Word Usage in German Texts on Women's Suffrage around 1900. Corpus Building, Lexical Documentation and the CLARIN-D Infrastructure

Anna Pfundt
Justus Liebig University
Giessen, Germany
`anna.pfundt@`
`germanistik.uni-gies-`
`sen.de`

Melanie Grumt Suárez
Eberhard Karls University
Tübingen, Germany
`melanie.grumt-suarez@`
`uni-tuebingen.de`

Thomas Gloning
Justus Liebig University
Giessen, Germany
`thomas.gloning@`
`germanistik.uni-gies-`
`sen.de`

Abstract

The paper presents core aspects of a project that examines word usage in the German controversy on women's suffrage around 1900. The investigation is based on a variety of written texts (including journal articles, books, and pamphlets) that began to condense in the 1880s and developed into a complex thematic network until the introduction of women's suffrage in 1918. The focus of this paper is the presentation of two infrastructure aspects. First, the corpus compilation (ongoing and already published to some extent) is accomplished in the CLARIN-D infrastructure component German Text Archive ("Deutsches Textarchiv", hereafter DTA). This project exemplifies a basic user need: to make new corpus texts available from the very beginning of a project which can then be analyzed for the project's purposes with the powerful search tool architecture of the DTA. This kind of strategy is a win-win situation for the author, for the infrastructure, and for the whole research community. Second, the dissertation and the text corpus will be accompanied by a lexical documentation on word usage in the discourse domain. This component relies heavily on standards and best practices adopted in the CLARIN-D community and in the ZHistLex-project. The results are to be published and hosted by one of the German CLARIN centers. This project is a use case with solutions that can be seen as generic for similar kinds of projects, be it dissertations, term papers or larger research projects.

1 The First Women's Movement in Germany and its discourse cosmos[1]

The so-called First Women's Movement in Germany[2] started around the middle of the 19th century and ended in the late 1920s. It was not a unified movement but rather a complex and dynamic configuration of groups with different positions and different strategies. They shared, however, the aim to improve the situation of women, to enhance the spectrum of their political rights and the possibility to participate in public life, to study at universities, to earn a living on their own, to choose specific occupations. There were other topics and aims, e.g. the professional situation of prostitutes, sexual morals and the situation of working class women, which were the main topics of specific groups. The improvement of the situation of working class women, e.g., was one of the aims of the proletarian women's movement with Clara Zetkin (1857-1933) as a leading figure. Here, the fight for women's rights is deeply embedded in the overall struggle for the rights of the working class. The bourgeois wing of the First Women's Movement

---

[1]By "discourse cosmos" we mean the complex and dynamically evolving ensemble of communicative contributions that belong to a discourse topic and its sub-topics which often are organized in specific "threads".

[2]See Frevert (1995); Gerhard (1990); Gerhard (2009); Hervé (1988); Rosenbusch (1998); Schaser (2006); Twellmann (1972).

in Germany had its leading figures as well, e.g. Helene Lange (1848-1930) or Anita Augspurg (1857-1943). From 1865 onwards, these groups began to become organized, e.g. in the "Allgemeiner Deutscher Frauenverein" and in other bodies of a more local or general nature. All these developments were part of a wider debate that started with the French Revolution and its ideas of equality including the equality of women and men. The German debates were related to debates in other languages and in other countries, both in respect of central ideas and in respect of textual interlinking.

The First Women's Movement brought about a dynamic discourse cosmos, the structure of which can be analyzed along main topics, discourse threads, actors, and a specific configuration of media and text types. Among the first topics to establish specific discourse threads were the issues of equal or at least improved education for women and the right to work for a living. Only later, by the end of the 19th century, the topics of political participation and specifically the right to vote brought about new discourse threads. The protagonists used a wide range of media, text types and oral genres for their communicative purposes, e.g. petitions to parliaments, position papers in support of petitions, printed pamphlets, journal articles, personal letters, open letters, speeches, conference discussions, and others. One of the most important linguistic research tasks is to show, how all these communicative genres and the linguistic means used in these debates were used in specific ways to support the relevant political aims.

Apart from argumentative strategies, word usage is one of the most important aspects of language use in the contributions to the discourse cosmos at hand. Like in modern discourse threads, e.g. those on abortion, nuclear energy or immigration, an important part of word usage in the discourse cosmos of the First Women's Movement is closely connected to specific positions, to communicative tasks, and to specific strategies. Word formation, metaphor, the use of topic-specific vocabulary or semantic innovation are but a few of the aspects that belong to the research tasks for the project that we will now describe.

2 Word usage in the German controversies over women's suffrage 1870-1918: A research project and its infrastructure support

The subject of Anna Pfundt's dissertation project[3] is the role of word usage in the discussions about women's suffrage from the 1870s until 1918. This discourse was taking place in a vast spectrum of texts (including journal articles, books, and controversy texts) that began to condense in the 1880s and then developed a complex discourse network until the introduction of women's suffrage in Germany in 1918. The discussion of this specific topic is an important part of a broad discourse on a widespread spectrum of controversial points brought about by the First Women's Movement in Germany, especially the aspect of political participation. It is an important aspect of this research task to include quite different positions that were promoted in the fight for women's suffrage by women groups of different convictions and strata, e.g. the champions of the proletarian party vs. different shades of bourgeois positions.

Word usage plays a central role in the constitution of points of view, in the formulation of views and their justification, as has often been the case in discourses about alternative and competitive word usages (see Pfundt, 2017; Gloning, 2012). The use of words is characterised by specific thematic profiles, but also by the controversial nature of the object and by aspects of historical development over several decades. In addition to the specific uses ("senses") of words and phrases, forms of word formation, metaphors, ad hoc uses, foreign-language expressions and functionally oriented vocabulary are also part of the study. Another important aspect for the project is to examine the connection of the use of words specific to different "camps" and their perspectives and opinions.

On the one hand, the research strategy is methodologically based on the work of the Düsseldorf School of discourse studies around Dietrich Busse, Georg Stötzel and Martin Wengeler (e.g. "Kontroverse Begriffe", ed. by Stötzel and Wengeler, 1995). In line with this methodology, the research is characterised by a discourse historiography in which the thematic developments of a discourse are reconstructed in narrative form in connection with the description of the lexical means used. On the other hand, the project of Anna Pfundt goes beyond this methodology in two respects: First, within the project the corpus texts are made available publicly (an ongoing task), second, word usage will be documented in a specific lexicographic component, which will be closely connected to the dissertation and to the corpus texts by way of structural or digital interlinking.

---

[3]See Pfundt (2017).

Here is an example for this kind of interlinking of dissertation, text corpus and lexical documentation: In the investigation, the narrative presentation of the discussion and the use of words anchored in it is based on a corpus of German texts on early women's suffrage, the texts of which are successively fed into the German Text Archive (DTA). By way of illustration we present a short example from an early text[4] by Hedwig Dohm (1831-1919), one of the earliest champions of women's suffrage in Germany, where she comments on the assumption of different spheres of men and women in a very critical way:

> O über dieses Geschwätz von der **Sphäre** des Weibes, den Millionen Frauen gegenüber, die auf Feld und Wiese, in Fabriken, auf den Straßen und in Bergwerken, hinter Ladentischen und in Bureaus im Schweiße ihres Angesichts ihr Brot erwerben.

> Wenn die Männer vom weiblichen Geschlecht sprechen, so haben sie dabei nur eine ganz bestimmte Klasse von Frauen im Sinn: Die Dame. Wie nach dem bekannten Ausspruch jenes bekannten österreichischen Edelmannes der Mensch erst bei dem Baron anfängt, so fängt bei den Männern das weibliche Geschlecht erst da an, wo es Toilette und Conversation macht und Hang zu Liebesintriguen und Theaterlogen verräth.

> Geht auf die Felder und in die Fabriken und predigt eure **Sphärentheorie** den Weibern, die die Mistgabel führen und denen, deren Rücken sich gekrümmt hat unter der Wucht centnerschwerer Lasten! (Dohm, 1876, p. 126-127)

In this passage, Hedwig Dohm harshly criticizes the assumption of different spheres of men and women, an assumption which was common in 19th century thinking and which nowadays is called "difference hypothesis" (Differenzannahme, Differenzhypothese).

This passage can illustrate three integrated components of the dissertation. First, the whole text of Hedwig Dohm is now available in CLARIN-D's German Text Archive (DTA) at the Berlin-Brandenburg Academy CLARIN center in a digital format. As such it is not only available for download and Open Access in different formats, it can also be used with the powerful search facilities that are provided by the DTA. Second, this passage will be commented on and analyzed in the dissertation, especially in the chapter on the "difference hypothesis" and its role in the debates on women's suffrage around 1900. Third, the component "Lexical documentation" will contain dictionary-like entries with lemmata like "Sphäre", "Sphärentheorie", "Sphärenanbeter", where the peculiar aspects of the usage of these words in their discourse contexts are analyzed and documented. We call this documentation "dictionary-like", because the organization follows the patterns of article structuring that is well-known from traditional dictionary making. There are, however, components that go well beyond traditional dictionary making, e.g. in providing an underlying "discourse ontology" by a specific system of markers, which allows to search for lexical items that are related e.g. to the "difference hypothesis". This third component, the extensive discourse dictionary, currently contains almost 1100 entries (see section 4, below).

The three components (dissertation/investigation, digital text corpus, digital discourse dictionary) and their interrelations are shown schematically[5] in the following figure (Wolff/Geyken/Gloning, 2015 fig. 6). The field in the upper part of the figure represents a (future) passage of the investigation component on the "difference hypothesis", below on the left is the relevant passage from one of the corpus texts (there will be others), below on the right is an example that represents the component "lexical documentation". The colors indicate structural connections that will be digitally implemented in a later stage of the project.

---

[4]http://www.deutschestextarchiv.de/book/show/dohm_frauenfrage_1876?p=134 (last access 2020-05-03).

[5]The figure is not a suggestion for screen design but rather a visualization of the structural connections between investigation, dictionary and corpus texts.

Figure 1: Interrelations between investigation, text corpus and discourse dictionary.

With regard to the text corpus component, a win-win situation arises for the author of the dissertation and for the CLARIN-D infrastructure component DTA. Every new contribution increases the textual basis, which can be analyzed with the powerful search tools[6] of the DTA. For the DTA, each of these texts is a thematic enrichment and – with regard to texts written by women – also a contribution to increasing the proportion of female authors. This is what we will now outline in more detail in the following section 3. In the subsequent chapter 4 we will expand on the infrastructure aspect of the component "lexical documentation".

3 Infrastructure I: The TdeF-Corpus in CLARIN's German Text Archive

The "Deutsches Textarchiv"/German Text Archive, henceforth DTA, is one of the most important infrastructure components within CLARIN-D.[7] It is hosted and managed by the Berlin-Brandenburg Academy of Sciences and Humanities and basically serves all research communities that work with texts from the 17th to the 19th centuries. These communities include German studies, especially literary studies, historical linguistics and lexicography, the study of intellectual history, discourse history, the history of terminologies to name but a few. The core corpus is balanced according to groups of text types and to subject fields over the decades. The organization of the core corpus is systematically documented on the DTA website.[8]

With respect to specific user needs and projects, the DTA allows the integration of new texts and text collections in a section with additions: "Erweiterungen des Deutschen Textarchivs" (DTAE)/Additions to the German Text Archive. Users can submit their texts together with relevant metadata which have to conform to the criteria of DTA's basic format of text encoding and meta data organization. These textual additions are ascribed to their producers which is an important point in respect of the metrics of academic success. But more importantly, these additional texts can be used as sub-corpora by using a

---

[6] http://www.deutschestextarchiv.de/doku/DDC-suche_hilfe (last access 2020-05-03).
http://odo.dwds.de/~moocow/software/ddc/querydoc.html (last access 2020-05-03).

[7] http://www.deutschestextarchiv.de/ (last access 2020-05-03).

[8] http://www.deutschestextarchiv.de/doku/textquellen#dta-kernkorpus (last access 2020-05-03).

specific element of the query syntax of the DTA. This option is currently used for building up a special corpus with German texts from the First Women's Movement.

This sub-corpus has the title "Texte der ersten Frauenbewegung"/German texts of the First Women's Movement, hereafter TdeF.[9] At present, the focus lies on texts that deal with the topic of women's suffrage, since this is the subject of Anna Pfundt's dissertation. It is planned, however, to subsequently enrich the textual substance over the years with contributions on other topics as well, e.g. education of women, access to university studies or the situation of working class women. Hence, the current TdeF corpus has a double function: it serves as an empirical basis for the investigations on word usage in Anna Pfundt's research project, but it is also the nucleus for a fully-fledged corpus of First Women's Movement texts in German.

At present, it includes 70 texts with 2481 pages, that are already published. This may not constitute a huge corpus yet, but it *is* a valuable contribution of texts written by women and dealing with highly important topics from the debates on the emancipation of women. The corpus is focused on the period 1870-1919, since the first texts on the issue of women's suffrage appeared around 1870 and women were granted the right to vote in 1918. The central media of the time were books, magazines and newspapers; therefore, the corpus texts are independent works in the form of books, brochures or lectures as well as magazine articles from the press organs of the First Women's Movement and newspaper articles. The spectrum of authors is manifold: opponents and supporters of the women's movement, individual persons and representatives of organizations, women and men (e.g. August Bebel) of different ages, some authors remain anonymous.

In order to create and prepare the texts as searchable full texts in a sub-corpus and to make them publicly accessible, they are digitized according to the specifications of the DTA base format.[10] The DTA base format has become a quasi-standard for the encoding of texts from the 17th century onwards. For lexical analysis, the texts in the sub-corpus TdeF can be addressed as a separate text group `for` queries. In the first step, the texts are either produced with ABBYY Recognition Server which is able to deal with blackletter typefaces, or they are keyed in manually depending on the quality and complexity of the images. Still in other cases transcriptions on the web can be used as a starting point. In all cases meticulous proofreading is necessary. In the second step, the texts are structurally annotated according to the DTA principles stated in the base format. As the third step, the relevant metadata are provided, equally following the DTA guidelines.

The following figure shows the start screen with Hedwig Dohm's text "Der Frauen Natur und Recht" (1876; 'On the nature and the rights of women'). From here, different user options are possible, ranging from download in different formats, online reading in different modes or using the search facilities and the visualization tools of the DTA environment.

---

[9]`http://www.deutschestextarchiv.de/doku/textquellen#tdef` (last access 2020-05-03).

[10]`http://www.deutschestextarchiv.de/doku/basisformat` (last access 2020-05-03). The DTA base format is now widely used within the community. The use of oXygen for work on historical texts is supported by specific DTA tutorials and documentation.
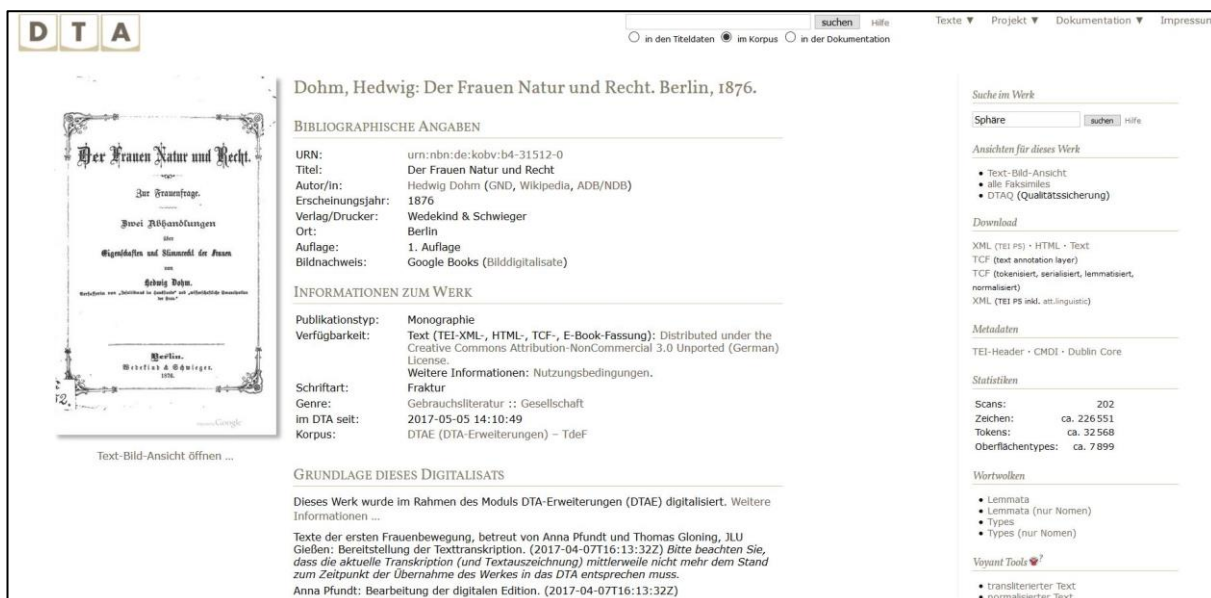
Figure 2: Hedwig Dohm: *Der Frauen Natur und Recht* (1876). DTAE start screen.
`http://www.deutschestextarchiv.de/book/show/dohm_frauenfrage_1876`

Once the texts are available in the DTA they can be addressed as a specific sub-corpus with a broad range of query options. The options that are most relevant for our research project on word usage include: the search for specific word-forms, the search for lemmatized words (which will provide all the relevant word-forms), the search for word-formations and their elements (e.g. word-formations with "Sphäre"), the search for derivational morphemes and the word-formations they produce, the search for different parts of speech. These options can be combined, e.g. adjectives formed with "-lich", they can also be combined with metadata, e.g. "Adjectives in the TdeF sub-corpus before 1900". An important feature is the opportunity to extract quotations with their metadata (author, work title, date), which are important components of the articles in the lexical documentation system.

As we have mentioned previously: providing source texts from a research project in an ongoing way for the DTA and thus building up a steadily growing project-specific sub-corpus *within* the DTA leads to a win/win/win-constellation for at least three parties: firstly, for the individual researcher and his or her project, who can take advantage of the powerful search functionalities of the DTA's query machine; secondly, for the DTA, because it steadily increases its holdings and gets further texts from specific areas (e.g. discourse topics, author groups, text types); thirdly, other researchers with different research interest can take advantage of the newly available texts as well, e.g. a person studying word formation in German adjectives or someone doing research on the evolution of text types, independent research questions that are not related to the First Women's Movement in Germany.

4 Infrastructure II: Lexical Documentation and CLARIN-D/ZHistLex

The digital lexical documentation system fulfils three specific functions within the overall structure of the project: single word documentation; providing an underlying discourse ontology across different dimensions; to allow for interlinking between the text of the investigation and the lexical documentation system.

Firstly, the lexical documentation system is supposed to document words and the usages of words that are relevant for the discourse topic under study (women's suffrage). As mentioned before, words like "Sphäre" oder "Sphärentheorie" are used in specific ways and for specific purposes in the debates. For this purpose, the traditional form of representation is the word article with its structured positions (e.g. lemma, explanation of meaning, quotations) as we know it from dictionaries. Hence, the structural backbone of the lexical documentation system consists of articles which document the use of words and their specific contribution to the discourse on women's suffrage. "To document" in this respect means to explain and to characterize, how the words are used, and to give quotations from the spectrum of

texts. For this task, the TEI Lex-0[11] annotation scheme is used. It is a standard which is currently adopted both in CLARIN-D contexts and in the ZHistLex[12] project which was funded by Germany's Federal Ministry of Education and Research (BMBF).[13] The article components and information positions include: lemma, definitions or other kinds of semantic explanation, comments on specific functions in the discourse, references to and comparisons with the information in the standard historical dictionaries of German, if applicable information on word history and/or borrowing, cross-references, finally a block with quotations together with their metadata. If there are different usages (senses), these components are repeated for each sense. The arrangement of the articles is strictly alphabetic, but it is clear that electronic access in a digital system will include other forms of access as well.

Secondly, the lexical documentation makes use of an underlying discourse ontology across different dimensions. The basic idea here is, that historical vocabularies are organized along different criteria and that they build up multi-dimensional structures that evolve over time (cf. Gloning, 2003). A word like "Sphärenanbeter" ('person that worships the idea that men and women have different spheres, tasks, and rights in life'), e.g. may be characterized in respect of its part of speech, its pattern of word formation, its communicative functions (referring to a type of person; expressing a negative/critical assessment), its place in the system of ideas of the First Women's Movement (here the so-called *Differenzhypothese*) etc. If these aspects are marked up in a systematic and consistent way, the documentation allows for combinatorial search, e.g. the search for words with which persons are negatively evaluated, or the search for compounds of the type N+N that have something to do with the difference hypothesis. For this task, the parts of the ontology that are specific to the discourse topic at hand, must be the result of the analysis of the texts. There are no preexistent discourse ontologies.

Thirdly, the components of the lexical documentation system are interrelated to specific passages of the investigation and its sub-chapters. Since it is the main objective of the research project to investigate different aspects of word usage in all its complexities, there will be specific chapters on these aspects of their relations, e.g. the role of word formation, the use of metaphor, the role of loanwords, the use of creative language use, the contribution of word usage to communicative aims of the authors, to name only the principal ones here. In the chapters of the dissertation it will not be possible to list all and to comment on all the words that are relevant for a point in question. For instance, in the chapter about the role of borrowing and loanwords it is not possible to list all the words and components from the Latin tradition. But it is possible in that chapter to provide the complex search string by which these words can be retrieved from the lexical documentation system.

Now we will give an example of the article structure and its TEI Lex-0 encoding. For this purpose, we present the articles "Sphäre" and "Sphärenanbeter", words that are used in the passage from Hedwig Dohm's book, quoted above. For the main part of the article the components are pretty straightforward and present no unexpected structures. The only peculiarity is the (mis-)use of the tag <usg> for implementing the discourse ontology and the analytical markup for different aspects of the word usages in question. We place this tag within the <sense> tag because the values of the analytical markup are quite often relative to a specific usage (sense) and not to the word with its multiple senses. Figure 3 below shows the article "Sphäre" in oXygen's author view:

---

[11]The last character is zero, which in the current font almost looks like an "o"; TEI Lex-0 is a new suggestion for a lexicographical baseline encoding format. See:
`https://dariah-eric.github.io/lexicalresources/pages/TEILex0/TEILex0.html` (last access 2020-05-03).

[12]`https://zhistlex.de` (last access 2020-05-03).

[13]`https://www.bmbf.de/en/index.html` (last access 2020-05-03).

Figure 3: Article "Sphäre"; oXygen's author view.

The following figure 4 shows the somewhat less complex article "Sphärenanbeter" in oXygen's text view with the XML encoding:

```
24083   <form type="lemma"><orth>Sphärenanbeter*</orth></form>
24084 ▽ <sense>
24085 ▽ <usg>Abwertende Bezeichnung für eine Person, die für die Auffassung entsprechend der
24086   Differenzhypothese eintritt, nach der dem männlichen und weiblichen Geschlecht jeweils
24087   unterschiedliche Bereiche der Tätigkeit bzw. Wirkung zugeordnet werden.</usg>
24088       <usg type="sichtweisen">Differenzhypothese</usg>
24089       <usg type="funktionen">Abwertung</usg>
24090       <usg type="wortbildung">N+N</usg>
24091 ▽ <cit type="quotation">
24092   <bibl><date>(1876)</date><author>Dohm</author><title>Natur</title><biblScope>127.10</biblScope></bibl>
24093 ▽ <quote>Geht auf die Felder und in die Fabriken und predigt eure Sphärentheorie den Weibern, die die
24094   Mistgabel führen und denen, deren Rücken sich gekrümmt hat unter der Wucht centnerschwerer Lasten!
24095   Könnt ihr allen Frauen ein behagliches Daheim schaffen und einen Mann, der für sie sorge? Nein - ihr
24096   könnt es nicht. Seid ihr <mentioned>Sphärenanbeter</mentioned> auch alle, alle verheirathet und habt
24097   ihr allesammt arme Mädchen geheirathet, um der Versorgung des weiblichen Geschlechts Rechnung zu
24098   tragen? Nein, ihr habt es nicht gethan. Nun denn, aus dem Wege mit Euch, Ihr Sphärenfabrikanten,
24099   gebt Raum und Luft für die Millionen, an Geist und Körper gesund gebornen Geschöpfe, die da
24100   verkümmern, weil sie Frauen sind!</quote>
24101   <bibl><biblScope>(127.4f.)</biblScope></bibl>
24102   </cit>
24103   </sense>
24104   </entry>
```

Figure 4: Article "Sphärenanbeter"; XML structure of the entry, oXygen's text view.

With this kind of analytical markup it is possible to search e.g. for entries that were marked in the following ways:

    <usg type="sichtweisen">Differenzhypothese</usg>
    <usg type="funktionen">Abwertung</usg>
    <usg type="wortbildung">N+N</usg>

To search, e.g., for all entries which are marked to belong to "sichtweise=Differenzhypothese" (point of view=difference hypothesis) one can use oXygen's xpath query syntax:[14]

    //entry[contains(.//usg[@type="sichtweisen"], "Differenzhypothese")]

---

[14]In a later stage the data will be hosted in a web environment that is able to process xPath-queries and to deliver the results in different formats. Furthermore, it has been pointed out that the "contains"-operator is not the ideal solution for our purpose, compared, e.g., to strict match options. Therefore, we will modify this practice.

The search result will include three entries, which refer to the articles "Sphäre", "Sphärenanbeter", and "Sphärenfabrikant", all related to the difference hypothesis, that is the point of view that women and men by nature belong to different spheres and therefore have different rights and obligations:

/TEI[1]/text[1]/body[1]/div[1]/entry[852]	"Sphäre"
/TEI[1]/text[1]/body[1]/div[1]/entry[853]	"Sphärenanbeter"
/TEI[1]/text[1]/body[1]/div[1]/entry[852]	"Sphärenfabrikant"

To sum up: The lexical documentation system on the one hand serves to analyze and to document individual word usages, on the other hand the system may be addressed from passages in the investigation in order to present further details, to provide the entire subset of relevant material and the textual quotations in their temporal order. The lexical entries are compiled to provide a pragmatic-semantic description of the individual lexical units and their specific uses in the discourse. Thus, the lexical documentation system provides a lexical inventory of the discourse. It is itself based on the TdeF-corpus. The lexical documentation system is to be published in the ZHistLex project by one of the CLARIN-D partners as well.

5   The Generic Potential of Our Infrastructure Strategies

We believe that the research project on word usage in the women's suffrage discourse has a generic potential in at least three important dimensions:

Firstly, we believe that the integrated coordination of three pillars in a project on lexical research is a fruitful strategy. It combines (a) overall analytical results on vocabulary structure and development (e.g. in a dissertation) with (b) the public presentation of the digital corpus texts that are the basis of these results and (c) a digital lexical documentation system that provides the details of word usage and lexical organisation.

Secondly, we believe that for texts from the 17th to the early 20th centuries the cooperation with CLARIN-D's German Text Archive (DTA) belongs to the best practices for building up a project specific digital corpus, which we have characterized as a win/win/win-situation for the research project, the DTA and also for other researchers with interests of their own.

Thirdly, the conception of the lexical documentation system with its underlying markup system for multiple dimensions of word usage and vocabulary structure allows to retrieve lexical material according to different criteria (e.g. PoS, communicative function, discourse role, connection to specific points of view in a discourse), to visualize different aspects of vocabulary structure and development, it provides the foundations for writing up the results on specific aspects of vocabulary structure and development in the book or article one has to write. Since there has been a long and controversial debate about the role of alphabetical order[15] and its alternatives[16] in the history of lexicography, the present suggestion for multi-dimensional organization and access via a multi-dimensional markup based on TEI Lex-0 and XPath may well contribute a low-level solution to this controversial point with a long history.

Currently, two other Gießen based projects adopt this kind of research architecture. Andre Pietsch works on word usage and vocabulary structure in texts related to early film and early cinema. Thomas Gloning works on the history and the development of German jazz vocabulary. [17] In respect of the jazz

---

[15]Jacob Grimm, one of the principal founders of German studies, wrote in the preface to vol. I. of the "Deutsches Wörterbuch" (1854, p. XI): "Nicht minder nothwendig ist dem wörterbuch die alphabetische ordnung und sowol die möglichkeit des vollen eintrags und der abfassung als die sicherheit und schnelle des gebrauchs hängen davon ab. wer reiche beiträge einschalten will, musz die stelle wohin vor augen haben und nicht unschlüssig herum zu suchen, ob das wort schon da sei oder fehle: die biene weisz genau die zelle, zu welcher sie honig einträgt. es würde die arbeit in den wörtern aufheben oder lähmen, wenn man den platz nicht kennt, aus dem sie zu holen sind. schon ihren eingeschränkten samlungen pflegten die alten diese alphabetfolge zum grunde zu legen und wer sie heute nicht handhabt, sondern aufhebt und stört, hat sich an der philologie versündigt." The preface is online via woerterbuchnetz.de:
`https://tinyurl.com/yaacpscv` or `http://woerterbuchnetz.de/cgi-bin/WBNetz/wbgui_py?sigle=DWB&mainmode=Vorworte&file=vor01_html#abs2` (p. XI) (last access 2020-05-08).
[16]E.g. "conceptual" systems like the one proposed by Hallig and von Wartburg (1952).
[17]This project has been presented at the Göttingen Conference on Historical Lexicography and slightly revised at the 2019 Germanistentag, both in September 2019. The slides are available here:
`https://zhistlex.de/folien/Gloning_2019_HistVok-Jazz_Saarbruecken.pptx`.

vocabulary project the lexical documentation system is the core component. The question, whether or not a digital text corpus is feasible at all, is not yet answered. Copyright questions are hard questions when it comes to texts from the 20th and the 21st centuries. Here, the expertise of CLARIN-D might be fruitful, too. Figure 5 shows an early version of the article "Baritonsaxofonist".

Figure 5: Current, preliminary state of article "Baritonsaxofonist"; Jazz Vocabulary Project.

One last point on the generic potential of multidimensional background markup. What has been suggested on forms of markup like "points of view=difference hypothesis" in the women's suffrage discourse can be adapted to dictionary retrodigitisation projects. While many traditional dictionaries have been digitized and made available in lexicographic portals like woerterbuchnetz.de, some of them are not yet available for digital access, e.g. the "Sudetendeutsches Wörterbuch".[18] Going through the articles of this dictionary one finds many articles that are closely related to a specific cultural background and to aspects of the form of life that is mirrored in the regional vocabulary structure, e.g. in the use of expression for church feasts ("Peter Ketten"), for plant and animal names (e.g. "Brotkäferlein"), evaluative expressions for persons ("Prahl-fotze" 'boasting female person') and others. It is obvious that such aspects of vocabulary structure would deserve additional markup in an electronic version.

6   Results and Open Questions

In this paper, we have portrayed a research project with a lexical focus that opens up a number of generic perspectives for the CLARIN-D infrastructure. Anna Pfundt's research project aims to analyze and describe the vocabulary structure and developments that are connected with the textual cosmos of the discourse on women's suffrage within the First Women's Movement in Germany around 1900. Her work comprises three major pillars: (1) the dissertation as an analytical prose text, which will be available both in print and in digital form; (2) a digital text corpus which is built up in an ongoing way within CLARIN's German Text Archive and which at present comprises around 70 full texts, (3) a digital lexical documentation system, which documents the details of word usages but which allows interlinking to other word articles, to the corpus texts and to the text of the dissertation. This lexical documentation component is to be published with one of the German CLARIN centers as well.

By now, it should have become clear that the integrated combination of (1) investigation, (2) digital text corpus and (3) digital lexical documentation system provides a fruitful architecture for research questions like word usage in the debates on women's suffrage within the First Women's Movement (or word usage related to early film and cinema; word usage related to the development of jazz in German). This research architecture provides generic solutions that are closely related to the CLARIN-D infrastructure: building up a project-specific sub-corpus within DTAE is certainly the most important aspect, the support in respect of lexical organization and the respective standards and best practices are equally important.

As a prototypical use case, the project shows how a lexical investigation can simultaneously benefit from the resources of the CLARIN-D infrastructure (here the DTA) and contribute to its further expansion. The compilation of the TdeF corpus is at the same time the textual basis for the research project

---

[18]https://www.uni-giessen.de/fbz/fb05/germanistik/forschung/sprache/sdwoerterbuch; https://de.wikipedia.org/wiki/Sudetendeutsches_W%C3%B6rterbuch (last access 2020-05-03).

on word usage in the discourse on women's suffrage and an important contribution to expand the holdings of the DTA in an important phase of German language and discourse history around 1900.

Finally, once a text is integrated in the DTA, it is also available for a broad spectrum of other research questions, other research methods and tools. Apart from the contributing project and the DTA, the win-win situation includes the whole research community and a plethora of possible research questions and methods.

One last important aspect of CLARIN-D's infrastructure must not be forgotten: the ongoing personal support we have experienced over the years, notably from the DTA team.

Acknowledgements

References

Hedwig Dohm. 1876. Der Frauen Natur und Recht. Wedekind & Schwieger, Berlin. http://deutschestextarchiv.de/book/show/dohm_frauenfrage_1876 (last access 2020-02-09).

Ute Frevert. 1995. „Mann und Weib, und Weib und Mann". Geschlechter-Differenzen in der Moderne. Beck, München.

Gerd Fritz. 2006. *Historische Semantik*. second, updated edition. Metzler, Stuttgart, Weimar.

Ute Gerhard. 1990. *Unerhört. Die Geschichte der deutschen Frauenbewegung*. Rowohlt. Reinbek near Hamburg.

Ute Gerhard. 2009. Frauenbewegung und Feminismus. Eine Geschichte seit 1789. Beck, München.

Thomas Gloning. 2003. Organisation und Entwicklung historischer Wortschätze. Lexikologische Konzeption und exemplarische Untersuchungen zum deutschen Wortschatz um 1600. Niemeyer, Tübingen.

Thomas Gloning. 2012. Diskursive Praktiken, Textorganisation und Wortgebrauch im Umkreis der ersten Frauenbewegung um 1900. *Historische Pragmatik*. Ed. Peter Ernst. De Gruyter, Berlin, New York. 127-146.

Thomas Gloning. 2013. Historischer Wortgebrauch und Themengeschichte. Grundfragen, Corpora, Dokumentationsformen. *Perspektiven einer corpusbasierten historischen Linguistik und Philologie. Internationale Tagung des Akademievorhabens „Altägyptisches Wörterbuch" an der Berlin-Brandenburgischen Akademie der Wissenschaften, 12.-13. Dezember 2014*. Ed. Ingelore Hafemann. Berlin. 317-370.

Florence Hervé. 1988. *Geschichte der deutschen Frauenbewegung*. Pahl-Rugenstein, Köln.

Rudi Keller. 1995. *Zeichentheorie. Zu einer Theorie semiotischen Wissens*. Francke, Tübingen, Basel.

Anna Pfundt. 2017. Frauenwahlrecht? Oder Damenwahlrecht? Oder doch ein allgemeines Wahlrecht? – Zum Wortgebrauch in der Diskussion um das Frauenwahlrecht um 1900. *Im Zentrum Sprache*. https://sprache.hypotheses.org/542 (last access 2019-08-23).

Ute Rosenbusch. 1998. Der Weg zum Frauenwahlrecht in Deutschland. Nomos-Verlag, Baden-Baden.

Angelika Schaser. 2006. *Frauenbewegung in Deutschland 1848-1933*. Wissenschaftliche Buchgesellschaft, Darmstadt.

Georg Stötzel and Martin Wengeler. 1995. *Kontroverse Begriffe. Geschichte des öffentlichen Sprachgebrauchs in der Bundesrepublik Deutschland*. De Gruyter, Berlin, New York.

Margrit Twellmann. 1972. Die Deutsche Frauenbewegung. Ihre Anfänge und erste Entwicklung 1843 – 1889. 2 Bände. Hain, Meisenheim am Glan.

Kerstin Wolff, Alexander Geyken, and Thomas Gloning. 2015. Kontroverse Kommunikation im Umkreis der ersten Frauenbewegung. Wie können digitale Ressourcen die sprachliche Untersuchung und die Ergebnisdokumentation verbessern? *Grenzen und Möglichkeiten der Digital Humanities*. Eds. Constanze Baum and Thomas Stäcker. Sonderband der Zeitschrift für digitale Geisteswissenschaften, 1. http://zfdg.de/sb001_010