



Project Number: **770299**

**NewsEye:  
A Digital Investigator for Historical Newspapers**

Research and Innovation Action  
Call H2020-SC-CULT-COOP-2016-2017

**D6.10: Contextualized Case Studies for academic use (d) (final)**

Due date of deliverable: M36 (30 April 2021)

Actual submission date: 18 April 2021

**Start date of project:** 1 May 2018

**Duration:** 45 months

Partner organization name in charge of deliverable: UIBK-ICH

<b>Project co-funded by the European Commission within Horizon 2020</b>		
<b>Dissemination Level</b>		
PU	Public	PU
PP	Restricted to other programme participants (including the Commission Services)	-
RE	Restricted to a group specified by the Consortium (including the Commission Services)	-
CO	Confidential, only for members of the Consortium (including the Commission Services)	-

## Revision History

Document administrative information	
<b>Project acronym:</b>	NewsEye
<b>Project number:</b>	770299
<b>Deliverable number:</b>	D6.10
<b>Deliverable full title:</b>	Contextualized Case Studies for academic use (d) (final)
<b>Deliverable short title:</b>	Contextualized Case Studies (final)
<b>Document identifier:</b>	NewsEye-T62-D610-ContextualizedCaseStudies-d-Submitted-v3.0
<b>Lead partner short name:</b>	UIBK-ICH
<b>Report version:</b>	V3.0
<b>Report preparation date:</b>	18.04.2021
<b>Dissemination level:</b>	PU
<b>Nature:</b>	Report
<b>Lead author:</b>	Jani Marjanen (UH-DH), Sarah Oberbichler (UIBK-ICH)
<b>Co-authors:</b>	Eva Pfanzelter (UIBK-ICH), Martin Gasteiner (UNIVIE), Stefan Hechl (UIBK-ICH)
<b>Internal reviewers:</b>	Hannu Toivonen (UH-CS), Marie-Eve Thérenty (UPVM)
<b>Status:</b>	Draft
	Final
	x Submitted

The NewsEye Consortium partner responsible for this deliverable has addressed all comments received, making changes as necessary. Changes to this document are detailed in the change log table below.

## Change Log

Date	Version	Editor	Summary of changes made
23/02/2021	0.1	Jani Marjanen (UH-DH)	First outline draft
25/02/2021	0.2	Jani Marjanen (UH-DH) and Sarah Oberbichler (UIBK-ICH)	First version
26/02/2021	0.3	Eva Pfanzelter (UIBK-ICH)	Restructuring and comments
02/03/2021	1.0	Jani Marjanen (UH-DH) and Sarah Oberbichler (UIBK-ICH)	Shortening, editing, harmonization, and submission to internal reviewers
09/03/2021	1.1	Hannu Toivonen (UH-CS)	Internal review
10/03/2021	1.2	Marie-Eve Thérenty (UPVM)	Internal review
31/03/2021	1.3	Jani Marjanen (UH-DH)	Revision and editing
08/04/2021	1.4	Stefan Hechl (UIBK-ICH)	Proofreading
08/04/2021	2.0	Sarah Oberbichler (UIBK-ICH)	Final version sent for quality management
18/04/2021	3.0	Antoine Doucet (ULR)	Minor adjustments and submission

## Executive summary

Within the project ‘NewsEye: A Digital Investigator for Historical Newspapers’, researchers from computer science and digital humanities collaborate with the three national libraries of Finland, France and Austria in order to develop methods and tools for effective exploration and exploitation of digital newspaper collections. To make these rich resources of cultural heritage better accessible by means of new technologies and ‘big data’ approaches, ‘close’ and ‘distant reading’ methods of digital humanities are being investigated and combined. The aim is to improve the ways researchers and experts, as well as the interested general public, study European cultural heritage.

This final deliverable, ‘Contextualized Case Studies for academic use (d)’, provides an overview of studies published or to be published in humanities or digital humanities journals and conference proceedings, as of M36 (April 2021). Many of the studies draw from methods developed in WP3, WP4 and WP5, as they use named entities, topic modelling, word embeddings and custom scripts in conducting analyses, but they also utilise hermeneutic interpretations of textual sources as well as other traditional humanities’ methods, and reflect on the workflows connecting multidisciplinary cultural heritage projects. The interplay between traditional humanities’ methods and methods drawn from natural language processing is also seen as providing a link to interface development and dissemination in WP7.

In this report, we discuss studies that have clear humanities’ research questions and that have been developed within the NewsEye project. Each study is presented briefly and followed by a discussion of its implications for method development and collaboration between the work packages. Although the studies vary in scope, ambition and methods, one combined conclusion is that studies aiming for humanities relevance cannot rely on one method only, but usually involve shifting the perspective and using different methods to examine different aspects of the topic at hand.

## Contents

<b>Executive Summary</b>	<b>3</b>
<b>1 Introduction</b>	<b>5</b>
<b>2 Published journal articles</b>	<b>6</b>
2.1 A National Public Sphere? Analyzing the Language, Location, and Form of Newspapers in Finland, 1771–1917 . . . . .	6
2.2 Digital interfaces of historical newspapers: opportunities, restrictions and recommendations	7
2.3 The expansion of isms, 1820–1917: Data-driven analysis of political language in digitized newspaper collections . . . . .	7
2.4 La propagation de l’épidémie : Traitement médiatique de la grippe « espagnole » dans la presse quotidienne héraultaise . . . . .	8
<b>3 Published conference papers</b>	<b>9</b>
3.1 The Helsinki Digital Humanities Hackathon: Two Perspectives on Multidisciplinary Historical Newspapers Research in a Hackathon Context . . . . .	9
3.2 Can Umlauts Ruin Your Research in Digitized Newspaper Collections? A NewsEye Case Study on ‘The Dark Sides of War’ (1914–1918) . . . . .	9

<b>4</b>	<b>Published academic texts for the general public</b>	<b>10</b>
4.1	Doing Historical Research With Digital Newspapers – Perspectives of DH Scholars . . . . .	10
4.2	Als eine andere Epidemie die Welt in Atem hielt: Die Spanischen Grippe 1918/19 in der österreichischen Presse . . . . .	11
<b>5</b>	<b>Scientific degree theses</b>	<b>11</b>
5.1	‘Wir dürfen wieder Österreicher sein!’ Die Rolle der Tagespresse in österreichischen Nation-Building-Prozessen 1945–1948 – eine quantitative Analyse ausgewählter digitaler Zeitungskorpora samt Vorschlägen zur didaktischen Umsetzung . . . . .	11
<b>6</b>	<b>Accepted journal articles</b>	<b>12</b>
6.1	A data-driven approach to studying changing vocabularies in historical newspaper collections . . . . .	12
6.2	Tracing Discourses in Digital Newspaper Collections: A Contribution to Digital Hermeneutics while Investigating ‘Return Migration’ in Historical Press Coverage . . . . .	13
6.3	Crossing or Intersecting the Emperor’s Desk with digitized Newspaper Data: Entity-source-networks in the late Habsburg Empire . . . . .	13
6.4	National Sentiment: Nation Building and Emotional Language in Nineteenth Century Finland . . . . .	14
6.5	Grippe-18/Covid-19: penser le monde d’hier et de demain, au prisme de la presse du XXe siècle . . . . .	14
6.6	From ‘German-speaking minority’ to ‘Austrian minority’ – South Tyrol in Austrian newspapers post 1945 . . . . .	15
<b>7</b>	<b>Accepted conference papers</b>	<b>16</b>
7.1	Topic Modelling Discourse Dynamics in Historical Newspapers . . . . .	16
<b>8</b>	<b>Journal articles and conference papers under review</b>	<b>16</b>
8.1	Les femmes journalistes dans L’Œuvre. Méthodologie d’une visibilisation grâce aux outils numériques . . . . .	16
8.2	International: From Legal to Civic Discourse and beyond in the Nineteenth Century . . . . .	17
8.3	Writing the Nation 1945–1948: The Austrian Daily Newspapers’ Contribution to Nation-Building . . . . .	17
8.4	Sound Impression – Opacity of the Voice – Construction of premises. Traces of radio in newspapers and vice versa . . . . .	18
8.5	Topic-specific corpus building: The creation of a representative newspaper corpus on the topic of return migration using LDA and JSD . . . . .	19
8.6	Integrated Interdisciplinary Workflows for Research on Historical Newspapers – Perspectives from Humanities Scholars, Computer Scientists and Librarians . . . . .	19
<b>9</b>	<b>Discussion</b>	<b>20</b>

## 1 Introduction

The studies conducted by the digital humanities scholars in WP6 are designed to contribute to discussions in historiography about change and continuity in European societies. On top of that, they provide feedback to the other work packages, so that the colleagues from the computer sciences can produce tools and methods better equipped for the study of newspapers as rich sources for exploring European heritage. Some of the studies use the NewsEye platform developed as part of WP7 (described in public Deliverables D7.8 and D6.9) in close collaboration with WP3 (semantic text enrichment), WP4 (dynamic text analysis) and WP5 (personal research assistant). Other studies are conducted in collaboration with researchers from the universities of Helsinki, La Rochelle, Innsbruck and Rostock, using custom tools for particular research questions, and yet others are studies that the humanities researchers have produced based on more traditional methods of qualitative historiography, such as discourse analysis or content analysis. This report gives an overview of these studies and briefly describes how they contribute to the method development in other work packages.

The studies are presented according to their current status, ranging from published texts to texts that have been accepted for publication and submitted texts that are still under review. Finally, we present how the texts are used in other forums, and discuss a few plans for the immediate future. Whenever the individual studies are available, a link is provided.

The publications are related to the four humanities cases studies on return migration, nationalism and revolutions, gender, and journalism. Apart from scholarly publications, which are the focus in this report, the case studies have also produced blog posts, teaching materials and other materials for public use, which are discussed in public Deliverables D6.11, D6.12, and D6.13. The case studies are described in detail on the project website and leaflets. Therefore we only give a short summary and information about methods used within these case studies:

- The case study on *migration* aims to examine language patterns and discourses on different migration movements, especially return migration movements. The case study uses digital methods such as topic modelling, word embeddings and collection building functions within the demonstrator to create and prepare subcorpora for further qualitative analysis (discourse and argumentation analysis).
- The case study on *nationalism* focuses on the changing language of nationhood in the late nineteenth and early twentieth century. It addresses the following questions: How did discourse on nationhood expand in the nineteenth century? How and why did the notion of nationalism become a central figure of thought in the early twentieth century? What role did patriotism play during the World War period? The case study combines qualitative reading of sources with frequency analysis and word embedding models to cluster key terms.
- The case study on *gender* aims to bring to light the evolution of the representation of women in the news from 1850 to 1950. The aim is to examine the feminine condition (the right to vote, the right to wear pants), as well as the female figures who infiltrated the journalistic space: female writers/journalists. What place is given to them in the newspapers? What role do newspapers play in the struggles of women? Digital tools (notably topic modelling and frequency tools) as well as named entities are very valuable in this case study for revealing the trends and key figures.
- The case study on *journalism* takes a closer look at the emerging profession of the journalist, their

practices, and the wider implications for society. The case study approaches two different areas: the relationship between the journalists and their audiences and the use of media, including their social function and impact, as well as their spheres of influence. The case study relies on a mix of close reading, frequency analysis and network analysis.

## 2 Published journal articles

### 2.1 A National Public Sphere? Analyzing the Language, Location, and Form of Newspapers in Finland, 1771–1917

Jani Marjanen, Ville Vaara, Antti Kanner, Hege Roivainen, Eetu Mäkelä, Leo Lahti, and Mikko Tolonen. “A National Public Sphere? Analyzing the Language, Location, and Form of Newspapers in Finland, 1771–1917”. In: *Journal of European Periodical Studies* 4.1 (June 2019), pp. 54–77. ISSN: 2506-6587. DOI: [10.21825/jeps.v4i1.10483](https://doi.org/10.21825/jeps.v4i1.10483)

The article is authored by Jani Marjanen and Mikko Tolonen from the UH-DH team along with Ville Vaara, Antti Kanner, Hege Roivainen, Eetu Mäkelä, and Leo Lahti as collaborators from the Helsinki Computational History Group. The article is published in *Journal of European Periodical Studies* and is available online as an open access publication.

**Abstract** During the period 1771–1917, newspapers developed as a mass medium in the Grand Duchy of Finland within two imperial configurations (Sweden until 1809 and Russia in 1809–1917), and in the two main languages — Swedish and Finnish. Finland is an ideal starting point for conducting comparative studies in that its bilingual profile already includes two linguistically separated public spheres that nonetheless were heavily connected. We use metadata and full text to analyse publication places, language, number of issues, number of words, newspaper size, and publishers, which we compare with existing scholarship on newspaper history and censorship, and thereby offer a more robust statistical analysis of newspaper publishing in Finland than has previously been possible. We specifically examine the interplay between the Swedish- and Finnish-language newspapers and show that, whereas the public discussions were inherently bilingual, the technological and journalistic developments advanced at a different pace in the two language forums. This analysis challenges the perception of a uniform public sphere in the country. In addition, we assess the development of the press in comparison with the production of books and periodicals, which points toward the specialization of newspapers as a medium in the period after 1860. This confirms some earlier findings about Finnish print production. We then show how this specialization came about through the establishment of forums for local debates that other less localised print media such as magazines and books could not provide.

**NewsEye context and implications** As the paper provides a birds-eye view of the transformation of Finnish public discourse in the nineteenth century, it is a necessary step to be done before other studies use the same data set. We envision similar metadata analyses using descriptive statistics for other countries with good metadata records of newspaper publishing. The paper highlights the need to better understand the language dynamics within a multilingual unit, as the different languages are by default separated into different corpora, but the relationship between the languages actually changed over time. It also provides background information for methods developed in WP3, WP4, and WP5, as they all rely on a subset of the full data and need to be able to assess the representativity of their data in different time periods.

## 2.2 Digital interfaces of historical newspapers: opportunities, restrictions and recommendations

Eva Pfanzelter, Sarah Oberbichler, Jani Marjanen, Pierre-Carl Langlais, and Stefan Hechl. “Digital interfaces of historical newspapers: opportunities, restrictions and recommendations”. In: *Journal of Data Mining and Digital Humanities* HistInformatics (Jan. 2021). Publisher: Episciences.org. URL: <https://hal.archives-ouvertes.fr/hal-02480654>

The article is authored by Eva Pfanzelter, Sarah Oberbichler, and Stefan Hechl from the UIBK-ICH team, Pierre-Carl Langlais from the UPVM team, as well as Jani Marjanen from the UH-DH team. It is published in *Journal of Data Mining and Digital Humanities* and is available online as an open access publication.

**Abstract** Many libraries offer free access to digitised, historical newspapers via user interfaces. After an initial period of search and filter options as the only features, the availability of more advanced tools and users’ desire for more options ushers in a period of interface improvement. However, this raises a number of open questions and challenges. While this paper does not provide definite answers to these questions, it gives an insight into the difficulties, challenges and risks of using interfaces to investigate historical newspapers. More importantly, it gives ideas and recommendations for the improvement of user interfaces and digital tools.

**NewsEye context and implications** The paper takes a closer look at available interfaces of digitised newspapers, as well as digital tools that will be part of such interfaces more and more in the future. The interfaces used were ANNO (ONB – Austrian National Library) and to a lesser extent ONB Labs, Digi and Korp (National Library of Finland), as well as Gallica and Retronews (BNF – National Library of France). It is a collaborative work of all DH teams in the NewsEye project and is a very shortened and revised version of the deliverable D6.2, which focused on the evaluation of user surveys conducted together with the national libraries involved in the project. Additionally, it looked at the testing of digital tools and interfaces. Researchers of the DH teams in Austria, Finland, and France experimented extensively with the existing newspaper search interfaces of the three national libraries, using the case studies they had already worked on for the report in M6. The conclusions and suggestions for improvement in this deliverable were used to further develop the demonstrator and methods from WP3 (semantic text enrichment), WP4 (dynamic text analysis) and WP5 (personal research assistant).

## 2.3 The expansion of isms, 1820–1917: Data-driven analysis of political language in digitized newspaper collections

Jani Marjanen, Jussi Kurunmäki, Lidia Pivovarova, and Elaine Zosa. “The expansion of isms, 1820–1917: Data-driven analysis of political language in digitized newspaper collections”. In: *Journal of Data Mining and Digital Humanities* HistInformatics (Dec. 2020). URL: <https://jdm.dh.episciences.org/6728>

The article is authored by Jani Marjanen from the UH-DH team, Elaine Zosa and Lidia Pivovarova from the UH-CS team, and Jussi Kurunmäki from Tampere University. It is published in *Journal of Data Mining and Digital Humanities* and is available online as an open access publication.

**Abstract** Words with the suffix -ism are reductionist terms that help us navigate complex social issues by using a simple one-word label for them. On the one hand, they are often associated with political

ideologies, but on the other hand, they are present in many other domains of language, especially culture, science, and religion. This has not always been the case. This paper studies isms in a historical record of digitised newspapers published in Finland from 1820 to 1917 to find out how the language of isms developed historically. We use diachronic word embeddings and affinity propagation clustering to trace how new isms entered the lexicon and how they relate to one another over time. We are able to show how they became more common and entered more and more domains. Still, the uses of isms as traditions for political action and thinking stand out in our analysis.

**NewsEye context and implications** The article provides a historical description of the expansion of isms as well as some methodological tools for similar cases. The methodological development in this case was drawn from the study “A data-driven approach to studying changing vocabularies in historical newspaper collections” (see Section 6.1). Also in this case, we were able to develop a multilingual analysis by comparing Finnish-language and Swedish-language data. Here, the ism itself worked as a node for translation. We further visualised the links between different timeslices using Sankey charts. Word embeddings are used extensively in WP5, so the study provides a useful case based on historical messy data. The visualizations with Sankey charts are possibly used in WP7 in the demonstrator. From a historical point of view, the case discusses themes that have been studied extensively with qualitative methods. It now provides a robust overview of the expansion of isms in political language, and thus provides a novel perspective in the study of past political discourse. The study is a collaboration between WP4, WP5, and WP6, with three out of four participants in Helsinki.

## 2.4 La propagation de l'épidémie : Traitement médiatique de la grippe « espagnole » dans la presse quotidienne héraultaise

Nejma Omari. “La propagation de l'épidémie: Traitement médiatique de la grippe espagnole dans la presse quotidienne héraultaise”. Français. In: *Études Héraultaises* (2020), p. 13. URL: <https://www.etudesheraultaises.fr/publi/la-propagation-de-lepidemie-traitement-mediastique-de-la-grippe-espagnole-dans-la-presse-quotidienne-heraultaise/>

The article is authored by Nejma Omari from the UPVM team. It is published in *Revue Études Héraultaises* and is available online.

**Abstract** In these troubled times, the world must face what political discourse has tended to call the ‘worst health crisis in a century’, referring to the so-called Spanish Flu of 1918. While it is not possible to confront influenza and COVID-19 from a cyclical point of view – the hygienic measures and the context of the Great War cannot be compared with our era – the analogy between these two pandemics can prove noteworthy when it comes to examining the management and media handling of the two crises (prophylactic measures, unpreparedness, communication and misinformation). The twentieth century press is a valuable source of information. Regional newspapers, in particular, allow us to discover how the inhabitants of Hérault lived through the epidemic. This article aims to trace the local history of the Spanish Flu as seen in the Montpellier daily newspapers ‘L’Éclair’ and ‘Le Petit Méridional’.

**NewsEye context and implications** The article illuminates the crises of the contemporary era by studying past experiences available to us thanks to the digitised historical press. This article follows from a journalistic text written together with Antoine Doucet for *The Conversation*<sup>1</sup> as well as several blog posts

<sup>1</sup> <https://theconversation.com/covid-19-et-grippe-espagnole-quand-la-presse-du-xx-siecle-rappelle-celle-de-2020-137035>



published on the NewsEye and Gallica websites. This article aims to complete the approach discussed in the context of the project by extending it to the regional press.

### 3 Published conference papers

#### 3.1 The Helsinki Digital Humanities Hackathon: Two Perspectives on Multidisciplinary Historical Newspapers Research in a Hackathon Context

Ruben Ros and Sarah Oberbichler. “The Helsinki Digital Humanities Hackathon: Two Perspectives on Multidisciplinary Historical Newspapers Research in a Hackathon Context”. In: *Digital Humanities in the Nordic Countries: Proceedings of the Digital Humanities in the Nordic Countries 5th Conference*. Vol. 2717. CEUR Workshop Proceedings. Riga, Latvia, 2020, pp. 66–74. URL: <http://ceur-ws.org/Vol-2717/paper07.pdf>

The paper is authored by Ruben Ros from the University of Utrecht (not a NewsEye partner) and Sarah Oberbichler from the UIBK-ICH team. It is a short reflection on the Helsinki Digital Humanities Hackathon in which several NewsEye members (CS and DH) were involved. This paper was presented at the DHN 2020 conference in Riga (online due to Covid-19). The paper was published in the conference proceedings and is available online as an open access publication.

**Abstract** This paper describes the 2019 edition of the Helsinki Digital Humanities Hackathon from the perspective of two of its participants. As (digital) historians, they were part of the group that investigated the history of medical advertisements in nineteenth-century British newspapers. The paper describes the research process, as well as the data and methods used during the research. The paper also considers the Hackathon as a laboratory for digital humanities research and reflects on the nature of the collaboration as experienced during this event. As such, the paper describes the challenges of multidisciplinary research and identifies the factors that hinder and foster collaboration in a digital humanities context.

**NewsEye context and implications** The aim of this paper is to underline the importance of collaborative and multidisciplinary work in the digital humanities and to reflect on experiences members made during the Digital Humanities Hackathon 2019. In addition, it points out the strengths and weaknesses of DH Hackathons. This paper is a collaborative work in the digital humanities and is thus relevant for further studies conducted in WP6. There are also further implications for method development (WP3-5) and dissemination (WP7) in NewsEye, as a separate NewsEye group is planned at the next Digital Humanities Hackathon (initially scheduled for May-June 2020, rescheduled as an online event in May 2021 due to the Covid-19 crisis). This group will specifically use named entities produced in WP3 and develop a comparative research case together with digital humanities students.

#### 3.2 Can Umlauts Ruin Your Research in Digitized Newspaper Collections? A NewsEye Case Study on ‘The Dark Sides of War’ (1914–1918)

Barbara Klaus. “Can Umlauts Ruin Your Research in Digitized Newspaper Collections? A NewsEye Case Study on ‘The Dark Sides of War’ (1914–1918)”. In: *Digital Humanities in the Nordic Countries: Proceedings of the Digital Humanities in the Nordic Countries 5th Conference*. Vol. 2612. CEUR Workshop Proceedings. Riga, Latvia, 2020, pp. 267–274. URL: <http://ceur-ws.org/Vol-2612/short12.pdf>

The paper is authored by Barbara Klaus from the UIBK-ICH team. It was presented at the DHN 2020 conference in Riga (online due to Covid-19) and is available as open access in the online conference proceedings.

**Abstract** Digitised newspaper collections offer several useful functions facilitating research. Still, the (automatic) research in these collections is limited. Based on the research conducted on the platform AustriaN Newspapers Online (ANNO) for the NewsEye case study ‘The dark sides of war’, the main challenges of working with digitised newspaper collections will be discussed in this paper. Especially two aspects – the fire catastrophe at the munitions factory Wöllersdorf (1918/09/18) in Lower Austria, and the Austrian press coverage of war widows during the First World War – will be used as specific examples. The discussed limitations include optical character recognition (OCR) quality, provided search functionalities and metadata, and others. Furthermore, possible improvements regarding these challenges, e.g. optical layout recognition (OLR), named-entity recognition (NER) and named-entity linking (NEL), will be presented in this paper.

**NewsEye context and implication** This paper seeks to reflect on the impact of optical character recognition (OCR) problems, missing search options and metadata using the case study ‘The dark sides of war’ and the ANNO interface for digital newspapers. It discusses necessary improvements concerning OCR, OLR, NER and NEL. This is important feedback for WP3, which develops text enrichment methods such as named entities recognition and linking, as well as WP2, which develops automatic text recognition (ATR).

## 4 Published academic texts for the general public

### 4.1 Doing Historical Research With Digital Newspapers – Perspectives of DH Scholars

Sarah Oberbichler, Eva Pfanzelter, Stefan Hechl, and J. Marjanen. “Doing historical research with digital newspapers - perspectives of DH scholars”. In: *EuropeanaTech Insight* 16: Newspapers (2020). URL: <https://pro.europeana.eu/page/issue-11-generous-interfaces>

**Abstract** The text discusses how historical scholarship has changed with the advent of digitised historical newspapers and pays particular attention to the advantages and new possibilities, but also some of the pitfalls that come with using with digitised sources. It dedicates sections to source preparation, cleaning and structuring of data, evaluating linguistic patterns in data, and the qualitative analysis of newspaper discourses. All of these sections highlight the opportunities, but also the difficulties, of integrating digital tools into humanities research. The dissemination of NewsEye results and findings (publications, notebooks, screencasts) also takes up a significant part of this paper.

**NewsEye context and implication** The text was solicited by EuropeanaTech editors as a part of a special issue that addresses the use and advantages, but also limitations, of digitised newspapers. As such, it presents work done in the NewsEye project in comparison with related projects and is reaching out to wider interested readership. EuropeanaTech is a multimedia online publication with a strong focus on visual inputs (including videos), which allowed us to share images as well as one of our NewsEye screencasts. The text in particular raises the question of collaboration between domain experts (historians) and computer scientists. In doing so, it especially targets what historians need to know when working with digitised historical newspapers.

## 4.2 Als eine andere Epidemie die Welt in Atem hielt: Die Spanischen Grippe 1918/19 in der österreichischen Presse

Sarah Oberbichler, Stefan Patrick Hechl, and Eva Pfanzerter. "Als eine andere Epidemie die Welt in Atem hielt: Die Spanische Grippe 1918/19 in der österreichischen Presse". In: *Tiroler Chronist - Fachblatt von und für Chronisten in Nord-, Süd- und Osttirol* 154 (2020), pp. 15–22

This paper is authored by Sarah Oberbichler, Stefan Hechl, Eva Pfanzerter from the UIBK-ICH team. It was originally published as NewsEye blog post<sup>2</sup>.

**Abstract** Starting from a highly topical theme, the Covid-19 crisis, this paper looks at the newspaper coverage of the Spanish Flu in comparison with the media reception of the current Corona pandemic. Using search functions and tools available in both the NewsEye platform and the ANNO interface, micro and macro analyses were conducted in order to find and compare recurring linguistic patterns, similar discourses, and additions. The paper also gives insights into search processes and methods and contributes to source and query criticism.

**NewsEye context and implication** This paper seeks to reach a broader audience by adding historical newspaper discourses to the contemporary media discourse on the Corona pandemic. It aims to reflect recurring linguistic patterns in times of comparable crisis (Spanish Flu and Covid-19), while reflecting the role of newspapers and linking to the NewsEye tools and platforms. It proves the usability of the developed tools by serving as best practice example and examines which further tools would be useful for humanities researchers. This is important feedback for WP3 and WP4, which are dealing with text enrichment and analysis tools.

## 5 Scientific degree theses

### 5.1 'Wir dürfen wieder Österreicher sein!' Die Rolle der Tagespresse in österreichischen Nation-Building-Prozessen 1945–1948 – eine quantitative Analyse ausgewählter digitaler Zeitungskorpora samt Vorschlägen zur didaktischen Umsetzung

Stefan Patrick Hechl. "'Wir dürfen wieder Österreicher sein!' Die Rolle der Tagespresse in österreichischen Nation-Building-Prozessen 1945–1948 – eine quantitative Analyse ausgewählter digitaler Zeitungskorpora samt Vorschlägen zur didaktischen Umsetzung". Master thesis. Jan. 26, 2021. DOI: [10.5281/zenodo.4468295](https://zenodo.org/record/4468295). URL: <https://zenodo.org/record/4468295#.YGXsb68zZPY>

This master thesis was written by Stefan Hechl and supervised by Eva Pfanzerter (both members of the UIBK-ICH team) at the University of Innsbruck (Institute of Contemporary History).

**Abstract** While the topic of Austrian national identity is well-researched, most studies focus on the years after the State Treaty in 1955 as well as the 1970s and 1980s. This thesis assesses processes of nation-building and the formation of Austrian national identity by using digital methods to analyse a large corpus of Austrian daily newspapers from 1945 to 1948, a period largely neglected by existing research. The research shows that while discourses on Austrian identity were not as explicit in the

---

<sup>2</sup> <https://www.newseye.eu/blog/news/epidemie-spanische-grippe/>

years 1945 - 1948 as they became after 1955, some important groundwork for the formation of national identity was laid in the immediate aftermath of World War Two. Especially the year 1945 is notable in this regard, while the following years gradually saw a return to a certain post-war daily routine with less discussion on topics connected to national identity. However, frequency analyses show that this topic closely correlates to various events related to negotiations regarding the Austrian state treaty, and that in general, the formation of Austrian national identity was strongly connected to external factors within the context of the beginning Cold War. The dimensions of the past and the future also played an important role in the discourses surrounding the early formation of post-war Austrian identity, as can be quantitatively deduced from the corpus. In this regard, the analysis was able to quantitatively prove some hypotheses previously formulated by scholars without access to the corpus, and to discover previously unknown quantitative textual patterns regarding the topic of Austrian national identity.

**NewsEye context and implications** The thesis focuses on distant reading methods such as diachronic frequency or word correspondence analysis to find (hidden) language patterns during the time of the re-building of Austrian national identity post-1945. The aim of this contribution is to prove existing hypotheses, as well as the discovery of previously unknown textual patterns and discourses.

## 6 Accepted journal articles

### 6.1 A data-driven approach to studying changing vocabularies in historical newspaper collections

Authors: Simon Hengchen, Jani Marjanen and Mikko Tolonen from the UH-DH team, and Ruben Ros from the University of Utrecht.

The article is accepted for publication in *Digital Scholarship in the Humanities* and should be published in the spring 2021.

**Abstract** Nation and nationhood are some of the most-studied concepts in intellectual history. At the same time, 'nation' and its historical usage are vague. The article develops a data-driven method using dependency parsing and neural word embeddings to help analyse some of the vagueness in the evolution of this concept. To do so, we create a large set of words pertaining to the topic of nation. Then, we train diachronic word embeddings and use them to quantify the strength of the semantic similarity between these words and create meaningful clusters, which are then aligned diachronically. To illustrate the robustness of the study across languages, time spans, as well as large datasets, we apply it to the entirety of five historical newspapers archives in Dutch, Swedish, Finnish, and English.

**NewsEye context and implication** The article studies the changing vocabulary of nationhood by developing data-driven and multilingual methods based on word embeddings to cluster the vocabulary relating to the adjective 'national'. Methods developed for this study have been further developed in WP4 and WP5, and have also been utilised in the article 'The expansion of isms' and 'International'. Most importantly, the case develops methods for multilingual analysis as the experience of the national became prominent everywhere in Europe at this time (but was by no means universal), and the focus on key adjectives as central nodes allows for a comparison across languages and nations.

## 6.2 Tracing Discourses in Digital Newspaper Collections: A Contribution to Digital Hermeneutics while Investigating ‘Return Migration’ in Historical Press Coverage

Authors: Sarah Oberbichler and Eva Pfanzelter (UIBK-ICH)

The article is accepted for publication in the volume *Digitised Newspapers – A New Eldorado for Historians? Tools, Methodology, Epistemology, and the Changing Practices of Writing History in the Context of Historical Newspapers Mass Digitization* and should be published by De Gruyter in the spring 2021.

**Abstract** Based on a case study on return migration – defined as the movement of emigrants, refugees, prisoners of war, and others back to their country of origin – this paper deals with a historical-critical approach to digital newspaper collections and interfaces. Using discourse-driven historical research questions to examine return migration, it highlights the usability and relevance of digital source criticism (OCR issues, digitization processes, the selection of newspaper titles, etc.), digital query criticism (selection of keywords or parts of keywords, consequences of length, polysemy, etc.), and interface criticism (possibilities to search, view, select, collect, visualise, contextualization efforts, metadata, etc.). We argue that due to the complexity of language coupled with non-transparent functionalities of newspaper interfaces, distorted, misunderstood, or misinterpreted results can occur. Some inherent biases in the digital collections themselves, as well as missing metadata and context can lead to distortions that seem trivial and banal only at first glance.

**NewsEye context and implication** Using a specific example – the topic of the return of emigrants to their place of origin between 1850 and 1950 – this paper contributes to the field of digital hermeneutics by addressing the importance of source, interface and tool criticism. It focuses on the newspaper interface (ANNO) of the Austrian National Library, reflects on the current status of this interface, and gives ideas of improvement with the help of digital methods. The conclusion and suggestions for improvement given in this paper, such as contextualised keyword searches, relative frequency graphs or greater transparency, should serve the further development of the demonstrator and methods developed in WP3 (text enrichment), WP4 (dynamic text analysis) and WP5 (personal research assistant). To convey the outcome of this study, results were presented in joint NewsEye meetings and workshops.

## 6.3 Crossing or Intersecting the Emperor’s Desk with digitized Newspaper Data: Entity-source-networks in the late Habsburg Empire

The article is accepted for publication in the edited volume *Digitised Newspapers – A New Eldorado for Historians? Tools, Methodology, Epistemology, and the Changing Practices of Writing History in the Context of Historical Newspapers Mass Digitization* and is expected to be published by De Gruyter in the spring 2021.

Author: Andreas Enderlin-Mahr, Martin Gasteiner (UNIVIE), and Janos Bekesi

**Abstract** The aim of the study is a reconstruction of the primary level of a network of governance in the decision-making process of the Habsburg Emperor Francis Joseph I. The network is produced on the basis of the RDF platform, network and entity data that are derived from NER in historical newspapers. Further enrichment is based on available digitised newspapers. The resulting platform allows for the addressing of issues regarding the relationship of data and entities, and shows the primarily domestic

focus of the Emperor's decisions. The findings of the study are visualised through network graphs, topological maps, and diagrams. The broader implications of the study are related to the interconnectedness of newspapers and government files as sources.

**NewsEye context and implication** The study particularly addresses journalistic practices and investigates newspapers' relationship to official state documents, thus shedding light on newspapers as source material. Based on newspaper data, the text shows how this source genre can be prototypically linked with other sources. The focus of this investigation is on the topics of media and the development of journalism in connection with legislation, and is thus thematically very closely linked to the media and journalism case study. In order to create the platform, the project is based on developments from the NewsEye project: the developed entity recognition model as well as the topic modelling module are used to explore networks and related topics.

## 6.4 National Sentiment: Nation Building and Emotional Language in Nineteenth Century Finland

Author: Jani Marjanen (UH-DH)

The article is accepted for publication in a volume titled *Lived Nation* edited by Ville Kivimäki, Sami Suonenjoki and Tanja Vahtikari, and is expected to be published by Palgrave Macmillan in the spring 2021.

**Abstract** During the course of the late eighteenth century and the early nineteenth century, the term 'national sentiment' was coined and established in several European languages. The emergence of the term in several different languages at roughly the same time is indicative of changes both in the experiences of nationhood and of emotion. This chapter explores the development of the term 'national sentiment' in Finnish public discourse, and argues that it was transformed during the course of the nineteenth century. Early in the century, it denoted an individualistic feeling that romantic intellectuals hoped people would turn to, whereas it later became a description of a collective emotion. It was used to describe the atmosphere among one of the nationalities in Finland in particular, or the Russian empire in general. In this process, the term became more restrictive and lost its links to performing emotions relating to the nation.

**NewsEye context and implications** The article pertains to the case study relating to nationalism. It uses the dataset provided by the National Library of Finland, and builds on close readings of newspaper material as well as frequency analyses of word use. The latter is an application of methodological principles discussed further in the article "Digital interfaces of historical newspapers" (see Section 2.2).

## 6.5 Grippe-18/Covid-19: penser le monde d'hier et de demain, au prisme de la presse du XXe siècle

Author: Nejma Omari (UPVM)

This article has been accepted for publication in a volume with the working title *Penser la Covid et penser le monde Réflexion critique sur les effets de la pandémie du printemps 2020* with Éditions Peisag.

**Abstract** Political discourse on the current pandemic recurringly refers to the so-called Spanish Flu of

1918 either directly or indirectly by talking about ‘the most serious health crisis in a century’. The analogy between these two pandemics is somewhat problematic, as contextual differences are enormous, but examining their political management and media coverage is useful. For this approach, the digital press of the twentieth century is a valuable source of information. This article studies the many similarities that are identifiable in the newspaper data: prophylactic measures, lack of preparation and excesses, communication and misinformation. Named entities, frequency analysis and distant search allow us to spot trends and debates about the ‘world after the pandemic’, which still resonate today, and can help us understand the crisis we are experiencing.

**NewsEye context and implications** This article, which examines the media treatment of the Spanish Flu epidemic, extends the reflections carried out within WP6. The many blog posts written about face mask, miracle cures, or curfew measure have enabled us to bring about an overall reflection on the usefulness of the digital press to better understand the current health crisis. The various digital tools used to carry out this article are also based on the experimental tools developed by NewsEye: the search by named entities, but also the searches for occurrences carried out on the platform were used to obtain results which were very useful for the realization of this article. This approach was supplemented by a more ‘traditional’ search on the BnF search engines.

## 6.6 From ‘German-speaking minority’ to ‘Austrian minority’ – South Tyrol in Austrian newspapers post 1945

Author: Sarah Oberbichler (UIBK-ICH)

This paper was presented at the German Studies Association Conference 2020<sup>3</sup> and has been accepted to be published in the peer-reviewed edited volume *Contemporary Austrian Studies*<sup>4</sup>, vol. 31 (end 2021/beginning 2022).

**Abstract** While Austria sought to reshape an ‘Austrian’ identity after the 1938-1945 period, the relationship of the German-speaking minority in South Tyrol to Austria and its continued existence also became an important point of discussion. Identity creation and the discourse about the German-speaking minority seem interwoven with how political parties spoke about the ‘lost’ minority in the South. For instance, some politicians, parties, interest groups, authorities, and the media in Austria began to use the term ‘Austrian minority’, thus replacing the previously common ‘German-speaking minority’ after 1945. The same applies to the rejection and critique of Germany’s interference in the so-called South Tyrolean question. The paper deals with the question of which discourses accompanied the South Tyrolean question in Austrian newspapers after 1945, and how the process of nation building framed these discourses. The aim is to examine and analyse these discourses in newspapers, using the online newspaper platform ANNO and methods from the interdisciplinary H2020 project ‘NewsEye: A Digital Investigator for Historical Newspapers’.

**NewsEye context and implications** The aim of this paper is to analyse the change of language and perception towards South Tyrol after 1945 using a combination of computer-aided methods (advanced frequency search and keyword suggestions) and manual analysis (content analysis) to examine the change of language and perception of South Tyrol after 1945. This paper uses methods from the NewsEye project (especially bigrams and step detection developed by WP4 at the University of Helsinki) in combination with qualitative analyses and close reading to examine and analyse discourses regarding

---

<sup>3</sup> <https://www.thegsa.org/conference/current-conference>

<sup>4</sup> <https://www.centeraustria.org/cas>

German-speaking minorities in South Tyrol after 1945. Bigrams clearly indicate that pairs of words that often occur together deliver meaningful results, as many expressed concepts are phrases and not just single words.

## 7 Accepted conference papers

### 7.1 Topic Modelling Discourse Dynamics in Historical Newspapers

Authors: Jani Marjanen, Mikko Tolonen and Simon Hengchen from UH-DH, and Elaine Zosa and Lidia Pivovarova from UH-CS.

This paper has been accepted for the DHN 2020 post-proceedings publication and will be published in the spring 2021.

**Abstract** This paper addresses methodological issues in diachronic data analysis for historical research. We apply two families of topic models (LDA and DTM) on a relatively large set of historical newspapers, with the aim of capturing and understanding discourse dynamics. Our case study focuses on newspapers and periodicals published in Finland between 1854 and 1917, but our method can easily be transposed to any diachronic data. Our main contributions are a) a combined sampling, training and inference procedure for applying topic models to huge and imbalanced diachronic text collections; b) a discussion on the differences between two topic models for this type of data; c) quantifying topic prominence for a period and thus a generalization of document-wise topic assignment to a discourse level; and d) a discussion of the role of humanistic interpretation with regard to analysing discourse dynamics through topic models.

**NewsEye context and implications** The study tests topic modelling methods developed in WP4 on a historical case. We used state-of-the-art methods in topic modelling with messy historical data. By testing both DTM and LDA on the corpus, the study provides a discussion on what topic models capture and how their interpretation can be used for assessing patterns in historical newspapers. We further experiment with different thresholds for topic prominence to test our hypothesis about disappearing discourses. The study was a collaboration between WP4, WP5, and WP6.

## 8 Journal articles and conference papers under review

### 8.1 Les femmes journalistes dans L'Œuvre. Méthodologie d'une visibilité grâce aux outils numériques

Author: Marie-Ève Thérénty (UPVM) submitted a paper for the journal *Clio*.<sup>5</sup>

**Abstract** A previous book (Marie-Eve Thérénty, *Femmes de presse, femmes de lettres. De Delphine de Girardin à Florence Aubenas*, CNRS éditions, 2019) shed light on the very limited participation of women in the daily newspapers of the early twentieth century. In this context, it seems that the newspaper *L'Œuvre* founded in 1904 by Gustave Téry functions both as a good ground to shed light on the various forms of women's participation in the newspaper, and also as an exception. The article will first recall the major forms of discrimination against women in the press and the articles that are more particularly

---

<sup>5</sup> <https://journals.openedition.org/cliio/>



reserved for them. Then, the article will study, thanks to the extremely fine reading allowed by digital tools, some forms of invisibility of women in the newspaper, perhaps more numerous than is generally believed, and sporadic, intermittent and poorly professionalised practices that may appear, for example, on the cinema page. Finally, the article will show that *L'Œuvre* constitutes an exceptional corpus, firstly by the proportionally very large number of women it employs, and secondly by its ability to attract the most famous among them. The article will focus on the innovations that the newspaper introduced. We will then try to explain the importance of *L'Œuvre* in the emergence of women's journalism.

**NewsEye context and implication** This paper explores the role of women in the editorial staff of the newspaper *L'Œuvre*, digitised and processed by the NewsEye project. It combines digital tools (keywords, named entities, frequencies) provided by WPs 2-5 within the NewsEye project, newspapers from Retronews and Gallica, and the study of archives (*BNF, Archives nationales*).

## 8.2 International: From Legal to Civic Discourse and beyond in the Nineteenth Century

Authors: Jani Marjanen from the UH-DH team, and Ruben Ros from the University of Luxemburg (previously University of Utrecht).

The article is part of an edited volume titled *Debating Internationalisms* edited by Pasi Ihalainen and Antero Holmila, and is currently under review with an academic publisher.

**Abstract** The chapter studies the conceptual history of 'international' from the first occurrence of the word in the late eighteenth century to its widespread use in the late nineteenth century. It shows how the concept was initially used in legal discourse, as a way to refer to the interaction between nation states. Qualitative and quantitative methods are used to show how in the mid-nineteenth century, the vocabulary of 'international' expanded to new domains of life. In conjunction with a national imaginary, an international one formed, primarily in the form of conferences, exhibitions and organizations that were explicitly branded as such. By the 1860s, the term gained an extended lexical application, which is demonstrated both qualitatively through examples as well as quantitatively by using word embeddings and affinity propagation clustering.

**NewsEye context and implications** The article traces the history of conceptualising things as 'international' in the late eighteenth and early nineteenth century. It connects to the case study on nationalism in WP6, and deploys the clustering methods developed in WP5 in order to grasp the lexical extension of key terms in politics. Rather than being a contribution to method development in the project, it is an example of how the project's methods are used independently to develop new cases that examine new humanities arguments.

## 8.3 Writing the Nation 1945–1948: The Austrian Daily Newspapers' Contribution to Nation-Building

Author: Stefan Hechl (UIBK-ICH)

This paper was presented in a seminar at the German Studies Association Conference 2020<sup>6</sup> and is

---

<sup>6</sup> <https://www.thegsa.org/conference/current-conference>

currently under review for publication in the *Journal of Austrian-American History*.<sup>7</sup>

**Abstract** The topic of Austrian national identity is well-researched – however, most studies focus on the years after the State Treaty in 1955 as well as the 1970s and 1980s. The immediate aftermath of World War II is most often neglected in this regard, a gap this research paper attempts to close. The paper assesses processes of nation-building and the formation of Austrian national identity by using digital methods such as diachronic frequency analyses or word correspondence analyses to examine a large corpus of Austrian daily newspapers from 1945 to 1948. It covers a wide ideological range from left to right and from pro-US to pro-Soviet, containing the complete text of all issues of five Austrian newspapers that appeared daily from April 1945 until December 1948. The digital newspapers collection was found in ANNO, converted to text using the open-source software Transkribus, and then analysed using a variety of third-party tools and improvised code, and in particular with methods and tools developed NewsEye.

**NewsEye context and implications** The paper focuses on distant reading methods to find (hidden) language patterns during the time of the re-building of Austrian national identity post 1945. The aim of this contribution is to prove existing hypotheses, as well as the discovery of previously unknown textual patterns and discourses. It uses digital methods such as diachronic frequency analyses, word correspondence analyses or topic models from WP4 (from the UH-CS group in Helsinki) as well as third-party tools and improvised code to discover textual trends (patterns). One aspect of this paper that sets it apart from others is the use of entire newspaper corpora rather than subcorpora based on articles.

## 8.4 Sound Impression – Opacity of the Voice – Construction of premises. Traces of radio in newspapers and vice versa

Author: Martin Gasteiner (UNIVIE)

The text is planned to be included in an edited volume titled 'Hearing is Believing', edited by Karin Moser for Vandenhoeck & Ruprecht. It is worth noting that the indicated title might be revised.

**Abstract** „Das Hören taucht unter in Schweigen und Taubheit.“ „Wenn wir Radio hören oder fernsehen, meinen wir, die Welt trete bei uns ein, in Person.“ Michel Serres. Die fünf Sinne. Frankfurt am Main 1993.

Based on a sociological model by Michel Serres, listening is understood as a social interaction that also finds its traces and reverberations in other media, and especially in the newspaper. In the process, these traces have an effect on the collective imagination and take root there. The influence of listening on the practised cultural techniques of writing and reading are explored on the basis of newspaper data. The focus of this article is the analysis of these traces, i.e. sources, left by radio programs in newspapers, and vice versa. It is precisely these intersections and transmissions that are crucial in media research and contexts and should point to settings in which radio was consumed and content was collectively imagined and discussed. For this source work, I also resort to forms of topic modelling, which I apply to the corpus of newspaper data as well as to transcribed machine-readable texts, and interweave them.

**NewsEye context and implication** The tools developed in WP4 and WP5, such as word frequency

---

<sup>7</sup> [https://www.psupress.org/Journals/jnls\\_JAAH.html](https://www.psupress.org/Journals/jnls_JAAH.html)

analysis and especially topic modelling, are used for this article. Additionally, WP3 provides evaluations with regard to named entity recognition, and thus contributes to giving the article the necessary contextualisation. It is the discovery and analysis of topic models and their possible implications that give the article the depth of focus it needs to embark on the scientific search for collective imagination.

## 8.5 Topic-specific corpus building: The creation of a representative newspaper corpus on the topic of return migration using LDA and JSD

Authors: Sarah Oberbichler and Eva Pfanzelter from the UIBK-ICH team.

Abstract has been accepted by the *Journal of Digital History*.<sup>8</sup>

**Abstract** The advance of digitization also calls for an update of the historical-critical method, the hermeneutic process that promises access to the past through the three steps of heuristics, source criticism and interpretation. Especially the second step, the demand for digital source (which often goes hand in hand with tool criticism) is frequently discussed. Other parts of the historical method, such as the relevance and critical assessment of adequate corpus creation in the heuristic research step, have so far not been considered enough. This paper addresses the motivation of corpus building for historical research and asks the question of whether text mining methods can support corpus building in order to distinguish between relevant and non-relevant articles on the topic of return migration. The aim is to highlight the necessity of corpus building and the critical evaluation in the heuristic research step.

**NewsEye context and implications** This paper shows how collections created in the NewsEye platform can be processed further outside the platform. In order to make this workflow available to other users and encourage them to improve the quality of corpora by using digital methods, Jupyter Notebooks were created. Their integration into publications is becoming increasingly important in the digital humanities, and will therefore also be part of this publication.

## 8.6 Integrated Interdisciplinary Workflows for Research on Historical Newspapers – Perspectives from Humanities Scholars, Computer Scientists and Librarians

Authors: Sarah Oberbichler and Evan Pfanzelter from UIBK-ICK, Mikko Tolonen and Jani Marjanen from UH-DH, Emanuela Boros and Antoine Doucet from ULR, Hannu Toivonen from UH-CS, and Juha Rautiainen from UH-NLF

The article is under review in the *Journal of the Association for Information Science and Technology*.<sup>9</sup>

**Abstract** This article considers the interdisciplinary opportunities and challenges of working with digital cultural heritage, such as digitised historical newspapers, and proposes an integrated digital hermeneutics workflow to combine purely disciplinary research approaches from computer science, humanities and library work. Common interests and motivations of the above-mentioned disciplines have resulted in interdisciplinary projects and collaborations such as the NewsEye project, which is working on novel solutions on how digital heritage data is (re)searched, accessed, used and analysed. We argue that

---

<sup>8</sup> <https://journalofdigitalhistory.org/en>

<sup>9</sup> <https://asistdl.onlinelibrary.wiley.com/journal/23301643>

collaboration of different disciplines can benefit from a good understanding of the workflows and traditions of each of the disciplines involved, but must find integrated approaches to successfully exploit the full potential of digitised sources. The paper furthermore provides an insight into digital tools, methods and hermeneutics in action, showing that integrated interdisciplinary research needs to build something in between the disciplines, while respecting and understanding each other's expertise and expectation.

**NewsEye context and implications** The article collects the experiences on collaborations within the NewsEye team, and suggests an integrated hermeneutics workflow in order to improve interdisciplinary collaboration. The aim of the paper is to contribute to a better understanding between different disciplines. This paper has brought together members from WPs 3-6, as well as a partner from the National Library in Finland, to reflect on interdisciplinary collaboration, to create an integrated digital hermeneutics workflow, and to showcase our project outcomes. Writing this paper has definitely led to more understanding of each other's work and workflows. We gained valuable insights into shared and not shared visions as well as the possibilities and limitations of each discipline.

## 9 Discussion

The studies discussed above show that WP6 has contributed to the publication goals set for the NewsEye project with, among others, ten journal articles and three conference papers published or forthcoming as of M36. More publications are likely to be added to the list in the coming months.

Many of the studies deal with issues of multilingualism in the past and develop language-agnostic methods as a way to support historical research. This is crucial for overall method development in the project and is important for the transferability of the methods to other historical and linguistic contexts.

The multilingual studies do, however, also show that language-agnostic methods are not always the best way to deal with historically shifting relationships between languages. For instance, the role of the German language in South Tyrol and its relationship to Austria and Germany on the one hand and Italy on the other does not boil down to language only: The languages carry many layers of affective and connotative meanings that are part of discourses on identity and belonging. The interplay between Swedish and Finnish in Finland or local variations of Romance languages in France highlight similar issues.

While many of our studies deal with cross-language comparison, we also see that connections across languages are – methodologically speaking – the most challenging topics. In the studies on vocabulary relating to nationhood and the development of isms, we use particular words as nodes for facilitating comparison across languages, which seems to work in quite a straightforward manner. However, in a study that attempted to trace discourses on refugees in Austria and Finland around World War I, and which was more ambitious methodologically speaking, identifying links by the means of translated key terms, locations, named entities and topic models proved to be much more difficult. Working on this study (which ultimately was not published) as a collaboration between WP4, WP5, and WP6 was a very useful experience in probing the limits of what can and what cannot produce reliable results with messy historical data.

All in all, the studies presented above showcase a wide range of methods used to answer humanities questions. It is clear to us that humanities questions in general profit from using a mix of methods even for simple and rather straightforward research questions. This experience suggest a shift from the idea

that humanities scholars can use a tool developed by computer scientist to answer their questions to the insight that humanities questions require a precise tailoring of analyses that match their questions better than off-the-shelf tools, or that humanities researchers are better off using many different tools simultaneously and producing an overall interpretation based on the results from multiple analyses. With this experience and with its international and multilingual outlook, the NewsEye project will continue to develop border-crossing approaches to European heritage.

## Published references as of April 2021

- [1] Jani Marjanen, Ville Vaara, Antti Kanner, Hege Roivainen, Eetu Mäkelä, Leo Lahti, and Mikko Tolonen. "A National Public Sphere? Analyzing the Language, Location, and Form of Newspapers in Finland, 1771–1917". In: *Journal of European Periodical Studies* 4.1 (June 2019), pp. 54–77. ISSN: 2506-6587. DOI: [10.21825/jeps.v4i1.10483](https://doi.org/10.21825/jeps.v4i1.10483).
- [2] Eva Pfanzelter, Sarah Oberbichler, Jani Marjanen, Pierre-Carl Langlais, and Stefan Hechl. "Digital interfaces of historical newspapers: opportunities, restrictions and recommendations". In: *Journal of Data Mining and Digital Humanities HistInformatics* (Jan. 2021). Publisher: Episciences.org. URL: <https://hal.archives-ouvertes.fr/hal-02480654>.
- [3] Jani Marjanen, Jussi Kurunmäki, Lidia Pivovarova, and Elaine Zosa. "The expansion of isms, 1820–1917: Data-driven analysis of political language in digitized newspaper collections". In: *Journal of Data Mining and Digital Humanities HistInformatics* (Dec. 2020). URL: <https://jdmhd.episciences.org/6728>.
- [4] Nejma Omari. "La propagation de l'épidémie: Traitement médiatique de la grippe espagnole dans la presse quotidienne héraultaise". Français. In: *Etudes Héraultaises* (2020), p. 13. URL: <https://www.etudesheraultaises.fr/publi/la-propagation-de-lepidemie-traitement-mediatique-de-la-grippe-espagnole-dans-la-presse-quotidienne-heraultaise/>.
- [5] Ruben Ros and Sarah Oberbichler. "The Helsinki Digital Humanities Hackathon: Two Perspectives on Multidisciplinary Historical Newspapers Research in a Hackathon Context". In: *Digital Humanities in the Nordic Countries: Proceedings of the Digital Humanities in the Nordic Countries 5th Conference*. Vol. 2717. CEUR Workshop Proceedings. Riga, Latvia, 2020, pp. 66–74. URL: <http://ceur-ws.org/Vol-2717/paper07.pdf>.
- [6] Barbara Klaus. "Can Umlauts Ruin Your Research in Digitized Newspaper Collections? A NewsEye Case Study on 'The Dark Sides of War' (1914–1918)". In: *Digital Humanities in the Nordic Countries: Proceedings of the Digital Humanities in the Nordic Countries 5th Conference*. Vol. 2612. CEUR Workshop Proceedings. Riga, Latvia, 2020, pp. 267–274. URL: <http://ceur-ws.org/Vol-2612/short12.pdf>.
- [7] Sarah Oberbichler, Eva Pfanzelter, Stefan Hechl, and J. Marjanen. "Doing historical research with digital newspapers - perspectives of DH scholars". In: *EuropeanaTech Insight 16: Newspapers* (2020). URL: <https://pro.europeana.eu/page/issue-11-generous-interfaces>.
- [8] Sarah Oberbichler, Stefan Patrick Hechl, and Eva Pfanzelter. "Als eine andere Epidemie die Welt in Atem hielt: Die Spanische Grippe 1918/19 in der österreichischen Presse". In: *Tiroler Chronist - Fachblatt von und für Chronisten in Nord-, Süd- und Osttirol* 154 (2020), pp. 15–22.
- [9] Stefan Patrick Hechl. "'Wir dürfen wieder Österreicher sein!' Die Rolle der Tagespresse in österreichischen Nation-Building-Prozessen 1945–1948 – eine quantitative Analyse ausgewählter digitaler Zeitungskorpora samt Vorschlägen zur didaktischen Umsetzung". Master thesis. Jan. 26, 2021. DOI: [10.5281/zenodo.4468295](https://doi.org/10.5281/zenodo.4468295). URL: <https://zenodo.org/record/4468295#.YGXsb68zZPY>.