

Research on dictionary use and the development of user-adapted views

Carolin Müller-Spitzer

Abstract. The development of user-adapted views of lexicographic data is frequently in demand by dictionary research on electronic reference works and hypertext information systems. In the printed dictionary it has been indispensable to develop a complete dictionary relative to a user group and using situations. In contrast, for any electronic presentation of lexicographic data there are possibilities to define user-specific views of an initially user-unspecific resource. However, research on the use of dictionaries in general, still has to answer several open questions as far as this subject is concerned. This paper will firstly provide an overview of the present state of research on dictionary use with respect to electronic lexicography. Subsequently, explanations of further prerequisites for a possible user-adapted access to data are followed, as exemplified by OWID, the Online Vocabulary Information System of the Institut für Deutsche Sprache. Finally, it will be outlined what results on the subject have been accomplished so far. Also the prospects of potential user-adapted presentations of lexicographic data will be highlighted.

1 The subject matter of this paper

Dictionaries are basic commodities or more precisely: “Der genuine Zweck eines Wörterbuchs besteht darin, daß es benutzt wird, um anhand lexikographischer Daten in den Teiltextrn mit äußerer Zugriffsstruktur [...] Informationen zu denjenigen Eigenschaftsausprägungen bei sprachlichen Ausdrücken zu erschließen, die zum jeweiligen Wörterbuchgegenstandsbereich gehören.”¹ (Wiegand 1998: 299) Therefore, it seems to be appropriate to compile dictionaries for specific user groups and particular using situations. This applies at least to printed dictionaries. In the field of electronic lexicography, the question of how to integrate a potential user conceptually must be addressed in a completely different way. Instead of adapting the lexicographic description to particular users and functions, one has to think about how far the user can define her/himself by a specific look-up situation as it occurs in a user-unspecific resource. It is possible in the digital dictionary to obtain information from the same data pool and to incorporate it in a suitable way typically relevant to a typical usage

1. In English: The genuine purpose of a dictionary is to be used in order to obtain information on those properties of linguistic expressions which belong to the domain of the dictionary; this information can be accessed via lexicographic data in partial texts with an outer access structure.

context. It is therefore possible to adapt lexicographical descriptions to specific user needs and using situation. Instead of the static entry found in a printed dictionary in which all potential information is closely compressed, the user will receive a digital hypertext dictionary entry “on demand” with the information relevant to the current user situation (cf. Storrer 2001: 64f.). However, user-adapted dictionaries have only so far been developed as prototypes (cf. Thielen et al. 1998). The fact that it has just been about prototypes up to now, however, is not so much due to it not being known which classes of data are typically relevant for which usual look-up situations. The more likely reason is that “a context adapting presentation of lexical information requires a linguistically motivated and finely-grained modelling of lexicographical data.” (Storrer 2001: 65).

Hence, in the project OWID (*Online-Wortschatz-Informationssystem Deutsch*, online-vocabulary-information system of German)² of the *Institut für Deutsche Sprache* (IDS), the lexicographic data available has been modelled in a very fine grained way in order to create a basis for a user-adapted presentation of data. However, the actual habits and needs when using electronic dictionaries need to be critically analysed first in order to achieve this goal adequately. Research on dictionaries has comparatively well established the question which types of data are predominantly used by which types of users in which using situations. However, such analyses have been performed on the basis of printed dictionaries for the most part. For a printed dictionary, it is sufficient to define the types of information and to include general guidelines for its layout. As far as a user-adapted presentation of a general lexicographic database in hypertextual environment is concerned, completely new questions have to be addressed which are only relevant for the electronic lexicography, i.e. general questions like: how do users navigate in electronic dictionaries? How do they use offered search options? When do clusters of specific word information stop to be user-friendly and when does unclearness set in?³ More specifically we need to ask: should a user (i.e. while using a dictionary) create a personal profile at the beginning of a session (e.g. user type: non-native speaker, situation of use: reception of a text) and should s/he navigate in all entries with this profile? Or is it more user-friendly to be able to change one’s profile and to look at the same entry with different profiles? Or is it better to let the user her/himself determine whether s/he wants to look at information on the meaning or on grammar instead of deriving a specific display of different kinds of information from her/his individual profile? Likewise, being able to provide the information with explicit headlines, which was not possible in a printed dictionary for lack of space, involves new challenges. For example, which types of terminology accommodate which user group the most has

2. See <http://www.owid.de>.

3. In addition, very general guidelines for the design of electronic media can be included, particularly web pages, but also special guidelines for online-dictionaries.(Cf. e.g. Almind 2005)

to be examined. All these questions show that research on use still has to answer many questions, especially for electronic dictionaries. Thus, in the following part, the present state of research on dictionary use – mainly with regard to electronic lexicography – will be outlined and some research methods applicable to the study of the use of electronic dictionaries will be described. In the third paragraph, OWID, an online dictionary portal, will be presented. On the one hand, the focus lies on the development of modelling as the basis of a user-adapted presentation and on the other hand on studies on the use of OWID which have already been accomplished. The paper finally concludes with an outlook on potential user-adapted presentations of lexicographic data in OWID.

2 On the present state of research on dictionary use

User research represents the main area of lexicographical research (cf. Wang 2001: 53). This statement might certainly overvalue the status of research on dictionary use as part of general meta-lexicographic research even though the situation has clearly improved over the past two decades: 25 years ago, one could say of the dictionary user: everyone talks about him, but no one knows him. Wiegand called him the unknown of the unknown. Today, one cannot speak of such a lexicographical “Yeti”. Not everything is known about him, but certainly more is now known (cf. Almind and Bergenholtz 2000: 259). In addition to research on the dictionary use of printed dictionaries (cf. i.e. Atkins 1998; Atkins and Varantola 1998; Ripfel 1989; Wiegand 1998), the use of electronic dictionaries has been increasingly examined. Infact, Engelberg and Lemnitzer still wrote in 2001 that there are hardly any studies about how the user’s behaviour is affected by innovations in the field of electronic lexicography (cf. Engelberg and Lemnitzer 2001: 71). However, in the meantime, this situation has improved to a small extent. Nevertheless, it is still true that research on use requires an enormous investment which seems to be justifiable only within the scope of academic research. (Cf. Lemnitzer 2001: 247)⁴ However, the users’ reactions have some influence on new editions of commercial dictionaries, insofar as they are reflected in the sales figures, the requirements of the book trade’s agents and in the incoming language queries (cf. Höhne 1991; Müller 1991). Academic dictionaries are not subject to this pressure to the same extent. For both academic as well as commercial lexicography, however, extensive research on the use is certainly a profitable undertaking.

Current research on use in the field of electronic dictionaries especially focuses on new opportunities of observing the behaviour of users in the internet, namely

4. For example, the study presented in Atkins (1998) was only possible by fundings gained from EU-RALEX and AILA.

through the analysis of log-files for electronic dictionaries; the reason being that this method can be accomplished with manageable investment. However, there is only a small number of projects reporting on how they put this opportunity into practice. "Although the proposal to draw upon log files in order to improve dictionaries was already expressed in the mid-1980s [...], and although numerous researchers have reiterated this idea in recent years [...], very few reports have been published of real world dictionaries actually making use of this strategy." (Schryver and Joffe 2004: 187) Exceptions are the study presented by Lemnitzer (2001) and the *Concept of Simultaneous Feedback* (i.e. Schryver and Prinsloo 2000) which was used, for example, for the online-dictionary "Sesotho sa Leboa Dictionary Project (SeDiPro)"⁵. However, more and more online dictionary projects make use of the analysis of log files like the last Euralex conference in Barcelona has shown (cf. i.e. Měchura 2008).

For both studies and methods respectively, it becomes obvious that research on use is especially interesting when the project examined can be improved immediately based on the results. This was much more difficult in the case of printed dictionaries as they can only be investigated when they have been completely published. More and more Internet dictionaries are not published once but in following parts so that insights into the user's behaviour can already be included in the working process. As De Schryver critically points out in his work "Simultaneous Feedback (SF)": "Feedback from the envisaged target user group is systematically and continuously obtained while compilation is still in progress." (Schryver and Joffe 2004: 187) Their idea is that a dictionary can be seen as a service for a "community", and therefore, users have the opportunity to point out missing entries which should then be added immediately. The concept of SF implies two elements: on the one hand the analysis of log-files which takes place manually even if perspectively this is intended differently: "ultimately, the idea is that an automated analysis of the log files will enable the dictionary to tailor itself to each and every particular user. At present, the analysis of the log files is still largely done manually, in part with the aim to draw up typical user profiles that will then be fed into the projected adaptive and intelligent dictionary of the future." (<http://tshwanedje.com/sf/>, 22.04.2008) On the other hand, the second element is the request to the users to react via e-mail, a procedure which produces good results according to De Schryver and Prinsloo. "On the whole, one observes a very good correspondence between the formal feedback through the online feedback forms, and the informal feedback obtained by means of an analysis of the log files. This is a satisfying sign indeed and indicates that modifying and adapting dictionary contents based on log-stat trends is a feasible strategy." (Schryver and Prinsloo 2000: 194) However, it is believed here that the results gained through examples of feedback and the incoming number of reactions via e-mail are

5. See <http://africanlanguages.com/sdp/>.

not entirely convincing to be considered as relevant enough a method of research on use. Also Lemnitzer reports that the request to react via e-mail was hardly used and that a third of all incoming e-mails consisted of letters with irrelevant information and personal complaints (cf. Lemnitzer 2001: 248).

The method of SF as well as the analysis of log-files in general are not necessarily methods of research on use, but initially just forms of receiving feedback concerning a particular dictionary with the goal of improving this dictionary. However, by generalising this log file and SF-data, results can be gained that can generally be further used for research on use. The question that arises here is which of the questions asked at the beginning can be answered through the analysis of log-files? The log files give rough information about the navigation of users such as: which search box did they use? Where did they click next? There is no information – depending on the technical composition of the page – about further navigation within the entry. The log-files also provide information about the following: what did the users search for? How did they spell the search word? This means that from the analysis of log-files, some indications for improving the search options can be derived (mainly from the analysis of unsuccessful search results) as well as indications for potential lacks of words in the dictionary. In the same way, ideas for improving the guidelines to the search functions can be derived from the analysis of unsuccessful searchings. In this context, a useful recommendation of Lemnitzer is that incorrect searches are not the mistake of the user, but an inadequacy of the user surface (cf. Lemnitzer 2001: 248). However, what cannot be interpreted from log-files are questions like: why are users not able to find a particular piece of linguistic information? Why did they actually search for a particular word? What information from the dictionary entry do they understand and what do they not? Generally speaking: the analysis of log-files is one of many alternatives in research on dictionary use. It should be used for any online-dictionary, but it does not replace other forms to examine dictionary use. Research on dictionary use in general should consist of different methods like written interview, observation, records and tests (cf. Wang 2001: 67; Albert and Koster 2002; Atkins and Varantola 1998).

What also has to be pointed out is that both of the dictionaries examined by Lemnitzer and De Schryver have a very flat microstructure. Lemnitzer himself admits that altogether, the results of the study are limited, which, among other things, is due to the flat structure of the dictionary data (cf. Lemnitzer 2001: 258). That is also a reason why the results are only partly applicable to dictionaries or lexicographic information systems respectively which have a much more elaborated and complex microstructure, such as *elxikol*(, a dictionary of contemporary German in OWID.

3 OWID – an example

OWID, a project of the IDS, is a lexicographic Internet portal which is currently being compiled and contains lexicographic-lexicological reference works of the IDS; perspectively, it is planned that data from external projects is added. Originally, OWID has its roots based in the IDS project *lexiko* (cf. Klosa 2008)⁶.

3.1 Outline of content

The main emphasis of OWID is on the presentation of corpus-based, academic lexicographical-lexicological works. The following dictionaries have been included in OWID:

- *lexiko*: This electronic dictionary consists of an index of about 300.000 short entries with information on spelling, syllabification, and inflection. In the near future, further information (e.g. on word formation) and corpus samples will be added for all lexemes. Furthermore, *lexiko* comprises over 900 fully elaborated entries of headwords which are highly frequent in the underlying corpus. These contain extensive semantic-pragmatic descriptions of lexical items in actual language use. The dictionary is being extended continuously by further elaborated entries (cf. Haß 2005; Klosa et al. 2006).
- *Neologismenwörterbuch* (Dictionary of Neologisms): This electronic dictionary describes about 800 neologisms of the 1990s (cf. Herberg et al. 2004). Neologisms of the 21st century will be added in the near future.
- *Diskurswörterbuch 1945-55* (Discourse Dictionary 1945-55): This dictionary is a reference work resulting from a larger study of lexemes that establish the notional area of “guilt” in the early post-war era in Germany (1945-55). It summarises the lexical and semantical results of this study in the form of about 80 lexical entries. (Cf. Kämper 2005)

Besides the above mentioned, the contents of the existing printed reference works “Handbuch Deutscher Kommunikationsverben” (Handbook of German Communication Verbs; Harras et al. 2004) as well as the “VALBU – Valenzwörterbuch deutscher Verben” (Valency Dictionary of German Verbs; Schumacher et al. 2004) will be revised and published electronically in OWID by spring next year.⁷

6. Until recently, the *lexiko*-project held a double function: on the one hand it was the lexicographic portal of the IDS, on the other hand a corpus-based dictionary of modern German is simultaneously being compiled in *lexiko*. In order to better differentiate between the overall portal and the dictionary *lexiko* itself, an independent portal-project named OWID was founded last year.

The consolidation of different lexicographic works under the same technological roof, brings about challenges of a very different kind. Various questions arise in the process: to what extent can and should the contents of each project be coordinated? How are the contents modelled and structured? How can common access structures to data be developed? The different works shown in OWID are independent projects as far as lexicographic content is concerned. Nevertheless, for potential users it is essential to develop as much common access structures on the particular contents as possible, which means to present more than a random collection of unrelated dictionaries and lexicographic resources, respectively. Therefore, it was necessary to maintain the independence of each individual dictionary project, while, at the same time, ensuring the integration of all the different data. One goal has always been the modelling of the data in such a way that a user-adapted presentation should be possible without changing the database.

3.2 Guidelines of data modelling

A lot of time has been invested on the data-modelling-level which represents the basis of all further possible data uses, by modelling the micro and content structures of each dictionary respectively according to one standardised concept so that the same phenomena are structured in the same way and that all the data is finely-grained structured such that one can access it targetted and specifically. On the basis of a new concept for specifically tailor-made modelling of lexicographic data (cf. Müller-Spitzer 2006, 2007a,b,c), a complex XML-DTD-Library was developed. Each individual information unit is tagged individually and named after its genuine purpose in terms of content. Thus, data modelling for OWID approaches the kind of modelling demanded by Storrer (2001: 61f.). Unlike leXeML, a suggestion for an XML-based tag-language for lexicographic data by Geeb (2001) aiming at the integration of the user in the modelling process, the lexicographic data here is initially modelled independently from potential situations of use in a way that they can be extracted both adaptively and flexibly from the database. This type of data modelling has been considered an innovative approach in general lexicographic practice by Kunze and Lemnitzer (2007: 85ff.) as well as Schlaps (2007).

We decided to use a specifically tailor-made modelling because the XML-structure also serves as a model for compiling the lexicographic entries in the XML-Editor. So the more customised with respect to the particular lexicographic project the XML-structure is, the less a lexicographer needs an additional manual for editing the entry structure. However, one could easily transform this specifically tailor-

7. VALBU will also be accessible via *grammis – the grammatical information system of the IDS*. (<http://hypermedia.ids-mannheim.de/grammis/>).

made structure into a standard one (like LMF or TEI). The XML detail of the entry “emailen” from the Dictionary of Neologisms shown in table 1 illustrates the tagging of information on valency and is an example of the overall granularity of tagging.

Table 1. XML detail of the entry “emailen”

```

<vb-valenz-neu>
<satzbauplan>
<satzbauplanA>jemand emailt (jemandem) (etwas)</satzbauplanA>
</satzbauplan>
<satzbauplan>
<satzbauplanA> jemand emailt (etwas) an jemanden</satzbauplanA>
</satzbauplan>
<satzbauplan>
<satzbauplanA>jemand emailt, dass [...]</satzbauplanA>
</satzbauplan>
<vb-komplemente-neu>
<subjekt-komp-neu obligatorisch="ja">
<nom-nominalphrase-neu/>
</subjekt-komp-neu>
<objekt-komp-vb obligatorisch="nein">
<dat-nominalphrase-vb/>
</objekt-komp-vb>
<objekt-komp-vb obligatorisch="nein">
<akk-nominalphrase-vb/>
<dass-satz-vb/>
</objekt-komp-vb>
<objekt-komp-vb obligatorisch="ja">
<prepositionalphrase-vb praeposition="an"/>
</objekt-komp-vb>
</vb-komplemente-neu>
</vb-valenz-neu>

```

In our internal editorial system, lexicographers are able to use this structure for advanced searches (with XPath expressions). For example, one can search for all regular verbs (*//vollverb*) which have obligatory object complements (*//objekt-komp-vb/@obligatorisch="ja"*), realised as a dative NP (*//dat-nominalphrase-vb*). In this example, the search results are entries from the *lexiko*- as well as from the neologism-dictionary (cf. fig. 1). It is one of *lexiko*'s aims to provide

these extended search options for users too.⁸

The screenshot shows the EDAS (Electronic Dictionary Administration System) interface. On the left, there is a sidebar with navigation options: 'Lemmasuche', 'Erweiterte Suche' (highlighted), 'Alphabetische Listen', 'Statistiken', 'Zuletzt bearbeitet', and 'Neuer Artikel'. The main content area is titled 'Erweiterte Suche' and contains three search criteria, each with a dropdown menu and a 'mit Inhalt' button. The criteria are: 'enthält //vollverb', 'enthält //objekt-komp-vb@obligatorisch="Ja"', and 'enthält //dat-nominalphrase-vb'. Below these, there is a 'Suche starten' button and a list of 25 search criteria: 1. überlängen, 2. abwählen, 3. antworten, 4. aufzwingen, 5. bewacht, 6. demonstrieren, 7. erraten, 8. erlöschten, 9. ermöglichen, 10. leiten, 11. laden, 12. malen, 13. stragehlt, 14. ermöglichen, 15. nachweisen, 16. gachelt, 17. veranlagten, 18. verbieten, 19. verbieten, 20. versagen, 21. versagen, 22. wätschen, 23. wünschen, 24. zugesellen, 25. zustimmen. At the bottom of the list, it says 'Suche beendet.'

Figure 1. Advanced search options for lexicographers

The prerequisites for a user-adapted presentation of the data are already created on the database level. However, this capability is not yet visible on the present user surface of OWID. What is already offered is access to the data of *lexiko* as well as to the Dictionary of Neologisms through “Extended Search Options”. However, these could also undergo further differentiation and enlargement. For further improvements of OWID, we want to know more precisely how users access and navigate through the portal and how such a user-adapted presentation of data could be well arranged. So far, we accomplished first user observations on it which shall be briefly described in the following.

3.3 Pilot studies on using OWID and Elexiko

Although analyses of Internet log-files have been carried out in OWID, their evaluation does not cover studies over a larger period of time. Primarily, they show us what users enter into the search box. Except freak values like “Wie+mache+ich+einen+Hühnersalat” or “Wie+lange+braucht+man+um+die+Erde+zu+umkreisen”⁹, log-

8. The development of the Electronic Dictionary Administration System (cf. fig. 1) is being developed by Roman Schneider, a researcher of the IDS.

9. In English: “How do I make a chicken salad”; “How long does it take to circle the earth”.

files show that most users search for single words or parts of larger expressions. Searchings with Boolean operators like “*lait*” are documented repeatedly, but they do not establish the majority of searches. This gives rise to the suspicion that the opportunity to search with Boolean operators should be explained at a more prominent place. Both, Scherer’s study (2008) as well as a short study at the University of Mannheim support this hypothesis which is briefly described below. Furthermore, it becomes apparent that users look up newer words like “Patchworkfamily” or proper names of e.g. music bands like “Silbermond”. Besides, there is a multitude of inconspicuous searches which cannot be assigned to a particular category. However, in searches like the orthographically aberrant “ddivergenzstellung” or “standart”, it becomes apparent that the search methods must be more tolerant to mistakes, just as search entries like “plumpen”, an inflected form of the adjective *plump*, show that ways of lemmatisation should be lodged in the search. Admittedly, the findings last mentioned are not new and were known before analysing log-files. However, these analyses show, for example, how many users look up information via the search box and who navigates via the list of headwords. Again, this is a very interesting information since the question whether a presentation of such a lemma-list is useful at all has been addressed repeatedly.

What cannot be concluded from log-files (in OWID) is which parts of the dictionary entry the user is looking at since this presentation is very interlaced. This is an important difference to the online-dictionary SeDiPro presented by De Schryver. Here, for example the entry “wish” consists of six equivalents with few and brief additional specifications, whereas in *lexiko*, the entry “wünschen” – as documented in printed pages – covers about 10 pages. Through log files it is possible to see that the user accesses a specific entry, but unfortunately (in our system) it cannot be seen whether the user navigates through specific senses and how s/he anticipates the given information. Through the analysis of the log-files alone in dictionaries like *lexiko*, it is not possible to obtain an insight into how a user navigates exactly within the information system. This illustrates the limits of a user observation of this kind: if one has a very flat-structured dictionary with little information on the search word, log-files pass on extensive information relative to the dictionary on to the lexicographer. If the microstructure is much more comprehensive, the analysis of log-files only provides limited results relative to the lexicographic product.

Furthermore, what is *not* searched for in OWID can be read from log-files. Interestingly, these are, for example, search words which indicate the demand for encyclopaedic information, which means there are hardly any searches for names of towns, for particular people etc. Presumably users seem to be aware of the fact that OWID is a linguistic reference book. Contrary to other studies (Lemnitzer 2001), hardly any searches for words from the so-called sexual and faecal vocabulary are documented. This might allow the assumption that OWID-users are not ordinary Internet users but people with more specific linguistic interests.

Scherer's study comprises a standardised online survey on OWID focussing on *ellexiko*. 58 people were questioned through a standardised online-questionnaire which was constructed according to the multiple-choice method. A single session lasted 45-60 minutes. The study filled an existing gap in terms of research on users of electronic dictionaries – albeit only in the form of a pilot study so far. As Scherer points out, online studies in the form of standardised surveys which ask for subjective perceptions of the layout and functionality of a dictionary as well as the user's own evaluation of the applicability of the dictionary, have still not been carried out to a large extent (cf. Scherer 2008: 95). The use of standardised questionnaires is primarily interesting since it allows the levying of a great number of test people, whereas observation studies and records can only take small numbers of test people into consideration, and its analysis is also much more expensive. At the same time, the results of standardised interrogations frequently include information one would like to clarify further by asking questions in an interview. For example, one statement to be assessed in Scherer's study (translated):

It is clear that the headword-list shown presents entries of all the four dictionaries.
 Strongly disagree 17%
 Disagree 37.7%
 Agree 39.6%
 Strongly agree 5.7%
 No statement 0%

In this case, it would be very informative to ask what is regarded unclear in the current presentation. (At the moment, the headwords of the different dictionaries are presented in different colours in OWID.) The result mentioned above is particularly difficult to assess in relation to the result of the next question:

Do you remember what the design of different colours of the headwords means?
 It categorises headwords by part of speech 5.7%
 It categorises headwords by morphology 1.9%
 It categorises headwords by the dictionaries available in OWID 64.2%
 I don't know 24.5%
 No statement 3.8%

Another smaller study on the use of OWID and *ellexiko* respectively was constructed as an observation study/interview: Within the scope of a seminar on "electronic lexicography" at the University of Mannheim, students were asked to look for five users who had to look up specific information in OWID/*ellexiko* according to a questionnaire. Altogether there were about 60 test people. One session lasted about 30-45 minutes. The students should not only record all the using-actions, but also ask further questions and note down the answers to them. Again, we were able

wünschen

Einzelbedeutungen

'ersehnen'

Mit **wünschen** bezeichnet man eine Handlung, bei der eine Person(en-gruppe) ein Anliegen hat und sich einen Gegenstand oder Sachverhalt ersehnt.

Passiv:	nicht bildbar
Satzbaupläne:	JEMAND wünscht (SICH) ETWAS

'verlangen'

Mit **wünschen** bezeichnet man eine Handlung, bei der eine Person(en-gruppe) einen Sachverhalt nachdrücklich verlangt, oft von einer anderen Person(en-gruppe).

Passiv:	bildbar
Satzbaupläne:	JEMAND wünscht ETWAS

'erhoffen'

Mit **wünschen** bezeichnet man eine Handlung, bei der eine Person(en-gruppe) ihr Anliegen ausspricht, sich etwas Bestimmtes (z. B. Erfolg, Gesundheit, Glück) für eine andere Person(en-gruppe) zu erhoffen.

Passiv:	bildbar
Satzbaupläne:	JEMAND wünscht (JEMANDEM) ETWAS

Figure 2. Draft of a possible user-adapted online presentation of the *lexiko*-entry “wünschen”

to learn from these tests. For example, the laypeople questioned (e.g. the students' sisters and brothers or parents) had difficulties in particular with the terminology on the user surface. At “lesartenbezogene Angaben”¹⁰, many assumed they would find information on pronunciation, for example. Besides, for some it was difficult to understand the fact that in *lexiko*, different batches of the lemma list are provided with very different information (i.e. that only some entries are fully elaborated).

So far, all of these pilot studies have not yet been backed up empirically in an adequate way in order to draw conclusions about the further development of user-adapted presentations. However, they give us a first impression about which field we should do further research in and towards where further developments should be directed to. Also important for OWID: the portal is currently being compiled which means the findings of the research on use can be included in the ongoing development of OWID.

3.4 Future prospects

Obviously, there is still a lot research work to be done regarding the use of electronic dictionaries. However, as said before, the question of what kind of information a dictionary user is searching for in a particular using situation is comparatively well

10. In English: Information related to particular “readings” (senses) of the headword.

The screenshot shows a web interface for the 'Meer' entry. At the top, there are two panels for 'Informationsarten' (Information types). The left panel, 'Lesartenübergreifende Angaben' (Cross-lesart information), includes checkboxes for 'Orthografie', 'Wortbildung', and 'Herkunft u. Wandel'. The right panel, 'Lesartenbezogene Angaben' (Lesart-specific information), includes checkboxes for 'Bedeutungserläuterung', 'Semantische Umgebung u. lexikalsche Mitspieler', 'Typische Verwendungen', 'Sinneverwandte Wörter', 'Besonderheiten des Gebrauchs', and 'Grammatik'. Below these panels, the entry for 'Meer' is displayed. The 'Orthografie' section shows 'Normgerechte Schreibung: Meer'. The entry is divided into two main sections: 'Lesart 'Gewässer'' and 'Lesart 'große Menge''. Each section has a 'Bedeutungserläuterung' (Meaning explanation) and 'Typische Verwendungen' (Typical uses). The 'Gewässer' section includes sub-sections for 'Meer mit Attribut', 'Meer in Verbalphrasen und Sätzen', and 'Meer als Attribut'. The 'große Menge' section includes sub-sections for 'Meer mit Attribut' and 'Meer als Attribut'.

Figure 3. Draft of a possible online view of *lexiko* with an information display for customising the microstructure dynamically (entry “Meer”)

investigated. Perspectively, this gives us the opportunity to outline potential user-adapted presentations in the near future, using *lexiko* as an example. It is important that these examples do not require any changes of the present database, but would be made possible solely by defining separate presentations through different XSLT-Stylesheets.

The entry “wünschen” from *lexiko* should serve as an example. In the current online-presentation, information on spelling and word formation is given on the first screen of the window that opens up by default and which summarises sense-unrelated details of a lexeme. Underneath, sense-related information is reached by clicking on links that represent individual senses.¹¹ Among this sense-related information, there is detailed information on the meaning, the semantic environment, on typical syntagmatic patterns, on sense-related paradigmatic words, such as synonyms and antonyms, on pragmatic and discursive features, and on grammar. One advantage of a user-adapted presentation certainly is that a potential user could create an individual profile at the beginning of an *lexiko* session. Supposing a non-native speaker of German in a situation of text reception requires an overview which is as brief as

11. See http://www.owid.de/pls/db/p4_anzeige.artikel?v_id=139627.

possible including information on grammar in cases where there is a difference in meaning. Taking “wünschen” as an example, the grammatical information on the single senses differs. In such a situation of use, any other information could be left out, so that the most important information can be received at first sight (cf. fig. 2).

Another possibility would be to use a separate part of the screen for a kind of form where specific types of information can be ticked in boxes in order to be seen in the frame below. A user could appoint, for example, that for *lexiko* only the meaning definition and typical syntagmatic patterns are shown. That is probably closest to the idea of individually customised microstructures (cf. Engelberg and Lemnitzer 2001: 224 and Schryver 2003: 178ff. and see a first draft in figure 3). However, one would have to find out through empirical research which types of users are really aware of what kind of information is most useful in each individual situation of use.

The challenge of transforming these very novel presentation alternatives adequately alongside existing essential user needs for an online vocabulary information system certainly remains one of the most important and most interesting challenges in the field of electronic lexicography and research on dictionaries. And it is academic lexicography that should play a key role in mastering this task. Or, like Atkins says: “If new methods of access (breaking the iron grip of the alphabet) and a hypertext approach to the data stored in the dictionary do not result in a product light years away from the printed dictionary, then we are evading the responsibilities of our profession.” (Atkins 1992, 521; cited from Schryver 2003: 144)

References

- Albert, Ruth and Cor. J. Koster (2002). *Empirie in Linguistik und Sprachlehrforschung. Ein methodologisches Arbeitsbuch*. Tübingen.
- Almind, Richard (2005). Designing internet dictionaries. *Hermes* 34:37–54.
- Almind, Richard and Henning Bergenholtz (2000). Die ästhetische dimension der lexikographie. In Ulla Fix and Hans Wellmann (eds.), *Bild im Text – Text und Bild*, 259–288, Heidelberg.
- Atkins, B. T. Sue (ed.) (1998). *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*. Tübingen: Max Niemeyer.
- Atkins, B. T. Sue and Krista Varantola (1998). Language learners using dictionaries: The final report on the euralex / aila research projekt on dictionary use. In B. T. Sue Atkins (ed.), *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*, 21–81, Tübingen.
- Engelberg, Stefan and Lothar Lemnitzer (2001). *Lexikographie und Wörterbuchbenutzung*. Tübingen: Stauffenburg.
- Geeb, Franziskus (2001). Sprache und datenverarbeitung. In *Sprache und Datenverarbeitung*, volume 2, 27–61.
- Harras, Gisela, Edeltraud Winkler, Sabine Erb, and Kristel Proost (2004). *Handbuch deutscher Kommunikationsverben. Teil 1: Wörterbuch (= Schriften des Instituts für Deutsche Sprache)*. Berlin / New York.: de Gruyter.
- Haß, Ulrike (ed.) (2005). *Grundfragen der elektronischen Lexikographie. Elexiko – das Online-Informationssystem zum deutschen Wortschatz*, volume 12 of *Schriften des Instituts für Deutsche Sprache*. Berlin / New York.: deGruyter.

- Herberg, Dieter, Michael Kinne, and Doris Steffens (2004). *Neuer Wortschatz. Neologismen der 90er Jahre im Deutschen. Unter Mitarbeit von Elke Tellenbach und Doris al-Wadi*. Berlin / New York: deGruyter.
- Höhne, Steffen (1991). Die rolle des wörterbuchs in der sprachberatung: Eine sekundäranalyse zur wörterbuchbenutzungsforschung. *Zeitschrift für Germanistische Linguistik* 19:293–321.
- Klosa, Annette (ed.) (2008). *Lexikografische Portale im Internet. (=OPAL-Sonderheft 1/2008)*. Mannheim: OPAL Sonderheft 1/2008, <http://www.ids-mannheim.de/pub/laufend/opal/>.
- Klosa, Annette, Ulrich Schnörch, and Petra Storzjohann (2006). Elexiko - a lexical and lexicological, corpus-based hypertext information system at the institut für deutsche sprache. In Carla Marellò et al. (ed.), *Proceedings of the 12th EURALEX International Congress (Atti del XII Congresso Internazionale di Lessicografia)*, EURALEX 2006, volume 1, 425–430, Turin, Italy.
- Kunze, Claudia and Lothar Lemnitzer (2007). *Computerlexikographie. Eine Einführung*. Tübingen.: Narr.
- Kämper, Heidrun (2005). *Der Schuldskurs in der frühen Nachkriegszeit. Ein Beitrag zur Geschichte des sprachlichen Umbruchs nach 1945*, volume 78 of *Studia Linguistica Germanica*. Berlin / New York: deGruyter.
- Lemnitzer, Lothar (2001). Das internet als medium für die wörterbuchbenutzungsforschung. In Ingrid Lemberg, Bernhard Schröder, and Angelika Storrer (eds.), *Chancen und Perspektiven computergestützter Lexikographie*, 247–254, Tübingen: Niemeyer.
- Měchura, Michal Boleslav (2008). Giving them what they want: Search strategies for electronic dictionaries. In *Proceedings of the 13th EURALEX International Congress. Euralex 2008*, Barcelona, Spain, cD-ROM.
- Müller, Wolfgang (1991). Einige problematische dudenbenutzungssituationen. ein florileg aus den sprachanfragen. In Gerhard Augst and Burkhard Schaefer (eds.), *Rechtschreibwörterbücher in der Diskussion*, 335–361, Frankfurt/M.
- Müller-Spitzer, Carolin (2006). Das konzept der inhaltsstruktur. ein ausschnitt aus einer neuen konzeption für die modellierung lexikografischer daten. In *OPAL - Online publizierte Arbeiten zur Linguistik 2*, <http://www.ids-mannheim.de/pub/laufend/opal/privat/opal06-2.html>.
- Müller-Spitzer, Carolin (2007a). Das elexiko-portal: Ein neuer zugang zu lexikografischen arbeiten am institut für deutsche sprache. In Georg Rehm, Andreas Witt, and Lothar Lemnitzer (eds.), *Datenstrukturen für linguistische Ressourcen und ihre Anwendungen*, 179–188.
- Müller-Spitzer, Carolin (2007b). *Der lexikografische Prozess. Konzeption für die Modellierung der Datenbasis*. Tübingen: Narr.
- Müller-Spitzer, Carolin (2007c). Vernetzungsstrukturen lexikografischer daten und ihre xml-basierte modellierung. In *Hermes*, volume 38, 137–171.
- Ripfel, Martha (1989). Ergebnisse einer befragung zur benutzung ein- und zweisprachiger wörterbücher. In *Lexicographica*, volume 5, 178–201.
- Scherer, Tanja (2008). *Umsetzung von Zugriffsstrukturen bei Online-Wörterbüchern*. Master's thesis, University Mannheim.
- Schlaps, Christiane (2007). *Grundfragen der elektronischen Lexikographie. Elexiko – das Online-Informationssystem zum deutschen Wortschatz*. Berlin, New York: deGruyter, short review, in: *Lexicographica* 22, 72-94.
- Schryver, Gilles-Maurice De (2003). Lexicographer's dreams in the electronic-dictionary age. *International Journal of Lexicography* 16(2):143–199.
- Schryver, Gilles Maurice De and David Joffe (2004). On how electronic dictionaries are really used. In Geoffrey Williams and Sandra Vessier (eds.), *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004*, volume 1, 187–196, Lorient, France.
- Schryver, Gilles Maurice De and Daan J. Prinsloo (2000). Dictionary-making process with the "simultaneous feedback" from the target users to the compilers. In Ulrich Heid, Stefan Evert, Egbert Lehmann, and Christian Rohrer (eds.), *Proceedings of the Ninth EURALEX International Congress. Euralex 2000*, volume 1, 197–209, Stuttgart, Germany.

- Schumacher, Helmut, Jacqueline Kubczak, and Renate Schmidt (2004). *VALBU – Valenzwörterbuch deutscher Verben*. Tübingen: Narr.
- Storrer, Angelika (2001). Digitale wörterbücher als hypertexte: Zur nutzung des hypertextkonzepts in der lexikographie. In Ingrid Lemberg, Bernhard Schröder, and Angelika Storrer (eds.), *Chancen und Perspektiven computergestützter Lexikographie*, 53–69, Tübingen: Niemeyer.
- Thielen, Christine, Elisabeth Breidt, and Helmut Feldweg (1998). Compass. ein intelligentes wörterbuchsystem für das lesen fremdsprachiger texte. In Angelika Storrer and Bettina Harriehausen (eds.), *Hypermedia für Lexikon und Grammatik*, 173–194, Tübingen: Narr.
- Wang, Weiwei (2001). *Zweisprachige Fachlexikographie. Benutzungsforschung, Typologie und mikrostrukturelle Konzeption*. Frankfurt/M.
- Wiegand, Herbert Ernst (1998). *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*, volume 1. Berlin / New York: deGruyter.