DHNB 2022 CONFERENCE

Digital Humanities in the Nordic and Baltic Countries 6th Conference

UPPSALA 15–18 MARCH 2022

BOOK OF ABSTRACTS

Karl Berglund, Matti La Mela, Inge Zwart & Clelia La Monica (eds.)
CDHU/Department of ALM, Uppsala University, March 2022

DHNB

UPPSALA UNIVERSITET

Contents

(Alphabetically by first author)

A-H 8

Anne Agersnap, Kirstine Helboe Johansen, Katrine Baunvig Frøkjær - The legacy of the Danish 'church fathers' N.F.S. Grundtvig and S.A. Kierkegaard in 11,955 contemporary sermons 8

Niko Marius Kustaa Aho, Anna Ristilä, Reima Välimäki, Mila Oiva, Aleksi Vesanto – Text re-use in large internet data: workflow and BLAST optimization 9

Karolina Andersdotter, Malin Nauwerck - Secretaries at work: accessing Astrid Lindgren's stenographed manuscripts through expert crowdsourcing....10

Everita Andronova, Anna Frīdenberga, Lauma Pretkalniņa, Renāte Siliņa-Piņķe, Elga Skrūzmane, Anta Trumpa, Pēteris Vanags - User-friendly search possibilities for the early Latvian texts: challenges posed by automatic conversion....12

Federico Aurora, Andrea Alessandro Gasparini, Alessandro Palumbo - ENCODE - Bridging the <gap> in ancient writing cultures, enhance competences in the digital era 13

Anda Baklāne, Valdis Saulespurēns - Detection of the target domain of conceptual metaphors in Latvian poetry: exploring the effectiveness of Word2vec and BERT models 15

Katrine Frøkjær Baunvig, Kristoffer Laigaard Nielbo – Mermaids are birds. Text Mining N.F.S. Grundtvig's Bestiary....16

Ümit Bedretdin, Pihla Toivanen - Supervised multi-class classification for humanities research: An article classifier using BERT and topic modelling 17

Jenny Bergenmar, Koraljka Golub, Siska Humlesjö - Queerlit database: making Swedish LGBTQI literature easily accessible....18

Karl Berglund, Mats Dahllöf - Books as muzak: Tracking streamed audiobook reading by the hour....20

Yngvil Beyer, Andre Kåsen - NorHand: a preliminary dataset for Norwegian HTR (1825-1940) .21

Magnus Breder Birkenes, Lars G. Johnsen, Andre Kåsen - Aligning the past and the present: two approaches to automatic modernization of historical texts 22

Robert Borges - MoReDaT: A Modular Remote Data-collection Toolkit for Linguistics Research 24

Kristin Charlotte Carlsson - Collaboration and the role of a metadata librarian in an e-infrastructure project at the University of Oslo 26

Steven Coats - The Corpus of British Isles Spoken English (CoBISE): A New Resource of Contemporary British and Irish Speech 27

Coppélie Cocq, Evelina Liliequist, Lacey Okonski - Protecting the Researcher in Digital Contexts....28

Alexander Conroy - Quantifying conceptual history?....29

Mats Dahllöf - Quotation and Narration in Contemporary Popular Fiction in Swedish – Stylometric Explorations....30

Milda Dailidėnaitė, Valts Ernštreits, Gunta Kļava - Latvian prefixes in Livonian and digital tools to sort them out....31

Jānis Daugavietis - Building Latvian DH tools and resources: rationale and typical practices....33

Katrin Dennerlein, Michael Huber - Reports on modelling dramatic metadata. With examples of the communicative relevance of female playwrights in the second half of the 18th century....34

Philip Diderichsen, Jens Bjerring-Hansen, Dorte Haltrup Hansen, Ross Deans Kristensen-Mclachlan - Mending Fractured Texts. A heuristic procedure for correcting OCR data....35

Alina El-Keilany, Thomas Schmidt, Christian Wolff - Distant Viewing of the Harry Potter Movies via Computer Vision....36

Antoinette Fage-Butler, Kristian Hvidtfelt Nielsen, Loni Ledderer, Marie Louise Tørring, Kristoffer Laigaard Nielbo - Exploring public trust and mistrust relating to the MMR vaccine in Danish newspapers using computational analysis and framing analysis....38

Olof Falk - Genre- and subject-based access to literary fiction in digital library catalogues: a brief overview....39

Elena Fernández Fernández - Social Acceleration: An Empirical approach using Computational Text Analysis and Newspapers in Spanish (1988-2018)....40

Rosemarie Fiebranz, Daniel Löwenborg - 18th-century village environment 3D visualized – the heritage site Ekeby hamlet....40

Rikard Friberg von Sydow - "Take these broken links" - Twitter, the Q-drops and the collapse of a digital ecosystem....42

Mats Fridlund - Digitizing Humanities Research in the GLAM: Institutionalizing Digital Humanities at the National Library of Sweden's KBLab, 2017-2021....43

Filip Ginter, Harri Kiiskinen, Jenna Kanerva, Li-Hsin Chang, Hannu Salmi - Deep Learning and Film History: Model explanation techniques in the analysis of temporality in Finnish fiction film metadata....45

Aleksandrs Gorbunovs - Peculiarities of e-learning objects that attract the learner: Towards model development through eye-tracking....45

Tamás Grósz, Noora Kallioniemi, Harri Kiiskinen, Kimmo Laine, Anssi Moisio, Tommi Römpötti, Anja Virkkunen, Hannu Salmi, Mikko Kurimo, Jorma Laaksonen - Tracing Signs of Urbanity in the Finnish Fiction Film of the 1950s: Towards a Multimodal Analysis of Audiovisual Data....46

David Håkansson, Carin Östman, Sara Stymne, Johan Svedjedal - How fiction made the Swedish language modern....48

Mika Hämäläinen, Tanja Säily, Daniela Landert -Data-Driven Neologism Mining in a TV Corpus 50

Fredrik Hanell, Pernilla Jonsson Severson - Netnography: two methodological issues and the consequences for teaching and practice....51

Trond Haugen - "The Peder Rafn-Project". Reading a 16th & 17th Century Collection of Danish-Norwegian and German Broadsheet ballads in Transkribus....53

Raphaela Heil, Fredrik Wahlberg - Machine learning based image restoration of archival images....54

Ulrike Henny-Krahmer, Robert Hesselbach - Computational stylistic analysis of literary aesthetics in Roberto Bolaño's 2666....55

Isto Huvila, Olle Sköld, Lisa Börjesson - Documenting and making sense of digital research processes: findings from an international survey of archaeologists 57

I-P 59

Jonas Ingvarsson, Daniel Brodén, Lina Samuelsson, Victor Wåhlstrand Skärström, Niklas Zechner - Between the interpretative and algorithmic: mixed methods and literary criticism 59

Gerth Jaanimäe - Challenges of normalizing historical texts written in a morphologically rich language 60

Anne Järvinen, Eetu Mäkelä - Detecting and analysing news flow dynamics and their changes in 20 years of Finnish news 61

Heidi Jauhiainen - Encoding Hieroglyphic Texts 62

Tommi Jauhiainen, Jussi Piitulainen, Erik Axelson, Krister Lindén - Language Identification as part of Text Corpus Creation Pipeline at the Language Bank of Finland 63

Ellert Þor Jóhannsson, Finnur Ingimundarson - Describing inflectional patterns of nouns in Old Icelandic 64

Marko Jouste, Jukka Mettovaara, Petter Morottaja, Niko Partanen - Archive infrastructure and spoken language corpora for Saami languages in Finland 65

Kati Kallio, Maciej Janicki, Eetu Mäkelä, Mari Sarv - Recognizing intertextuality in the digital corpus of Finnic oral poetry 67

Andra Kalnača, Tatjana Pakalne - Assigning meaning to novel productively formed complex words in actual language use: a case of the Latvian agentive suffix -tāj- 68

Almazhan Kapan, Suphan Kirmizialtin, Rhythm Kukreja, David Joseph Wrisley - Fine Tuning NER with spaCy for Transliterated Entities Found in Digital Collections From the Multilingual Arabian/Persian Gulf 69

Heidi Karlsen, Lars Johnsen - A Digital Discourse Analysis of the Norwegian National Library's Collections - The Idea of a Feminine Essence in the First Half of the Twentieth Century 71

Maria Kazakova - The Proposal of an Algorithm for the Studies in Humanities based on TikTok-material 72

Joonas Kesäniemi, Mikko Koho, Eero Hyvönen, Esko Ikkala - Using Wikibase for Managing Cultural Heritage Linked Open Data Based on CIDOC-CRM 74

Heikki Keskustalo, Laura Korkeamäki, Kimmo Kettunen, Elina Late, Sanna Kumpulainen - "What is Missing from our Palette?": Methodological Learning Experiences from a Digital Humanities Research Project 75

Kimmo Kettunen - Geographic Space in Pentti Haanpää's Novel Korpisotaa – where does the War Happen? 76

Mikko Koho, Heikki Rantala, Eero Hyvönen - Digital Humanities and Military History: Analyzing Casualties of the WarSampo Knowledge Graph 77

Heikki Kokko, Tuula Pääkkönen - Translocalis Project: Making of the new digital cultural heritage from the forgotten 19th century historical material 79

Mogens Kragsig Jensen, Jakob Povl Holck, Evgenios Vlachos - Structuring the Past using Text Mining. Occupational Perspectives on Dansk Folkeblad 1838 and 1840. 80

Clelia R. LaMonica - Using online legal databases in English for Specific Purposes 81

Rafael Leal, Heikki Rantala, Mikko Koho, Esko Ikkala, Minna Tamper, Markus Merenmies, Eero Hyvönen - WarMemoirSampo: A Semantic Portal for War Veteran Interview Videos.....83

Peter Shaff Leonard - Rummet som icke är: Generative models and latent space in visual collections. 84

Petri Leskinen, Heikki Rantala, Eero Hyvönen - Analyzing the Lives of Finnish Academic People 1640–1899 in Nordic and Baltic Countries: AcademySampo Data Service and Portal 85

Thea Lindquist, Erik Radio - Metadata Tools for Bibliographic Data Science.....86

Eetu Mäkelä, Pihla Toivanen - Analyzing the representation of politicians in the media – results and methodological concerns 86

Johan Malmstedt – Collecting Silences: A comparative analysis of silence in Swedish radio from P1 and P3, 1980 - 1989 88

Jani Marjanen, Antti Kanner, Eetu Mäkelä - Using a bigram model for semantic change to study Finnishness in early newspapers 89

Benjamin George Martin - Digital Analysis of Global Debates: Text Mining the UNESCO Courier, 1948-2011 90

Antonina Martynenko - Reading the unreadable: towards formal distinctions between 19th-century Russian women's poetry and one failed hoax 91

Inés Matres - Practices of looking from the photo archive for a postphotographic age 92

Haralds Matulis, Sanita Reinsone, Ilze Ļaksa-Timinska - Automatic Detection of Dates in the Corpus of Diaries.....94

Florian Meier - Towards Contributionhip Attribution in the Trykkefrihedsskrifter: A Stylistic Analysis of the Danish Freedom of the Press Writings' Main Writers 95

Aleksi Nicolas Moine - An Exploration of the Mythical Networks of Northern Karelian Incantations: Epistemological and Methodological Issues.....96

Liisa Maria Näpärä - Connecting researchers and digital collections at the National Library of Finland 97

Michael Neiß - 3D laser scanning as a tool for artefact studies 98

Seraina Nett, Rune Rattenborg, Carolin Johansson, Gustav Ryberg Smidt, Jakob Andersson - Here, There, and Everywhere: A Global Heritage Perspective on Cuneiform Culture.....99

Kristoffer Nielbo, Rebekah Baglini, Andreas Roepstorff - Information Decoupling as a Pandemic Signature in News Media 100

Lars Oestreicher, Jan von Bonsdorff - From Visual Forms to Metaphors - Targeting Cultural Competence in Image Analysis 102

Patrik Öhberg, Daniel Brodén, Mats Fridlund, Victor Wåhlstrand Skärström, Magnus P. Ängsal - A Unifying or Divisive Fear? Terrorism in Swedish Public Opinion and Parliamentary Motions 1986–2020 103

Emily Sofi Öhman - SELF & FEIL: Reusing emotion lexicons for multilingual emotion detection in interdisciplinary projects 104

Teemu Tapani Oivo - Environmental knowledge, media producers and technology, in Finnish-Russian border regional media flows 106

Arttu Oksanen, Minna Tamper, Eero Hyvönen, Jouni Tuominen, Henna Ylimaa, Katja Löytynoja, Matti Kokkonen, Aki Hietanen - A Tool for Pseudonymization of Textual Documents for Digital Humanities Research and Publication 107

Eljas Oksanen, Heikki Rantala, Jouni Tuominen, Michael Lewis, David Wigg-Wolf, Frida Ehrnsten, Eero Hyvönen - Digital Humanities Solutions for pan-European Numismatic and Archaeological Heritage Based on Linked Open Data 108

Siim Orasmaa, Kristjan Poska, Kadri Muischnek, Anna Edela - Named Entity Recognition in 19th Century Communal Court Minute Books 110

Jacob Orrje - A digital history of a scientific academy. Exploring the actors of the Royal Swedish Academy of Sciences 1742–1826 111

Petri Paju, Hannu Salmi, Heli Rantala, Patrik Lundell, Jani Marjanen, Aleksi Vesanto - Textual Migration across the Baltic Sea: Creating a database of text reuse between Finland and Sweden 112

Niko Partanen, Rogier Blokland, Michael Rießler, Jack Rueter - Transforming archived resources with language technology: From manuscripts to language documentation 113

Natalia Perkova, Kirill Kozhanov - Towards the corpus of Latvian Romani texts: deciphering the manuscripts from Jānis Leimanis' archive 114

Ginta Perle-Sile, Sanita Reinsone - Transcription as a Tool for Deep Reading and Teaching of Folklore 115

Nadezhda Povroznik - Museum Digital Identity: Building a New Vision of Museum Functions in Virtual Environments 116

Q-Ö....118

Heikki Rantala, Esko Ikkala, Mikko Koho, Ville Rohiola, Eljas Oksanen, Jouni Tuominen, Eero Hyvönen - FindSampo: A Linked Data Based Service for Analyzing and Disseminating Archaeological Finds 118

Krista Stinne Greve Rasmussen, Jon Tafdrup, Kim Steen Ravn, Katrine Frøkjær Baunvig - The case for scholarly editions 119

Annika Rockenberger - Digital Humanities in the Nordic and Baltic Countries - A Living Bibliography 120

Anne Grethe Sæbø - Serving as a Method's Partner to the Researcher in Complex Digital Times: A vision for a new era for the University of Oslo Library 122

Kirsi Sandberg, Mykola Andrushchenko, Risto Turunen, Jani Marjanen, Jussi Kurunmäki, Jaakko Peltonen, Timo Nummenmaa, Jyrki Nummenmaa - Analyzing temporalities in parliamentary speech about ideologies using dependency parsed data....123

Thomas Schmidt, Sabrina Hartl, Konstantin Kulik, Vera Wittmann - Systematic Evaluation of Annotation Tools and Analysis of Annotation Behavior for Three Annotation Tasks....124

Maria Skeppstedt, Rickard Domeij, Gunnar Eriksson, Jenny Öqvist - Digital humanities for the spreadsheet nerd: Presenting the output of a topic modelling tool as tabular data 125

Bo Ærenlund Sørensen, Lars Kjær - Looking for dangerous liquids in Chinese literature: a programmatic approach 126

Polina Staroverova, Natalia Perkova - Towards improving the OCR quality of 18th century Russian pre-reform books and periodicals....128

Emil Stjernholm - Distant Reading Televised Information: Exploring the Communication of Swedish Government Agencies, 1978–2020 129

Jana Sverdljuk, Lars Johnsen, Magnus Birkenes - Merging Digital Humanities and Discourse Analysis in the Study of COVID-19 Vaccine Distribution in Norwegian Newspapers.....130

Minna Tamper, Jouni Tuominen, Eero Hyvönen - Extending the Finnish Linked Data Infrastructure with Natural Language Processing Services in FIN-CLARIAH 131

Iiro Lassi Ilmari Tiihonen, Yann Ryan, Lidia Pivovarova, Aatu Liimatta, Tanja Säily, Mikko Tolonen - Distinguishing Discourses: a Data-Driven Analysis of Works and Publishing Networks of the Scottish Enlightenment.....133

Pihla Toivanen - A workflow for selecting the research material: Using BERT and active learning in catching the phenomenon.....134

Jouni Tuominen, Mikko Koho, Ilona Pikkanen, Senka Drobac, Johanna Enqvist, Eero Hyvönen, Matti La Mela, Petri Leskinen, Hanna-Leena Paloposki, Heikki Rantala - Constellations of Correspondence: a Linked Data Service and Portal for Studying Large and Small Networks of Epistolary Exchange in the Grand Duchy of Finland.....135

Risto Juhani Turunen, Ilari Taskinen, Lauri Uusitalo, Ville Kivimäki - Mining Emotions from the Finnish War Letter Collection, 1939-1944.....137

Raf Van Rooy, Xander Feys, Maxime Maleux, Andy Peetermans The rocky road to DaLeT: Pitfalls and successes in developing a database of the Trilingual College of Leuven (1517–1578).....138

Joshua Wilbur - The usefulness of “small data” in capturing syntactic change in an under-documented endangered language.....139

(1) veicinā-tāj-s ‘promoter, facilitator (may refer to any of the above)’
SPECIALIZATION: occupation, specialized tool, machine, technology vs. non-specialized performer of an action, etc.;
(2) uzlabo-tāj-s ‘enhancer, additive, improver with object’, rakstī-tāj-s ‘writer'
EVENT: event-related vs. non-event-related;
(3) turē-tāj-s ‘holder (may be event or non-event related)’
SEMANTIC ROLE: agent, experiencer, etc.;
(4) meklē-tāj-s ‘seeker, searcher’ – slimo-tāj-s ‘smb who is ill’
TYPE OF ACTION: single, specific action vs. repeated, habitual actions, prolonged state/ status based on actions vs. hypothetical action.
(5) staigā-tāj-s ‘walker (may refer to any of the above)’

Bibliography

Alexiadou, A. 2017. Nominal Derivation. The Oxford Handbook of Derivational Morphology. Lieber, R., Štekauer, P. (eds.). Oxford: OUP.

Kalnača, A., Lokmane, I. 2021. Latvian Grammar. Riga: ULP.

Nau, N. 2013. Latvian agent nouns: their meaning, grammar, and use. Baltic Linguistics 4, 79–13.

Soida, E. 2009. Vārddarināšana [Word-Formation]. Riga: ULP.

Štekauer, P., Valera, S., Körtvélyessy, L. 2012. Word-formation in the world’s languages: A typological survey. Cambridge: CUP.

Roy, I., Soare, E. 2013. Event related nominalizations. Categorization and Category Change, Cambridge Scholars Publishing, 123–152.

The Balanced Corpus of Modern Latvian 2018 (LVK2018). http://www.korpuss.lv/id/LVK2018

Almazhan Kapan[2], Suphan Kirmizialtin[1], Rhythm Kukreja[1], David Joseph Wrisley[1]

Fine Tuning NER with spaCy for Transliterated Entities Found in Digital Collections From the Multilingual Arabian/Persian Gulf

Keywords *Named Entity Recognition, Gulf Studies, Colonial Archives, Transliterated Names, custom models*

Contribution short paper

Affiliation 1: NYU Abu Dhabi, United Arab Emirates;
2: NYU Shanghai, China

Abstract

Searchable, transcribed cultural heritage text collections have become an important part of digital GLAM. With the democratization of handwritten text recognition (HTR) platforms, this trend of studying and reusing more texts from archives will no doubt continue. The situation presents an ethical dilemma for computational study, however, since archival materials, particularly those of an intercultural nature, or those written in metropolitan languages about the colonized world are poorly served by text processing methods. Processes like named entity recognition (NER), which allow for the semi-automated annotation of transcribed archives, hold much promise for network and spatial analysis of the sources of the historical humanities, and yet state-of-the-art NER systems (e.g. NLTK) are generally trained on English-language corpora with a metropolitan, media focus and, thus, do not accurately identify non-English entities and labels. They do not offer much in terms of tag customization to move beyond western cultural notions of an entity, and NER is further complicated by inconsistent transliteration practices prevalent in Orientalist scholarship.

Indeed, initiatives exist which begin to tackle the inequities within the field of digital textual analysis. For example, the ongoing #NewNLP workshop (https://newnlp.princeton.edu/about/) is fostering the development of language resources for various world languages. We would like to underscore, however, that the inequities are still present for NLP approaches to English corpora historical in nature and originating from non-metropolitan environments. Our paper reports on collaborative work attempting to close this gap for collections in English, containing transliterated names coming from Arabic-speaking or adjacent Muslim cultures. We present an approach to extracting a variety of named entities (NE) from unstructured historical datasets from open digital collections dealing with a space of the informal British empire—the nineteenth- and early twentieth-century Persian Gulf region. The sources are largely concerned with people, places, tribes and transactions in the region, yet models in state-of-the-art NER systems function with limited sets of tags and they do not capture many of the entities of interest to the historian and do not perform well with entities transliterated from other languages. We build custom spaCy-based NER models trained on domain-specific annotated datasets: the correspondence ledgers of the British Colonial Residency in Bushire and the encyclopedic Gazetteer of the Persian Gulf, Central Arabia and Oman attributed to John Gordon Lorimer. We also extend the set of named entity labels provided by spaCy and focus on detecting entities of non-Western origin, particularly from Arabic and Farsi. We test and compare performance of the blank, pre-trained and merged spaCy-based models and suggest further improvements. Our study makes an intervention into thinking beyond Western cultural notions of the entity in digital historical research.