Re-Reading Lists in Historical Newspapers:
Digital Insights into an Overlooked Text Type

Nina C. Rastinger
Austrian Centre for Digital Humanities and
Cultural Heritage, Austrian Academy of
Sciences
`ninaclaudia.rastinger@oeaw.ac.at`

Abstract

The paper presents an ongoing doctoral project dedicated to periodically published lists in historical newspapers between 1600 and 1850. By employing approaches from Corpus Linguistics and Digital Humanities, the project aims to locate the studied 'small' texts within existing digital resources, analyse them with regard to their textual characteristics and evaluate their potentials and challenges for automated information extraction. The article primarily focuses on two key aspects: firstly, on search strategies for locating lists in digital newspaper corpora and collections, and secondly, on a case study into lists of arriving persons published in the *Wien[n]erisches Diarium* between 1703 and 1725. These empirical investigations reveal that periodically published lists form a central and frequent component of early modern newspapers and offer numerous potentials for Digital Humanities research due to their textual features, such as periodicity, repetitiveness or inherent (semi-)structuredness. In this regard, the paper identifies the overlooked newspaper text type as a data treasure awaiting discovery and underscores the need to investigate 'small' newspaper texts on a large scale.

1 Introduction

The creation and use of lists constitute central components of everyday life and function as fundamental cultural techniques (cf., e.g., Vismann, 2000: 20; Adelmann, 2021: 26). Subsequently, we are highly proficient in dealing with lists, as Esposito (2017: 356) summarises: "Lists are easy to write and easy to read." According to Goody (1977) and Waldispühl (2019: 197–198), this self-evidence of lists is rooted in a long-standing tradition that began with the emergence of writing systems and has since continued through different eras and writing cultures, constituting a historical and cultural continuum of the text form. Paradoxically, it seems to be precisely the omnipresence of lists that has led to them being neglected in research (cf. Waldispühl, 2019: 198; Schaffrick & Werber, 2017: 303). In addition, lists closely align with Hausendorf's (2009) concept of 'small' texts and are, in prototypical cases, perceived as 'small' in multiple senses:

1) size: lists as short and/or small-scale texts;
2) complexity: lists as simple, non-complex texts;
3) functionality: lists as praxis-oriented texts;
4) design: lists as stereotypical, template-like texts;
5) ambition: lists as unambitious, unelaborate texts.

These five textual features, which are tightly intertwined with each other, might have further contributed to the oversight of lists: as 'small' and ubiquitous texts, lists tend to have been underestimated or deemed irrelevant for scholarly research.

A systematic exception is currently only formed by literary lists, i.e., lists in and around literature, as discussed, among others, by Belknap (2004), Jullien (2004), Mainberger (2003, 2017, 2018), Fludernik (2016), von Contzen (2018, 2017a, 2017b, 2016), Rüggemeier (2020), Barton et al.

(2022) and Chinca et al. (2022). Also, lists that originate from modern, digital contexts can be identified as a growing research subject. Esposito (2017), Adelmann (2021), and Esposito and Stark (2019), for instance, deal with ratings and rankings in popular culture and the World Wide Web, Diederichsen (2021) provides an insight into playlists, Temmerman and Vandenabeele (2018) analyse the new journalistic genre of 'listicles', and Bubenhofer (2020) discusses the use of lists as a form of visualisation in digital linguistics.

In the overall perspective, however, the gap prevails: contrary to their frequency, historical non-literary or 'practical' lists have only been considered as a text type to be investigated in isolated cases (e.g., Doležalová, 2009; Ledin, 2015; Waldispühl, 2019). This 'oversight' particularly concerns studies on early modern press products, as a high number of historical newspapers contain periodically published lists, e.g. arrival lists (i.e. lists of persons who have arrived in a certain city), death lists, marriage lists, birth lists or price lists (cf. Figure 1). Such lists were published recurrently over a certain period, appearing monthly, weekly or, in many cases, in every issue of a periodical. Hence, unlike lists printed as parts of news articles, advertisements, or other sections, they constitute a distinct component and specific text type of historical newspapers.
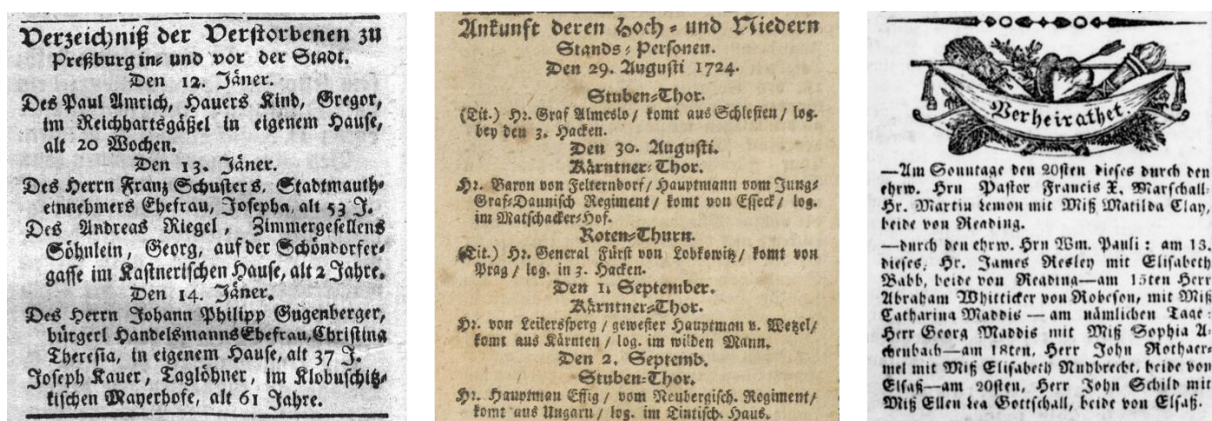

Figure 1. Death list from the *Preßburger Zeitung* (16.01.1765: 8), arrival list from the *Wien[n]erisches Diarium* (02.09.1724: 8), and marriage list from *Der liberale beobachter und Berks, Montgomery und Schuylkill Caunties allgemeine anzeiger* (29.09.1840: 3)

Nevertheless, in contrast to other newspaper text types, like news articles (e.g., Haß-Zumkehr, 1998; Pfefferkorn et al., 2017), advertisements (e.g., Bendel, 1998; Kurzmann, 1999; Farø, 2005; Ehrenheim, 2011) or reader's letters (e.g., Fix, 2008: 273–333; Fix, 2011), periodically published lists have so far neither been discussed theoretically nor examined empirically for their textual properties. Furthermore, they are missing from general overviews of (historical) newspaper text types (cf. e.g., Dovifat, 1976; Brand & Schulze, 1993; Straßner, 2000; Lenk & Chesterman, 2005) and are partly already left out during the (full text) digitisation of periodicals, which makes them all the more invisible to the scientific and public eye.

Given the lack of research, the dissertation project discussed here is dedicated to the identification and analysis of periodically published lists in German-language newspapers published between 1600 and 1850. Specifically, approaches from Corpus Linguistics and Digital Humanities are employed to pursue three key research interests: first, the text type should be identified in existing digital corpora and text collections (cf. Chapter 2), second, it should be analysed in regard to its textual characteristics (cf. Chapter 3), and third, it should be explored with respect to its challenges and potentials for the field of Digital Humanities, especially focusing on (semi-)automatic approaches to text analysis and information extraction (cf. Chapter 4). This three-step process of the dissertation, both in terms of content and practical procedure, will be detailed in the following chapters.

2 Looking for lists: identification of 'small' texts in historical newspapers

In the last decades, newspapers and journals have undergone mass digitisation (cf., e.g., Nicholson, 2013; Blome, 2018; Ehrmann et al., 2023). As a result, users can nowadays access facsimiles and/or full texts of (German) historical newspapers via various online resources, such as *AustriaN Newspapers Online* (= ANNO, Österreichische Nationalbibliothek), *Chronicling America* (Library of Congress),

digiPress (Bayerische Staatsbibliothek), *Digitales Forum für Mittel- und Osteuropa* (= DIFMOE, Digitales Forum Mittel- und Osteuropa e.V.), *impresso* (Clematide et al.), *Teßmann digital* (Landesbibliothek Dr. Friedrich Teßmann), *zeit.punkt NRW* (Universitäts- und Landesbibliothek Bonn et al.), and many other corpora and text collections. Within the CLARIN infrastructure, there are also multiple resources that offer (often already structurally or semantically annotated) full texts of historical newspapers, for example, *Deutsches Textarchiv* (= DTA, Berlin-Brandenburgische Akademie der Wissenschaften), *HIST – Archive of the Historical Corpora* (Institut für Deutsche Sprache) and *Wien[n]erisches DIGITARIUM* (Resch & Kampkaspar). Concentrating on printed German-language newspapers from 1600–1850, the dissertation project uses (a selection of) the named resources as a starting point to locate lists in digitised press texts.

One challenge in this endeavour clearly lies in the amount of unknown factors when searching for a specific text type rather than a concrete linguistic unit, entity or topic. The fact that both the lexical content and the (typographical) structure of periodically published lists may have varied over time and between different newspapers (cf. Figure 1) prevents the use of a universal approach that encompasses all cases. Instead, to cover and uncover a wide range of lists, the doctoral project employs a mixed methods approach that combines multiple search strategies. To be precise, four different strategies are presently being used in parallel to look for lists in digital newspaper corpora and collections, namely (1) the reuse of existing annotations, (2) full text search in combination with the calculation of a word/phrase reuse ratio, (3) the training of layout recognition models and (4) the close reading of random samples.

2.1 Reuse of existing annotations

Some of the available newspaper corpora and collections do not only offer facsimiles and plain full texts, but already include structural and/or semantic-markup. A small subset of these resources also account for lists in their annotation systems. The W*ienerisches DIGITARIUM* (Resch & Kampkaspar), created at the Austrian Centre for Digital Humanities and Cultural Heritage, for instance, contains over 300 full text issues of the historical *Wiener Zeitung* in XML/TEI-P5 encoding. Within this data set, which is downloadable through the CLARIN-B centre ARCHE (ACDH-CH, https://hdl.handle.net/21.11115/0000-000F-746A-8), both singular and periodically published lists were annotated with the <list> element and further distinguished into separate entries through the <item> element (cf. Resch et al., 2023). As a result of this mark-up, users are able to systematically retrieve lists by querying for certain structural elements, with help of the Oxygen XML Editor and XPath, Beautiful Soup or other tools able to parse XML.

Although such a reuse of pre-existing annotations is highly efficient, it is rarely applicable in practice, since, at present, resources providing structural and/or semantically annotated full texts of historical German-language newspapers form an exception within the digital landscape. Instead, the majority of resources publicly available offers transcripts which were created automatically through Optical Character Recognition (OCR) and not corrected or edited further. This pragmatic decision in favor of quantity over quality makes sense considering the vast amounts of newspapers that have been preserved from the past, but, at the same time, requires users to adapt their approach to a resource accordingly. In the end, multiple factors may hinder the successful text recognition through OCR and lead to a high error rate, such as low image quality, (changes between) certain print types (cf. Kampkaspar, 2019, about OCR and Fraktur; Rastinger, in print, about OCR and print type changes between Fraktur and Antiqua), or the incorrect recognition of text regions during layout analysis (e.g., due to complex list layouts). OCR errors, in turn, not only complicate a newspaper's enrichment through annotations, but may also distort search processes and skew analyses, as shown, among others, by Torget (2023). Hence, even when deploying separate search strategies besides the reuse of existing annotations, one essentially always needs to take into account potential limitations due to restricted full text quality.

2.2 Full text search and reuse ratio

It is especially important to address the quality of full texts when using the full text search at hand in most digital corpora and resources to date. One strategy adopted here to deal with this limitation is the calculation of a reuse ratio of certain words and/or phrases (potentially) tied to periodically published lists. This approach makes use of the observation that periodically published lists tend to involve word and/or phrase reuse in multiple ways, namely (1) within a certain issue, (2) over multiple issues and (3)

between different newspapers. To give a concrete example, the death list published in the *Tiroler Zeitung* under the title *Allhier Verstorbene in und vor der Stadt* 'deceased persons here in and outside the city' repeatedly includes the term *J.* [Jahr] 'year' as an age measure, uses similar list titles and subtitles over several issues and shares certain key words (e.g., *Verstorbene* 'deceased') with periodically published lists in other newspapers, for instance the death lists of the *Bozner Zeitung* (list title: *Verstorbene in Bozen*), the *Klagenfurter Zeitung* (list title: *Verzeichniß der hier Verstorbenen*), and the *Tiroler Zeitung* (list title: *Verstorbene*).[2] In the doctoral project, such inter- and intratextual overlaps are utilised by querying for both accumulations of certain linguistic elements within a single newspaper page and patterns of word/phrase reuse across entire newspapers and newspaper collections. Both strategies should – when carried out on large datasets – help with spotting periodical text reuse and mitigating potential OCR errors.

On a practical level, this search strategy can be demonstrated using an example: a lemma that repeatedly appears within the list titles gathered so far is the noun *Civilstand* 'civil status'. In the context of early modern newspapers, this term seems to serve as an overarching concept to group together birth, marriage, and death records, and can be linked to the French *état civil*, the 'management of population lists' (*Führung der Bevölkerungslisten*), according to Herders Conversationslexikon (1854–1857). The noun was queried for in a selected collection, namely the newspaper portal *zeit.punktNRW*, funded by the state of North Rhine-Westphalia. This digital resource provides access to over 400 local German newspapers, among which 112 periodicals fall within the study period from 1600 to 1850. The approach started with compiling a list of these 112 historical newspapers. Subsequently, a web crawler, created with the python library Beautiful Soup, was employed to perform automated full text searches within the periodicals and capture the total number of pages per newspaper that include the exact term *Civilstand*. Additionally, a second search was carried out for the highly frequent conjunction *und* 'and', as to estimate the total number of (searchable) pages per newspaper. The collected counts were then used to calculate an approximate percentage of pages within one newspaper that contained the word *Civilstand*. Also, to get a better overview of the results and easily compare them with each other, every queried newspaper was given a number and was visualised as part of a heatmap:
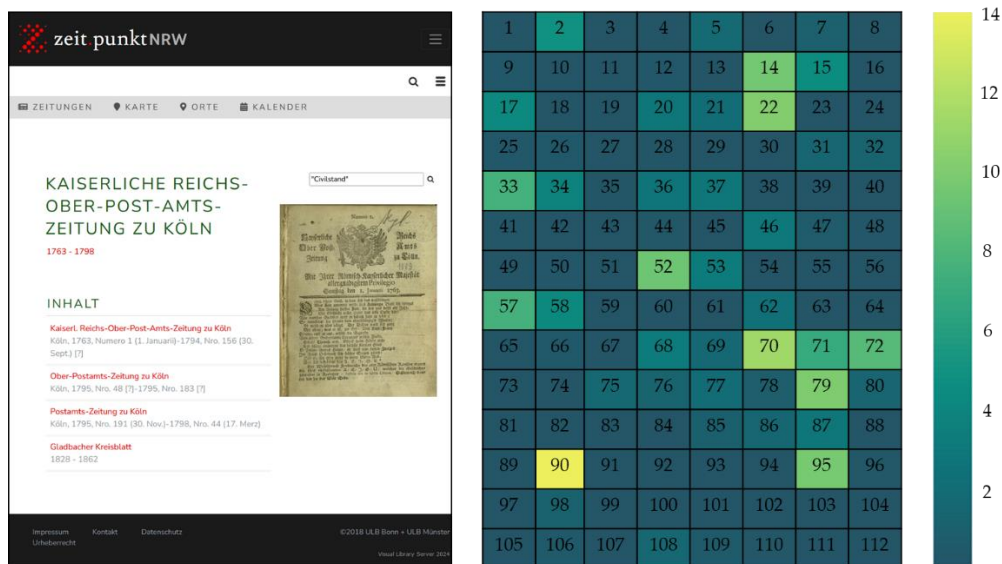

Figure 2. Search interface of an exemplary newspaper in *zeit.punktNRW* (left) and heatmap of the reuse ratio of *Civilstand* within 112 German newspapers from 1600–1850 (right)

In theory, a higher percentage – visualised in the heatmap in Figure 2 as a lighter color – suggests a greater likelihood of a certain newspaper to contain periodically published texts including the term *Civilstand*. As can be seen in Figure 2, multiple newspapers (e.g., 90 = *Täglicher Anzeiger für Berg und Mark*, 70 = *Neue Rheinische Zeitung*) available through *zeit.punktNRW* showcase such a high

---

[2] This should not lead to the erroneous assumption that specific types of lists, such as death lists, are homogenous in their nomenclature. Various alternative titles exist, e.g., *Starb:* 'Died:', *Beerdigte* 'Buried persons', *Sterbefälle* 'Cases of death', or *Todesfälle* 'Cases of death'.

probability. When taking a closer look into all newspapers with a ratio of at least 4 % (= 16 cases), it was found that each of them actually involved periodical lists on locals birth, marriages and deaths. Among the identified text series are, for example, *Civilstand der Stadt Luxemburg* published in the *Luxemburger Wort*, *Civilstand der Stadt Aachen* as part of the *Echo der Gegenwart*, *Civilstand der Bürgermeisterei Kempen* printed in the *Kempener Kreisblatt*, and *Civilstand der Stadt Neuwied* featured in the *Neuwiedische Nachrichten*. From a methodological view point, these and further successful findings speak for adopting a word reuse ratio as another strategy to systematically identify lists in historical newspapers and simultaneously investigate the spatial and temporal developments in their titling.

2.3 Layout recognition

At the same time, two approaches are still unlikely to be suited for all types of reoccuring lists – which is why it makes sense to expand one's perspective from the linguistic-content level to the visual-typographic level. Looking at their text surface, lists published in historical newspapers tendentially involve more 'white space' than other text types, like news or official announcements, which is usually due to more line breaks, indentations, subheadings, etc. This visual nature of lists is also addressed in research literature, for example when Belknap (2004: 15) defines a list as "a <u>formally</u> organized block of information that is composed of a set of members" [emphasis: N.C.R] or when Bubenhofer (2020: 31) locates lists in the transitional area between diagram and text because they relate linguistic signs to an arrangement in space.

Approaching the topic from the field of Digital Humanities, this characteristic of lists can be leveraged by training and using list-specific layout recognition models. Currently, this approach has been explored through training a layout recognition model on the layout of the historical *Wiener Zeitung* of the late 18th century. Using the transcription platform Transkribus (READ-COOP), 24 issues and over 300 pages of the Austrian periodical were structurally annotated, meaning all text regions were identified and categorised. The latter was done according to a tag set differentiating, among other elements, between paragraphs, headings (cf. pink marking in Figure 3), figures, seperators, catch words, and – of course – lists (cf. lilac marking in Figure 3). Based on these structural tags, a field model was trained, reaching an accuracy (mAP) of 88,06 %.
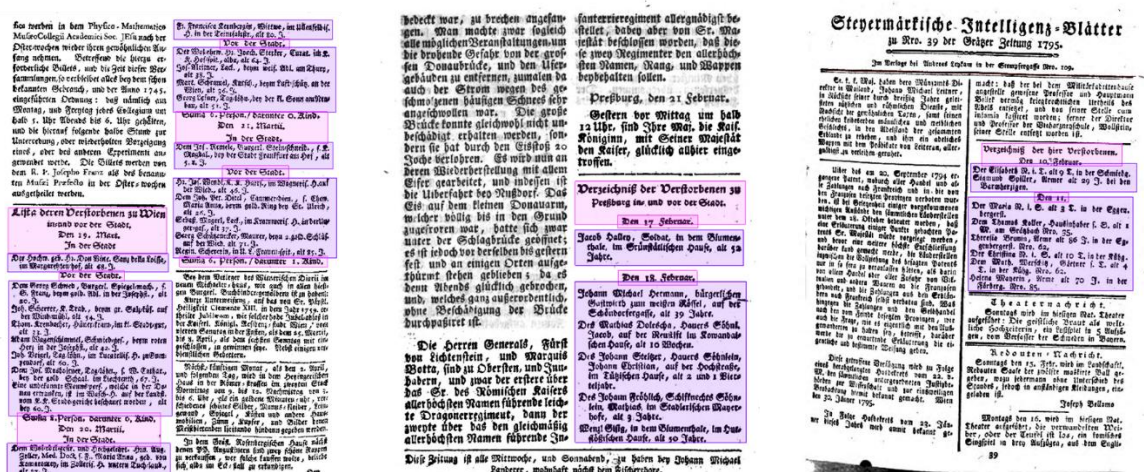

Figure 3. Exemplary results of using layout analysis for list detection, in the *Wiener Zeitung* (left), *Preßburger Zeitung* (middle) and *Grätzer Zeitung* (right)

As illustrated in Figure 3, the resulting layout recognition model cannot only support the automatic identification and annotation of lists within the *Wiener Zeitung* (left), but may also be applied to search through further newspapers with a similar layout, such as the *Preßburger Zeitung* (middle), available through *ANNO* and *DIFMOE*, or the *Grätzer Zeitung* (right), accessible in *ANNO*. By focusing on the visual level of periodically published lists, the approach enables the efficient search of large image datasets and contributes further to the systematic discovery of the 'small' texts in historical newspapers.

2.4 Towards an open-ended 'list of lists'

The three search strategies presented so far can be viewed as forms of distant reading (cf. Moretti 2016) and are complemented by the more traditional approach of close reading. Due to the amount of data available, it makes sense to select certain resources, especially prioritising those that are less accessible with distant methods, and draw random, time-distributed samples per newspaper. For assembling these samples, it has proven beneficial that multiple newspaper collections (e.g. *ANNO*, *digiPress*, *zeit.punktNRW*) offer access to an IIIF Image API with which facsimiles can systematically be exported into various viewers (e.g., *Mirador*, *Universal Viewer*) and tools (e.g., *Transkribus*). Based on the premise that the investigated lists are in particular characterised by their periodicity, systematic close reading should assist in identifying additional representatives of the newspaper text type not yet (un)covered through other approaches. In combination, the four search strategies presented both complement and inform each other, as each new (type of) list found with one method might again provide input for another approach.

The ultimate goal of this iterative and ongoing process of looking for lists is to create an open-ended 'list of lists', i.e. a database on lists which were periodically circulated through historical newspapers. Especially in the absence of research literature (cf. Chapter 1), such a compilation allows for a first-time overview of the examined 'small' texts, provides insights into their publication through time and space, and may possibly even shed light on larger scale trends and shifts within the early modern print market. Moreover, a systematic identification of periodically published lists serves as a foundational step for corpus-based analysis, as it supplies the necessary knowledge and texts for the assembly of a representative analysis corpus.

While the envisioned 'list of lists' is currently still being compiled, a preliminary interim status can already be provided: with the help of the methodic repertoire discussed in this chapter, eleven digital corpora and collections that contain German printed newspapers from 1600 to 1850 have already been (partly) queried so far. Concretely, the search has been completed for six resources (*Chronicling America*, *Darmstädter Tagblatt*, *DTA*, *impresso*, *Teßmann digital*, *Wienerisches DIGITARIUM*) and started for five resources (*ANNO*, *DIFMOE*, *digiPress*, *ZEFYS*, *zeit.punktNRW*).Within these corpora and collections, 192 periodically published lists were identified so far, which originate from 62 different newspapers and cover almost the whole 18th and early 19th century, with the earliest finding dating back to 1703.[3] In parallel and even though only German-language newspapers were being taken into account, the publishing locations of the found texts are scattered all over Central Europe and the text type can even be observed in periodicals published in 19th century North America. To sum it up, these initial findings already indicate the high frequency of periodically published lists in historical newspapers – and make the need for research in this area all the more clear.

3 Looking at lists: digital analysis of textual characteristics

Systematically collecting periodic lists might reveal diachronic patterns within the early modern newspaper landscape. The temporal and spatial distribution of different list types or 'paradigms', for instance, could be highly informative and forms one of the doctoral projects' core interests. In this context, the concept of the 'list paradigm' refers to a unifying aspect to which all elements of a list are related and through which they become part of a whole – a notion that is, even if not mentioned explicitly, evident in many definitions of lists, for instance by Tankard (2006) and Mainberger (2003):

(1) "A list is a written or printed series of names, dates, numbers, or items, organized according to some need or principle." (Tankard, 2006: 339)
(2) "Enumerations name distinct elements and equalize them under a thematic or formal aspect." (Mainberger, 2003: 7)

It is also true for the lists identified so far that multiple connecting themes reoccur. Currently, the most frequent list types are death lists, found in 26 newspapers (e.g., *Wiener Zeitung*, *Tiroler Zeitung*), and arrival lists, present in 17 newspapers (e.g., *Luxemburger Wort*, *Hamburgischer Unpartheyischer*

---

[3] As of now, no lists from the 17th century have been found, which may however be attributed to the choice of corpora and collections examined so far as they mainly contain newspapers from the 18th century onwards.

Correspondent). These are closely followed by price lists (16 newspapers, e.g., *Der Vaterlandsfreund*, *Innsbrucker Wochenblatt*), and marriage lists (13 newspapers, e.g., *Der Morgenstern, Gülich and bergische wöchentliche Nachrichten*). Moreover, one repeatedly encounters theatre performance lists (12 newspapers, e.g, *Wiener Zeitung, K. K. priv. Prager Zeitung*), birth lists (10 newspapers, e.g., *Bozner Zeitung, Rheinischer Merkur*) and various other types (e.g., departure lists, newspaper agent lists). Although this enumeration is of preliminary nature, it already showcases that periodically published lists in historical newspapers constitute not only a common, but also a manifold phenomenon, which could potentially be relevant to multiple disciplines.

Besides differentiating between different types of lists, the doctoral project addresses the text type's diversity by quantitatively and qualitatively inspecting its textual characteristics. Based on the 'list of lists' currently in compilation (cf. Chapter 2.4), an analysis corpus as representable as possible is going to be created, that should cover different list types, newspapers, and time periods, making them comparable to each other. The focus of the corpus linguistic analysis will then be put on textual patterns on various levels of periodically published lists, namely on the level of (1) their selection and organisation of entries, (2) their typographic structure, (3) their linguistic features and (4) their pragmatic context. On each level, selected aspects are empirically examined, e.g., the inclusion of 'empty' items (1, e.g. *Niemand* 'no one' in arrival lists), (the number of) applied ordering principles (1, e.g., hierarchical, alphabetical, chronological, geographical), typographic resources used for text structure (2, e.g., indentations, blockings, bullet points), abbreviation density (3), morpho-syntactic completeness or 'grammaticality' (3, e.g., use of verbs, interpunctuation), and text purpose (4, e.g., information, legitimation).

On a practical level, the analysis will again be carried out by integrating close and distant reading approaches, resulting in a form of 'blended reading' (Lemke & Wiedemann, 2015) or 'scalable reading' (Mueller, 2020). Analogous to Google Earth, this approach should allow for different perspectives on a singular research subject, as Mueller (2020) notes: "you can zoom in and out of things and discover that different properties of phenomena are revealed by looking at them from different distances". In this way, the doctoral project aims to obtain a first, multi-faceted picture of lists published in historical newspapers and a better understanding of their textual characteristics and diachronic development.

4 Case study: digitisation, annotation and mapping of arrival lists

The third and final aspect of the doctoral project focuses on the potentials and limitations of the newspaper text type within the field of Digital Humanities. This part of the investigation is grounded in the idea that the studied lists might be especially well-suited for semi-automatic information extraction approaches due to their periodicity, repetitive nature, semi-structured format, and high density of named entities. This theory is put into practice through a concrete use case, namely the City of Vienna funded project "Visiting Vienna – digital approaches to the (semi-)automatic analysis of the arrival lists found in the *Wien[n]erisches Diarium*" (PI: Nina C. Rastinger, 2022–2023).

As the project title reveals, the focus is put on arrival lists, namely the ones printed in the *Wien[n]erisches Diarium* (since 1780: *Wiener Zeitung*) from August 1703 to March 1725. Published under the title *Ankunft deren Hoch= und Niedern Stands=Personen*[4] 'arrival of high- and low-ranking persons of social standing', these texts were disseminated with every issue of the semi-weekly appearing Austrian newspaper and document the arrival of upper class persons, such as aristocrats, secretaries, clerics or couriers. Both Johann Baptist Schönwetter and Johann Peter van Ghelen who consecutively served as publishers of the *Diarium* at the time had access to this information through an imperial privilege (Mader-Kratky et al., 2019: 99). The preserved arrival lists are distinguished by their rich detail, providing not only personal information about the individuals arriving, such as names, professions, titles, and affiliations, but also detailed spatial and temporal data about their taken travel routes. Typically, this encompasses the place of departure, the date of arrival, the city gate used to enter Vienna, the accommodation inside or around the city, and/or the final destination. An exemplary item containing (almost) all of these components is depicted in Figure 4:

---

[4] Here, one exemplary spelling is displayed; overall, the list title is represented in over 60 different writing variants.
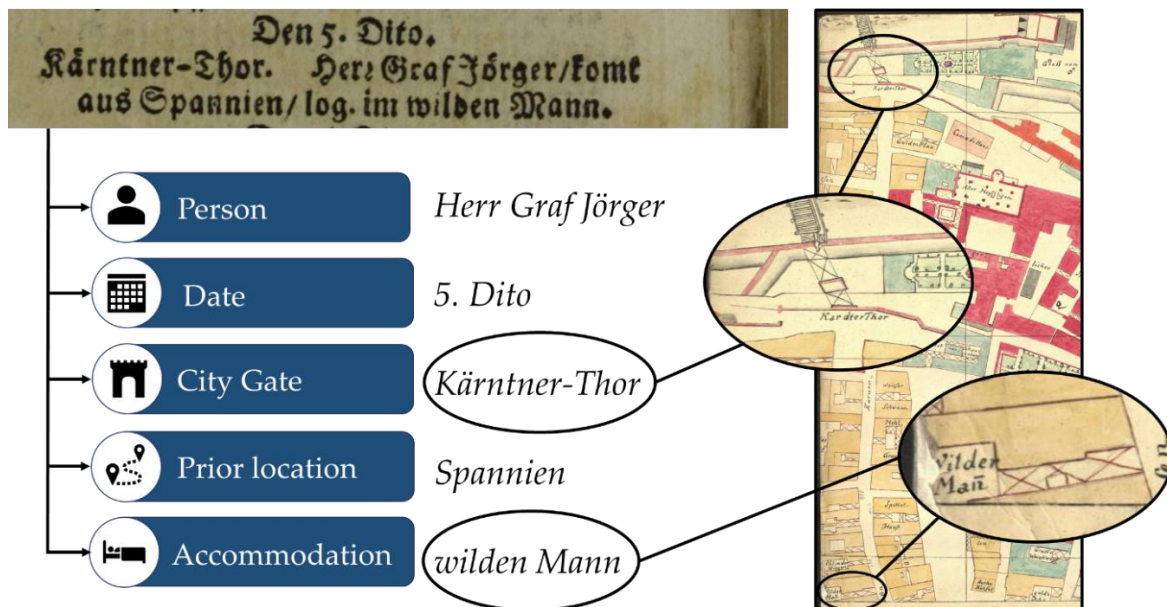
Figure 4. Graphical representation of NER and mapping of an exemplary list entry (*Wienerisches Diarium*, 07.01.1711: 7) on a replication of the Steinhausenplan (1710) by Schimmer (1847)

The illustration further portrays part of the steps taken over the course of the case study: First, reliable full texts were created (cf. Chapter 4.1), secondly, these texts were semantically enriched through named entity recognition (NER) (cf. Chapter 4.2), thirdly, the identified geographical entities were mapped on historical city plans of Vienna (cf. Chapter 4.3), and fourthly, the resulting data will be published online and archived in a long-time repository, enabling its safe storage and reuse in future research (cf. Chapter 5). With the help of this four-step workflow, the doctoral project leverages the information density of the arrival lists to trace past movements in and around an early modern city, while also exploring the potentials and limitations of the newspaper text type for (semi-)automatic information extraction processes.

4.1 Full text digitisation

To effectively analyse the Viennese arrival lists with the help of digital methods, it is essential to first establish a comprehensive and high-quality full text dataset. The newspaper *Wienerisches Diarium*, in which the periodical texts were published, is presently accessible via two digital platforms, namely *AustriaN Newspapers Online* (*ANNO*), hosted at the Austrian National Library, and the afore-mentioned *Wienerisches DIGITARIUM*, developed at the Austrian Centre for Digital Humanities and Cultural Heritage. As detailed in Chapter 2.1, the *DIGITARIUM* provides a selection of 18th century newspaper issues in a reliable, annotated full text format plus facsimiles, among which 76 issues include arrival lists. In comparison, *ANNO*, which also provides both image and text, hosts a more extensive, though also not exhaustive,[5] collection of 18th century arrival lists. The limitation is that these full texts were produced automatically using OCR without further error handling and thus partially exhibit a high Character Error Rate (CER). This issue seems to be particularly significant for the arrival lists, which are prone to have more complex layouts than standard news texts. An analysis of a balanced, random sample of ten lists from *ANNO*, encompassing around 1,500 tokens, revealed a CER of 50 %, indicating that, on average, every second character of the arrival lists was incorrectly recognised. Consequently, the full texts offered in *ANNO* are not suitable for automated NLP processes that rely on token-level accuracy and necessitate high precision, such as named entity recognition (cf. also Müller, 2016).

---

[5] Currently missing are, for instance, the volumes for the years 1717 and 1718, for which the arrival lists thus have to be collected and processed from other sources, e.g., the microfilms available in the Wienbibliothek im Rathaus (https://www.wienbibliothek.at). This is presently being done, with future plans to add these lists to the corpus as well.

Against this backdrop, all arrival lists were first collected as facsimiles from *ANNO*, grouped by year, pre-processed and uploaded to the transcription platform *Transkribus* (READ-COOP), where they underwent (semi-)automatic layout and text recognition. In the latter step, the HTR model "German Fraktur 18th Century – WrDiarium_M9" (Resch & Kampkaspar, 2020), that was developed during the creation of the *Wienerisches DIGITARIUM* (Resch & Kampkaspar), was applied. Since this model was trained on the 18th century *Wiener Zeitung*, achieves a low CER of 0.8 % and is publicly available, it is ideally suited for the text recognition of the arrival lists. Additionally, to ensure the high accuracy of the transcriptions, all texts were manually inspected and corrected where necessary. This digitisation process has presently been completed for almost 1,500 arrival lists from the *Wien[n]erisches Diarium*, resulting in a dataset of around 176,500 tokens which spans from August 1703 to March 1725.

4.2 Named entity recognition (NER)

High quality full texts pose a fundamental basis for reliable results from NLP tasks (cf. Ehrmann et al., 2023; Torget, 2023). However, even when given without OCR errors and other artificial noise, historical newspaper texts tend to present difficulties for NER due to their high graphematical variation and language dynamics, in addition to a scarcity of resources and other obstacles (cf. Ehrmann et al., 2023; González-Gallardo et al., 2023). In light of these challenges, multiple approaches were explored and compared throughout the case study, namely (1) the reuse of existing NER models for historical German, (2) a rule-based pipeline using spaCy's EntityRuler, and (3) the use of Large Language Models (LLMs), concretely GPT-3.5 (OpenAI). Based on a small (!), time-distributed gold standard (10 lists, 83 items, 1474 tokens),[6] it has been shown that the out-of-the-box use of existing NER models for historic German[7] yielded the worst results for the arrival lists of the *Wien[n]erisches Diarium*, achieving only low F1-scores ranging from 0.17 to 0.42. These poor results might be attributed to the fact that NER models are typically trained on running text, such as news articles, and thus struggle with the minimal or absent syntactic context of lists.

In comparison, slightly better results were obtained by creating project-specific RegEx-based rules for NER, as a relatively small set of rules was able to identify the majority of all entities in the gold standard. However, the creation of individual rules has proven to be both time-consuming and inflexible, especially considering the diachronic nature of the dataset and the planned future reuse of the workflow in other datasets. The NER strategy that ultimately achieved the best results was the use of the LLM GPT-3.5, accessed via the OpenAI API. The input and output of the model were optimised using Promptify, an open-source Python library dedicated to prompt engineering. With its help, a one-shot approach was taken in which the LLM was provided with one annotated example, i.e., one arrival list with nine entries. While all instructional prompts on the NER task were presented in English, the labels were given in German and set in close accordance to the textual structure of the studied lists. Besides refering to PERSON (person) and DATUM (date) as common NER categories, places were differentiated into STADTTOR (city gate), HERKUNFTSORT (place of departure), UNTERKUNFTSORT (place of accommodation) and ZIELORT (destination). Lastly, a postprocessing and feedback loop was integrated into the workflow to determine the correct positions of the entities and detect possible hallucinations, i.e. named entities that differ in orthographical or grammatical form in input versus output. As can be seen in the exemplary output in Figure 5, this approach yielded very good results:

---

[6] The gold standard was annotated with CATMA 6 (Gius et al., 2022) and exported with GitMA (Vauth et al., 2022).
[7] All models used are publically available on HuggingFace. Among the models tested were the flair models *flair-de-ner*, *flair-historic-ner-onb*, and *flair-historic-ner-lft* as well as the spaCy models *de_RTA_NER* and *easy_fnhd*.

Figure 5. Annotations of an exemplary arrival list (*Wienerisches Diarium*, 28.07.1708), visualised with displaCy

A first systematic evaluation with the previously mentioned gold standard (F1-score: 0.99) as well as a cursory inspection of the data indicate a high potential of LLM-based NER for historical, semi-structured texts. In particular, the comparatively easy adaptation to the text material (e.g. via free choice of labels, annotated examples) has shown to be of significant advantage in context of the use case. While pre-trained models come with a fixed set of categories, general-purpose LLMs, like GPT-3.5, offer a free choice of labels and thus allow for more fine-grained entity types and a certain degree of specialisation without having to create extensive training material. The same is true for the use of one- or few-shot approaches, which seem to work especially well in combination with inherently structured input, such as lists.

These factors might also explain why the results presented in this paper differ from what González-Gallardo et al. (2023) have observed in their investigation on ChatGPT-based NER in historical newspapers texts: The authors report F1-scores from 0.278 to 0.794, depending on the evaluated dataset and evaluation mode, and primarily emphasise the limitations posed by using the LLM. This differing outcome might be due to one or more of the following differences between the NER approach employed by González-Gallardo et al. (2023) and the one taken here: (1) ChatGPT web interface vs. API, (2) zero-shot vs. one-shot approach, (3) coarse-grained NER with high-level entity types vs. fine-grained NER with highly text-specific entity types, and (4) multiple, mainly unstructured and heterogenous newspaper text types vs. one semi-structured and homogenous newspaper text type. As González-Gallardo et al. (2023) state themselves, "the capacity of ChatGPT of identifying named entities is really dependent on the dataset and the type of entities."

In case of the arrival lists published in the *Wien[n]erisches Diarium*, a GPT-based approach has proven highly fruitful: In total, over 39,500 named entities could be identified and extracted from 1,447 texts. Besides including over 10,100 mentions of persons and over 4,200 cases of dates, this data set is especially rich on geographical entities and contains around 5,200 instances of city gates, 10,000 instances of points of departure, 9,400 instances of accommodations and 500 instances of destinations.

4.3 Spatio-temporal mapping on historical city maps

Having such an extensive data set on early modern named entities allows for various quantitative analyses. What seemed particularly promising due to the multitude of geographic entities, was mapping the data on historical city maps of Vienna. To do so, it was first neccessary to deduplicate (cf. 3), disambiguate (cf. 4) and geocode (cf. 3, 4) the identified entities, as illustrated here for places of accommodations:

(3) *7. Saullen*, *7. Säulen*, *7. Saulen. auf dem Neuenmarck*, *sieben Saüllen*, … > ‚Sieben Säulen'
     (16,37047294; 48,20650766)
(4) *Weissen Wolffen, weissen Wolff* (16,3771622; 48,21009885) ≠ *weißen Wolfis. Hauß, Weissenwolffischen Hauß* (16,36914529; 48,21192704)

To this end, multiple external knowledge resources were consulted, both automatically through distance measures (e.g., Levenshtein distance) and manually: (1) the platform *Wien Geschichte Wiki* (= WGW,

City of Vienna), (2) the prosopographical portal *ViecPro* (Romberg et al.), (3) a house directory published by Joseph Anton von Trattner (1773),[8] and (4) a house sign register from the year 1795.[9]

In addition to reusing existing digital resources, a new gazetteer was created on basis of the so-called Steinhausenplan, a city map of Vienna created by Werner Arnold Steinhausen in 1710. The plan, which has previously been georeferenced by the City of Vienna, is characterised by its high accuracy and a large number of labels. To make this data automatically processable, all labels found on the historic map were transcribed and categorised in regard to their readability. Moreover, as far as possible, each location was supplemented with information from the book *Schatz / Schutz / und Schantz Deß Ertz=Herzogthumbs Oesterreich / Das ist Ein sehr genaue / und ordentliche Beschreibung [...] der berümten Haubt= und Kayserl. Residentz=Statt Wienn [...]*[10] (1701), written by the imperial mail carrier Johann Jordan to allow for better navigation inside early modern Vienna. Combined, the resulting gazetteer consists of 1,469 data points, with 957 locations (65 %) including information from both the Steinhausenplan and Johann Jordan's house register.

Utilising both the external and internal resources mentioned, it was so far possible to precisely locate 71 % of all mentions of accommodations and 55 % of all distinct accommodation places.[11] Furthermore, 2 % respectively 13 % of all buildings used for lodging could be located approximately, i.e. linked to a certain street, square or neighboring house. These merged success rates of 73 % and 68 % result in almost 7.000 spatio-temporal data points, which were visualised and analysed with the open-source software QGIS, as demonstrated in Figure 6:
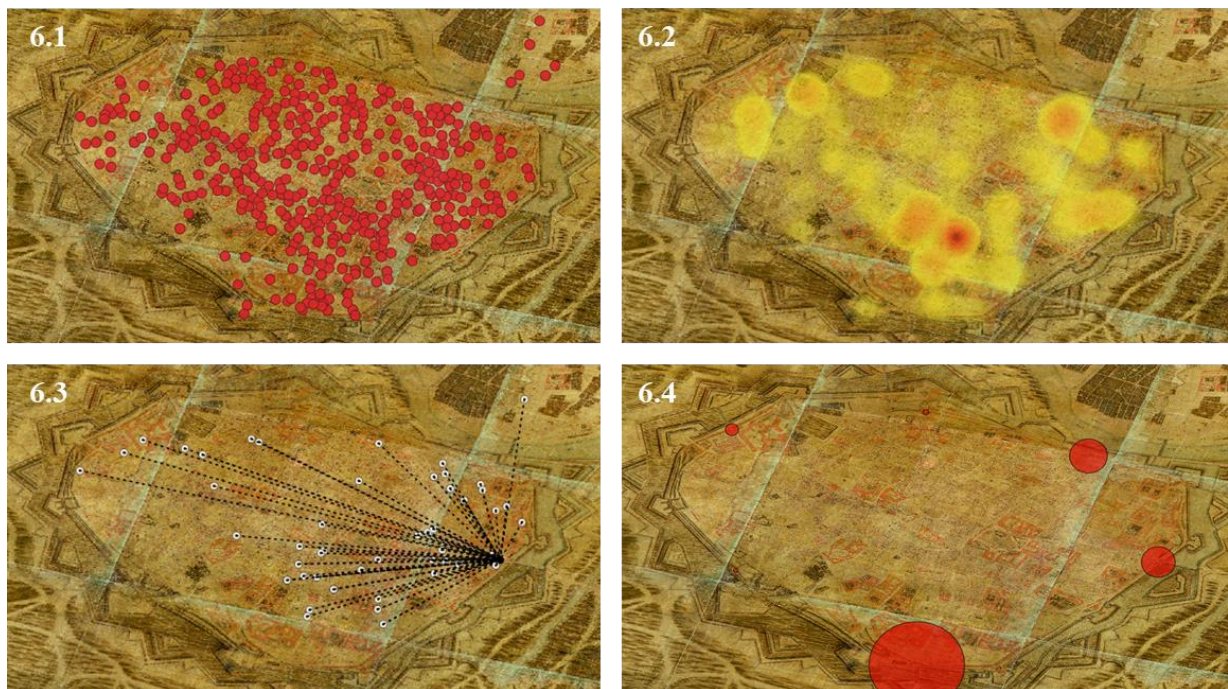

Figure 6. Four exemplary visualisations of arrival list data on the Steinhausenplan (1710) with QGIS

Like the exemplary visualisations in Figure 6 illustrate, mapping early modern arrival lists enables various insights into the urban history of a city, in this case of Vienna. It becomes possible, for example, to understand in which parts of the city (high-ranking) individuals found accommodations (cf. 6.1),

---

[8] *Verzeichniß der in der k.k. Haupt= und Residenz=Stadt Wien befindlichen Gassen Hausinnhabern deren Schildern und numerirten Häusern* 'Directory of the Alleys, House Owners, Their Signs, and Numbered Houses in the Imperial and Royal Capital and Residence City of Vienna'

[9] *Wiener Schildregister, oder Anweisung, wie man sich auf der Stelle helfen kann, wenn man in Wien den Schild eines Hauses oder eines Kaufmannsgewölbes in und vor der Stadt suchen, und ihn finden will* 'Viennese Sign Register, or Instructions on How to Immediately Assist Oneself in Locating and Finding the Sign of a House or a Merchant's Vault in and Around the City of Vienna'

[10] 'Treasure / Protection / and Bulwark of the Archduchy of Austria / that is a very precise and orderly Description […] of the famous Capital and Imperial Residence City Vienna […]'

[11] In the second percentage, cases requiring precise identification of an arriving individual to determine their lodging are not included. This applies, for example, when someone stays with a relative such as father, mother, or brother-in-law.

which inns or private houses were particularly popular (cf. 'hot spots' on the heatmap of 6.2), where individuals went after entering through a specific gate (cf. visitors entering through Stubentor from August 1719 to August 1720 in 6.3), and how city gates differed in their frequency of use (cf. 6.4). These findings exemplify the broader utility of periodically published lists in historical research, as the presented workflow could potentially also be applied to other cities or other types of lists. Moreover, they demonstrate that including a 'small', previously undervalued text type can offer new perspectives and foster the development of innovative research questions and workflows.

5 Conclusion: lists in historical newspapers as 'small' texts with great potential

The paper has showcased a variety of potentials that lists found in historical newspapers have to offer for (Digital) Humanities research: they constitute a frequent, both spatially and temporally distributed phenomenon, enable diachronic analyses due to their periodicity, and are relevant for various disciplines due to their manifold forms and contents. Adding to this, they showcase a high density of named entities, a repetitiveness that promises to be useful for the study of language change, and an inherent semi-structuredness that forms an optimal prerequisite for automatic annotation approaches. In summary, periodically published lists prove to be 'small' texts with great potential and a valuable data treasure for the (Digital) Humanities.

To lift this treasure trove to some extent, the doctoral project aims, parallel to its threefold approach (cf. Chapters 2–4) and in alignment with CLARIN's mission of accessability, to make the developed resources as publicly available as possible. Specifically, the created datasets (e.g., facsimiles, full texts, NER data of arrival lists; gazetteer for early modern Vienna; analysis corpus) should be (re-)integrated into the CLARIN infrastructure by archiving them in the long-term repository and CLARIN-B centre ARCHE (ACDH-CH). Additionally, it is planned to publish parts of the developed workflow, such as the scripts created for NER, to facilitate future adaptation and reuse for other (semi-structured) historical texts.

Moreover, the open-ended 'list of lists', which is currently being compiled and should, among other things, include information on a list's title, its publishing medium and its digitisation status, may also serve as a valuable resource. In the end, periodically published lists in historical newspapers are not limited to text linguistic purposes, but may, depending on their content, also be of great interest for researchers in other fields, such as onomastics (e.g., toponyms, family names in lists), lexicology and orthography (e.g., repetition of the same lemmas over time), social and medical history (e.g., arrival lists, death lists), prosopography (lists of persons in general), economics (e.g., price lists), or theatre studies (e.g., lists of theatre performances). A publicly accessible 'list of lists' could thus serve as a central reference point for interested scholars and would certainly further contribute to the overall vision of the doctoral project: by empirically investigating periodically published lists in historical newspapers on a large scale, it aims to deepen our understanding of the underrepresented 'small' texts as well as to make them both (more) visible and (re-)usable in Digital Humanities research.

Acknowledgements

References

Adelmann, R. (2021). *Listen und Rankings*. Transcript.

ARCHE – A Resource Centre for Humanities. Austrian Centre for Digital Humanities and Cultural Heritage. https://arche.acdh.oeaw.ac.at (04.04.2024).

AustriaN Newspapers Online (ANNO). Austrian National Library. https://anno.onb.ac.at (04.04.2024).

Barton, R. A., Böckling, J., Link, S., & Rüggemeier, A. (2022). *Forms of List-Making: Epistemic, Literary, and Visual Enumeration*. Palgrave Macmillan.

Beautiful Soup. https://pypi.org/project/beautifulsoup4 (04.04.2024).

Belknap, R. E. (2004). *The List. The Uses and Pleasures of Cataloguing*. Yale University Press.

Bendel, S. (1998). *Werbeanzeigen von 1622–1798: Entstehung und Entwicklung einer Textsorte*. Niemeyer.

Blome, A. (2018). Zeitungen. In: Busse, L., Enderle, W., Hohls, R., Meyer, T., Prellwitz, J., & Schuhmann, A. (eds.), *Clio Guide – Ein Handbuch zu digitalen Ressourcen für die Geschichtswissenschaften*. 2nd edition. DOI: 10.18452/19244

Brand, P., & Schulze, V. (eds.) (1993). *Die Zeitung: ein Handbuch*. Institut für Medienentwicklung und Kommunikation.

Bubenhofer, N. (2020). *Visuelle Linguistik: Zur Genese, Funktion und Kategorisierung von Diagrammen in der Sprachwissenschaft*. De Gruyter.

Chinca, M., Eikelmann, M., Stolz, M., & Young, C. (eds.) (2022). *Sammeln als literarische Praxis im Mittelalter und in der Frühen Neuzeit. XXVI. Anglo-German Colloquium, Ascona 2019*. Narr Francke Attempto.

von Contzen, E. (2018). Experience, Affect, and Literary Lists. *Partial Answers* 16(2), 315–327.

von Contzen, E. (2017a). Die Affordanzen der Liste. *Zeitschrift für Literaturwissenschaft und Linguistik* 47, 317–326.

von Contzen, E. (2017b). Grenzfälle des Erzählens: Die Liste als einfache Form. In: Koschorke, A. (ed.), *Komplexität und Einfachheit. DFG-Symposion 2015*. J. B. Metzler, 221–239.

von Contzen, E. (2016). The Limits of Narration: Lists and Literary History. *Style* 50(3), 241–260.

Deutsches Textarchiv (DTA). Berlin-Brandenburgische Akademie der Wissenschaften. http://www.deutschestextarchiv.de (04.04.2024).

Diederichsen, D. (2021) (ed.). *Listen to lists*. Spector Books.

digiPress. Bayerische Staatsbibliothek. https://digipress.digitale-sammlungen.de (04.04.2024).

Digitales Forum Mittel- und Osteuropa (DiFMOE). Digitales Forum Mittel- und Osteuropa e.V. https://www.difmoe.eu (04.04.2024).

Doležalová, L. (ed.) (2009). *The charm of a list: From the Sumerians to computerised data processing*. Cambridge Scholars Publishing.

Dovifat, E. (1976). *Zeitungslehre*. 2 volumes. 6th edition, ed. by J. Wilke. De Gruyter.

Ehrenheim, A. (2011). *Das Textdesign der Stellenanzeige. Linguistisch und interdisziplinär*. Peter Lang.

Ehrmann, M., Hamdi, A., Pontes, E. L., Romanello, M., & Doucet, A. (2023). Named Entity Recognition and Classification in Historical Documents: A Survey. *ACM Computing Surveys* 56(2). https://doi.org/10.1145/3604931

Esposito, E. (2017). Organizing without Understanding: Lists in Ancient and in Digital Cultures. *Zeitschrift für Literaturwissenschaft und Linguistik* 47(3), 351–359.

Esposito, E., & Stark, D. (2019). What's Observed in a Rating? Rankings as Orientation in the Face of Uncertainty. *Theory, Culture & Society* 36(4), 3–26.

Farø, K. (2005). „Parkettfähige Deutsche sucht jantefreien Dänen". Eine kontrastive Untersuchung zu Lexik, Phraseologie und Interkultur in der Textsorte ‚Kontaktanzeige'. In: Lenk, H. E. H., & Chesterman, A. (eds.), *Pressetextsorten im Vergleich*. Olms, 261–282.

Fix, U. (2011). Leserbriefe. Die mediale Konstruktion von Diskursgemeinschaften. In: Held, G., & Stöckl, H. (eds.), *Pressetextsorten jenseits der "News": medienlinguistische Perspektiven auf journalistische Kreativität*. Peter Lang, 139–156.

Fix, U. (2008). *Texte und Textsorten – sprachliche, kommunikative und kulturelle Phänomene*. Frank & Timme.

Flair. https://github.com/flairNLP/flair (04.04.2024).

Fludernik, M. (2016). Descriptive Lists and List Descriptions. *Style* 50(3), 309–326.

Gius, E., Meister, J. C., Meister, M., Petris, M., Bruck, C., Jacke, J., Schumacher, M., Gerstorfer, D., Flüh, M., & Horstmann, J. (2022). *CATMA (6.5.0)*. https://doi.org/10.5281/zenodo.6419805

González-Gallardo, C.-E., Boros, E., Girdhar, N., Hamdi, A., Moreno, J. G., & Doucet, A. (2023). Yes but.. Can ChatGPT Identify Entities in Historical Documents? https://doi.org/10.48550/arXiv.2303.17322

Goody, J. (1977). *The domestication of the savage mind*. Cambridge University Press.

Haß-Zumkehr, U. (1998). *„Wie glaubwürdige Nachrichten versichert haben". Formulierungstraditionen in Zeitungsnachrichten des 17. bis 20. Jahrhunderts*. Narr.

Hausendorf, H. (2009). Kleine Texte. Über Randerscheinungen von Textualität. *Germanistik in der Schweiz. Onlinezeitschrift der Schweizerischen Akademischen Gesellschaft für Germanistik* 6, 5–19.

Herders Conversations-Lexikon (1854–1857). Digitised version (01/23). Wörterbuchnetz des Trier Center for Digital Humanities. https://www.woerterbuchnetz.de/Herder (04.04.2024).

HIST – Archive of the Historical Corpora. Leibniz-Institut für Deutsche Sprache. https://www2.ids-mannheim.de/cosmas2 (04.04.2024).

HuggingFace. https://huggingface.co (04.04.2024).

Impresso – Media Monitoring of the Past. https://impresso-project.ch/app (04.04.2024).

Jullien, François (ed.) (2004). *Die Kunst, Listen zu erstellen*. Merve.

Kampkaspar, D. (2019). Das DIGITARIUM – Volltexterstellung und Nutzungsmöglichkeiten. *Wiener Geschichtsblätter* 74(2), 131–135.

Kurzmann, K. (1999). *Entwicklungstendenzen in der Syntax der Anzeigenwerbung, untersucht an der Zeitung "Die Presse" (1850–1998)*. Master Thesis. University of Vienna.

Ledin, Per (2015). Listans och tabellens semiotik. *Sakprosa* 7(1), 1–25.

Lemke, M., & Wiedemann, G. (2015) (eds.). *Text Mining in den Sozialwissenschaften. Grundlagen und Anwendungen zwischen qualitativer und quantitativer Diskursanalyse*. Springer.

Lenk, H. E. H., & Chesterman, A. (eds.) (2005). *Pressetextsorten im Vergleich*. Olms.

Chronicling America – Historic American Newspapers. Library of Congress. https://www.loc.gov/collections/chronicling-america (04.04.2024).

Mader-Kratky, A., Resch, C., & Scheutz, M. (2019). Das Wien[n]erische Diarium im 18. Jahrhundert. Neue Sichtweisen auf ein Periodikum im Zeitalter der Digitalisierung. *Wiener Geschichtsblätter* 74(2), 93–113.

Mainberger, S. (2003). *Die Kunst des Aufzählens: Elemente zu einer Poetik des Enumerativen*. De Gruyter.

Mainberger, S. (2017). Exotisch – endotisch oder Georges Perec lernt von Sei Shonagon: Überlegungen zu Listen, Literatur und Ethnologie. *Zeitschrift für Literaturwissenschaft und Linguistik* 47(3), 327–350.

Mainberger, S. (2018). A List Is (Not) a List: Some (Non-Conclusive) Remarks on Lists and Literature. *Literatur in Wissenschaft und Unterricht* 1, 8–21.

Mirador. https://projectmirador.org/ (04.04.2024).

Moretti, F. (2013). *Distant Reading*. Verso.

Mueller, M. (2020). *Scalable Reading*. https://sites.northwestern.edu/scalablereading/2020/04/26/scalable-reading (04.04.2024).

Müller, C. (2016). ANNO – Der digitale Zeitungslesesaal der Österreichischen Nationalbibliothek. *Bibliothek: Forschung und Praxis* 40(1), 83–89.

Nicholson, B. (2013). The Digital Turn. Exploring the methodological possibilities of digital newspaper archives. *Media History* 19(1), 59–73.

OpenAI API. OpenAI. https://openai.com/blog/openai-api (04.04.2024).

Oxygen XML editor. Syncro Soft. https://www.oxygenxml.com (04.04.2024).

Pfefferkorn, O., Riecke, J., & Schuster, B.-M. (2017). *Die Zeitung als Medium in der neueren Sprachgeschichte. Korpora – Analyse – Wirkung*. De Gruyter.

Promptify. Promptslab. https://github.com/promptslab/Promptify (04.04.2024).

QGIS. https://www.qgis.org (04.04.2024).

Rastinger, N. C. (in print). An der Schnittstelle von Sprache und Bild: Drucktypenwechsel zwischen Antiqua und Fraktur in historischen Periodika. In: Rammerstorfer, L., & Stiemer, H. (eds.): *Re-Writing the Past in the Digital Age: Perspektiven und Methoden der Journalforschung*. De Gruyter.

Resch, C., & Kampkaspar, D. (eds.). *Wien[n]erisches DIGITARIUM*. https://digitarium.acdh.oeaw.ac.at (04.04.2024).

Resch, C., & Kampkaspar, D. (2020). *German Fraktur 18th Century – WrDiarium_M9*. https://readcoop.eu/de/modelle/german-fraktur-18th-century/

Romberg, M., Kaiser, M., & Schlögl, M. (eds.). *ViecPro*. https://viecpro.oeaw.ac.at (04.04.2024).

Rüggemeier, A. (2020). Life Writing and the Poetics of List-Making: On the Manifestations, Effects, and Possible Uses of Lists in Life Writing. *Auto/Biography Studies* 36, 183–194.

Schaffrick, M., & Werber, N. (2017). Die Liste, paradigmatisch. *Zeitschrift für Literaturwissenschaft und Linguistik* 47, 303–316.

SpaCy. https://spacy.io (04.04.2024).

Stäcker, T., Müller, M., & Kampkaspar, D. (eds.). *Eine Darmstädter Zeitung in drei Jahrhunderten – Digitalisierung des Darmstädter Tagblatts (1740–1986).* https://www.ulb.tu-darmstadt.de/forschen_publizieren/forschen/darmstaedter_tagblatt.de.jsp (04.04.2024).

Straßner, E. (2000). *Journalistische Texte.* Niemeyer.

Tankard, P. (2006). Reading lists. *Prose Studies* 28(3), 337–360.

Temmerman, M., & Vandenabelee, D. (2018). Listicles: A journalistic genre for digital media. *Language and the New (Instant) Media*, 149–158.

Teßmann digital. Landesbibliothek Dr. Friedrich Teßmann. https://digital.tessmann.it (04.04.2024).

Torget, A. J. (2023). Mapping Texts: Examining the Effects of OCR Noise on Historical Newspaper Collections. In: Bunout, E., Ehrmann, M., & Clavert, Frédéric (eds.), *Digitised Newspapers – A New Eldorado for Historians? Reflections on Tools, Methods and Epistemology*. De Gruyter, 47–66.

Transkribus. READ-COOP. https://www.transkribus.org (04.04.2024).

von Trattner, J. A. (ed.) (1773). *Verzeichniß der in der k.k. Haupt= und Residenz=Stadt Wien befindlichen Gassen Hausinnhabern deren Schildern und numerirten Häusern.* Wienbibliothek im Rathaus. https://resolver.obvsg.at/urn:nbn:at:AT-WBR-8402

Universal Viewer. https://universalviewer.io (04.04.2024).

Vauth, M., Meister, M., Hatzel, H. O., Gerstorfer, D., & Gius, E. (2022). *GitMA (1.4.9).* https://doi.org/10.5281/zenodo.6330464

Verlag der Expedition des Wienerjournals (ed.) (est. 1795). *Wiener Schildregister, oder Anweisung, wie man sich auf der Stelle helfen kann, wenn man in Wien den Schild eines Hauses oder eines Kaufmannsgewölbes in und vor der Stadt suchen, und ihn finden will.* Wienbibliothek im Rathaus. https://resolver.obvsg.at/urn:nbn:at:AT-WBR-8761

Vismann, C. (2000). *Akten. Medientechnik und Recht.* Fischer.

Waldispühl, M. (2019). Die Liste als Ordnungsmedium im mittelalterlichen Libri vitae. *Zeitschrift für Literaturwissenschaft und Linguistik* 49(2), 197–218.

Wien Geschichte Wiki. City of Vienna. https://www.geschichtewiki.wien.gv.at (04.04.2024).

Zeit.punktNRW. Universitäts- und Landesbibliothek Bonn, Universitäts- und Landesbibliothek Münster, Universitäts- und Landesbibliothek Düsseldorf, Hochschulbibliothekszentrum des Landes Nordrhein-Westfalen. https://zeitpunkt.nrw (04.04.2024).

Zeitungsinformationssystem (ZEFYS). Staatsbibliothek zu Berlin. https://zefys.staatsbibliothek-berlin.de (04.04.2024).