen, 1985); and reducing technological anxiety (Saadé & Kira, 2009).

I will also discuss how this project is one small portion of a four-year $700,000 Andrew Mellon Foundation grant aimed at exploring and developing the digital humanities at St. Olaf College. One unique piece of this project is the emphasis given to faculty-student collaboration; it simultaneously funds faculty to explore new ways of teaching and new lines of inquiry for research while also enables students to learn digital research methodologies relevant to careers in the humanities and humanistic social sciences.

I will conclude with the preliminary results of an intensive research project conducted on student use and perception of the portal which seeks to complete the theory to practice and back to theory cycle, again a student-faculty research collaboration.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: applied linguistics, foreign language learning

Automated Improvement of Search in Low Quality OCR Using Word2Vec

Thomas Egense
Statsbiblioteket, Denmark

In the Danish Newspaper Archive[1] you can search and view 26 million newspaper pages. The search engine[2] uses OCR (optical character recognition) from scanned pages but often the software converting the scanned images to text makes reading errors. As a result the search engine will miss matching words due to OCR error. Since many of our newspapers are old and the scans/microfilms is also low quality, the resulting OCR constitutes a substantial problem. In addition, the OCR converter performs poorly with old font types such as fraktur.

One way to find OCR errors is by using the unsupervised Word2Vec[3] learning algorithm. This algorithm identifies words that appear in similar contexts. For a corpus with perfect spelling the algorithm will detect similar words synonyms, conjugations, declensions etc. In the case of a corpus with OCR errors the Word2Vec algorithm will find the misspellings of a given word either from bad OCR or in some cases journalists. A given word appears in similar contexts despite its misspellings and is identified by its context. For this to work the Word2Vec algorithm requires a huge corpus and for the newspapers we had 140GB of raw text.

Given the words returned by Word2Vec we use a Danish dictionary to remove the same word in different grammatical forms. The remaining words are filtered by a similarity measure using an extended version of Levenshtein distance taking the length of the word and an idempotent normalization taking frequent one and two character OCR errors into account.

Example: Let's say you use the Word2Vec to find words for banana and it returns: *hanana, bananas, apple, orange. Remove bananas* using the (English) dictionary since this is not an OCR error. For the three remaining words only hanana is close to banana and it is thus the only misspelling of banana found in this example. The Word2Vec algorithm does not know how a words is spelled/misspelled, it only uses the semantic and syntactic context.

This method is not an automatic OCR error corrector and cannot output the corrected OCR. But when searching it will appear as if you are searching in an OCR corrected text corpus. Single word searches on the full corpus give an increase from 3 % to 20 % in the number of results returned. Preliminary tests on the full corpus shows only relative few false positives among the additional results returned, thus increasing recall substantionally without a decline in precision.

The advantage of this approach is a quick win with minimum impact on a search engine [2] based on low quality OCR. The

algorithm generates a text file with synonyms that can be used by the search engine. Not only single words but also phrase search with highlighting works out of the box. An OCR correction demo[4] using Word2Vec on the Danish newspaper corpus is available on the Labs[5] pages of The State And University Library, Denmark.

Topics: Nordic Textual Resources and Practices
Keywords: text, corpora, NLP, OCR

References

[1] Mediestream, The Danish digitized newspaper archive. http://www2.statsbiblioteket.dk/mediestream/avis
[2] SOLR or Elasticsearch etc.
[3] Mikolov et al., Efficient Estimation of Word Representations in Vector Space https://arxiv.org/abs/1301.3781
[4] OCR error detection demo (change word parameter in URL) http://labs.statsbiblioteket.dk/dsc/ocr_fixer.jsp?word=statsminister
[5] Labs for State And University Library, Denmark http://www.statsbiblioteket.dk/sblabs/

Reading Moravian Lives: Overcoming Challenges in Transcribing and Digitizing Archival Memoirs

Katherine Faull
Bucknell University, United States of America
Trausti Dagsson
University of Gothenburg, Sweden
Michael McGuire
Bucknell University, United States of America

The Moravian Lives project aims to digitize, transcribe, and publish for analysis more than 60,000 manuscript and print memoirs, written by members of the Moravian Church between 1750-2012. These memoirs are housed in archives throughout the world, making it difficult for scholars to engage with them as an entire corpus. Furthermore, of the 18th-century memoirs, over 90 % are in manuscript form. As project collaborators establish the foundations of a massive digital archive that houses facsimiles of the memoirs, we wrestle with how best to publish the memoirs in machine-readable format: existing optical character recognition (OCR) software does not reliably manage 18th century German script; in addition, the volume of pages to be transcribed challenges traditional transcription capabilities. Research teams at Bucknell and the University of Gothenburg in Sweden are collaborating to develop a suite of tools that will support large-scale controlled crowdsourcing of transcription and exportation of text and data sets to support a wide range of research needs by scholars in fields ranging from autobiography to theology, religious history, social history, historical and computational linguistics, and gender studies. In this paper, Katie Faull and Trausti Dagsson will discuss the challenges we face as we establish best practice for developing an interactive platform for editing and accessing this critically significant collection.

Topics: Nordic Textual Resources and Practices
Keywords: transcription, digital history, autobiography, metadata, Moravian

Bibliography

"Doing DH in the Classroom: Transforming the Humanities Curriculum through Digital Engagement" (with Diane Jakacki) Doing Digital Humanities: Practice, Training and Research. Richard J. Lane, Raymond Siemens, and Constance Crompton, eds. Abington, UK: Routledge. Forthcoming.
"Reifying the Maker as Humanist" (with Diane Jakacki and John Hunter). Making Humanities Matter, Jentery Sayers, ed. Minneapolis, MN: U. of Minnesota Press. Forthcoming.
Faull, Katherine (with Diane Jakacki). "Digital Learning in an Undergraduate

Context: Promoting Long Term Student-Faculty Collaboration." Digital Scholarship in the Humanities. DOI: http://dx.doi.org/10.1093
"Anna Nitschmann" Pietismus Handbuch, ed. Wolfgang Breul, Mohr Siebeck Verlag. Forthcoming.

Senses and Emotion of Early-modern and Modern Handicrafts – Digital History Approach

Johanna Ilmakunnas
University of Turku, Finland

The proposed paper explores handicrafts (embroidery, plain sewing, shellwork, papercuts, silhouettes, woodturning etc.) in Europe, c. 1700–1850 and sensory and emotional practices linked to them. The paper discusses how manual work can be found from the wealth of sources, both digital and non-digital, both textual, visual and material. The paper aims also to explore what possibilities and restrictions historians may encounter while using digitized museum collections as source material. The paper will discuss the possibilities of exploring before relatively closed museum collections of objects and potentiality for novel approaches digitized collections offer for history research. It also discusses the opportunities text and image recognition brings to a subject that has been little researched despite important recent work on handiwork and research projects digitizing sources (e.g. project 'Lady's Magazine: Understanding the Emergence of a Genre' at the University of Kent). Furthermore, restrictions such as insufficient information on images, inadequate meta data or strict copyright regulations will be discussed.

The paper presents a new research project that explores handiwork done by European elites. It is part of a larger project on work and profession of early-modern European elites, lead by prof. Johanna Ilmakunnas. Within the project, citizen science and crowdsourcing will be used especially when collecting visual and material interpretations of early-modern and modern handiwork. Furthermore, the project will apply text recognition tools (Transkribus) developed within the EU H2020 project 'READ – Recognition and Enrichment of Archival Documents'.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: material culture, elites, 18th century, 19th century, Europe

Bibliography

Johanna Ilmakunnas & Jon Stobart (eds), A Taste for Luxury in Early Modern Europe: Display, Acquisition and Boundaries. Bloomsbury 2017.
Johanna Ilmakunnas, Marjatta Rahikainen & Kirsi Vainio-Korhonen (eds), Early Professional Women in Northern Europe, c. 1650–1850. Routledge 2017.
Johanna Ilmakunnas, 'Embroidering Women & Turning Men: Handiwork, Gender and Emotions in Sweden and Finland, c. 1720–1820', Scandinavian Journal of History, Special Issue on Gender, Material Culture and Emotions in Scandinavian History. 41:3 (2016), pp. 306–331. DOI:10.1080/03468755.2016.1179831.
Johanna Ilmakunnas, Joutilaat ja ahkerat: Kirjoituksia 1700-luvun Euroopasta. Siltala Publishing 2016, 272 p. [Idle and industrious: Writings from eighteenth-century Europe]

Reading Through the Machines: Epistemology, Media Archeology and the Digital Humanities

Jonas Ingvarsson
University of Skövde, Sweden

In this presentation I approach the more abstract relations between art and digital culture, the dimension of »digital epistemology», where »the digital» is regarded not as a

set of technologies, structures or gadgets, but rather as *a lens* (Lindhé 2013; O'Gorman 2006), through which we focus on culture, history and our own contemporary times.

Initially, this has meant to relate literary texts to digital culture and history even though these texts not explicitly mentions digital culture, computers or networks, or are published on a digital platform. By performing these readings, I have also found it productive to relate the *forms* of the digital to early modern aesthetic genres, many of which – for example the emblem (Daly 1979; Agrell 1994; Manning 2002) and the cabinets of curiosities (Bredekamp 1995) – were regarded not only as genres but as *modes of thought*. The entrance into digital culture, and digital aesthetics therefore also becomes a historical tool. Moreover, the connections between our own digital age and early modern modes of thought could foster a new understanding of our own technological times.

This short presentation will introduce some of the critical perspectives I have probed in an ongoing research project. While discussing these perspectives, I use »mode of thought» as an epistemological concept, and »lens» as the driving metaphor. As a background, though, I should mention professor Alan Liu's short paper on the notion of »the epistemology of the digital» (Liu, 2014). Liu identifies a few important fields where digital environments could or should influence the academic curriculum in general and the Humanities in particular. The point of Liu's text is that digital knowledge is not a concern only related to digital objects and electronic culture, big data, the digitalization of the cultural heritage and new positivist trends in its wake – no, digital knowledge should announce an epistemic shift for the academic practice as such. The aim of Liu's »provocation more than a prescription» is to challenge the basic structures of knowledge distribution and production within the academic field.

In this presentation, I will narrow down these challenges to a few more concrete aspects of how digital epistemology can inform the analyses of literature and cultural artifacts. Digital epistemology in this mode functions as a multifocal lens by which we zoom in and explore the digital not only as technology, object or network, but as a critical concept and historical facticity in the reflection upon our cultural environments. In this presentation, then, I intend to propose a few intersecting – and heuristic – approaches to digital epistemology:

1. *Relating literary texts and artworks to digital culture and digital history. That is*: What does it mean to relate cultural artifacts to the communicational and organizational logic that has been put forward – in different ways – by digital technology since the 1950's?

2. *Reading analogue literature and art as if they were electronic texts. That is*: What happens if we analyze for example a print novel in terms of embodiment, processes, performativity, materiality and even »software», or other »buzz concepts» in the analytic tradition of electronic texts and digital culture? Will this encourage a focus not on what an artwork *mean* but what it *does*?

3. Juxtaposing expressions of digital culture with early modern modes of thought. That is: How does social media as Instagram, Facebook and Twitter relate to the Salon Culture? How does computer games and web pages relate to the aesthetics of the Renaissance emblem? How does the result of Internet search engines relate to the Cabinets of Curiosities, or to the archival »principle of pertinence» (sort by subject rather than provenance)?

These lines of digital epistemology do have one thing in common: the digital is seen as a *mode of thought,* rather than as a set of gadgets, machines or electronic networks. The concept of digital epistemology suggests that the humanities curriculum should be revised, since «the digital» – understood as a perspective, or a set of *lenses* – shifts our focus in the treatment of contemporary culture as well as of historical topics and aesthetics.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: digital epistemology, media archaeology, hypertext theory, game philosophy

Organizational and Educational Issues in Representing History through a Series of Data Sprints on Visual Data from an API

Lars Kjær
Ditte Laursen
Stig Svenningsen
Mette Kia Krabbe Meyer
The National Library, Denmark

While archives and libraries have made digital data available in dissemination platforms for decades, with access to one single object at a time, they have little – but a growing – experience in making data available as datasets through API's and making them available in user friendly ways. Correspondingly, students and researchers in the field of humanities have little – but a growing – understanding of using digital data from API's. In this presentation, we will present the results of a university and library collaboration on making available and bring into play data through an API, in a series of data sprints. We base our presentation on interviews with participants, on analysis of the products that they made doing the data sprints, and on our own experiences as organizers and data providers.

A data sprint is in our definition an intensive period where a group of people work with selected data by collecting, refining, analyzing and visualizing it to solve a problem, to create insights, and to learn about a topic. About 50 students and researchers from Copenhagen University and Copenhagen IT University joined the exploration of the material in three data sprints during autumn 2016 (http://kub.kb.dk/humlab/datasprint). The participants had very different skills within humanities and IT. For instance, some were experts on the subject colonial history, others had technical programming skills, and others just had an interest in combining and learning about using digital data in new ways. In turn, we as organizers brought a variety of competences, ie. in the archival material, in visual culture, in spatial humanities, in technology and in running data sprints.

The data involved were maps, images and metadata related to the former Danish colonies. 2017 is the centenary of the sale of the Danish West Indies and the material raises a range of possible questions like: What is drawn, surveyed and photographed? What is the origin and the context of the material and how did it found its way to the collections of the library? How do we communicate them in today's postcolonial society?

While we are still running the events and processing interviews and experiences when writing this abstract, preliminary results suggest that a data sprint is a suitable format for creating an interdisciplinary and cross-material framework for releasing the potential of digital humanities in relation to digitized cultural heritage in archives and national libraries. However, there is also a need for improving access as well as interoperability to the digital data held by the library and other cultural institutions. Moreover, there is a need for setting up boot camps/workshops prior to data sprint events to strengthen digital skills among the participants, such as Tableau, OpenRefine, Python and Geographic Information Systems (GIS).

On a broader canvas, this study provides empirical evidence of organizational barriers and possibilities for archives and libraries of making digital data available in new ways, as well as support for recent discussions on educational issues in balancing a strong theoretical and methodological grounding in humanities with an understanding of technology.

Topics: Nordic Textual Resources and Practices
Keywords: data sprints, API, visual data, open data

Bibliography

http://kub.kb.dk/humlab/datasprint

The Afterlife of Early Modern Portraiture in Digitized Museum Collections: Discovering Conventions and Forgotten Images

Charlotta Krispinsson
Stockholm University, Sweden

The aim of this paper is to discuss how digitized museum collections of early modern portraiture added analytical possibilities to my recently finished PhD project (Historiska porträtt som kunskapskälla: Samlingar, arkiv och konsthistorieskrivning, Nordic Academic Press, 2016).

A methodological point of departure for the project was to treat early modern portraiture as a material as well as mental category of images. My interest was to study the modern reception history of this category, ca 1880-1945. For this reason, the investigation started with a need to take stock of the characteristics of early modern painted portraits. The previous research on early modern portraiture is vast, but is often characterised by an aesthetic and art theoretical focus on singular works that do not reflect the historical artistic production of portraits in different medias as whole.

The Swedish national portrait collection (part of the collection of Nationalmuseum) consists of ca 3 000 objects. It is, together with the collections of the National Portrait Gallery in London, one of the largest collections of portraiture in Europe. Together they comprise a large quantity of portraits, selected mainly according to the name of the depicted subject (and not according to the artistic merits of the portrait painter). Browsing these digitized museum collections of early modern portraiture thus made it possible to better detect visual conventions characteristic of the historical production as whole.

Early modern portraiture and contemporary selfies are both images of individuals where identity reflected through stereotypical expressions of self is key. To continue this comparison, digitized national portrait collections provides a centralized data mass of images, similar to the data set of 3 200 selfies provided by one of most well-known projects in digital humanities today, SelfieCity, coordinated by media theorist Lev Manovich.

In my presentation, I would like to compare methods and outcomes between my project and the methodological foundation of SelfieCity, and also expand upon how digitized museum collections could provide new opportunities to art historical research. Scanning through the afterlife of early modern portraits in digitized national portrait collections provided different kind of historical insights than close readings of a few, select portraits could. It showed how the typical kind of early modern portraiture put on display in art museums today (chosen for originality, artistic quality, or the works position in the history of art) need to be regarded as rare exceptions in the total production of portraits, just as the iconic selfie is a rare exception to the big data of quickly forgotten, digital images.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: art history, SelfieCity, early modern portraiture

Bibliography

Historiska porträtt som kunskapskälla: Samlingar, arkiv, konsthistorieskrivning, diss. Stockholm, Nordic Academic Press, Lund 2016.

”Collecting Faces: Art History and the Epistemology of Portraiture. The Case of the Swedish Portrait Archive”, Sensorium Journal, no. 1, 2016.

”Collection BIOMUS / Museum Fantasies”, How to gather? Acting in a Center in a City in the Heart of the Island of Eurasia, utst. katalog, Moscow Biennale Art Foundation, Moska 2015.

”Aby Warburg’s Legacy and the Concept of Image Vehicles. ”Bilderfahrzeuge”: On the Migration of Images, Forms and Ideas. London 13-14 March 2015”, Konsthistorisk tidskrift, nr. 4, vol. 84, 2015.

”The Challenge of the Object. CIHA:s (Congrès International d'Histoire de l'Art) 33:e internationella kongress för konstvetare. Nürnberg 15–20 juli 2012.”, Konsthistorisk tidskrift, vol: 81, 2012:3.
”Catharina Nolin: En svensk lustgårdskonst - Lars Israel Wahlman som trädgårdsarkitekt, Stockholm 2008”, Konsthistorisk tidskrift, vol. 80, 2011:1.
”Lars Nilsson och svensk postmodernism före 1987”, Valör 2011:1.

Málið.is: An Icelandic Web Portal for Dissemination of Information on Language and Usage

Ari Páll Kristinsson
Halldóra Jónsdóttir
The Árni Magnússon Institute for Icelandic Studies, Iceland

A new web portal on the Icelandic language, and language use, was opened in Iceland in November 2016. The users of málið.is only need this single web address in order to access abundant reliable and authoritative information, guidance, help and advice on the Icelandic language, its use and nuances, historically and contemporarily. This concerns e.g. orthographical matters, grammatical issues such as inflections, grammatical agreement, and a variety of other questions of syntax, word formation, semantics, the lexicon, the history and etymologies of particular lexical entities, phraseology, synonymity, terminologies and translations of technical vocabulary, and many questions of language and usage.

Previously, these resources were accessible via a variety of different formats, user interfaces, web addresses, search methods and functions, which caused problems for many users as they were typically not aware of all possibilities.

Among the principal target groups of málið.is are students, and writers of non-fiction, while the web portal is nonetheless designed to serve the Icelandic speaking public in general. Málið.is strives for plain and non-technical exposition and conciseness, whenever possible.

A major challenge in the process of creating and launching málið.is was the different nature and content of the various language resources on the one hand, and the different motivations and expectations of individual users on the other hand. Some of the data are explicitly of a prescriptive nature, while others have primarily descriptive function. As we do not expect users (unless those who have linguistic training) to be immediately familiar with this fundamental distinction, we realized that this could perhaps lead to misinterpretation of the data presented. However, since málið.is facilitates the comparison between the two data types, our conclusion is that users will be able to acknowledge the distinction. Indeed, one theoretical contribution of málið.is is that it highlights the difference between descriptive and prescriptive language resources, for the benefit of students and researchers.

The name of the web portal is its web address: málið.is, which translates as 'the language.is'. The functions of this portal are in many ways similar to the Danish web portal sproget.dk. Indeed, the Danish portal, initiated and operated by our colleagues at the Danish Language Council and Society for Danish Language and Literature, served as a model and an inspiration as we were planning this Icelandic web portal, at The Árni Magnússon Institute for Icelandic Studies. Thus, málið.is is an example of fruitful Nordic cooperation in the field of digital humanities. The two portals differ in some details, e.g. in that málið.is primarily focusses on its source data bases and search results, while the sproget.dk main site also offers the user a variety of links, games, suggestions etc. The team behind the planning of málið.is is not convinced that it is feasible to add much material of this type on the website. Another difference worth mentioning is that while sproget.dk e.g. explicitly comments that the spelling of the ODS is not necessarily in harmony with

modern spelling rules, málið.is leaves the interpretation of the data to the user.

Topics: Nordic Textual Resources and Practices
Keywords: web portal, Icelandic, language resources, Nordic cooperation, dissemination of knowledge

Bibliography (selected)

Kristinsson, Ari Páll. 2016. Language in public administration in present-day Iceland: some challenges for majority language management. In: Language use in public administration. Theory and practice in the European states. Pirkko Nuolijärvi & Gerhard Stickel eds. European Federation of National Institutions for Language. Budapest: Research Institute for Linguistics, Hungarian Academy of Sciences. Pp. 83-92.

Kristinsson, Ari Páll. 2016. English Language as 'Fatal Gadget' in Iceland. In: Why English? Confronting the Hydra. Pauline Bunce, Robert Phillipson, Vaughan Rapatahana & Ruanni Tupas eds. Bristol: Multilingual Matters. Pp. 118–128.

Kristinsson, Ari Páll. 2016. Om følgerne af leksikalsk purisme i Island. [On the consequences of lexical purism in Iceland] Dansk Noter 1/2016:40–44.

Kristinsson, Ari Páll. 2016. Editor of Orð og tunga 18. [Orð og tunga is a peer-reviewed journal on language and linguistics, published annually by the Árni Magnússon Institute for Icelandic Studies.]

[Re]use of Medieval Paintings in the Network Society: A Study of Ethics

Pakhee Kumar
IMT School of Advanced Studies Lucca, Italy

The internet society is a "network society" (Castells, 2014) characterised by quickness of information. It is an also an "era of blurring boundaries between interpersonal and mass, professional and amateur, bottom-up and top-down communications" (Shifman, 2014). In this society, the cultural participants are not interested in being passive consumers (Kolb, 2005) of culture, rather they recreate culture by reusing, remixing, and recirculating it. One of the intrinsic character of the internet society is using images to convey sentiments, often ranging from cynicism to humorous, often on contemporary issues, referred as meme (refer Fig 1). In fact, this simplified, clear and concise way of expressing complex sentiments is an essential and indispensable part of the contemporary digital culture.

Figure 1. *Memes created by Medieval Reactions using painitings.*

The word meme was introduced by Dawkins (1989, s. 92) to explain a concept of culture. He noted that meme is the new replicator, a noun that conveys the idea of a unit of cultural transmission, or a unit of imitation. Further, Oxford dictionary defines meme as "a humorous image, video, piece of text, etc., that is copied (often with slight variations) and spread rapidly by Internet users" (meme, nd) to spread particular idea (Colin & Knobel, 2007).

The creation of meme does not require any particular artistic skills, only connection to the internet and hence, can be created, circulated and consumed by anyone. This reflects the freedom to participate envisioned by Berners-Lee (2010, s. 82) that "people must be able to put anything on the Web, no matter what computer they have, software they use, or human language they speak and regardless of they have a wired or wireless connection". In this process, not only the original context of the paintings/artwork is lost but also the meaning of the painting is altered. However, this does not diminish the popularity of such images. Shifman (2014) raised a few question regarding this issue: how did such bizarre piece of culture become so successful? Why are so many people investing so much effort inventing it? Why do some of these amateur imitations attract millions of viewers?

One of the possible reason for their popularity may be that it is minimalistic, therefore, an easy way of catching attention. Moreover, the relationship to contemporary issues further adds to its popularity. Lastly, the attempts to humor-ise even the immoral situations may also be the reason of its popularity. Indeed, every age looks at the past in a different way. However, the question is whether this creation and consumption degenerates the original content or enhances it by utilising amateur and untrained yet skilful people.

This paper will examine the reuse of medieval paintings in the internet age. It will examine various typologies of reuse to represent the contemporary sentiments. Lastly, the paper will examine ethics related to circulation of such images in the internet.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: digital culture, meme, ethics

Digitization of Literary Fiction. Example of Jan Potocki's The Manuscript Found in Saragossa

Rafał Kur
Jagiellonian University, Poland

The Manuscript Found in Saragossa (original title Manuscrit trouvé à Saragosse) was written by Jan Potocki in the years 1979-1805. The work consists of several plots making up different stories. The thick web of connections between places, plots and protagonists in The Manuscript Found in Saragossa while partly following the story within a story formula, goes beyond it. However, it more resembles a tangle or a maze. While it is in fact one story, it is told in several dozen ways and it is filled with quotes and repetitions. This kind of composition, recorded on paper, in which one story includes another one, while within the second one emerges a third, still obscures from the reader the web of internal connections between the narrators, characters, events and places.

That is why a Krakow literary community with the help of IT specialists and graphic designers created a reinterpretation of the work adding a visual layer. The completed work was made available on a website. Owing to the project, the book may be read anew, discovering even deeper the talent and the imagination of Potocki, overwhelmed by the sheer number of pages in the traditional printed form.

Digital text is a web and a database, a space in which distant elements are only one click away from each other. Each of the 66 days-chapters was given a plaque, owing to which we will not get lost in the labyrinth of plots and characters, and we will be able to follow individual plots in a free order without the fear of missing a part of the story. The only required tool is a web

browser. A clear and simple graphical interface leads the reader wherever one wishes. While the visual setting creates a unique atmosphere. The Manuscript Found in Saragossa highlights the vividness of the form of the story, but the creators used additionally the iconography of the film adaptation of the novel (The Saragossa Manuscript, 1965, directed by Wojciech Has), that is the gothic, picturesque, grotesque and vintage elements.

I chose the example of the work of Potocki, since it is an interesting, fresh electronic adaptation of a novel and it fits perfectly the literary and digital game of associations. An equally good material could be the stories "The Garden of Forking Paths" of Borges (El jardín de los senderos que se bifurcan, 1941), "Hopscotch" by Cortazar (Rayuela, 1963), or "Life a User's Manual" by Perec (La vie mode d'emploi, 1978).

Usually this kind of novels finishes with the syndrome of tiredness with ever new, budding stories and the confusion of names of characters and names of the novel's locations. This type of books is not easy for an average reader. That is why, while presenting the example of the prose of count Potocki in a new digital setting, I would also like to show the method of refreshing of literary texts. The visualisation and the interface, while being basic tools for this type of novel, are becoming increasingly widespread.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: digitization, visualisation of literary narrative, internet, eighteenth-century literature

Multidisciplinary Terminology Work in the Humanities: New Form of Collaborative Writing

Tiina Mirjami Käkelä-Puumala
University of Turku, Finland

In my paper, I'll present a multidisciplinary terminology project that started in February 2015 in the The Bank of Finnish Terminology in Arts and Sciences (BTA). The BTA was founded in 2011 as a permanent open access termbase for all fields of research in Finland. One of the main goals was to create a collaborative environment for experts from different fields. Term entries in the BTA are written by experts, but the termbase uses a Semantic MediaWiki platform, which offers all registered users the possibility to participate in the discussion about scientific terms. The Bank of Finnish Terminology has been funded by the University of Helsinki and the Academy of Finland.

In the beginning of 2015 I started with a couple of researchers a special working group within the BTA that focused on terms that were used across different humanities disciplines (terms like representation, sign, performance, discourse, affect, realism, critique, text, code, expression, etc.). The group consisted of experts from aesthetics, linguistics, literary studies, philosophy, semiotics and theatre studies. Our goal was to write collaboratively definitions and descriptions for multidisciplinary terms in the humanities. Our work represents a new form of collaborative writing made possible by digitalization. Firstly, it exceeds disciplinary boundaries that have usually been very strong in terminology work and provides via hyperlinks much more information of the multidisciplinary use of scientific terms. Secondly, the collaborative writing process and multidisciplinary approach creates a new way of depicting and understanding conceptual history, which is essential for both research and higher education (not to mention the general public). Thirdly, the collaborative work represents a form of academic communication that is ongoing, self-correcting, and not confined to the conditions of predigital academic publishing.

Links
http://tieteentermipankki.fi/wiki/Termipankki:Etusivu/en (in English)
http://tieteentermipankki.fi/wiki/Termipankki:Etusivu (in Finnish)
http://tieteentermipankki.fi/wiki/Monitieteinen_termity%C3%B6 (in Finnish)

Topics: Nordic Textual Resources and Practices
Keywords: humanities, terminology, multi-disciplinary

Bibliography

"Interdisciplinary Terminological Work, Family Resemblance and Interdisciplinary Concept Analysis." Markku Roinila & Tiina Käkelä-Puumala (Presentation at conference Crossing Borders 2015). https://www.academia.edu/20941996/Interdisciplinary_Terminological_Work_Family_Resemblance_and_Interdisciplinary_Concept_Analysis

(forthcoming 2017)"This Land Is My Land, This Land Also Is My Land": Real Estate Narratives in Pynchon's Fiction" Textual Practice 1:2017

"Postmodern ghosts and the politics of invisible life." Death in Literature. Sari Kivistö and Outi Hakola (eds.). Cambrige Scholars Publishing, 2014.

Towards a Reader-friendly Digital Scholarly Edition

Sebastian Köhler
Society of Swedish Literature in Finland

Looking at digital scholarly editions today you will usually find that "digital" implies "to be used in a desktop environment". Digital scholarly editions are seldom adapted, let alone optimized for small screen devices like smartphones and tablets, and often have features superfluous to users who just want to read the "plain" text. Consequently, digital scholarly editions are generally not expressly reader-friendly in a non-scholarly sense and fail to meet the needs of a wider public.

In this paper I will present the Digital Edition 2 platform of the Society of Swedish Literature in Finland, with special consideration of its lightweight user interface, targeted primarily at read- rather than research-oriented users, as well as smartphone and tablet use.

The platform has been in development since October 2016 and utilises strictly open source software in order to facilitate long term maintenance. It is implemented on a mobile first approach as a progressive web app built on the AngularJS 2 and Ionic 2 frameworks. A RESTful API handles communication between the backend and the user interface. The platform is initially built to host two digital critical editions, Zacharias Topelius Skrifter and Henry Parlands Skrifter; however, it is intended as a generic platform able to accommodate future scholarly editions of other types as well. The live demonstration of the platform will showcase material from Zacharias Topelius Skrifter.

The responsive user interface of the platform enables presenting the digital edition in two modes depending on media and user choice: basic and advanced mode. The basic mode revolves around the reading text and a minimal set of paratextual materials. The idea is to display the text in a more accessible form to primarily support reading, rather than studying. Thus, for instance, annotations are available, but variants, facsimiles and transcribed manuscripts are not. These appear only in the advanced mode. The basic mode is essentially a stripped down version of the full digital edition, featuring a limited critical apparatus with some, but not all aspects of the history and transmission of the text, as well as limited scholarly paratexts, search options and tools.

This restricted scope combined with the fact that the basic mode is first and foremost intended for access on mobile devices, though also accessible on PCs and laptops, should provide a reading experience suitable for a wider audience than that of the traditional scholarly edition. This includes, among others, students, teachers, non-editor scholars and people passionate about literature in general.

For research-oriented users the advanced mode with the complete set of features will be available in desktop environments.

The critical and annotated edition of the writings of Zacharias Topelius currently comprises six volumes, the first published in 2010. Thus far the digital edition contains

the equivalent of about 4,000 pages of text by Topelius, 400 pages of introductions by editors and 10,000 annotations. It is freely accessible at topelius.fi.

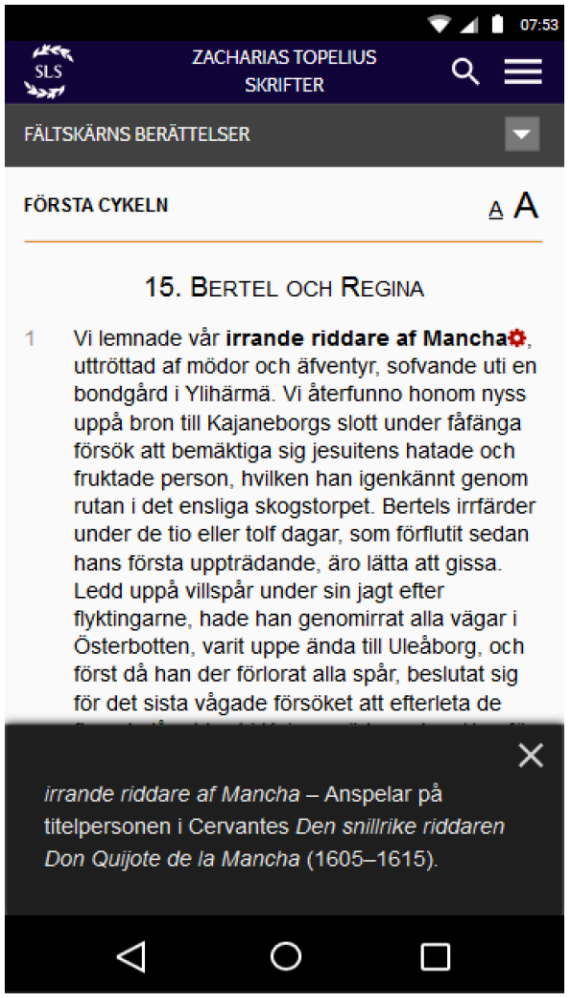

Figure 1. *A mock-up of the basic mode on a smartphone, displaying an annotation to the novel Fältskärns berättelser by Topelius.*

Topics: Nordic Textual Resources and Practices; The Digital, the Humanities, and the Philosophies of Technology
Keywords: digital scholarly edition, progressive web app, reader-friendly

Bibliography

Köhler, Sebastian, Boel Hackman & Carola Herberts, Kommentar till Edith Södergrans Dikter och aforismer. Varia, SSLS 563:3, Helsingfors: Svenska litteratursällskapet i Finland 2016

Köhler, Sebastian, "'Det gjelder å feste blikket'. Kampen mot nihilismen i Karl Ove Knausgårds Min kamp", Norsk litterær årbok 2014, red. Heming Gujord & Per Arne Michelsen, Oslo: Det Norske Samlaget 2014, s. 212–226

Towards a Digital Edition of the Codex Regius of the Prose Edda: Philosophy, Method, and Some Innovative Tools

Michael John MacPherson
University of Iceland

The Codex Regius of the Prose Edda (GKS 2367 4to, or R) is the subject of a new project based at the Árni Magnússon Institute in Iceland. One of the aims of this project is to produce a multi-level, fully lemmatized, and morphologically analyzed digital edition of the manuscript in TEI-XML. The purpose of this talk is to address the motivation for such an edition in the context of current textual research on the Prose Edda, and to present some tools developed internally which are intended to make the edition more flexible while reducing the resources required.

Recent publications have increased our understanding of the prehistory of two other main manuscripts of the Prose Edda, Codex Wormianus and Codex Upsaliensis (Johansson 1997 and Mårtensson 2013), and one of the proposed outcomes of the project is to perform an analogous study of R. The proposed method, modeled on these earlier publications, involves an investigation into the palaeographic, graphemic, and orthographic norm upheld by its main scribe. Deviations from this norm can sometimes be explained as influence from the scribe's exemplar, allowing us to reconstruct the prehistory of the manuscript. In an attempt to move towards a more quantitative and

reproducible approach, this investigation will leverage the digital edition as a source of information about the scribe's norm. A close transcription policy was developed to account for significant variation at the palaeographic level, with the main effort of transcription dedicated to capturing this variation.

This type of close transcription is often time-consuming. A novel tool was developed leveraging open source grapheme-to-phoneme (G2P) software. G2P is commonly applied in text-to-speech problems, where computers need to guess the best pronunciation of a word based on its orthography. Instead, models of the scribe's practice were trained using a sample of the close transcription and existing normalized sources. Close transcriptions were then generated for the remaining unseen text using the existing normalized sources. This method generates entirely correct words 70 % of the time, with most of the wrong words being only off by one or two letters. The generated text was then incorporated into the workflow of the project's transcribers.

This is then followed up with a grapho-phonetic markup based on Alex Speed Kjeldsen's work on Icelandic Original Charters Online (forthcoming) and in his doctoral thesis (2010). This involves mapping each character in the transcription with a corresponding theoretical etymological phonetic value based on our understanding of Old Norse language history. This allows for the exploration of grapho-phonetic relationships first implemented with success by Weinstock in 1967. Further tools are then developed for the automatic generation of transcriptions according to multiple diplomatization and normalization schemes. It also allows for modernization, granting access to the natural language processing toolkit IceNLP for lemmatization and morphological analysis.

The result is a highly flexible digital edition designed from the ground up to describe the habits of its main scribe, allowing for quantitative queries of philological criteria which can be easily reproduced by future researchers.

Topics: Nordic Textual Resources and Practices, The Digital, the Humanities, and the Philosophies of Technology
Keywords: digital philology, prose Edda, linguistics, TEI, grapheme-to-phoneme

Bibliography

Johansson, Karl G.. Studier I Codex Wormianus: Skrifttradition och avskriftsverksamhet vid ett isländskt skriptorium under 1300-talet. Göteborg: Novum Grafiska AB, 1997.

Kjeldsen, Alex Speed. Et Mørt håndskrift og dets skrivere: Filologiske studier i kongesagahåndskriftet Morkinskinna. PhD thesis, University of Copenhagen, 2010.

———. Icelandic Original Charters Online. Forthcoming.

Mårtensson, Lasse. Skrivaren och förlagan: Norm och normbrott I Codex Upsaliensis av Snorra Edda. Oslo: Novus AS, 2013.

Weinstock, John Martin. A Graphemic-Phonemic Study of the Icelandic Manuscript AM 677 4to B. PhD thesis, University of Wisconsin, 1967.

Contributing to Nordic Cultural Commons through Hackathons

Sanna-Maria Marttila
Aalto University, Finland

Digitalization has affected nearly all aspects of our society, albeit in different ways. For cultural and memory institutions, it has created enormous potential to expand public access to their (digital) holdings and establish and renew collaborative relationships with their visitors. Along with the digitizing of cultural heritage, new digital tools are also creating novel ways for people to access, appropriate and reinvent culture. Despite these developments, cultural and memory institutions are not providing as much access as they could to their digitized collections (Bellini, et al. 2014), nor are they

creating good conditions for people's creative re-use activities (Terras, 2015). This short paper explores how cultural hackathons can enhance creative re-use of digital holdings of memory and cultural institutions, and contribute to co-designing, building and sustaining of open cultural commons.

Governmental bodies, businesses and cultural institutions alike are hosting hackathons to stimulate innovation with their digital offerings and resources. This emerging approach has become an effective and favourite way to encourage exploration and creativity with digital technologies (Briscoe and Mulligan 2014). A hackathon is often described as a problem-solving event through intensive software programming and development in a short period of time (Topi and Tucker 2014). It can also refer to a competition where participants can pitch and develop their ideas and prototypes together with others. Often these events draw together software developers and designers from various fields to collaborate either in teams, or working together solving a specific problem, idea or theme. This has also been seen as a challenge of hackathons, as prototypes rarely are developed into finalized products that could generate revenue or monetary business value (Komssi et al. 2014).

The empirical material is based on long-term engagement and action research on designing and organizing cultural hackathons in Finland, Denmark and Sweden in the recent years, and personal reflections on these experiences. Through these case studies on the Hack4FI, Hack4DK and Hack4Heritage hackathons focusing on creative re-use of digital cultural heritage materials, this article explores the application of hackathon as a approach and way to engage people in public matters (such as discussion on intellectual property rights) and in building shared cultural common-pool resources. Through the critical analysis, the author reflects if and how these arranged events can support social and digital innovation, and creation of new services, tools and practices. The paper sheds light on the strategies and tactics organizing and facilitating hackathons, and furthermore discuss how they can contribute to the building and sustaining open cultural commons.

Topics: Nordic Textual Resources and Practices
Keywords: cultural commons, open digital heritage, hackathon

References

Bellini, F., Passani, A., Spagnoli, F., Crombie, D., & Ioannidis, G. (2014). MAXICULTURE: Assessing the Impact of EU Projects in the Digital Cultural Heritage Domain.

Briscoe, G., & Mulligan, C. (2014). Digital Innovation: The Hackathon Phenomenon.

Terras, M. (2015). Opening Access to collections: the making and using of open digitised cultural content, Online Information Review, Vol. 39 Iss: 5, pp.733 – 752.

Topi, H., and Tucker, A. (2014). Computing Handbook, Third Edition: Information Systems and Information Technology. CRC Press.

Young People's Historical Thinking in the Face of Digitized Sources

Åsa Olovsson
Uppsala university, Sweden

The framework for my research is a database within a project called Gender and Work (GaW) at Uppsala university, which presents how men and women made a living during the period 1550-1800. The sources here are mainly court records from different parts of Sweden. Besides gender and work, the database provides a wide base of information enabling different themes of interest for young people. Some examples are sexuality, relations, marriage and children. I believe this kind of modern digital technology may prove to be essential for the development of

history education. With the help from available primary sources and usable methods, the students can achieve a nuanced view on history as such and proper scientific thinking, including methodology. The aim for the project I am planning in the field of digital humanities is to produce concrete tools for history teaching and learning in the upper secondary school in Sweden. The curriculum clearly marks scientific thinking as key. Thus the use of primary sources in the classroom should be a fundamental part of history education.

Design study methodology is suitable for this project, since it provides the opportunity to conduct research in real life classroom situations. Within this method, it is possible to develop concrete tools.These tools could be scaffolding of different kinds - assisting and supporting students in their learning process. For example, I will create complete paths or activities for the mentioned themes. Above that, study handbook is necessary for teachers and students, with instructions on how to read and interpret sources. Further, glossary which explains concepts from the current theme. The scaffolding could be written, but it is also possible to use multimedia and make tutorial videos.

I am especially interested in exploring the students' historical thinking and what meaning they make of history, their history. How will that meaning evolve under exposure of authentic sources from the digital archive? In order to know if this method works, a comparison it is necessary with a control group working in a more traditional way. Surveys or interviews are necessary before and after working with the database working to follow the student's progress. Working with primary sources from the database will make the students active learners instead of passive recipients. They will make them mediate the digitized sources and through them I believe they will become co-creators, see new perspectives and create meaning. Preliminary research questions are How can GaW be used as a learning tool? How does teaching with digitized sources affect students reasoning and learning in history. A related question is whether the students are influenced or not by the sources they work with. Another feature is if the past, present and future become visible through the studied phenomenon. Finally I will find out how the students' expressions may be explained. Hopefully this way of working will activate students narrative about the past and through that also develop their historical thinking.

The search for relevant theories is still in progress. Since I believe that this form of teaching may engage the students' emotional sides, theories concerning emotion and historical learning could be useful.To my knowledge so far, this study will take me to mainly unknown territory. Only a few national studies have had specific focus on what happens when students are exposed to digitized sources. The use of the database GaW in history education has never been studied from a didactic perspective.

Topics: Nordic Textual Resources and Practices
Keywords: primary sources, digital archives, design study, cultural heritage, gender and work

References

Brush, Thomas A and Saye, John W. "A Summary of Research Exploring Hard and Soft Scaffolding for Teachers and Students Using a Multimedia Supported Learning Environment" *http://www.ncolr.org/jiol/issues/pdf/1.2.3.pdf*

Lévesque, S. (2007) Can Computational Technology Improve Students' Historical Thinking? Experience from the Virtual Historian© with Grade 10 Students. " Journal of the Ontario History and Social Science Teachers Association , (Printemps):

Nygren, Thomas, Sandberg Karin & Vikström, Lotta "Digitala primärkällor i historieundervisningen: En utmaning för elevers historiska tänkande och historiska empati" [Digital primary sources in history education: A challenge for students' historical thinking and historical empathy] *Nordidactica*, 2014, 2

Nygren, Thomas och Vikström, Lotta (2013) "Treading Old Paths in New Ways" in *Education Sciences*
Sandberg, Karin (2014) *Möte med det förflutna Digitaliserade primärkällor i historieundervisningen Lic.avh. Umeå: Umeå universitet*
Shavelson, Richard J., Phillips, D. C., Towne, Lisa and Feuer, Michael J., *On the Science of Education Design Studies*

Sixties Biopoetics: A Media Archaeological Reading of Digital Infrastructure

Jesper Olsson
Linköping University, Sweden

Through the rise of planetary scale computing and global digital infrastructure during the last decades (Cf. Bratton 2016, Gabrys 2016, Starosielski 2015, and others), a new ecology of nature and culture, bodies, protocols, and machines has emerged. Including everything from the internet of things, i.e. sensor topographies, smart fridges, and wearables, to underwater cables from the 19th century and distant server halls this process has had radical epistemic, economic, aesthetic, political, and social consequences. Not least, it has challenged and dissolved the charged boundaries between humans and their surroundings, necessitating an analysis that thinks and analyzes 'natureculture' (Haraway 2003) as always intertwined and merged. Accordingly, water, air, minerals, plants, animals, cellphones, optical fiber, pads, pods, and satellite technologies are part of one ecology – one world, many forms, to paraphrase Gilles Deleuze.

However, this tangible transformation did not take place in an instant. It has a long and dwindling material history. In this paper I will try to disentangle some of the strands of this history by returning to artistic and literary practices of the late 1960s and early 1970s. Focusing on some aesthetic experiments and poetic speculations of the period, I hope to shed light on the contemporary digital ecology and its larger cultural implications. Specifically, I will look into a series of essays (under the rubric 'Semicolon') and some text-sound poetic experiments by the Swedish poets and composers Lars-Gunnar Bodin and Bengt Emil Johnson, in which they approach what might be called a 'bio-poetics', bringing bodies and machines in closer contact and trying to re-articulate, even dissolve, the mediating moments in the assemblage artist–artwork/art event–viewer/reader/listener. I will also bring up and discuss the somewhat later 'bio-music' of Manfred Eaton, partly taking its cue from some works by the American composer Alvin Lucier, not least, perhaps, 'Music for Solo Performer' (1965), in which EEG electrodes were attached to the skull of the composer and performer and connected to a sound system, which then generated sound and music.

In these works, a different ecology of natucultures is imagined and explored, partly through an artistic misuse of technologies, which 'prehends' (to use A. N. Whitehead's concept) a contemporary media ecological formation. My aim is, thus, to explore how a seemingly marginal cultural practice, such as avant-garde poetry, can function as a media archaeological probe or platform for analyzing and experiencing some of the other, submerged layers and temporalities of our brand new machine park and the various effects it displays.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: media archaeology, media ecology, infrastructure, natureculture, poetry

Bibliography

Brattton, Benjamin, The Stack, MIT Press 2016
Gabrys, Jennifer, Program Earth: Environmental Sensing Technology and the Making of a Computational Planet, University of Minnesota Presss 2016
Haraway, Donna, The Companion Species Manifesto, University of Chicago Press 2003
Starosielski, Nicole, The Undersea Network, Duke UP 2015

Mapping Letters Across Editions

Vemund Olstad
Directorate for Cultural Heritage, Norway
Hilde Bøe
Munch Museum, Norway

Kartfesting av utgitte brev

I løpet av de seneste 20-30 årene har man i de nordiske landene publisert (eller satt i gang publiseringsarbeid av) en rekke digitale tekstkritiske utgaver av viktige forfattere og kulturpersonligheter. Av de større prosjektene kan nevnes:

* Henrik Ibsens skrifter (Norge)
* eMunch (Norge)
* Ludvig Holbergs skrifter (Danmark/ Norge)
* Grundvigs Værker (Danmark)
* The Linnean Correspondence (Sverige)
* Zacharias Topelius skrifter (Finland)

Edisjonsfilologimiljøet i Norden er relativt lite og oversiktlig, hvilket gjør at man i stor grad har kunnet utveksle erfaringer på tvers av både prosjekt og landegrenser. En konsekvens av dette er at grunnlagsmaterialet for de forskjellige utgavene i veldig stor grad benytter seg av samme kodestandard og format (TEI XML). Måten grunnlagsdataene presenteres ut til brukerne på varierer, naturlig nok, fra prosjekt til prosjekt – og vi ønsker å se nærmere på muligheten til å presentere materiale fra flere nordiske utgaveprosjekt sammen i en kartbasert løsning, ved å sammenstille utvalgte data og metadata fra forskjellige utgaveprosjekt.

I perioden frem til DHN 2017 vil Riksantikvaren, som en av de deltagende etatene i K-labsamarbeidet (http://www.riksantikvaren.no/Veiledning/Data-og-tjenester/K-lab), sammen med Munch-museet arbeide med å georeferere brev fra utgaveprosjektetene eMunch og Henrik Ibsens skrifter for å kunne plassere disse på et interaktivt kart. Tanken bak dette pilotprosjektet er kunne lage til en kartbasert inngang til større nordiske brevsamlinger, der man ved å navigere rundt i kartet skal kunne se hvor brev er skrevet, hvem som skrev de, få opp grunnleggende informasjon om brevet og så kunne gå videre til mer detaljert informasjon på prosjektenes utgavesider.

Til dette arbeidet ønsker vi å benytte oss av kartprogramvaren som ble utviklet i forbindelse med prosjektet Kultur- og naturreise. Kildekode og beskrivelse er tilgjengelig her: http://knreise.no/demonstratorer/. Denne kartløsningen administreres og videreutvikles av K-lab, og den vil bli videreutviklet i 2017 – blant annet med en tidslinjefunksjonalitet, som vil gi oss muligheten til å kunne gjøre et kronologisk utvalg for de stedfestede brevene. En veldig tidlig prototyp er tilgjengelig på http://knreise.github.io/demonstratorer/demonstratorer/historiskeBrev.html.

I tillegg til det praktiske arbeidet med å formidle brev i kart, vil vi forsøke å se litt nærmere på hvordan utgaveprosjekter jobber med å tilrettelegge sine grunnlagsdata for andre brukere. En viktig forutsetning for at vi skal kunne samle utgavedata i en sentralisert løsning er at de er tilgjengelige via en tjeneste det går an å hente data fra. Dette kan være i form av eksterne endepunkt (rest API / sparql endepunkt eksempelvis), eller ved at man laster opp data til en aggregatortjeneste (Norvegiana / K-Samsök / Europeana). Hvor flinke er utgaveprosjekt til å tenke på denne typen etterbruk? Hva kan gjøres for å samordne innsatsen på dette området?

Topics: Nordic Textual Resources and Practices
Keywords: maps, geotags, letters

The Battle of the Text – Quantitative Methodologies in Literary Studies

Julia Pennlert
Umeå University, Sweden

It is often claimed that our digital present time gives the literary scholar possibilities to question and reconfigure what it is to read or analyze a literary text. This statement is part of a larger discourse that emphasize that

our digital time, is as a time of change. Due to the fact that online publication venues have become a vital part of literary culture, or by projects that digitize literary texts by presenting them in online archives such as the Swedish Litteraturbanken (litteraturbanken.se) - the literary text is attached to others in a network of literary publications. The notion of a digital text is often explained as the main reason for why the literary scholar needs to address methodological issues. As a result, the literary scholar is part of the discourse that underlines change, and as a consequence the researcher has to adjust or develop new methods to read, analyze or study a certain text.

During the last decade literary studies is characterized by a methodological turn, especially within in the field of digital humanities. Several theorists have presented 'new' types of reading for example "distant reading" (Moretti), "macro-analysis" (Jockers) or "hyper-reading" (Hayles). These new forms of studying and analyzing texts have been discussed and criticized. These methods present a new optic, or gaze, to study or analyze a certain text. In Literary Studies in the Digital Age (2013), Tanya Clement describes the computer-assisted method as a way for the researcher to get an overview of a vast material by using a "magnifying glass upside down."

However, these methodologies can be compared to historical equivalent discussions that highlight what a literary scholar (should) study and how a reading of a text should be performed. In a Swedish research context technological tools in literary studies is especially discussed during the 1960's and 1970's. During this time several suggestions on how the researcher can adopt their methodologies are presented, in for example Litteraturvetenskap – Nya Mål och Metoder (1966), or Forskningsfält och metoder inom Litteraturvetenskap (1970).

In my paper I will compare the methodological turn within digital humanities with historical examples and argue for why it is important to trace historical similar movements and descriptions in to our present digital time. I will also present an alternative method, a combination of methodologies that can be used as a productive way out of the sometimes polemical discussions on what and how literary studies can be conducted.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: distant reading, quantatitve analysis, statistical readings, literary methodologies

Bibliography

"Textuella bataljer - om kvantitativa metoder i litteraturvetenskapens tjänst" (artikel kommande, i antologin *Kvantitativa Metoder inom Humaniora och samhällsvetenskap*)

Spatial Humanities and the Norwegian Folklore Archive

Kristina Skåden
University of Oslo, Norway

This papers idea is to present the ongoing work on "spatial Humanities" at The Norwegian Folklore archive and the Department of Cultural Studies and Oriental Languages (IKOS) at the University of Oslo. The main focus will be on how Spatial Humanities may be of interest for education and research in the field of cultural history and museology.

The spring term 2017, the department will start up two new and innovative projects: Firstly, a course on MA-level "Cultural heritage production, Eilert Sundt and Digital Humanities":

http://www.uio.no/studier/emner/hf/ikos/KULH4015/kulh4015var2017.html

In 2017 is it 200 years sins the important cultural and the social scientist Eilert Sundt was born. This event is the starting point for the education in critical cultural heritage production. The aim is that the students, by a theoretical and practical digital humanity-approach will produce an Eilert Sund-map. Some que questions the students will work on: How will the mapping of different quali-

tative sources enrich the understanding of Eilert Sunds research? What kind of space is produced be this mapping practice?

Secondly, the IKOS-department will develop a digital mapping tool for use in research, education and communication. This project is in progress, and it will therefore be interesting to discus with Nordic colleges, different opportunities and pitfalls.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: spatial humanities, mapping, cultural heritage

How to Study Online Popular Discourse on Otherness – Public User Interfaces to Online Discussion Forum Materials

Jaakko Suominen
Elina Vaahensalo
University of Turku, Finland

Suomi24 (suomi24.fi, established in 1998) is Finland's biggest online discussion forum and leading topic-centric social media, and one of the largest non-English online discussion forum in the world. According to TNS Metric service, over 80 % of Finns visit the site monthly, at least when searching information with Google or with other search engines on various topics. Citizen Mindscapes research initiative has collaborated with Aller Media, the owner of Suomi24, as well as with FIN-CLARIN, and opened the discussion forum posts for research use. The data, available e.g. via KORP user interface (https://korp.csc.fi/), consist of over 2 billion words, 53 million comments and almost 7 million threads and covers online discussions over 15 years. Thus, the data gives opportunities to longitudinal studies considering very many aspects, not only focusing on questions on online cultures but also questions on the change of Finnish society in general.

However, there are also other ways to study, at least partially, the Suomi24 discussions, not only with data stored in the Finnish Language Bank and opened for the researchers. This methodological paper examines critically possibilities to search and browse Suomi24 discussions not only with above mentioned KORP interface as well as with Suomi24's own search engine, Google search, Internet Archive, and with the National Library's collection of Finnish websites. We ask here how the different user interfaces affect to the ways of finding discussions, contextualizing them and use them in other ways in digital cultural research. We use the study of analyzing online popular discourses on otherness as the methodological case example. The study introduced here, is part of "Citizen Mindscapes – Detecting Social, Emotional and National Dynamics in Social Media" project funded by the Academy of Finland Digital Humanities Research Programme (http://www.aka.fi/digihum).

Topics: Nordic Textual Resources and Practices, The Digital, the Humanities, and the Philosophies of Technology
Keywords: online discussion forums, methodology, search engines, contextualization

Bibliography

Suominen, Jaakko (2016): "How to Present the History of Digital Games: Enthusiast, Emancipatory, Genealogical and Pathological Approaches." Games & Culture, Published online before print, June 20, 2016, doi: 10.1177/1555412016653341

Suominen, Jaakko (2016): "Helposti ja halvalla? Nettikyselyt kyselyaineiston kokoamisessa." Korkiakangas, Pirjo, Olsson, Pia, Ruotsala, Helena, Åström, Anna-Maria (toim.): Kirjoittamalla kerrotut – kansatieteelliset kyselyt tiedon lähteinä. Ethnos-toimite 19. Ethnos ry., Helsinki, 103–152. [Easy and Cheap? Online surveys in cultural studies]

Suominen, Jaakko & Sivula, Anna (2016): "Digisyntyisten ilmiöiden histori-

antutkimus." In Elo, Kimmo (toim.): Digitaalinen humanismi ja historiatieteet. Historia Mirabilis 12. Turun Historiallinen Yhdistys, Turku, 96–130. [Historical Research of Born Digital Phenomena]
Suominen, Jaakko – Saarikoski, Petri – Turtiainen, Riikka – Östman, Sari (2016, accepted): "Survival of the Most Flexible? National social media services in global competition: The Finnish Case." In Goggin, Gerard & McLelland, Mark (Eds.), Routledge Companion to Global Internet Histories, forthcoming. Routledge, London.

Socio-Economic Relations in Ptolemaic Pathyris: A Network Analytical Approach to a Bilingual Community

Lena Tambs
University of Cologne, Germany

Sometime between 165 and 161 BCE, a subdivision of the larger military camp of Krokodilopolis was established at Pathyris, c. 30 km South of Thebes in Ancient Egypt. Following Upper Egyptian practice, the community mainly consisted of local soldiers and their families (Vandorpe 2011: 295-296). However, progressive efforts to Hellenize the region soon led to Pathyris evolving into a bi-cultural society, with co-existing Egyptian and Greek languages, institutions and practices.

From the time of its establishment until its abandonment in 88 BCE, the structural and cultural complexity of the Pathyrite community can be studied in some detail. This is made possible by a considerable amount of surviving documentary sources. To date, a total of 21 Greek-Demotic archives have been reconstructed (Vandorpe 1994; Vandorpe & Waebens 2009), providing detailed information about the camp, its inhabitants and their affairs.

Previous work on the archives as well as recent development of open access databases such as Trismegistos (TM)[12] enables large scale and systematic examination of the texts. Despite enormous potential of the sources, methods more traditionally used in the field of Egyptology generally fall short in terms of comprehending and embracing the diversity and complexity they represent. For big data projects such as this, newly developed digital tools can help unlocking the potential embedded in the source material.

Particularly relevant for the current project is 'Social Network Analysis' (SNA), aspects of which have been fruitfully applied to written sources from ancient Egypt in the past (e.g. Ruffini 2008; Broux 2015; Broux & Depauw 2015; Cline & Cline 2015; TM Networks).[13] Within a network perspective, ancient societies can be conceptualised as dynamic 'whole-networks' (Marsden 2005: 8) that are structurally composed by complex systems of overlapping, collaborating, and competing sub-networks. My working hypothesis is that using the network analytical software 'Gephi' in employing various analytical tools embedded in SNA to map a high number of specific relations, will facilitate subsequent analysis and interpretation of emerging patterns of socio-economic connectivity in Pathyris.

The current talk provides an outline of the project's main objectives, theoretical approaches and applied methodologies. Arguing for the applicability and usefulness of formal SNA, not only as a powerful visualisation tool but also a multi-functional digital toolbox and interactive interface for analysis and hypothesis testing, examples will be drawn from a case study of the 'Archive of Horos, son of Nechouthes' (TM Arch 106).

Topics: Visual and Multisensory Representations of Past and Present
Keywords: texts, ancient Egypt, social network analysis, Gephi 0.9.1

[12] http://trismegistos.org
[13] http://trismegistos.org/network

Bibliography

Broux, Y. 2015, 'Graeco-Egyptian Naming Practices: A Network Perspective', in: *Greek, Roman and Byzantine Studies*, vol. 55, pp. 706-720

Broux, Y. & M. Depauw 2015, 'Developing Onomastic Gazetteers and Prosopographies for the Ancient World through Named Entity Recognition and Graph Vistualization: Some Examples from Trismegistos People', in: *Social Informatics. SocInfo 2014 International Workshops, Barcelona, Spain, November 10, 2014. Revised Selected Papers*, Aiello, L. M. & D. McFarland (eds.), pp. 304-313

Cline, D. H. & E. H: Cline 2015, 'Text Messages, Tablets and Social Networks: The "Small World" of the Amarna Letters', in: *There and Back Again – the Crossroads II: Proceedings of an International Conference held in Prague, September 15-18, 2014*, Mynářová, J., Pavel, O. & P. Pavúk (eds.), pp. 17-44

Marsden, P. V. 2005, 'Recent Developments in Network Measurement', in: *Models and Methods in Social Network Analysis*, Carrington, P. J. et al. (eds.), Structural Analysis in the Social Sciences, vol. 27, Cambridge: Cambridge University Press, pp. 8-30

Ruffini, G. R. 2008, '*Social Networks in Byzantine Egypt*', Cambridge: Cambridge University Press

Vandorpe, K. 2011, 'A Successful, but fragile biculturalism. The Hellenization process in the Upper Egyptian town of Pathyris under Ptolemy VI and VII', in; *Ägypten swischen innerem Zwist und äusserem Druck: Die Zeit Ptolemaios' VI. Bis VIII. Internationales Symposion Heidelberg 16.- 19.9.2007*, Jördens, A. & J. F. Quack (eds.), Philippika 45, Wiesbaden: Harrassowitz Verlag

Vandorpe, K. 1994, 'Museum Archaeology or How to Reconstruct Pathyris Archives', in; *Acta Demotica: Acts of the Fifth International Conference for Demotists, Pisa, 4th – 8th September 1993*, Bresciani, E. (ed.), EVO 17, pp. 289-300

Vandorpe, K. & S. Waebens 2009, *Reconstructing Pathyris' Archives. A Multicultural Community in Hellenistic Egypt*, Collectanea Hellenistica III, Brussel: Koninklijke Vlaamse Academie van Belgie & l'Union Academique Internationale

Combining Data Sources for Language Variation Studies and Data Visualization

Kristel Uiboaed
Eleri Aedmaa
Maarja-Liisa Pilvik
University of Tartu, Estonia

Mapping and cartographic visualization is an essential component of dialectology, additionally to several other fields of the humanities. The current paper introduces an ongoing project on digitizing, combining and visualizing data of different types and sources for linguistic research purposes. In the presentation, we introduce applied methods, tools and basic workflow of our project "Spatial Data in Linguistics" (http://rurake.keeleressursid.ee/).

The initial idea of the project was to digitize the maps in the only existing atlases on Estonian dialects (Saareste 1938, 1941, 1955) and make them publicly available. The digitization was necessary for presenting and analyzing the old atlas data with contemporary methods and tools, thus enabling queries and a wider range of visualization options, among other things. We, therefore, created a new resource for automatic processing of old dialectological data and made it available for other research purposes as well.

We proceeded with combining the digitized atlas data with the data from the Estonian Dialect Corpus (CED). The comparison and simultaneous analysis of different data sources make it possible to shed light on the studied phenomenon from different perspectives, thereby creating a deeper understanding of the spread and actual frequency of linguistic material. These kinds of combined data are not only of interest to

linguists but can be made use of in other areas of the humanities as well as they convey information about history, ethnography etc. Making the available data reusable and accessible to different disciplines is an important facet of modern research practice and should be encouraged. It is also necessary to share and develop the tools and techniques for working with this data. We therefore also introduce the tools we have used in our project and demonstrate how GIS and R can be combined to present spatial data, and how R can be applied for producing interactive applications of data visualization. Employing such widely-used software as GIS applications and R makes our contributions generalizable and usable also for other researchers of various fields.

Topics: Nordic Textual Resources and Practices
Keywords: textual data processing, spatial data, corpus linguistics, data visualization

References

CED = Corpus of Estonian Dialects <http://www.murre.ut.ee/estonian-dialect-corpus/>

QGIS Development Team 2016, QGIS Geographic Information System. Open Source Geospatial Foundation. URL http://www.qgis.org/.

R Development Core Team 2016, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing., R version 3.2.4, Austria, Vienna, http://www.r- project.org/.

Saareste, Andrus 1938, Eesti murdeatlas. I vihik = Atlas des parlers estoniens. I fascicule. [Atlas of Estonian Dialects. I part] Tartu: Eesti Kirjanduse Selts.

Saareste, Andrus 1941, Eesti murdeatlas. II vihik = Atlas der parlers estoniens. II fascicule. [Atlas of Estonian Dialects. II part] Tartu: Teaduslik Kirjandus.

Saareste, Andrus 1955, Petit atlas des parlers estoniens: Väike eesti murdeatlas. [Small atlas of Estonian Dialects] Uppsala: Almqvist & Wiksell.

Places and Journeys of the Contemporary Norwegian Novel: A Pilot Study

Kim Tallerås, Tonje Vold & David Massey
Oslo and Akershus University College of Applied Sciences, Norway

In *Atlas of the European Novel 1800-1900* (1998), Franco Moretti investigates the European novel from the point of view of maps. What does geography and settings mean in these storylines? Moretti's provoking and compelling idea is that "each space determines, or at least encourages, its own kind of story" (p. 70). A corresponding study has not been conducted in the Norwegian context although Norwegian literature typically is very conscious of geography, and the meanings of regional and local specifics, in a thinly populated country of mountains and valleys, fjords and a long coastline.

The National Library of Norway have digitized and made available large amounts of Norwegian literature, which represent a promising basis for a large-scale automated analysis of geographical information. How can this digitized collection be used in order to investigate Moretti's idea further through a 'distant' reading of Norwegian novels? This research question calls for an interdisciplinary approach and methods from the digital humanities. In this short paper, we introduce a pilot study for a Norwegian "Atlas"-project through focusing on places and journeys in contemporary Norwegian novels. The study includes a test conducted on a limited selection of digitized novels, in order to discover challenges and opportunities for an automated analysis. The thematic limitations of journeys and contemporary literature reflect a methodological need for a consistent and comparable selection, but also enable certain perspectives on e.g. gender, urbanism and environmental criticism we wanted to include in the project.

In the test, a simple schema of entity and relationship types were developed experi-

mentally based on text snippets from the sample corpus. Then, three annotators used Brat[14] and the schema to annotate geographical entities and contextual information found in the sample novels. Eventually, the resulting annotations were analyzed, to see i) what types and extent of information we find in the novels and ii) which textual features (hypernyms etc.) that could be used as evidence in an automated processing.

The project's long-term goal is twofolded, on the one hand to contribute to literary studies: What does geography and settings (fjords and valleys, small towns and cities (etc.), and the journeys between them) signify in contemporary Norwegian novels? On the other, we want to investigate and contribute to digital humanities methods that can be used in order to exploit the newly digitized Norwegian text corpus.

Topics: Nordic Textual Resources and Practices
Keywords: literary studies, entity recognition, information extraction, annotation, visualisation

References

Moretti, Franco (1998). *Atlas of the European Novel 1800–1900*. London: Verso

The Use of Medical Visualisation in Cultural Heritage Exhibitions

Karin Wagner
Gothenburg University, Sweden

This paper deals with how medical imaging is used in visualising cultural heritage, taking the British Museum's exhibition Ancient lives, new discoveries (2104-15) as its case study. In the exhibition, eight exemplars from the museum's collection of mummies were on display together with medical visualisations, that hade been composed into interactive displays. The technology used in the exhibition was similar to the technology that has been developed for the virtual autopsy table at the Center for Medical Image Science and Visualization (CMIV) at Linköping University, Sweden. Huge data sets generated by computer tomography (CT) were used to create three-dimensional images of the mummies, revealing different layers of the body: skin, muscles, organs, and the bone structure as well as the cartonnage case and the wrappings that surrounded the mummies. Also revealed were amulets and other objects hidden beneath the wrappings. Some of these amulets had been 3D-printed and replicas were displayed side by side with the screen-based visualisations. To facilitate for the visitors to interpret the images, colour coding was used. The surface layer of the visualisation was coloured blue, and inside the mummies embalming tools had been coloured green and organs had been given different blue nuance. The 3D-printed replicas of the amulets were of white plastic material, because it was not possible to decide what metal the original amulets were made of, although it was probably gold or silver. Using white was a way of indicating uncertainty and keeping to the facts. All the visualisations in the exhibition were life-size, as it was considered important for the understanding of the images. This paper will discuss the importance of colour coding in visualisations used in cultural heritage exhibitions. What does colour mean in this context and how does it relate to colour coding in medical visualisations meant for an audience of medical professionals? How does scale influence our understanding of reproductions of objects? We are used to seeing art and cultural heritage artefacts reproduced in smaller scale in books, but with digital visualisations the possibility to offer life-size reproductions of objects museum has been greatly increased. The potential of 3D-printing and the tactile dimension this can add to cultural heritage exhibitions will also be explored, and how different types of visualisations can work together.

[14] http://brat.nlplab.org/

Topics: Visual and Multisensory Representations of Past and Present
Keywords: medical visualisation, colour coding, scale, 3D-printing

Bibliography

Wagner, K. 2015 "Reading packages: social semiotics on the shelf", Visual Communication, vol. 14, no. 2, pp. 193-220.

Visualizing the Landscape of Contemporary Norwegian Novels

Miroslav Zumrík
Slovak Academy of Sciences, Slovak Republic

I would like to present my research idea, which deals with visualization of contemporary Norwegian novel production in a given time period (a year, a decade), that is, with detecting common features, tendencies and extremities with respect to narrative structure of novels in question. My project could thus be seen as an enhancement of what is already being done in the series of Samlaget's Norwegian Literary Yearbook (Norsk Litterær Årbok) in the section "A year in the novel" (Romanåret). I would argue that a computationally aided and statistically evaluated analysis of a representative set of novels could provide literary scholars with a suitable empirical background for making statements and formulating hypothesis on narrative/stylistic tendencies employed by contemporary Norwegian novelists, such as use and distributional schemes of narrators and tenses. One of the aims of the presentation, which will focus mostly on the new and also on the ideas developed in my Phd-work, is to address and find research fellow(s) in Norway that would be willing to cooperate on starting project of visualization of narrative features in contemporary Norwegian novels.

As a theoretical background for my project, I would like to employ the theory of fictional worlds, as created by the Czech-Canadian scholar Lubomír Doležel. The theory aims at discovering narrative patterns in a given text and thus reconstructing "fictional world" as a linguistically constructed semiotic object. Doležel starts with looking at the "texture" of the text in question, that is, at the distribution of its linguistic features (tenses, persons/narrators, chapters, paragraphs, etc.) with emerging regularities and irregularities. This basic structure can be further analyzed as a way of expressing text's extensional (themes, events, motifs, characters) and intensional structure (rendered by Doležel's "functions" of authentication and saturation). In the end, one arrives at the stylistic pattern(s) or "shape" of a given narrative text/fictional world. I would like to stress that such analysis does not rely on semantic annotation and does not require in-depth plot segmentation. This next step would require much more time and effort, given that discerning temporal structure (with respect to categories proposed by Genette), or key motifs/events is already a question of interpretation. Doležel's theory, as I understand it, complies with some recent research tendencies within narratology – the interest for the peculiar, experimental, "unnatural" narratives (Hansen – Iversen – Nielsen – Reiter (eds.) 2011, Alber – Hansen (eds.) 2014), the use of computational methods and tools on narrative texts (Weixler – Werner (eds.) 2015, Brunner 2015, Gius 2015), reconstructing the "shape" of fictional worlds (Pettersson 2016). This theory I already applied in my PhD thesis, where I dealt with the novel "In the Shadow of Singularity" (2013) by the Norwegian writer Thure Erik Lund. The novel makes extensive use of a disembodied "we" narrator from the "posthuman" future, re-telling the history of both mankind and the author of the novel, in which it appears, and stating that it had "used" the writer as a vehicle in order to "protrude" back to the "presingular" time.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: contemporary Norwegian novel, narratology, theory of fictional worlds

POSTERS

Interdisciplinary Collaboration for Making Cultural Heritage Accessible for Research

Johanna Berg
The Swedish National Archives
Rickard Domeij
Swedish Language Council
Jens Edlund
KTH Royal Institute of Technology, Sweden
Gunnar Eriksson
Swedish Language Council
David House
Zofia Malisz
KTH Royal Institute of Technology, Sweden
Susanne Nylund Skog
Jenny Öqvist
The Folklore Archives, Sweden

In this poster, we will present plans and initial experiences of collaborating between disciplines within the newly started project Tilltal, as well as studying users and research activities related to the use of memory archives for research.

Currently, the large amounts of recorded speech available at Swedish memory institutions are rarely used due to the lack of effective methods for handling archival sounding material. The aim of the project Tilltal is to examine how speech technology methods can make speech recordings at public memory institutions more accessible to researchers. The project explores how speech technology methods and tools can be adapted and developed to process large amounts of historical voice recordings from the archives of the Institute for Language and Folklore (ISOF). To make this possible, language technologists, SSH researchers and data holders will work in close cooperation within the project.

In a workshop we had in 2015, the idea was to put together groups of three partners: an SSH researcher, a speech technologist, and a data holder. We wanted to take the SSH researchers' current work procedure as the starting point and give them the opportunity to describe their work in small groups, allowing data holders and speech technologists to suggest ways in which the research process could be facilitated by large speech data sets and speech technology. The hands-on task was to come up with suggestions for research projects. As a result, the three sub-studies within the Tilltal project were conceived. They examine how speech technology can be used to investigate research questions within three disciplines: folklore, dialectology and conversation research.

Along with the three sub-studies, a usage study will be performed that applies activity theory to survey the research activities surrounding the archival materials, cf. Nardi 1996. Considering the needs of the researchers, we will model characteristic situations of use following ideas in Hansen et al. 2014, propose language technology solutions and assess their usefulness in practice by means of use cases, in spirit of Jacobson et al. 1992, 2011.

The long-term goal of the Tilltal project is to make the Swedish speech archives more accessible in general, and to SSH researchers in particular. We hope to achieve this not only by describing methods by which speech technology can be used to reach SSH research goals, but also by providing examples of fruitful interdisciplinary collaborations. In the poster we will present our plans, experiences and results so far regarding interdisciplinary collaboration, surveying of research activities and use case modelling.

The project is a collaboration between Digisam, The Institute of Language and Folklore (ISOF), the Royal Institute of Technology (KTH) and Sweclarin. It is funded by the Swedish Foundation for Humanities and Social Sciences from 2017 to 2020.

Topics: Nordic Textual Resources and Practices
Keywords: speech technology, folklore, dialectology, conversation analysis, user studies

References

Berg, Johanna, Rickard Domeij, Jens Edlund, Gunnar Eriksson, David House, Zofia Malisz, Susanne Nylund Skog and Jenny Öqvist (forthcoming). Till-Tal – making cultural heritage accessible for speech research. In: *Selected Papers from the CLARIN 2016 Conference, Aix-en-Provence, 25-28 October 2016*. Linköping Electronic Conference Proceedings.

Hansen, Preben, Anni Järvelin, Gunnar Eriksson, Jussi Karlgren (2014). *A Use Case Framework for Information Access Evaluation*. I: Paltoglou, Georgios, Loizides, Fernando, Hansen, Preben (red.), Professional Search in the Modern World: COST Action IC1002 on Multilingual and Multifaceted Interactive Information Access (ss. 6-22). Springer.

Jacobson, I., Christerson, M., Jonsson, P., and Overgaard, G 1(992). *Object-Oriented Soft-ware Engineering: A Use Case Driven Approach*. Addison-Wesley.

Jacobson, Ivar, Ian Spence, Kurt Bittner (2011). Use Case 2.0: *The Guide to Succeeding with Use Cases*. Ivar Jacobson International.

Nardi, B. A. (1996). *Context and consciousness: Activity theory and human-computer interaction*. MIT Press.

Mapping Language Vitality

Coppélie Cocq
Umeå University, Sweden

This poster will present an ongoing project aiming at visualizing urban linguistic landscapes through digital mapping in the purpose of approaching representations and conditions for multilingualism in Sweden.

Languages available in our environment in the form of words and images, and displayed in public places have been the focus of scrutiny within a rapidly growing research area called *Linguistic Landscape Studies* (see for instance Blommaert 2013; Shohamy et.al. 2008; Shohamy & Gorter, 2010).

A linguistic landscape is shaped by the combination of different forms of official and less official signs, i.e. "road signs, advertising billboards, street names, place names, commercial shop signs, and public signs on government buildings [in a given] territory, region, or urban agglomeration" (Landry & Bourhis, 1997:25). Linguistic Landscape Studies give attention to the consequences and impact that the visibility and materialization of languages can have not only by having an informative and symbolic function, but also their impact on language vitality (Landry & Bourhis, 1997:45). Attitudes toward a language, in relation to visibility and use in public spaces, influence the language's prerequisites and conditions for acquisition and revitalization (Grenoble & Whaley 2006; Hyltenstam 1991).

One of the focuses of the project *Mapping Language Vitality* that will be illustrated and described in the poster is the case of official minority and Indigenous languages in Sweden. The historical hierarchical relationship between these languages and majority languages, and recent changes in minority politics motivate this choice of focus. Also, the practice of naming places has an additional dimension in Indigenous contexts: to name a place in the Indigenous language of the inhabitants is described a pert of a decolonization project (Tuhiwai Smith 2012: 158).

In the poster, we propose to present a pilot study conducted in 2015-2016 in Umeå, the prototype (deep map) that has been developed, and the preliminary results upon which we plan to pursue the project. In the pilot study, about 400 linguistic expressions were photographed in Umeå's cityscape, of which about 150 were coded. The linguistic expressions consist of fixed signs and signboards, posters, temporary vernacular signs etc. A visualization was created by producing a digital map in order to see where and when languages, for example Ume Sami or Finnish were materialized in the city. The photographs were coded and assigned characteristics such as language, position, type of sign, sender, addressee etc. and placed geographically on an interactive map with filterable categories, i.e. that enables the user

to navigate through layers based of the data linked to the images.

In the next step of the project, the same model and tools are used and further developed to include and analyze a larger set of data (photos) covering several geographical areas. The digital map is central in order to visualize connections between languages (as they materialize landscape) and other layers of information. Also, this form of digital visualization enables us to explore how and to what extent different languages coexist, and examine how this looks in relation to the majority language Swedish and other socio-cultural and socio-linguistic factors.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: languages, revitalisation, mapping, linguistic landscape

References

Blommaert, Jan. 2013. Ethnography, Superdiversity and Linguistic Landscapes. Chronicles of Complexity. Toronto: Multilingual Matters.

Grenoble, Lenore A., and Linsay J. Whaley. 2006. Saving Languages : An Introduction to Language Revitalization. Cambridge: Cambridge University Press.

Hyltenstam, Kenneth. 1996. Tvåspråkighet Med Förhinder? : Invandrar- Och Minoritetsundervisning. Lund: Studentlitteratur.

Kasanga, L. 2012. Mapping the linguistic landscape of a commercial neighbourhood in Central Phnom Penh. Journal of Multilingual and Multicultural Development 33(6):1-15

Landry, R., & Bourhis, R. Y. 1997. Linguistic Landscape and Ethnolinguistic Vitality: An Empirical Study. Journal of Language and Social Psychology, 16(1), 23–49.

Shohamy, Elana & Gorter, Durk. 2008. Linguistic Landscape. Expanding the Scenery. Taylor and Francis.

Shohamy, E. et al. 2010. Linguistic Landscape in the city. Multilingual matters.

Tuhiwai Smith, Linda. 2012. Decolonizing Methodologies: Research and Indigenous. Peoples. Zed Books.

Bibliography

Recent publications:

Cocq, C. Turning the Inside Out: Social Media and the Broadcasting of Indigenous Discourse. (European Journal of Comunication, Accepted for publication). Co-author: Lindgren, Simon.

Cocq, C. Narrating Climate Change: Conventionalized Narratives in Concordance and Conflict. (Narrative Works, accepted for publication 2016). Co-author: Andersson, Daniel.

Cocq, C. Exploitations or Preservation? Your choice! Digital modes of expressions for perceptions on nature and the land. I: Communicating environment. How different communication forums react to ecological dangers. Red: Heike Graf. Cambridge: Open Book Publishers. 2016.

Cocq, C. Reading small data in indigenous contexts: ethical perspectives. I: Research Methods for Reading Digital Data in the Digital Humanities, Eds. Gabriele Griffin and Matt Hayler. Edinburgh University Press. 2016.

Cocq, C. Mobile Technology in Indigenous Landscapes. In: Indigenous People and Mobile Technologies. Red. Laurel Evelyn Dyson, Stephen Grant and Max Hendriks. Routledge, New York and Milton Park, UK. Pp147-159, 2016.

Cocq, C. Indigenous voices on the web: Folksonomies and Endangered Languages Journal of American Folklore, Vol. 128, no 509, 2015. pp. 273-285.

Enemies of Books

Olof Gunnar Essvik
Gothenburg University, Sweden

I download the book from the Internet, https://books.google.com, Enemies of books, written by William Blades, published in 1881. A book on the decay of books. The enemies of the physical book – fire, water, gas, the bookworm, dirt, bigotry etc. A digitized book with no identity. Black letters on a white background. I buy a copy of the book from 1881. A yellowed, and stained copy. The book be-

ars traces of a former owner, one Dr. Sarolea. Newspaper cuttings on his death and his extensive book collection. I compare the two texts, the original and the digitized copy. Using my computer I create a tool for binding books. I combine and modify traditional tools that have been used for hundreds of years and print out the components on a 3D-printer. The next day I print out another copy of the digitized book, and using the 3D-printed tool I make an exact replica of the book from 1881, in its original design. I construct a manual describing the process and upload the files to the Internet.

This is a description of a project, which I begun in 2014 as an artistic development project financed by the University of Gothenburg. A project exploring digitization, human traces and the unique copy, and at the same time an act of resistance against digitization and technology. The outcome of the project was a book, presented together with the tool used to produce it.

The project has been presented at art-museum, conferences and universities, both as workshops as well as performance-lectures. In 2016/2017 I will publish two new books within the project about code, marbling and chance (a chapter in the book DATA BROWSER 06, Autonomedia (will be released late 2016)) and another book about shadow libraries and pirated books (Georges Perec, The Machine, Rojal förlag 2017).

I would like to present this project as an experimental poster together with objects from the project. Objects such as coded marbleized papers, 3d printed bookmaking tools, books and posters describing the process of bookbinding using digital tools. I have also done performances during conferences where I make calendars for 100 years. There are numbers of options that could be discussed depending on space and your interest.

Read more and see pictures at: http://www.rojal.se/theenemiesofbooks/
The upcoming article about marbled books and code could be found here: http://www.rojal.se/chanceexecution

Topics: Visual and Multisensory Representations of Past and Present
Keywords: bookbinding, shadow libraries, calenders, 3d printed, objects, time, 100 years

Bibliography

Books
Essvik, Olle & Nordqvist Joel, Den här datorn, 2015, Rojal Förlag, Gothenburg, Sweden
Essvik, Olle & Nordqvist Joel, Virtuella Utopier, 2015, Rojal Förlag, Gothenburg, Sweden
Essvik, Olle, Enemies of Books, 2014, Rojal Förlag, Gothenburg, Sweden
Articles/papers
Essvik, Olle, Chance Execution, Data Browser 06, Executing Practices, (http://www.data-browser.net/06/), Autonomedia, New York, USA (upcoming release late 2016)
Essvik, Olle The museum: as a game, Art and Game Obstruction, Skövde Högskola 2016, Rojal Förlag, Gothenburg Sweden (upcoming release late 2016)
Essvik, Olle, For Years to Come 2015 , In: PARSE Conference, Nov 4-5, The 1st PARSE Biennial Research Conference on TIME.(Conference paper)

Working with Digital Newspapers

Katrine Gasser
Mogens Vestergaard Kjeldsen
Royal Danish Library

Introduction

Since 2014 the Royal Danish Library, Aarhus (RDL) in Denmark has been digitizing historic newspapers. Until date more than 25 million newspaper pages have been scanned and OCR processed. This has not only generated one of the biggest digital newspaper archives in the world, but also a huge amount of text documenting the Danish history and language from the 1800s up until our time.

As a means of inspiration for researchers in the digital humanities, we at RDL find it highly relevant to show some of the possibil-

ities given when taking a digital research approach to the newspaper archive. There are many ways of exploring and approaching a text corpus like this. We have selected visualization, mapping and OCR correction as the starting points. This has produced the beta version tools Smurf, Dots and W2C, which are described in detail below. We find all the tools suitable for implementation with other text corpora than the digital newspaper archive.

Smurf

Smurf is a tool that visualizes how the use of words/phrases in Danish newspapers has evolved since the 18th century. The visualization consists of a graph with a timeline (X) and (Y) which represents the occurrence of the searched word or terms as represented in the digitized newspapers. With Smurf you can search for words or phrases and see graphs, do multiple searches and compare graphs. Clicking a graph point will take you directly to the source newspaper.

Smurf has been available for the public since mid-September 2016 and has been presented for various researchers but also for groups of students (history / literature). The tool's full potential still needs to be explored. Link:

http://labs.statsbiblioteket.dk/smurf

Dots

Dots is a visualization tool based on the newspaper corpus (1800–2013) and a map from Kortforsyningen which contains coordinates and names of cities in Denmark. Dot visualizes the occurrences of words from the newspapers on a map. Dots has a timeline where it is possible to limit the search to a specific period. Furthermore, the timeline makes it possible to track the geographic presence of a word or sentence over time.

Dots is a fairly newly developed tool and will be public in Spring 2017.

Word2vec

Word2Vec is a high-dimensional word embedding tool based on an unsupervised machine learning algorithm using a simple neural network. It maps each unique word in a large text corpus to a vector. The vector representation of the words reflects semantic properties of the words. Words that appear in the same context will be close in the vector-space (similar words). But distance between words can also be used to find analogies. The word2vec tool features several corpora including a very large one based on the digital newspaper archive.

Currently, RDL is examining if word2Vec can be used to correct OCR scanned text by comparing a list of words from word2vec with words from a Danish dictionary. Link:

http://labs.statsbiblioteket.dk/dsc/

Future research options

RDL holds various collections and archives including the Danish Web Archive. We imagine the tools presented above have potential to be used to explore content in the archive.

The Danish Web Archive

Netarkivet.dk is the Danish web archive. It contains the Danish part of the Internet from July 2005 onwards. Due to Danish laws on personal data and data protection, access to netarkivet.dk is restricted to researchers with permission for relevant research projects. The archive may be accessed at http://netarkivet.dk/.

Key learnings

One of the key learnings so far is the value of incorporating other digital sources. As an example the STO dictionary (Den store danske SprogTeknologiske Ordbase) (Braasch & Olsen 2004) delivered from Centre for Language Technology, University of Copenhagen is essential in making the tool Word2Vec useful when trying to identify errors in the OCR generated text. The tools presented have been customized to the library's newspaper archive and are not open source.

Topics: Nordic Textual Resources and Practices
Keywords: text, mapping, corpora, archive

References

Braasch, Anna & Olsen, Sussi. 2004: STO: A Danish Lexicon Resource - Ready for Applications. In Proceedings of the Fourth International Conference on Langauge Resources and Evaluation, vol IV. Lisbon, pp.1079-1082.

Towards a Material Politics of Intensity – Mimetic, Virtual and Anarchistic Assemblages of Becoming-Non-Human/Machine in Minecraft

Marleena Huuhka
University of Tampere, Finland

My poster for the Nordic Digital Humanities Conference will present my ongoing PhD project titled "Towards a Material Politics of Intensity – Mimetic, Virtual and Anarchistic Assemblages of Becoming-Non-Human/ Machine in Minecraft".

My doctoral thesis examines Minecraft and other such sand box building games as material, mimetic, virtual, nomadic and anarchistic performance rhizomes and locations of becoming-something created in cooperation with human and non-human agencies. Human agents involved are for example human players, human spectators and human game designers. Non-human agents include game devices, pixels, electricity, programming language, game avatars and virtual game environments. My research is located at the intersections of performance research, new materialist philosophy, and game research.

My research deconstructs the subject-object dichotomy between human and non-human agents. In virtual game performances non-human agents participate in the production of the performances together with the human player. The avatars movements, though orchestrated by human hands, are the results of human/non-human cooperation. The performance is thus constructed in a shared process of different yet equally important agents. My claim is that in this process the human agent merges into a greater agential ensemble of non-human quality. This approach stems from the work started by Gilles Deleuze and Félix Guattari, and more recently continued by philosophers and media theorists such as Jane Bennett, Jussi Parikka, and Rosi Braidotti.

The performance research aspect of my thesis is to produce an-archic performances and performance spaces – as introduced by director and theatre theorist Antonin Artaud – in video game context. For Artaud theatre is "the sense of gratuitous urgency with which they are driven to perform useless acts of no present advantage." The Artaudian performance is by nature anarchic/anarchistic: it does not concede to negotiate with power structures or to affect via political channels, rather it works primarily through performative demonstrations, which gain their quality outside the norm system.

This kind of performativity can be created in video games through the strategies of counterplay. Counterplay (Nakamura & Wirman 2005, Apperley 2010) means ways of playing, in which the player searches the virtual environment for ways of being and acting unthought-of or unintended by the developers of the game. Usually these practices go against the set goals or intended uses of the game in question. Counterplay is thus gratuitous action done purely for the sake of itself. The combination of Artaud's theory with game theory provides a fresh angle of approach to games as performances.

My thesis concentrates on following questions: 1) in what performative assemblages do the human and nonhuman agents participate in video games, and what kind of meanings are constructed in these assemblages; 2) what kind of performative subjectivities are created through the potentialities of mimesis, virtuality, and becoming-something; and 3) does the Artaudian, anarchistic practice of counterplay open up possibilities for nomadic existence, and if so, why?

Video games have a growing influence on our thinking, societies and cultures. The importance of transdisciplinary research is thus greater than ever. My research opens up new spaces of possibilities by combining performance research, game research and new materialist philosophy. I suggest new ways of participation, resistance and being that transcend the boundaries of species and materialities. The approaches participate in the discussions in the field of digital humanities, and link the pleasure of game play with the critique of hypercapitalism.

In my poster I will present the above introduced key concepts of my research, accompanied with practical examples from my own game play experiences.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: theatre, game research, new materialism, Minecraft

References

Artaud, Antonin 1983. *Kohti kriittistä teatteria.* Delfiinikirjat, Otava.

Apperley, Thimas H. 2010. *Gaming Rhythms: Play and Counterplay from the Situated to the Global.* Institute of Network Cultures.

Bennett, Jane 2010. *Vibrant Matter. A political ecology of things.* Duke University Press.

Braidotti, Rosi 1994 *Nomadic Subjects.* Columbia University Press.

Braidotti, Rosi 2013. *The Posthuman.* Polity.

Deleuze, Gilles & Guattari, Félix 1987 (trans. Massumi, Brian) *A Thousand Plateaus. Capitalism and Schizophrenia (Mille Plateaux. Capitalisme et Schiziphrénie).* The University of Minnesota Press.

Dolphijn, Rick & Van Der Tuin, Iris 2012. *New Materialism: Interviews & Cartographies.* Open Humanities Press.

Galloway, Alexander R. 2006. *Gaming. Essays on Algorithmic Culture.* University of Minnesota Press.

Nakamura, Rika & Wirman, Hanna 2005. "Girlish Counter-Playing Tactics." in *Game Studies,* volume 5, issue 1. http://www.gamestudies.org/0501/nakamura_wirman/

Parikka, Jussi 2014. *The Anthrobscene.* The University of Minnesota Press.

Parikka, Jussi 2015. A Geology of Media. The University of Minnesota Press.

Bibliography

2017 Journeys in Intensity — Human and Non-human Co-agency, Neuropower and Counter-Play in Minecraft in RE-CONFIGURING HUMAN AND NON-HUMAN: TEXTS, IMAGES AND BEYOND. Eds. Karkulehto, Koistinen & Varis. Peer reviewed.

2016 Experience the Wild – Non-human Agency and Performership in Video Games in NÄYTTÄMÖ JA TUTKIMUS 6. Eds. Arlander, Gröndahl, Kinnunen & Silde. Peer reviewed.

2015 Labyrinth – Perspectives on Games and Performances in ESITYSTUTKIMUS. Eds. Arlander, Erkkilä, Riikonen & Saarikoski. Partuuna. With Marjukka Lampo.

The Cultural Heritage HPC Cluster

Per Møldrup-Dalum
State and University Library, Denmark

The Danish e-Infrastructure Cooperation (DeIC) has been charged with spreading High-Performance Computing (HPC) to new research areas, such as the humanities and social science areas. In order to respond to this, DeIC and the State and University Library have agreed to establish the DeIC National Cultural Heritage Cluster, State and University Library.

The cultural heritage cluster applies state-of-the-art technologies within data science, and for the first time ever facilitates quantitative research projects on the digital Danish cultural heritage – e.g. radio and TV programmes, websites and historical newspapers.

Collections Available to Research Projects
The State and University Library and the Royal Library together are responsible for

collecting and preserving Danish cultural heritage, including the digital cultural heritage. This digital cultural heritage is divided into numerous collections, each with its own properties, formats and possibilities. Examples of collections that are now made available to researchers include radio/TV, the Netarchive and the Danish Newspaper Collection.

The radio/TV collection contains more than 1 million hours of TV broadcasts and more than 1.5 million hours of radio programmes broadcast on Danish channels from the 1980s until today. The collection's data are made accessible as audio and video files. The collection also contains large amounts of metadata, such as programme titles, broadcast times and subtitles, depending on the epoch from which the material originates. Read more at mediestream.dk.

The Netarchive contains more than 800 TB data, corresponding to more than 20 billion objects gath-ered from the Danish part of the Internet from 2005 until today. This archive also contains both data and metadata, and both are made available to research projects. The Netarchive is a joint national project between the Royal Library and the State and University Library, and you can read more at netarkivet.dk.

The digital newspaper collection contains 25 million newspaper pages from the 1700s until today. All of these pages are stored as image files along with a large amount of metadata and optical character recognition data (OCR).

The Cultural Heritage Cluster is also available to research projects that bring their own data.

Platform

The Cultural Heritage Cluster is to support new areas and methodologies, particularly within digital humanities. It was therefore decided to design a system that would make it easier easy to conduct well-established analyses without having to compromise in relation to advanced and be-spoke methods.

The Cultural Heritage Cluster is making IBM's BigInsights platform available to research projects. This platform consists of the Open Data Platform (ODPi) and commercial products. ODPi features most of the current Hadoop technologies e.g. Spark and MapReduce.

BigInsights adds a number of commercial applications to ODPi, most prominently BigSheets and Text Analytics. BigSheets uses the spreadsheet metaphor and makes it easy to get started analysing billions of rows of structured data. Text Analytics is a browser-based work area for analysis of unstructured data e.g. text corpora and it comes with a number of complete modules for e.g. POS, NER, and sentiment analysis.

RStudio and Jupyter technologies will also be available for performing more programmatically and advanced analyses ensuring that the system can scale to arbitrarily complex research projects.

Pilot Projects

In the first phase, three pilot projects will utilise the system's new facilities. The State and University Library in collaboration with the DeIC eScience centre of competence will make facilities available and offer training in use of the system to the researchers working on these projects free of charge. In 2017 and 2018, DeIC and the State and University Library will offer further, fully financed pilot projects through open project invitations.

In the course of 2017, it will be possible to buy calculation time and consultancy assistance under a transparent price model, which will be developed in connection with the first pilot projects.

The three planned pilot projects are:

* Probing a Nation's Web Domain, run by Professor Niels Brügger from Aarhus University. The project will analyse the Danish part of the Internet as it has developed from 2005 until today. Their data source will primarily be metadata from the Netarchive.

* Digital Footprints Research Group, run by Anja Bechmann, Aarhus University. This project will analyse data from social media. The data source will be both the project's own data and data from the Netarchive.

* A project run by Sabine Kirchmeier-Andersen from the Danish Language

Council's research institute. This project will analyse the development in the Danes' language usage on the social media, and the data source will be the Netarchive.

Topics: Nordic Textual Resources and Practices
Keywords: HPC, big data, distant reading, quantitative research, EDA

Collecting Speech Data over the Internet

Tommi Nieminen
University of Eastern Finland
Tommi Kurki
University of Turku, Finland

Collecting speech data for linguistic purposes has always been a notoriously tedious and labour-oriented process. Even when it is possible to gather the informants all in one place, the recording studio, the fine-tuning of the equipment, test recordings, and other initial organizing of the session takes so much time that even in optimal situation it is rarely possible to have one to two hour recordings of more than a handful of informants during one workday. And when the research is targetting the areal or social variation of speech it is hardly ever possible to have the informants come to the researcher but the other way around. Thus in many cases it takes one to two workdays per an informant, which means that gathering of speech corpora require considerable investments in work time, and that the build up of large corpora tends to be extremely slow.

This was the dilemma we created the Prosovar project, or more fully "The dialectal and social variation of Finnish prosody", to solve; the project was funded by the Kone Foundation from 2013 to 2015. In it, we experimented on gathering the speech data in the Internet using Web 2.0 techniques: the informants sat comfortably in their homes and used their own computers or mobile devices to connect to our site (https://puhu.utu.fi/) to record their own speech samples. Our intention was to crowdsource this most tedious phase of the process to the informants themselves as fully as possible.

The site hosted several different kinds of tasks for the informants. Some of them just prompted the user what to read; this of course has a seriously deteriorating effect on the naturalness of the informants's speech. In order to collect more spontaneous speech, other tasks gave the informants just barest instructions of what to do. For instance, there was a task where the informant was shown two pictures with minor differences; their task was to spot the differences and report them verbally. In another task, the informant was shown a map of an imaginary city and asked to guide a stranger from one point to another. In still another, the informant was to assume the role of a buyer in a marketplace and ask for some berries from the salesman. And so on.

Of course in order to count as crowdsourcing instead of being only transfer of responsibilities from the researchers to the informants, it is necessary to offer some real baits for the users. Our bait was to create some game-likeness in the site. The informants could for example try and recognize the dialects of other informants – i.e., the could listen to other informants' speech samples and hazard a guess. This game-like nature was however, never fully quite realized during the Prosovar's lifetime.

Now that the project is almost over (the site will close in December 2016) we are in a position to report what you can and cannot do in this way. First of all, we now know that it is fully possible and extremely time-saving thing to do. It is obvious that the Web 2.0 and its facilities are here to stay even when the researcher is interested in speech data.

There are nevertheless several obstacles. First of all, since the recordings are fully carried out by the informants, it is not possible to control the recording settings (in any other way than counselling the users). This tends to leave us to the mercy of the whims of the web browsers and their plugins and add-ons. Secondly, because of this and other considerations, the quality of the speech data

is very variable. In this case, we were more lucky than others might be since the features of speech we were interested in, the prosodic features, are more robust than some others. Still, since the recording almost always involved (lossy) packing, the spectral information in the sound data is always somewhat corrupted. Thirdly, there is no obvious technological candidate for implementing the recording. We used Flash although it is quickly being phased out. For security reasons, mobile devices often block the access to the recording device totally when called from a remote connection. This creates further problems we hope to be able to solve in the future.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: speech corpora, web 2.0, crowdsourcing

Bibliography

Małisz, Zofia & O'Dell, Michael & Nieminen, Tommi & Wagner, Petra 2016: Perspectives on speech timing: Coupled oscillator modeling of Polish and Finnish. Phonetica 73: 233–259.

O'Dell, Michael L. & Nieminen, Tommi & Lennes, Mietta 2015: Hazard regression for modeling conversational silence. – ICPhS 2015: Proceedings. 18th International Congress of Phonetic Sciences, 10–14 Aug 2015 Secc Glasgow Scotland UK.

Nieminen, Tommi & O'Dell, Michael L. 2013: Visualizing speech rhythm: A survey of alternatives. – Eva-Liina Asu & Pärtel Lippus (toim.), Nordic Prosody: Proceedings of the XIth Conference, Tartu 2012. Peter Lang, Frankfurt am Main. 265–274.

Staging the Medieval Religious Play in Virtual Reality

Annika Rockenberger
University of Oslo, Norway

Late Medieval Germany has seen the emergence of the so-called *religious play* as the predominant 'dramatic' form in an institutional context, while the theatre tradition of Greek and Latin Antiquity had been discontinued. Contents of these plays were mostly taken from the New and Old Testament (incl. Apocrypha) as well as hagiography. However, settings and themes from the secular sphere were often included as social satire and for comic relief. Performed during Christian Holidays, and often intertwined with liturgy and church parades, their venues and 'stages' were either set in(side) sacral spaces (churches or other religious buildings) or within close proximity: like central markets or town squares. Religious plays are often believed to have been performed over the course of several days, employing *multiple setting* ('Simultanbühne'), that allowed for a non-chronological as well as a perpetual acting and a non-stationary, oscillating focus of the audience.

Both the older and the more current scholarly editions of these religious plays often do not take their unique character into account when it comes to performance and setting (Auditor 2009). Instead, as a result of anachronistic projections the antique theatre or more modern forms of drama are evoked by mode of representation which has lead to inadequate, misinformed, or flawed interpretations of single plays and an overall misconception of medieval dramatic forms. (Wolf 2004, De Marco 2006, Schulze 2012)

Against this, I propose to take full advantage of current technological possibilities such as *Virtual Reality* and *3D-modelling* to create a probabilistic model of the medieval multi-sensory, multi-setting 'stage' that allows to test common assumptions and new hypotheses about setting, artistic perfor-

mance, mise-en-scène, audience-performer-interaction and participation in religious plays.

My corpus consists of Medieval German plays from late C14th to C16th, most of which are accessible in scholarly editions (Bergmann 1986). I will extract staging information from the plays' paratexts and relevant contexts, especially of plays where church buildings or staging spaces are known or can be inferred from historical sources; in single cases these historical buildings are still 'intact'. This material will be the starting point from which I create a probabilistic model – an experiential analogy (Foka, Arvidsson 2016) – of the performance space using 3D-modelling. Further, I will make use of a VR engine (UNREAL Engine), following a set of pre-preparations, to re-enact single scenes or entire plays.

Since we know very little about the *How* of performing a medieval religious play but can infer some information from spatial setting and artistic motion sequences and patterns, the modelling of the performances, the audience-performer-interaction and especially the multiple setting will have to be experimental. The following questions serve a heuristic function:

(1) Given an exclusively 'inside church-building' venue (as accounted for in most of the shorter plays), where and how would the multiple setting have to be installed to ensure (a) feasible stage arrangement (size, shape, hight and number of scene-space(s) for n performers with at least minimum visibility and audibility for the audience), (b) feasible scene arrangement ('storyline', narration, salvation-historical 'logic' in relation to spatial, movemental, and sensory confinement of audience), (c) feasible time or duration arrangement (within/together with/in addition to a liturgical performance or parade; in relation to seasonal constraints (daylight/temperature/duties/etc.); in relation to physical constraints (endurance of performers, audience; attention span, sensory overload); (d) in regard to additional, more general environmental and circumstantial constraints like lighting, acoustics, but also ephemeral things like sounds, noise, and smells.

(2) Given a 'mixed' venue making use of both church buildings and open spaces (markets, town squares etc.), how would multiple setting, stage, scene and time arrangement have to be done differently? What other general constraints apply here?

(3) Given the great number of text lines per play and taking into account the aforementioned considerations, I believe that the manuscripts do not provide 'the text as it was to be performed' but rather serve as a complete compilation of possible scenes related to a specific Holiday of which a stage director had to pick those scenes he deemed relevant and fitting for a concrete performance within the spatial, physical, and time constraints of the location.

With the proposed poster, I aim to visualize these guiding questions and how VR and 3D-modelling can help answering them. As an example I chose the well-known *Neustifter-Innsbrucker Osterspiel* from 1391 which is believed to originate in Southern Thuringia, Germany, possibly in the town of Schmalkalden, where the medieval church building has survived and can thus be sampled for testing simple VR modelling using Google Cardboard and photo spheres.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: virtual reality, middle high German, medieval times, religious play, liturgical performance

Bibliography

(German) Medieval Religious Plays

Hofmeister, Wernfried, Cora Dietl, and Astrid Böhm, eds. Das Geistliche Spiel Des Europäischen Spätmittelalters. Wiesbaden: Reichert, 2015. Print. Jahrbuch Der Oswald-von-Wolkenstein-Gesellschaft / Oswald-von-Wolkenstein-Gesellschaft. - Wiesbaden : Reichert, 1981- 15.

Mattern, Tanja. 'Liturgy and Performance in Northern Germany: Two Easter Plays from Wienhausen'. A Companion to

Mysticism and Devotion in Northern Germany in the Late Middle Ages. Ed. Elisabeth Andersen, Henrike Lähnemann, and Anne Simon. Leiden: Brill, 2014. 285–315. Print.
Schulze, Ursula. Geistliche Spiele Im Mittelalter Und in Der Frühen Neuzeit: Von Der Liturgischen Feier Zum Schauspiel ; Eine Einführung. Berlin: Schmidt, 2012. Print.
Prosser-Schell, Michael. Szenische Gestaltungen Christlicher Feste: Beiträge Aus Dem Karpatenbecken Und Aus Deutschland. Münster [u.a.]: Waxmann, 2011. Print. Schriftenreihe Des Johannes-Künzig-Instituts / Johannes-Künzig-Institut Für Ostdeutsche Volkskunde. - Freiburg, Br : Johannes-Künzig-Inst. Für Ostdeutsche Volkskunde, 1998 13.
Auditor, Anne. 'Die "Innsbrucker Spielhandschrift". Überlegungen zu einer Neuedition'. 'Texte zum Sprechen bringen'. Philologie und Interpretation. Festschrift für Paul Sappler. Ed. Christiane Ackermann and Ulrich Barton. Tübingen: Max Niemeyer Verlag, 2009. 297–305. Print.
Bergmann, Rolf. Katalog Der Deutschsprachigen Geistlichen Spiele Und Marienklagen Des Mittelalters. München: Beck [in Komm.], 1986. Web. Veröffentlichungen Der Kommission Für Deutsche Literatur Des Mittelalters Der Bayerischen Akademie Der Wissenschaften.
Ogden, Dunbar H. The Staging of Drama in the Medieval Church. Newark, Del.: Univ. of Delaware Press, 2002. Print.
Hoffmann, Yvonne. Festtagsgeschehen Und Formgenese in Den Gewölben Der Spätgotik. Mannheim: Waldkirch, 2008. Print.
De Marco, Barbara, ed. Performance in the Middel Ages and Renaissance. Binghamton: Center for Medieval and Renaissance studies, 2006. Print. Mediaevalia : An Interdisciplinary Journal of Medieval Studies Worldwide. - Binghamton, NY : Center, 1975- 1.
Wolf, Klaus. 'Für eine neue Form der Kommentierung geistlicher Spiele. Die Frankfurter Spiele als Beispiel der Rekonstruktion von Aufführungswirklichkeit'. Ritual und Inszenierung. Geistliches und weltliches Drama des Mittelalters und der frühen Neuzeit. Ed. Hans-Joachim Ziegeler. Tübingen: N.p., 2004. 273–312. Print.
3D-modelling, Virtual Reality, Augmented Reality in Historical and Archeological Research
Greengrass, Mark, and Lorna M. Hughes, eds. The Virtual Representation of the Past. Aldershot: Ashgate, 2008. Print. Digital Research in the Arts and Humanities.
Carter, Brian Wilson. 'The Evolution of Virtual Harlem: Bringing the Jazz Age to Life'. Digital Humanities 2016: Conference Abstracts. Kraków: N.p., 2016. 143–147. Web.
Foka, Anna, and Viktor Arvidsson. 'Experiential Analogies: A Sonic Digital Ekphrasis as a Digital Humanities Project'. Digital Humanities Quarterly 10.2 (2016): n. pag. Web. 28 Oct. 2016.
Scheuermann, L., L. Jantke, and W. Scheuermann. 'Erlebter Raum Im Rom Der Späten Republik - Eine Digitale Forschungsumgebung'. Digital Humanities 2016: Conference Abstracts. Kraków: N.p., 2016. 670–671. Web.

Use of Digital Methods to Switch Identity-related Properties

Jon Svensson
Roger Mähler
Umeå University, Sweden
Mats Deutschmann
Örebro University, Sweden
Anders Steinvall
Satish Patel
Umeå University, Sweden

It has long been observed that language is at the heart of mechanisms leading to stereotyping and inequality. In fact, language is a major factor in our evaluation of others, and it has experimentally been demonstrated that individuals are judged in terms of intellect and other character traits on the basis of their language output alone. Thus, awareness of such mechanisms is of crucial importance in education, especially in the training of groups who will be working with people in their future profession; groups such as teachers, police, psychologists, nurses. Although some courses deal with the consequences of linguistic stereotyping on a theoretical level, there is a need to provide students with a deeper understanding of how they themselves are affected by such processes, so that this knowledge can have an impact on their future practices.

The RAVE research project at Umeå University addresses exactly this issue by exploring and developing (pedagogical) methods for revealing sociolinguistic stereotyping in regard to identity-related properties such as gender, age, physical appearance, ethnicity, etcetera. The main approach is the use of digital matched-guise testing techniques with the ultimate goal to create an online, packaged and battle-tested, method available for public use. The project is however not only devoted to creating the product itself but also to testing and evaluating the effectiveness of various digital methods and configurations with respect to raising awareness of linguistic stereotypes, that is, to answer questions such as what is the best script, the best media, the best setup, the proper length, the appropriate content, etcetera.

The method relies to a great extent on a treatment where two groups of test subjects (i.e. students) are exposed to a scripted dialogue between two characters, let's say "Terry" and "Robin", in two different versions in which each character is assigned presumed stereotypical properties. In one version, for example, "Terry" may sound like a man, while the other recording has been manipulated for pitch and timbre so that "Terry" sounds like a woman. After the exposure the test, subjects are presented with a survey where they are asked to respond to questions related to linguistic behaviour and character traits of the interlocutors. The responses of the two sub-groups are then compared and followed up in a debriefing session, where issues such as stereotype effects are discussed.

The project produces the two property-bent versions based on a single recording, and the switch of the property (for instance gender or age) is done using digital methods. The reason for this procedure is to minimize the number of uncontrolled parameters in the experiment. It is a very difficult, if not an impossible, task to transform these identity-related aspects of a voice recording, such as gender or accent, into a "perfect" voice - a voice that is opposite in the specific aspect, but equivalent in all other aspects, and doing so without changing other properties in the process or introducing artificial artifacts.

This project doesn't strive for perfection. Instead the focus is on the perceived credibility of the scripted dialogue. Various kinds of techniques, for instance the use of audio-visual cues, can be used to both distract the test subject from the "artificial feeling", as well as enforce the target property. For instance, to enforce the gender, we can use visual cues, switching images between a man and a woman. We can also add distractions that lessen the listeners' focus on the speaker. It is also possible to use scrambling techniques, for instance by setting up the dia-

logue as a low-quality phone call or a Skype session.

This poster session will present the experiences gained and lessons learned so far in the in the ongoing project. We will give an overview of the various methods used and tested so far, starting from methods used in prior projects with rather simple, low-quality, gender morphed voices in Second Life enforced with avatars, to more sophisticated qualitative attempts made by audio experts and with the use of sophisticated software.

The focus will be on the credibility aspects and the methods to determine if a certain dialogue is perceived as credible or not. The presentation covers aspects such as selection of narrators, use of actors, use of standard audio manipulation software as well as dedicated phonetic software such as PRAAT, the use of speech synthesizers, and morphing towards a reference (or imposter) voice. The poster will be supplemented by a mp3 player and a headset where visitors can listen to samples.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: language, stereotyping, voice-morphing, match-guise technique, perceived credibility

Prozhito: Private Diaries Database

Nataliya Tyshkevich
Ivan Drapkin
National Research University, Moscow, Russian Federation

Despite the continuing interest in ego- and microhistory research, with particular focus on collective memory (Burns 2006), scholars still tend to neglect personal writings. The most valuable of them are private diaries with entries, usually tied to a chronological line, representing personal narratives of people of different ages and cultural and social levels. There are many projects based on materials of personal diaries of one person, such as the diaries of George Orwell (George Orwell Diaries 1938–1942). At the same time multi-author online-diaries corpora are rather popular, see for example, (Teddiman 2009). However there exist neither relevant diary subcorpora of any national European corpora (such as British National Corpora), nor special multilingual databases of personal diaries.

We are first to present a global database of private diaries, tied to a chronological line, representing personal narratives of people of different ages and cultural and social levels. "Prozhito" is a non-profit project, which blends the structural experience of blog platforms and archival tradition of curating personal writings. Our database includes 400 diverse non-authorized diaries or 150 000 entries from the XIX-XX centuries in Russian and Ukrainian with a possibility to multilingual expanding. A researcher can work not only with particular texts but with the whole collection of diary entries, using complex search queries by author's gender and age, journal types (f.e, war, tourist, dream etc.) and filtering results by exact dates and places of records.

The first version of the site was opened April 24, 2015 and contained 100 journals and 30,000 diary entries, collected from public internet sources. As a result of collaborations with the Russian media audience of Prozhito groups in social networks has

grown several times and there was a steady stream of volunteers. In November 2016 we launched a new version, focused primarily on the ordinary users. The main features of the last version are intelligent search tools, a developed system of classification (genre, author, language, geographical origin) and multilingual architecture.

On the home page users can use simple tool to search across notes of all diaries by keyword and dates and observe collections of notes and random quotes. Extended search panel on the Diary page provides parameters for selecting authors and searching notes by author's name, last name, age, notes tags or keywords etc. On the Diary page users can observe all list of uploaded diaries, cathegorized by languages. Every diary author has his own profile page with information about author and his diary.

From the beginning, the project was designed as a heavily visited web-platform with unlimited scalability. Architecture of the current version combines several databases: MySql, Sphinx, Redis. In MySql we store all the basic entities: a person, diary, enries, comments, thematic tags, copyright status, preview, etc.

Search by various parameters, and the main morphological search and implemented through the Sphinx system. Sphinx indexes all entries words, diary dates and other parameters.

A feature of the loaded data is that not all known records have the exact dating of writing. To include these entries in the sample, data for each entry on the stage of the diary is automatically determined by the estimated date period in which it is written.

Currently published private diaries bibliography, made by participants of the project, contains more than 1,500 units. Project participants identify existing publications associated with the heirs, the owners or publishers for electronic text, search for text copy, scan and recognize the book.

In addition to working with the already published material Prozhito implements its own publishing program.

Work with texts from family collections needs to strike a balance between personal and public to avoid publishing the information that could possibly compromise third parties. This is one of the most difficult tasks of private archives publishing, the key to which can only be found in close cooperation with the heirs and administrators of family archives. In Prozhito the manuscript owners (person or family) continue to participate in its preparation for publication and control the text on all the steps of its transformation from the manuscript to the machine-readable database unit. They have the right to exclude fragments, considered unappropriate due to ethical reasons.

Working with a family history often activates intrafamily communication, but the information, stored in the family archives, is of interest not only for the family members. Prozhito project allows any user to explore the diaries data and gives huge research material for researchers of everyday life.

Topics: Nordic Textual Resources and Practices
Keywords: private writing, database, corpora, private diaries

Creating Children's Books in the Context of Pokémon Go, Museums and Cultural Heritage

Lars Vipsjö
University of Skövde, Sweden

The poster will show the contents of *Kiras och Luppes Bestiarium* (Kiras and Luppes Bestiary) - a children's book series set in Skaraborg cultural environments. The books are supported by an Augmented Reality app: KLUB Bestiarium, which can be downloaded for free from App Store or Google Play. Users of the AR app will see the mythical beings from the books appear also above the books in 3D. This happens when you hold the mobile phone's or reading pads camera over letter shapes (drop caps/anfangs) found on certain pages.

The stories are designed together with museum educators and written and illustra-

ted by game development students at the University of Skövde. The books are published in cooperation with Lokrantz book publishers in Lidköping and the app company Solutions Skovde. The project was first presented to the public at the Book and Library Fair in Gothenburg in autumn 2016. Two books are published and several more are under construction. The friends Kira and Luppe are in the books helping the trollsresearcher Lovis save mythical beings from the evil ringmaster, who is actually also a troll, a mountain king. The ringmaster wants to force other mythical creatures to perform at his circus.

The idea is that the fairytale figures through the app also will be portrayed and "come to life" at the local heritage sites and museums where the stories are unfolding. Mythical beings, when found, are scanned and saved in the app's bestiary, where there users can find facts to read about them.

Idea and outcome targets: Because the books and the app's content is produced together with children and museum educators (target group 7-11 years) the material is thought to be useful in teaching situations. Within museums, libraries and schools, the books and app. can be tailored to local requirements. Kiras and Luppes Bestiary (KLUB) is a part of the project Kastis - Kulturarv och spelteknologi i Skaraborg (Cultural heritage and gaming technology in Skaraborg) supported by Skaraborgs kommunalförbund (Skaraborg municipal association), the University of Skövde and a number of Skaraborg Municipalities. Some of the books in KLUB has also received funding in the form of reading promotion support from the Västra Götaland region. KASTiS is a sub-regional collaborative and knowledge platform for the sustainable use of game technology and interactive media in Skaraborg. The project started August 1, 2015 and ends on July 31, 2018.

Topics: Visual and Multisensory Representations of Past and Present
Keywords: augmented children books, cultural heritage

Bibliography

Vipsjö, Lars och Johansson, Therese (2016). Kiras och Luppes Bestiarium. Jättinnan. Lidköping: Lokrantz förlag. ISBN 978-91-98351323

Vipsjö, Lars och Hansson, Matilda (2016), Harmannen. Lidköping: Lokrantz förlag. ISBN 978-91-983513-0-9

Arvas, Maina (2015). "Lars Vipsjö om onda stereotyper". Intervju under temat "Detalj" i Tecknaren #5 2015, Stockholm: Föreningen Svenska Tecknare.

"Scarred and evil – A villain stereotype that does not inspire empathy when he loses" (2015) In Westin, J., Foka, A. and Chapman, A. (eds) Challenge the past / diversify the future – proceedings March 19-21 2015 Gothenburg: University of Gothenburg. http://hdl.handle.net/2077/38407

Vipsjö, Lars och Bergsten, Kevin (2014) Tecknad karaktär – anatomi, fysionomi och psykologi. Lund: Studentlitteratur. ISBN 978-91-44-08482-4

From Online Research Ethics to Researching Online Ethics

Sari Östman & Riikka Turtiainen
University of Turku, Finland

Along with the rise of a research field called digital humanities, online specific research ethics plays an especially significant role. Research on the same (Internet related) topic is usually multidisciplinary, and understanding research ethics even inside the same research community may vary essentially. It is important to recognise and pay attention to online specific contexts as well as the researcher's own disciplinary background.

In our poster, we will present a model which we have under developing process: this fourfold table will help researchers in positioning themselves as ethical actors on multidisciplinary online-related fields. The positioning is based on the researchers'

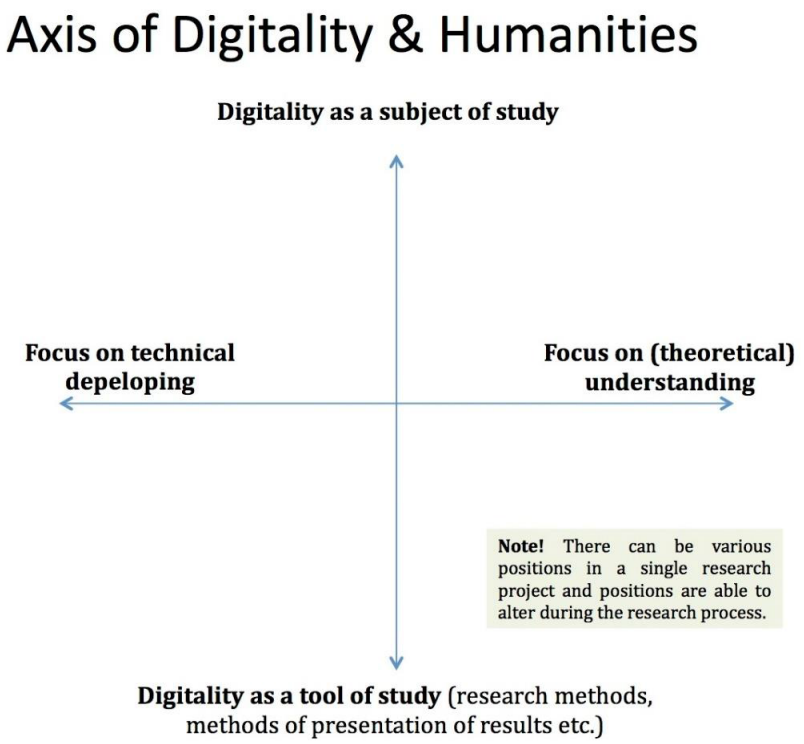

Figure 1. *Axis of Digitality & Humanities*. Suominen & Haverinen 2015.

topics and their disciplinary backgrounds in relation to the position of the internet in their study.

The model is based on Jaakko Suominen's and Anna Haverinen's *Axis of Digitality & Humanities*, which is a tool for positioning yourself on the field of Digital Humanities as such. (Figure 1. Suominen & Haverinen 2015; translated S. Östman & R. Turtiainen, 2016, p. 3.)

Our model for ethical positioning builds on basis of this fourfold table: we suggest that researchers would ask themselves − in addition to their subjects and backgrounds − whether the internet is *a tool, the source* or *the subject* to their study. It could be just one or even all three; in latter case we would consider internet as *research environment* of the study. On the basis of the position of the internet, the individual take to research and e.g. the disciplinary background, we suggest the researchers would locate themselves in a coordination system like the one in Figure 2 (see next page) in order to see the relevance of such ethical matters, which are characteristic for humanistic research.

In Figure 2 (see following page), we have situated 15 informants to our model. They are all researchers in a multidisciplinary consortium project, which studies Suomi24, Finland's oldest and largest online discussion forum. They answered to a survey, which mapped their backgrounds and current research as well as their understanding of research ethics. According to their answers, we have defined their positions in the coordination. Preliminary results suggested that researchers from quantitatively and theoretically oriented disciplines on average see research ethics less as a reflective, analytical process and more so as concerning copyright laws, for example. On the contrary, empirically and understandingly oriented (often cultural studies -based) researchers tended to be more analytical, versatile, and reflective in their ethical views. Humanistic researchers also had more education about the subject (Östman & Turtiainen 2016, 72–73).

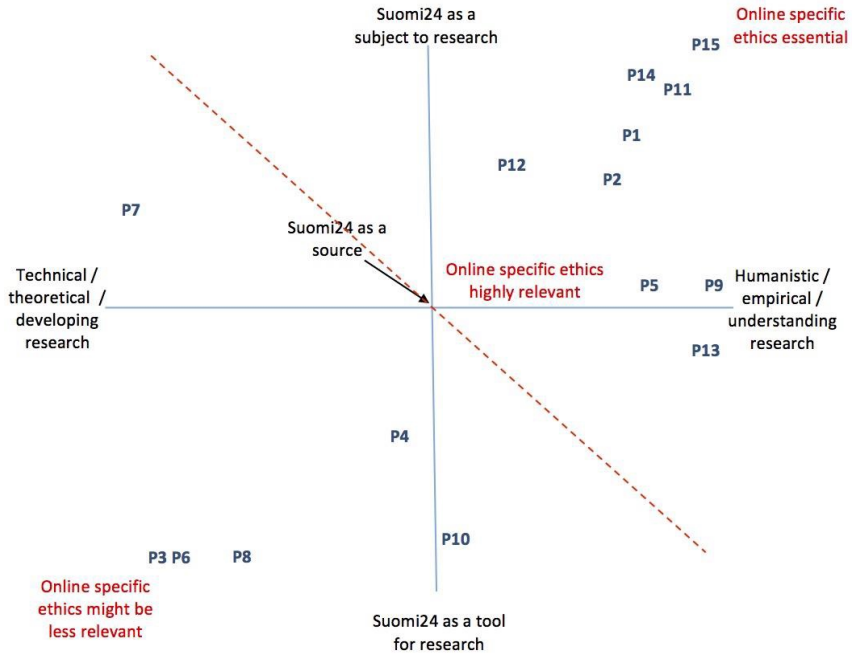

Figure 2. *Model for ethical positioning.* Östman & Turtiainen 2016, 72.

We are currently further studying the multidisciplinary research ethics and further developing the model for ethical positioning. The questions are, among others, following:

1. Which kind of ethical questions are relevant among different disciplines taking part to multidisciplinary consortium?m hans
2. Which kind of multidisciplinary ethical guidelines could be provided as a result of studying online research ethics in such consortium?

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: digital humanities, digital culture, research ethics, online research ethics

Literature

Suominen, J., & Haverinen, A. (2015). Koodaamisen ja kirjoittamisen vuoropuhelu?—Mitä on digitaalinen humanistinen tutkimus. *Ennen ja nyt.* Retrieved from http://www.ennenjanyt.net/2015/02/koodaamisen-ja-kirjoittamisen-vuoropuhelu-mita-on-digitaalinen-humanistinen-tutkimus/#identifier_14_1502

Östman, S. & Turtiainen, R. (2016). From Research Ethics to Researching Ethics in an Online Specific Context. In *Media and Communication*, vol 4. iss. 4. pp. 66–74. Retrieved from http://www.cogitatiopress.com/ojs/index.php/mediaandcommunication/article/view/571

Bibliography

Östman, S., & Turtiainen, R. (2016a). From Research Ethics to Researching Ethics in an Online Specific Context. In Media and Communication, vol 4. iss. 4. pp. 66¬–74. Retrieved from http://www.cogitatiopress.com/ojs/index.php/mediaandcommunication/article/view/571

Östman, S., & Turtiainen, R. (2016b). Understanding Ethics in Digital Humanities. Guidelines and tools for conducting research in online contexts. In Folklore Fellows Communications nr´. 310 (eds. Pekka Hakamies & Anne Heimo). (In publishing process.)

PRE-CONFERENCE WORKSHOPS

Higher Education Programs in Digital Humanities: Challenges and Perspectives

Koraljka Golub
Linnaeus University, Sweden
Jenny Bergenmar
University of Gothenburg
Isto Huvila
Uppsala University, Sweden
Marcelo Milrad
Linnaeus University, Sweden
Mikko Tolonen
University of Helsinki, Finland

Introduction

Different aspects related to higher education programs in Digital Humanities (DH), whether, what and how they should be organized, are currently discussed at many higher education institutions in Nordic countries and beyond. In recent years the establishment of new educational programs under the title of Digital Humanities, for example in the USA, UK and Germany, are an indication of a perceived need for developing such specific curricula. DARIAH-EU has a dedicated research and education centre under the title of Virtual Competency Centre (VCC) Research and Education Liaison (http://www.dariah.eu/activities/ research-and-education.html). DARIAH-EU also runs a registry of Digital Humanities education in Europe (http://dh-registry.de .dariah.eu) which, as of 10 January 2017, lists currently active 17 Bachelor degrees, 38 Master degrees, and 8 individual courses. The University of Stuttgart and the University of Trier are just two examples that run programs under the actual title of Digital Humanities. Similarly, EADH (European Association for Digital Humanities) provides a list of education programs, courses and seminars in Europe (http://eadh.org/ education) and names: 7 undergraduate programs and courses, all with terms like Digital Humanities, Humanities Computing and related in the title; 20 postgraduate ones with a more mixed array of titles; and, 4 PhD programs, all with the title of Digital Humanities or very similar (University College London, King's College London, a cluster of 4 Irish universities, and University of Passau).

In the Nordic countries similar efforts are underway at the University of Gothenburg (http://lir.gu.se/english/education/masters-second-cycle/master-s-programme-in-digital- humanities), which is launching a Master in Digital Humanities in autumn 2017. The University of Helsinki (https:// www.helsinki.fi/en/researchgroups/helsinki -digital-humanities) is also offering a set of courses in Digital Humanities. Linnaeus University (https://lnu.se/digihum/) aims towards developing an international distance Master program in Digital Humanities, with a pilot program starting in the autumn of 2017. At the same time, at other universities, courses in digital methods and topics have been integrated as a part of existing and new programs as specific compulsory and elective modules, or by including Digital Humanities related topics and perspectives as a part of other courses.

However, what a dedicated course, module or program in the field of Digital Humanities should cover is not always clear. There is a considerable variation between different offerings including diverse content and approaches. The vast range of disciplines, fields, areas and topics relevant to Digital Humanities present a challenge as to what to include in a dedicated program, how to address the different challenges related to bringing together different disciplinary traditions and methods, and how to accommodate professional, infrastructural and academic requirements for such initiatives. Moreover, there are several challenges associated with what is expected from the outcomes of these new educational programs and efforts. Which job positions and tasks could a graduate Digital Humanist take on after completion of a Digital Humanities program? Is there a need for Digital Humanists as such or should education in all humanities subjects be more inclusive of digital technology-related, cross-disciplinary and cross-sectorial topics? If the latter is the case, do we need entire programs or could the alter-

native of focusing on dedicated modules or individual courses address existing and emerging needs of both the academic and the non-academic spheres? Furthermore, if both approaches were deemed to have their merits, how do they differ, overlap and complement each other in the context of educating future researchers and professionals for different sectors of the society?

The aim of this proposed workshop at DHN 2017 is to bring together scholars, educators and others interested in different aspects of Digital Humanities education to explore the current potential and challenges and opportunities related to the teaching and learning of Digital Humanities. The workshop will provide an opportunity to share experiences, discuss existing programs, modules and courses in Digital Humanities, research and development activities, evaluation approaches, lessons learned, and findings. A further objective is to systematically engage in discussions in common areas of interest with selected related communities and to investigate potential co-operation and concrete collaborative activities.

The workshop will allow major established programs and initiatives to report results, newcomers to interact with established people in the field in order to allow the entire community to critically discuss topical issues. The DHN venue encourages participation by Digital Humanities teachers, researchers and developers from different perspectives (reflecting the different conference threads). As the first workshop on education at DHN, it may set the path for future workshops at the annual DHN conferences in order to establish and provide a regular forum for discussions on education in Digital Humanities in Nordic countries and beyond.

Workshop themes

The proposed workshop will have three themes as the main focus, together with topical presentations arising from the workshop CfP. The main themes are:

1. Existing programs, modules or individual courses in Digital Humanities: design, target student groups, content, job market, evaluation, experiences and lessons learned.

2. Currently developed programs, modules or individual courses in Digital Humanities: approaches to the design, target student groups and related issues.

3. Cross-disciplinary and cross-sectorial collaboration in Digital Humanities education.

Workshop structure

Indicative agenda structure, covering approximately 4 hours:

Session 1: Welcome, introduction, mutual presentations (30 min);

Session 2: Presentations on the main themes (90 min);

Session 3: Directed discussion emerging from the main session 30 min);

Session 4: Presentation and discussion of submitted papers on timely and related topics according to the CfP (60 min);

Session 5: Concluding discussion, including options for co-operation (30 min).

Audience

The intended audience includes: teachers and managers at existing and developing Digital Humanities programs; researchers working with topics in Digital Humanities education; professionals who are interested in taking a Digital Humanities program, modules, or courses.

Number of participants: 20

Register to: koraljka.golub@lnu.se

Data Management for Humanities Scholars – An Introduction to Data Management Plans and the Cultural Heritage Data Reuse Charter

Marie Puren
Charles Riondet
INRIA, France

With the growth of the Open Science movement in the past few years, researchers have been increasingly encouraged by their home institutions, their funders, and by society at large, to share the data they produce. Significantly, the Horizon 2020 Research and Innovation Programme has undertaken to open the research data produced by H2020 funded projects. A new model of data sharing is emerging, and the challenges this new model raise are impacting more and more dramatically the research ecosystem.

Rather than seeing in it an additional constraint, scholars can benefit from the advantages that this model of openness offers. Sharing their data allows them to collaborate with fellow researchers within the same discipline or with colleagues from other disciplines, to reduce costs by avoiding duplication of data collection, to make easier validation of results, and to increase the impact and visibility of their research outputs.

Opening research data induces not only a change in mentality, but also a change in work methods. Data management has to be seen as the baseline of the research lifecycle. In this regard, it should be thought of as early as possible in a research project, and should be flexible enough to evolve all along the project. For researchers, this practice supposes to plan and decide how data will be collected, organised, managed, stored, preserved and shared during a research project, and after the project is completed. These requirements can best be addressed by setting up so called data management plans. This method is fairly new to most Humanities scholars, although it is a key element of good data management. Data management plans (DMP) have in particular the great advantage that they take into account the fact that data has a longer lifespan than the research project that creates them. DMP are conceived and applied in order to ensure that data will be preserved and useful both now and in the future, for both their creators and their reusers. Besides, in order to support open access for research data, several funders make data sharing mandatory, and their applicants must thus provide a data management plan to do so.

Data Management Plans are not simply management tools at project level, they also allow a broader reflexion on research data in the Humanities on a larger scale. Although they can apply to any data for any research field, we have chosen to make their benefit easier to grasp by addressing one specific use case - but a use case that applies to a wide range of Humanities research projects. The focus here will be on the reuse, for research purposes, of data emanating from Cultural Heritage Institutions. In this specific situation, there is often a lack of a clear policy on interactions between institutions and scholars. Therefore researchers encounter difficulties to develop a clear data management policy for their research projects in connection with Cultural Heritage data.

The Cultural Heritage Data Reuse Charter we are currently developing in the context of DARIAH-EU and other research infrastructures tackles this issue by offering an online environment dedicated to all actors taking part in scholarly reuse of digital data generated by Cultural Heritage Institutions. The Charter online environment will allow the main actors to declare general principles (common work ethics), and more broadly to express their position on all the relevant information needed to understand how a given dataset can be reused. Institutions will be able to declare their collections; researchers their research interests and existing publications so that these are connected together. The Charter will also help document the knowledge generation process and, consequently, increase the quality of

data and metadata accessible to research. Signing the Charter will also imply making a statement about the technical quality of the data to be reused, or the data derived by such a reuse. More broadly, the Charter will offer a concrete implementation framework for the FAIR principles (make the data findable, accessible, interoperable and reusable). Finally, clarifying the reuse conditions of cultural heritage data, and by that also the relationships between scholars and GLAM institutions, will enable to widen the cooperation opportunities.

Within the framework of this future online environment, CHI and scholars will be able to explicit their constraints concerning data reuse. The Charter will not only allow CHI to clarify their policy on data reuse and enable researchers to have a precise overview of their rights, it will also allow CHI and researchers to handle easily the digital data they produced and therefore help them to define their strategy on data management. In other words, the Charter will strongly connect with data management planning, whose main goal consists of clearly stating the data policy of a research project, and will be an essential asset for data management planning for research on Cultural Heritage.

Workshop provisional program

We expect the workshop to last about three hours. Detailed presentations will be accompanied by open discussion, where we would like to take advantage of the presence of DH researchers and representatives of Cultural Heritage Institutions to engage in a fruitful exchange.

The workshop will be divided in three sub-sessions:

1) *Data management for researchers: Overview and challenges* (60 min)

In this session, we will discuss the new model of data sharing that is actually emerging as described above. Participants will also get an overview of research data management and data management planning. Data management can offer many advantages, like higher quality data, increased visibility and better citation rate. In this approach, research data is an asset and a resource that can be shared with mutual benefits for the person who share the data and the one who collect the data.

2) *Sharing research data: methods, tools and benefits* (presentation + hands-on session: 60 min)

Sharing their research data allows the researchers to organise and retrieve them effectively, to ensure their security, to collaborate with fellow researchers within the same discipline or from other disciplines, to reduce costs by avoiding duplication of data collection, to make easier validation of results, to increase the impact and visibility of their research outputs. Many are still reluctant to share their data, but, fortunately, data sharing is gradually evolving towards a greater openness with the movement for Open Science and the development of Open Access. However, researchers need to be aware of the benefits of sharing their research data, because sharing (or not) rests most of the time on the shoulders of the researchers who decide whether and how to share their data.

3) *A future pan-european framework for exchanging information about Cultural Data reuse: the Charter online environment* (presentation and discussion: 60 min)

In this session, we will present the prototype of a future online environment dedicated to Cultural Heritage data reuse. By taking into account the longer lifespan of Cultural Heritage data, this future tool will offer many valuable elements (e.g. documentation, guidelines, list of services) that could be used to easily create data management plans:

* Long-term and persistent access to metadata, texts, images;
* Licensing of the content;
* Formats and standards;
* Dissemination of both CHI information and research (visibility of the work of all stakeholders);
* Retro-provision (communicating enrichments based on CHI data to the CHI they originally emanate from);
* Quality control at all levels according to appropriate standards.

* This session will be dedicated to discuss the features that could be offered by this online environment, regarding data management planning and improved cooperation between relevant actors (Cultural Heritage institutions, researchers, data centers, infrastructures and other facilities).

Approximate number of participants: 20.
Authors: Marie Puren and Charles Riondet, Ph.D., are junior researchers in Digital Humanities at the French Institute for Research in Computer Science and Automation (INRIA) in Paris. They currently work on the creation of a Data Management Plan for the PARTHENOS H2020 project. Marie Puren also contributes to the IPERION H2020 project, especially by upgrading its Data Management Plan. Charles Riondet is also involved in H2020 EHRI project as a metadata and standards specialist.

Topics: The Digital, the Humanities, and the Philosophies of Technology
Keywords: data management, research data, reuse, cultural heritage institutions, cooperation

Bibliography

Principles of the Data Reuse Charter
Laurent Romary, Mike Mertens, Anne Baillot, "Data fluidity in DARIAH – pushing the agenda forward", BIBLIOTHEK Forschung und Praxis, De Gruyter, 2016, 39 (3), pp.350-357. <hal-01285917v2>
Background information on Open Access to publications
Murray-Rust, Peter, "Open Data in Science", Serials Review, Vol 34, No 1. Accessed October 28, 2016. https://goo.gl/9ZqdiQ

Data management and curation
Ray, Joyce, Putting Museums in the Data Curation Picture, Springer, 2014.
University College London, Advancing Research and Practice in Digital Curation and Publishing. Summary Report and Recommendations of the Workshop on Next Steps in Research, Education and Practice, 2010. Accessed October 30, 2016. https://goo.gl/hqKuKQ
Data management planning
Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age, National Academy of Sciences, Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age, National Academy Press, 2009. https://goo.gl/URJglu
Licensing
Europeana, The Europeana Licensing Framework. Accessed October 28, 2016. https://goo.gl/947T4z
Standards
Laurent Romary, "Stabilizing knowledge through standards - A perspective for the humanities",Going Digital: Evolutionary and Revolutionary Aspects of Digitization, Science History Publications, 2011. <inria-00531019>

Developing a Repository and Suite of Tools for Scandinavian Literature

Mads Rosendahl Thomsen
Aarhus University, Denmark
Timothy R Tangherlini
UCLA, United States of America
Kristoffer Laigaard Nielbo
Aarhus University, Denmark

The goal of the workshop is to set benchmarks for the further development of a machine-readable corpus of Scandinavian literary texts which is part of a project that is the continuation of two Carnegie-Mellon Foundation sponsored conferences on computational approaches to Scandinavian literature. A third conference is planned for UCLA in November 2017.

At the workshop the practical implementation of the following goals will be discussed:

1) a preprocessed benchmark corpus of selected literary texts in the Scandinavian language;

2) a wider machine readable Scandinavian corpus. The corpora will be assembled from DSL, Litteraturbanken and Norwegian libraries;

3) a portfolio of tools with documentation.

The workshop is focused on aligning the needs of literary scholars with the technical solutions that can be developed by the core group members.

Professor with Special Responsibilities Mads Rosendahl Thomsen (madsrt@cc.au.dk, Aarhus University), Professor Timothy Tangherlini (tango@humnet.ucla.edu, UCLA) and Associate Professor Kristoffer L. Nielbo (kln@cas.au.dk, Aarhus U) will chair the workshop. We expect to attract 10-12 other participants from Scandinavian and the US, including scholars from Gothenburg University and Oslo University who have taken part in prior meetings.

Topics: Nordic Textual Resources and Practices
Keywords: Nordic literature, text mining, corpora building

Transkribus: Handwritten Text Recognition Technology for Historical Documents

Louise Seaward
University College London, United Kingdom
Maria Kallio
National Archives, Finland

Transkribus (https://transkribus.eu/ Transkribus/) is a platform for the automated recognition, transcription and searching of handwritten historical documents. Transkribus is part of the EU-funded Recognition and Enrichment of Archival Documents (READ) (http://read.transkribus.eu/) project. The core mission of the READ project is to make archival material more accessible through the development and dissemination of Handwritten Text Recognition (HTR) and other cutting-edge technologies.

The workshop is aimed at researchers and students who are interested in the transcription, searching and publishing of historical documents. It will introduce participants to the technology behind the READ project and demonstrate the Transkribus transcription platform. Our team has already conducted over 20 similar workshops over the course of the past year, including several sessions with digital humanities scholars and students.

Transkribus can be freely downloaded from the Transkribus website. Participants will be instructed to create a Transkribus account and install Transkribus on their laptops in advance of the workshop. They will also be asked to upload a few images of historical documents to Transkribus prior to the session. They should bring their laptops along to the workshop.

The workshop will consist of four parts:

1. *Introduction to Handwritten Text Recognition (HTR) technology* (20 min)
The introduction to this workshop will explain how new algorithms and technologies are making it possible for computer software to process handwritten text. Handwritten Text Recognition (HTR) technology works differently from Optical Character Recognition (OCR) for printed texts (Leifert et al., 2016). Rather than focusing on individual characters, HTR engines process the entire image of a word or line, scanning it in various directions and then putting this data into a sequence. This introduction will outline the workings of HTR technology and show examples of the successful automatic transcription and searching of historical documents. The latest experiments demonstrate that Transkribus can automatically generate transcripts with a Character Error Rate of 5 %. This means that 95 % of the characters in the transcript would be correct.

2. *Overview of the READ project* (20 min)
This presentation will give an overview of the READ project and the specific tools it is

creating. Computer scientists working on READ are developing HTR technology using thousands of manuscript pages with varying dates, styles, languages and layouts. Testing the technology on a large and diverse data set will make it possible for computers to automatically transcribe and search any kind of handwritten document, from the Middle Ages to the present day, from old Greek to modern English. This research has huge implications for the accessibility of the written records of human history. The READ project is making this technology available through the Transkribus platform but also developing other tools designed to make it easier for archivists, researchers and the public to work with historical documents. The workshop leaders will present prototypes of some of these tools. These include a system of automatic writer identification, an e-learning app to enable users to train themselves to read a particular style of writing, a mobile app to allow users to digitise and process documents in the archives and a crowdsourcing platform where volunteers can transcribe with the assistance of HTR technology. These tools will be open source and are designed to be used and adapted by other institutions and projects.

3. *Introduction to Transkribus* (20 min)
HTR technology is made available through the Transkribus platform, which is programmed with JAVA and SWT (Mühlberger et al.) A transcription of a handwritten document can be undertaken in Transkribus for two main purposes. The first is a simple transcription – this allows users to train the HTR engine to automatically read historical papers. The second is an advanced transcription – this allows users to create a transcription of a document which may serve as the basis of a digital edition. This presentation will explain both uses of Transkribus.

HTR engines are based on algorithms of machine learning. The technology needs to be trained by being shown examples of at least 30 pages of transcribed material. This helps it to understand the patterns which make up words and characters. This training material is known as 'ground truth' (Zagoris et al., 2012, Gatos et al., 2014). The workshop leaders will demonstrate how 'ground truth' training data can be prepared using Transkribus.

Transkribus can also be used simply for transcription. This presentation will explain how to create a rich transcription of a document in the platform, using structural mark-up, tagging, document metadata and an editorial declaration.

4. *Working independently with Transkribus*
(120 min)
In the last part of the workshop, the participants will be able to try out the functions of Transkribus on their own laptops. They will be supported by the workshop leaders who will explain the different elements of the platform and then give participants the chance to practice each function for themselves. The workshop leaders will circulate around the room to answer any questions.

The workshop leaders will demonstrate the following tasks. After each demonstration, participants will be given 10-15 minutes to practice what they have learned.

* Document management – how to upload, view, save, move and export documents in standard formats (PDF, TEI, docx, PAGE XML)

* User management – how to allow specific users to view and edit documents

* Layout analysis – how to segment your documents to create training data for the HTR engines

* Transcription – how to create a rich transcript with tags and mark-up

* HTR – how to apply HTR models to automatically generate transcripts, how to conduct a keyword search of your documents, how to assess the accuracy of automatically generated transcripts

The workshop will close with a Question and Answer session where participants can clarify anything they are unsure about. They will also have the opportunity to provide feedback on the Transkribus tool via our user survey.

Number of participants: 15
Participants will need to bring their own laptops on and install Transkribus (https://transkribus.eu/Transkribus/) before attending the workshop.

Topics: Nordic Textual Resources and Practices
Keywords: digitisation, handwritten text recognition, digital scholarly editing, crowdsourcing, OCR

References

Leifert, G., Strauß, T., Grüning, T., and Labahn, R., 'Cells in Multidimensional Recurrent Neural Networks' (2016), https://arXiv.org/abs/1412.2620v02

Mühlberger, G., Colutto, S., Kahle, P., 'Handwritten Text Recognition (HTR) of Historical Documents as a Shared Task for Archivists, Computer Scientists and Humanities Scholars. The Model of a Transcription & Recognition Platform (TRP)' (pre-print)

Gatos, B., Louloudis, G., Causer, T., Grint, K., Romero, V., Sánchez, J.A., Toselli, A.H., and Vidal, E., 'Ground-Truth Production in the tranScriptorium Project', Document Analysis Systems (DAS), 2014 11th IAPR International Workshop on Document Analysis Systems (2014), 237-244

Stamatopoulos, N., and Gatos, B., 'Goal-oriented performance evaluation methodology for page segmentation techniques', 13th International Conference on Document Analysis and Recognition (ICDAR) (2015), 281-285.

Konstantinos, Z., Pratikakis, I., Antonacopoulos, A., Gatos, B., and Papamarkos, N., 'Handwritten and Machine Printed Text Separation in Document Images Using the Bag of Visual Words Paradigm", in: Frontiers in Handwriting Recognition (ICFHR), 2012 International Conference, Bari (2012), 103-108. DOI: 10.1109/ICFHR.2012.207.

Contact information: The workshop will be delivered by Louise Seaward (University College London) and Maria Kallio (National Archives Finland). The contact is Louise Seaward.

Dr Louise Seaward, Bentham Project, Faculty of Laws, University College London, Bidborough House, 38-50 Bidborough Street, London, WC1H 9BT
louise.seaward@ucl.ac.uk
+44 020 3108 8397

Bibliography

Seaward, L. 'The Small Republic and the Great Power: Censorship between Geneva and France in the later Eighteenth Century', The Library: Transactions of The Bibliographical Society, Forthcoming

Seaward, L. (2014) 'The Société typographique de Neuchâtel (STN) and the Politics of the Book Trade in late Eighteenth-Century Europe, 1769-1789', European History Quarterly, vol. 44 (3), pp. 439-479

Seaward, L. (2014), 'Censorship through Cooperation: The Société typographique de Neuchâtel (STN) and the French Government, 1769-89', French History, vol. 28 (1), pp. 23-42