Volker Harm

WORTGESCHICHTE DIGITAL: A HISTORICAL DICTIONARY OF NEW HIGH GERMAN

Abstract   *Wortgeschichte digital* ('digital word history') is a new historical dictionary of New High German, the most recent period of German reaching from approximately 1600 AD up to the present. By contrast to many historical dictionaries, *Wortgeschichte digital* has a narrated text – a "word history" – at the core of its entries. The motivation for choosing this format rather than traditional microstructures is briefly outlined. Special emphasis it put on the way these word histories interact with other components of the dictionary, notably with the quotation section. As *Wortgeschichte digital* is an online-only project, visualizations play an important role for the design of the dictionary. Two examples are presented: first, the "quotation navigator" which is relevant for the microstructure of the entries, and, second, a timeline ("Zeitstrahl") which is part of the macrostructure as it gives access to the lemma inventory from a diachronic point of view.

Keywords Historical lexicography; word history; quotations; visualizations

1.   Introduction: How to tell a word's history

When historical lexicography started during the 19th century, words were very often imagined as living beings whose biography had to be told in a dictionary – "every word should be made to tell its own history", as Herbert Coleridge has put it in a famous statement which sketched the new lexicographical enterprise that later became famous as the *Oxford English Dictionary* (cited from Mugglestone 2016, p. 556). Thus, dictionaries like the OED or the *Deutsches Wörterbuch* by the brothers Grimm are not only designed for looking up words, but also for reading and studying a word's history. Jacob Grimm, in his preface to the first volume of his *Deutsches Wörterbuch*, has made this quite clear: "The dictionary could be read at home, with pleasure and sometimes with devotion, too" ([1]DWB 1, p. xiii).[1]

While working on the last fascicles of the second edition of the *Deutsches Wörterbuch* in 2016, we started thinking about new ways of doing historical lexicography in the internet age. We always had those quotations in mind: Our dictionary should tell the history of words much in the sense of Coleridge and Grimm, and a user should be given the opportunity to learn about a word's history by reading in our dictionary. However, to really meet these requirements traditional entry structures seemed insufficient to us. In our view, the major shortcomings in these structures are their opacity, the isolated treatment of individual words due to the alphabetical order, and, most importantly, their lack of a genuinely historical approach. These traditional structures are opaque as many users do not know how to extract useful information out of a list of definitions and quotations which are amassed in such an entry, often without comments or explanations. These structures prove insufficient in many cases as traditional dictionaries do not account for the fact that many developments can only be understood when words are considered within the context of their paradigmatic relations – an insight well known to linguistics ever since Trier (1931). And after all, these structures do not offer a historical approach as they provide nothing more than a

---

[1]   Translation by the author (original: "so könnte das wörterbuch zum hausbedarf, und mit verlangen, oft mit andacht gelesen werden").

chronological arrangement of word meanings and quotations. But, as every historian knows, history is much more than chronology: instead of cataloguing word senses by their earliest attestation, one has to explain contexts, to show connections, and to sketch developments. Thus, the so-called "historical principle" of lexicography (Considine 2016, p. 163) has to be reinterpreted and adapted to our present needs and research questions.

2.   *Wortgeschichte digital:* An online-only dictionary with narrated entries

In order to be clear about the aim of the new dictionary the project was termed *Wortgeschichte digital* ("digital word history", WGd), which is a nod of course to the great lexicographers of the 19th century whose intention was to provide word histories in their dictionaries. The inclusion of the adjective *digital* is not just a concession to fashion. Rather, the attribute points to an important hallmark of the project: It is, at least as far as I can see, the first online-only dictionary in the field of German language history. There is no retro-digitized dictionary from which the work starts, nor is there any print edition as a supplementary product. The decision to work digitally from the very beginnings is fundamental and has far-reaching consequences for the design of our dictionary. I will come back to this later.

Next to its digital nativity, the most outstanding feature of WGd is its style of presentation which at first sight, is not digital at all, but deeply rooted in the era of printing: Rather than using traditional microstructures consisting of definitions, usage labels, and quotations we opted for narrative entries, i. e. for articles written as continuous texts. In this respect, WGd has more in common with, e. g., Alain Rey's *Dictionnaire historique de la langue française* (DHLF) or Raymond Williams's *Key Words* than with the OED or the DWB. There are mainly two closely connected reasons for this decision: the one has to do with historical lexicology as the object of the dictionary, and the other with the dictionary's users.

The first motivation for turning to a narrative style is that the realities of word history are "complex and often messy" (Durkin 2016, p. 252). There is a number of issues which the lexicographer has to cope with: The word's origin and development has to be determined against the background of a sometimes scarce attestation, the successive emergence of new meanings out of older ones has to be reconstructed while accounting for the whole scale of the word's diatopic, diastratic, and diaphasic variation, and more than often the specific historical setting of a word usage has to be determined in order to properly contextualize the semantic changes. And after all, the mechanisms that brought about the lexical innovations – mostly metaphors and metonymies – have to be identified, which is not always a straightforward task. All this makes it difficult to deal with word histories within the constrictions of traditional entry structures. Those structures tend to prefer clear-cut distinctions between senses, each illustrated by well-fitting historical quotations. But when it comes to history, transitions are more relevant than well-established usages, and rather than the 'good' quotations it is the ambiguous attestations of a word which very often play a crucial role as bridging contexts for the rise of new meanings. In sum, the high flexibility which is necessary for dealing with the vagueness of many historical developments and for really living up to the complexities of historical lexicology can best be achieved in a continuous text.

As indicated above, there is another reason for turning to narrative entries. This style of representation is not only appropriate to the matter of word history itself, it is also more

easily comprehensible for the reader. The enumeration of senses, often forced into intricate semantic hierarchies, and large sections with quotations are unsatisfying for many users, especially if they are not used to reading historical texts. If readers expect answers to their questions into words and their developments, these answers should be spelled out as clearly as possible rather than to be hidden in the quotation section or in a labyrinthine entry structure. If dictionaries do not meet these demands, people interested in language will turn to other, less reliable, resources which can easily be found in the vastness of the internet.

Next to the decision to publish an online-only narrated dictionary, there is another feature which sets WGd apart from current historical dictionaries: its onomasiological approach. Onomasiology is relevant for the project design both with respect to the overall lemma selection and the individual entries. As for the selection of lemmata, WGd starts from conceptual domains such as, e.g., society, politics, economy, technology, communication, traffic, and everyday culture, thereby preferring those domains of vocabulary which underwent significant change from 1600 to the present. These domains are dealt with in individual work periods of several years. At present, WGd is dealing with the large area of *politics and society* for which five years are scheduled. The onomasiological approach has the advantage of offering a more economical way of describing the vocabulary, because instead of arbitrarily shifting from topic to topic following the contingencies of the alphabet, the lexicographers can focus on selected domains on which they work for a certain time. On the side of the user, this has the advantage of a far more coherent picture of the dictionary entries which are more or less from the same mould (for more detail on the onomasiological approach see Harm 2022, p. 175–177).

Onomasiology is also relevant on the level of the individual entry, although to a lesser extent. Here, next to the standard entries which are mostly dedicated to single words, one can also find synoptical entries describing word fields. One example is the field of terms for the upper classes containing lemmata such as *Beaumonde, die oberen Zehntausend,* or *Jetset.* The summarizing entry accounts for the interdependencies between the words and their developments and gives an overall picture of the ways a certain domain has been lexicalized over time.

3. Word history and quotation evidence in WGd

When it comes to telling word history in the sense of Coleridge and to compiling a dictionary as a "Lesebuch" in the vein of Jacob Grimm, the advantages of a narrated dictionary text are obvious. But there is a significant disadvantage of this dictionary type as the example of Rey's and Williams's dictionaries shows, namely the absence of quotation evidence. Whereas quotations are certainly not of primary interest for the ordinary reader, they play a crucial role for historical lexicography as a branch of philology. As far as dictionaries have scientific aspirations they cannot go without quotation evidence: quotations from historical sources are indispensable not only because they illustrate the word meanings in question, but also, and most importantly, because they provide the means for critically checking the lexicographer's hypotheses.

Most obviously, it is difficult to integrate a larger amount of quotations (including their bibliographic references) into a continuous text. An easy way to solve this problem would be to drastically shorten the quotations so as not to interrupt the text flow. But quotations which are too short are of little use because they miss their fundamental illustrative func-

tion. In order to provide a solution to this issue, the DHLF, e. g., inserts merely a date into the text as an indication that there is an attestation of the word or word usage in question somewhere in the dictionary archives. In this case, the reader has to blindly trust the information he is given. *Trübners Deutsches Wörterbuch*, another prominent example of narrative lexicography, tried to solve the problem in a different manner: Selected quotations are given in a footnote appendix to each entry. Hereby, the reader gets access to at least some of the evidence the entry is based on, but the amount and quality of the information provided this way remains far too scarce altogether. Thus, in a printed dictionary, writing a continuous text and presenting the quotation evidence in a satisfactory manner is nearly impossible.
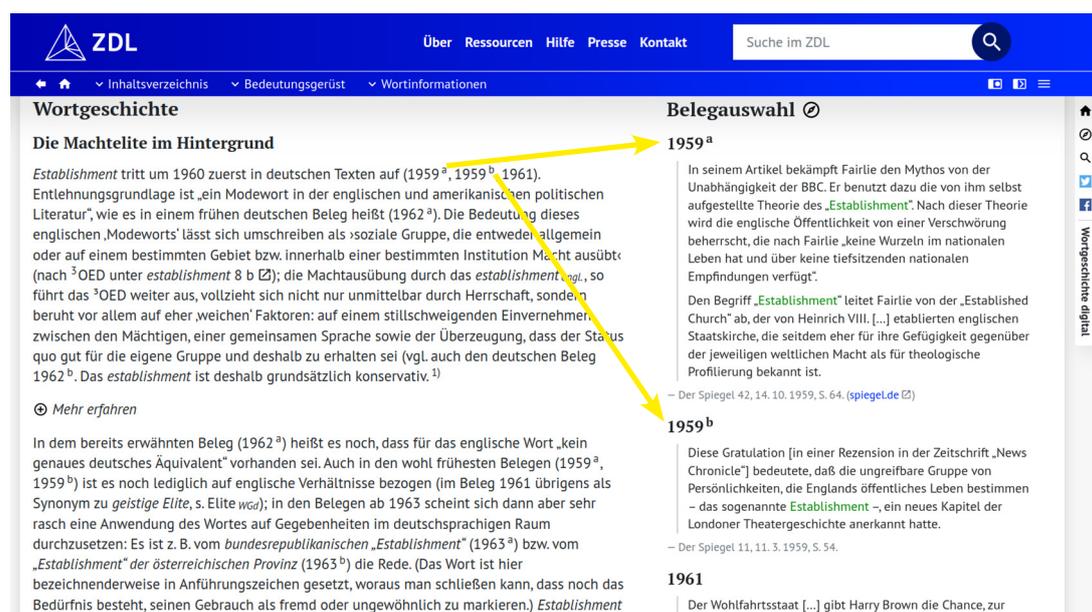
Fig. 1: Word history and quotation selection

It is only in a digital dictionary where both demands – presenting a readable text while not omitting the quotations – can be met. Therefore, WGd has developed a solution termed "movable quotation section" (see Fig. 1). The basic structure of this feature is as follows: the text and the quotation section are presented alongside on the screen, whereby the text (as the most interesting component) occupies the largest part and the quotation section is given on the right side. In order to show the relationship between text and quotations both are mutually connected by links (symbolized by the arrows in Fig. 1). In the text, the quotations are indicated by the year of publication of the source; by clicking on the date, the relevant quotation will be highlighted and scrolled automatically into view. The link is working the other way round too: Clicking on the quotation (i. e. the year index) will lead the reader to the place in the text where the quotation is dealt with. Whenever possible, the quotations themselves are connected with their source in the corpus[2] or on the internet by an external link.

---

[2]   WGd mainly relies on current reference corpora of German such as Deutsches Textarchiv, a large text collection consisting of about 1500 sources from 1600 to 1900, and the DWDS-Korpus ranging from about 1900 up to the present, cf. https://www.deutschestextarchiv.de and https://www.dwds.de. Next to these corpora which are the standard references for the project, a number of other resources is used: library collections such as https://anno.onb.ac.at/, e. g., or the corpora provided by the *Leibniz-Institut für Deutsche Sprache* in Mannheim (IDS). – As larger internet-corpora like DTA and DWDS

By the joint presentation of both the word history and the quotation evidence it is based on at least two different goals can be reached. First, a broader audience can be addressed as both experts and laymen can find what they want: The professional reader has all the material at his disposal in order to critically examine the entries and to draw his own conclusions, and non-professionals will not be disturbed or distracted by information which is of no use to him unless he wishes to go into further detail. Second, interlinking word history and quotations has positive consequences for the lexicographical structure itself. In traditional dictionaries such as DWB, it is not always clear what lexicographical information a quotation is meant to illustrate or corroborate. In the mode of presentation used in WGd, however, the relationship is unambiguous since quotation and dictionary description are always closely tied together.

4.   The microstructure of WGd entries

The ideal users of WGd are those who have a basic interest in and perhaps some knowledge of word history and who are willing to take the time to read such an entry in its entirety. But this is not the only audience we have in mind. As normally dictionaries are the only interface between linguistics and a broader audience, we see it as our duty to address a wide array of readers. Therefore, the dictionary has to accommodate its structure to an audience consisting not only of experts. In order to make the entries more easily accessible, it is necessary to provide different modes of reading. Next to a close reading of the whole entry, the dictionary has to offer valuable information also for those who just want to get a short answer to a very specific question and to those who just like to browse through the dictionary.

In order to enable users to browse through the entries and to focus on individual sections of the entry, a special introductory component has been created which is located on the top of each entry: the so-called "orientation section" ("Orientierungsbereich"). As shown in Figure 2, this section includes:

a) a short abstract of the word history ("in short" / "Kurz gefasst")
b) a table of contents ("Inhaltsverzeichnis")
c) an overview of all the word meanings relevant for the entry ("Bedeutungsgerüst" / "relevant senses").
d) a section with additional information on the word ("Wortinformationen" with sense relations, collocations, word formations)

The components b) to d) are linked to the relevant heading or passage in the text. So, the user has direct access to the piece of information he might be interested in. This way, a selective reading of the entries is possible.

---

already exist there is no need for WGd to build up an extra text collection for the project. New digital sources which are of special interest to WGd can be implemented by the collaboration with the *Zentrum für digitale Lexikographie der deutschen Sprache* (ZDL), the overarching institution the project is part of.
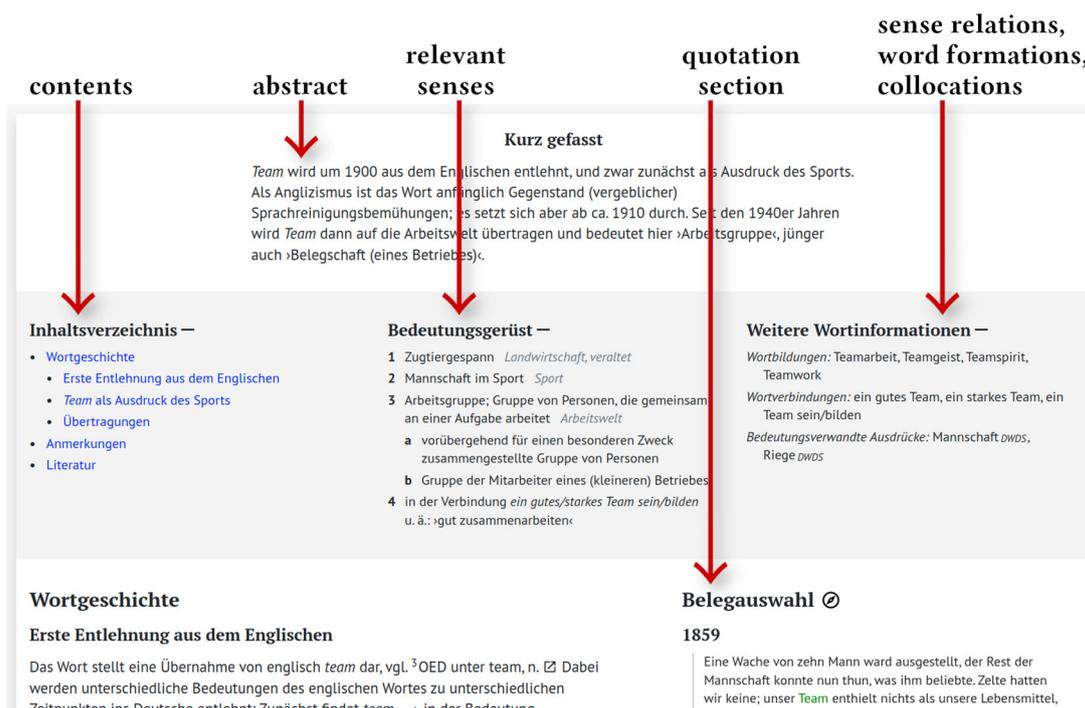
Fig. 2: Orientational section ("Orientierungsbereich")

As already mentioned, the quotation section normally includes a large number of quotations. In order to give an overview and to make this section more informative, a special feature has been introduced, the so-called "quotation navigator" ("Belegnavigator"; see Fig. 3.). This visualization can be opened by clicking a special icon. It is a timeline along which the quotations selected for the entry are listed as bullet points. These bullets are clickable, so that the reader can navigate to a quotation which he might find interesting: If one is interested in the 18th century, e.g., or in neologisms of the seventies, the bullets provide direct access to quotations from that period, and since the quotations themselves are interlinked with the passage where they are dealt with, the reader can also use this tool for a selective reading of the word history. The distribution of the bullets (i.e. the quotations they symbolize) can also give insights into historical facts which are not necessarily highlighted in the entry. Most notably, the hotspots of a word history, the periods which are critical for a word's semantic development may become visible by the distribution of the bullets. Figure 3, e.g., tells us that for the word *Establishment*, a loan from English, the first decade of its attestation (1959 to 1970) appears to be the most dynamic whereas since then no significant semantic innovations seem to have happened, at least according to the lexicographer who wrote the entry and selected the quotations. The suggestions provided by the quotation timeline might not always prove true, but, in any case, the tool can be seen as an invitation to the reader to deal with the content of the entry, to follow his own interests and even to draw his own conclusions. Most crucially, by the "quotation navigator" a playful component is introduced into the dictionary which provides alternative access to historical semantics.
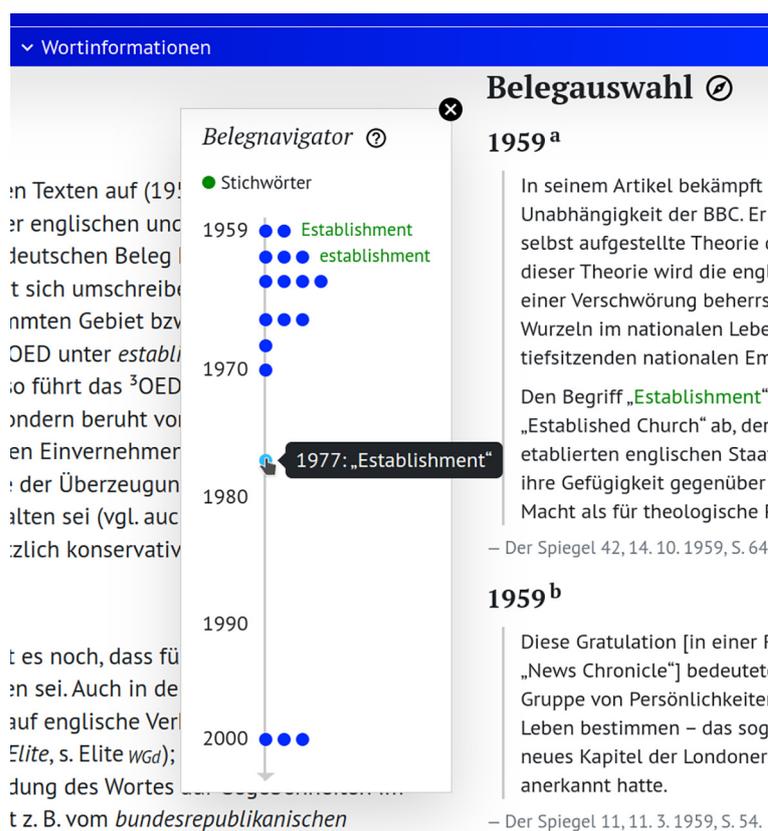
Fig. 3:   Quotation navigator

5.   Macrostructures: the WGd timeline for lemma selection

Since in online dictionaries search windows offer very easy ways to look up words, macro-structures seem no longer necessary: A user who has just to type in what he wants to know does not need a structure leading him through the dictionary. Whereas it goes by itself that every online dictionary has to be searchable via a window, WGd aims at more: It seeks to arouse people's interest in things which they were not interested in or did not even know before. In order to reach this goal, it does not suffice to adopt the usual ways of public relations (notably Twitter and YouTube). The dictionary itself has to provide attractive tools enabling users to make unexpected discoveries. For this reason, WGd has a homepage including not only a search window and multifaceted filters, but also a complete lemma list with three optional arrangements:

a) by alphabet
b) by so-called "cross-reference clusters" which mirror the (mainly semantic) relations between the lemmata
c) a timeline ("Zeitstrahl")

As WGd is a historical dictionary the timeline is the most important tool for retrieving information. Therefore, it will be briefly illustrated here (for a detailed account of the cross-reference cluster see Dorn in this volume). If the user opts for the timeline view by clicking the relevant icon, all lemmata will be presented in chronological order according to their oldest attestation in the dictionary (which, by the way, is not necessarily identical to the earliest attestation in German altogether, because WGd is a dictionary covering New High German, the most recent language period reaching from approximately 1600 to the present). An extract of the timeline is given in Figure 4.
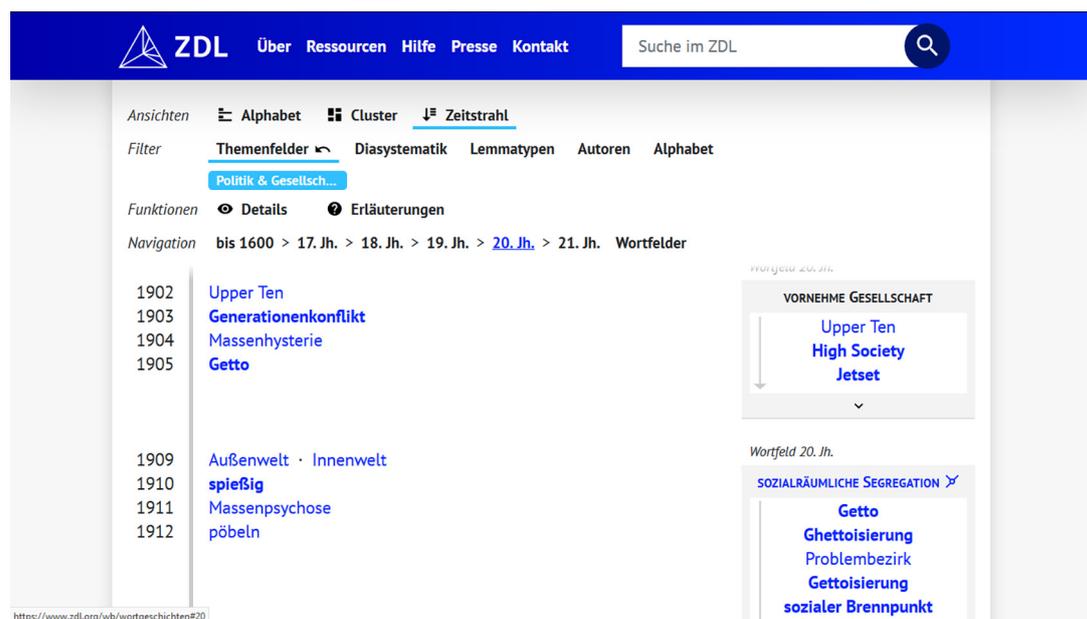

Fig. 4:   WGd timeline ("Zeitstrahl")

On the left-hand part of the screen, the lemmata are listed in chronological order (the bold ones are so-called "Hauptlemmata"/"principal lemmata", which have an entry of their own, whereas the other lemmata are treated together with principal lemmata in a single entry). The chronological presentation offers the possibility of browsing through the lemma inventory following a personal interest in certain periods: Which words came about in the year of my birth? Which words emerged during the Second World War? It lies in the nature of such questions that they do not always have meaningful answers. Such a timeline, though, gives the reader the opportunity to explore the vocabulary on his own and to detect unexpected combinations and connections. However this may be judged, browsing through a timeline is certainly more informative than going through the alphabet.

As mentioned in section 4, WGd includes synoptical entries in addition to the standard entries. The lemmata dealt with in these entries can be found in the boxes on the right side (see again Fig. 4). By ticking the arrow below, the window opens and all relevant lemmata are displayed in chronological order (those in blue are from the relevant century which is selected). The lemma overview presented in these boxes, as simple as it is, can already reveal interesting facts about the evolution of the word field in question. In the case of the field "upper class terms" in Figure 5, e.g., it is obvious at first glance that the older words are mostly loans from French, whereas the words attested since the 20th century are of English origin. What is most important about this visualization is that every lemma appearing in the

timeline is linked with the respective entry. So, the reader is invited to directly go to the entry and read the text.

The lemma timeline as well as the quotation navigator explained in the previous section are created automatically. Nevertheless, these visualizations always rely on individual decisions made by lexicographers on the basis of their interpretation of the data. By integrating a larger number of these individual interpretations into a whole picture new perspectives and constellations become visible which reach beyond the single entries.
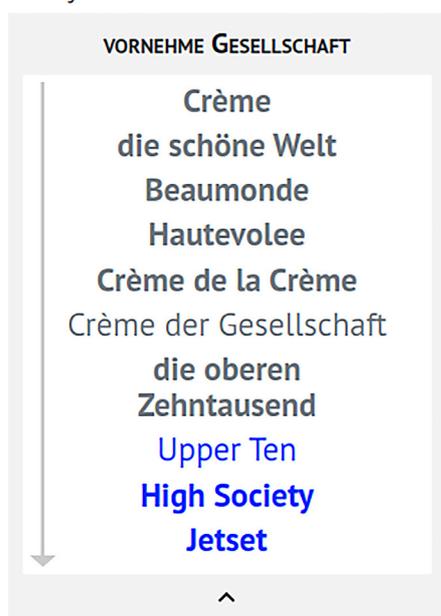

Fig. 5: Word field "upper class" in the timeline presentation

6.   Conclusion

Wortgeschichte digital represents a significant departure from previous historical dictionaries as it has opted for "word histories", i e. for a narrative format of representation, and for an onomasiological approach to lemma selection. Being an online-only dictionary from scratch, it makes extensive use of automatized visualizations which are complementary to the word histories. These visualizations offer new ways of accessing lexicographical information to the reader, who can more easily browse both through an individual entry and the dictionary as a whole. As we hope, these tools as well as the narrative entry format will make word history more attractive to the broad audience interested in words and language. But as important as digital working methods and visualizations are for the project, at its core, *Wortgeschichte digital* is a philological enterprise. The use of online corpora and digital tools does not change the fact that the interpretation of historical data is the cornerstone of the work on the project. In this respect, dictionary making has remained almost the same from the days of Jacob Grimm up to now: The aim of historical lexicography is still to write biographies of words, and this still needs highly skilled biographers in the first place.

References

Considine, J. (2016): Historical dictionaries: history and development, current issues. In: Durkin, Ph. (ed.): The Oxford handbook of lexicography. Oxford, pp. 136–175.

DHLF: Rey, A.: Dictionnaire historique de la langue française. 1992, vols. 1–2. Paris.

Durkin, Ph. (2016): Etymology, word history, and the grouping and division of material in historical dictionaries. In: Durkin, Ph. (ed.): The Oxford handbook of lexicography. Oxford, pp. 236–252.

[1]DWB: Grimm, J./Grimm, W.: Deutsches Wörterbuch, vols. 1–16 [1–32]. Leipzig (1854–1971). https://www.woerterbuchnetz.de/DWB1. (last access: 25-03-2022)

[2]DWB (1960–2018): Grimm, J./Grimm, W. Deutsches Wörterbuch. Neubearbeitung, vols. 1–9. Leipzig/Stuttgart. https://www.dwds.de/d/wb-2dwb, https://www.woerterbuchnetz.de/DWB2. (last access: 25-03-2022).

Harm, Volker (2021): Beyond the "Grimm": German historical lexicography after *Deutsches Wörterbuch.* In: Van de Velde, H./Dolezal, F. (eds.): Broadening perspectives in the history of dictionaries and word studies. Newcastle upon Tyne, p. 117–134.

Harm, Volker (2022): *Wortgeschichte digital.* Ein neues Wörterbuch zur Geschichte des neuhochdeutschen Wortschatzes. In: Diehl, G./Harm, V. (eds.): Historische Lexikographie des Deutschen. Perspektiven eines Forschungsfeldes im digitalen Zeitalter. (= Lexicographica. Series Maior 161). Berlin/Boston, p. 173–191.

Mugglestone, L. (2016): Description and prescription in dictionaries. In: Durkin, Ph. (ed.): The Oxford handbook of lexicography. Oxford, pp. 546–560.

Trier, J. (1931): Der deutsche Wortschatz im Sinnbezirk des Verstandes. Die Geschichte eines sprachlichen Feldes. Bd. I: Von den Anfängen bis zum Beginn des 13. Jahrhunderts. Heidelberg.

Trübner: Trübners Deutsches Wörterbuch. Im Auftrag der Arbeitsgemeinschaft für deutsche Wortforschung bearbeitet von Alfred Götze, vols. 1–5. Berlin 1939–1957.

Williams, R. (1976): Keywords. A vocabulary of culture and society. New York.

Wortgeschichte digital. https://www.zdl.org/wb/wortgeschichten. (last access: 24-07-2022).

Contact information

Volker Harm
Zentrum für digitale Lexikographie der deutschen Sprache (ZDL), Akademie der Wissenschaften zu Göttingen
vharm@uni-goettingen.de