OWID – A Dictionary Net for Corpus-Based Lexicography of Contemporary German

Carolin Müller-Spitzer
Institut für Deutsche Sprache, Mannheim

The Online-Wortschatz-Informationssystem Deutsch (OWID; Online German Lexical Information System) is a lexicographic Internet portal for various electronic dictionary resources that are being compiled at the Institute for the German Language (Institut für Deutsche Sprache, IDS). The main emphasis of OWID is on academic lexicographic resources of contemporary German. Presently, the following dictionaries are included in OWID: a dictionary of contemporary German called elexiko, a dictionary of neologisms, a small dictionary of collocations, and a discourse dictionary covering the lexemes that establish the discourse about "guilt" in the early post-war era 1945-1955. In the near future (2010/2011), several additional dictionaries will be published in OWID: a Textbook of German Communication Verbs, a Valency Dictionary of German Verbs, two further discourse dictionaries – one about the "democracy" discourse around 1968, the other covering the keywords of the German reunification 1989/1990. Moreover, 300 entries from a corpus-based project on proverbs will be integrated into OWID. Thereby, OWID is a constantly growing resource for academic lexicographic work of the German language.

Altogether, OWID is a special kind of dictionary portal owing to its content and its design, namely the integration of the various dictionaries, the access possibilities and the presentation features. With OWID, we try to establish a dictionary net where the different resources are jointly accessible not only by headwords, but also on the microstructural level. Prerequisite for these common access- and navigation-possibilities across the various dictionaries is the same concept for the lexicographic data model which we put into practice in OWID. Data from all dictionaries in OWID are structured according to a tailor-made, fine-granular, XML-based data model. In this data model, similar content is modelled similarly, dictionary related differences are preserved.

The main tasks for the future are to enhance OWID with further dictionary resources, to improve the inner access structures so that they exhaust the possibilities of the data model, and to customize the layout of the dictionaries as well as the search options according to the user's needs.

1. Resources in OWID

The *Online-Wortschatz-Informationssystem Deutsch* (OWID; Online German Lexical Information System) is a lexicographic Internet portal for various electronic dictionary resources that are being compiled at the Institute for German Language (Institut für Deutsche Sprache, IDS). The main emphasis of OWID is on academic lexicographic resources of contemporary German. Presently, the following dictionaries are included in OWID:

> *elexiko* consists of an index of about 300,000 short entries with information on spelling, spelling variation, and word division. Furthermore, *elexiko* contains more than 1,000 fully elaborated entries of high-frequency headwords, focussing on extensive semantic-pragmatic descriptions of lexical items in actual language use. The dictionary is continuously extended by further elaborated entries.

> The *Neologismenwörterbuch* (Dictionary of Neologisms) describes in detail about 900 new words and new meanings of established words, added to the German vocabulary since the 1990s. This dictionary is also constantly updated.

> *Feste Wortverbindungen* (Collocations Online) (2007ff.) publishes the research results of the project *Usuelle Wortverbindungen* (Fixed Multiword Expressions).

The *Diskurswörterbuch 1945-55* (Discourse Dictionary 1945-55) (2007) is a reference work covering the lexemes that establish the discourse about "guilt" in the early post-war era 1945-1955.

In the near future (2010/2011), several additional dictionaries will be published in OWID: the *Handbuch deutscher Kommunikationsverben* (Textbook of German Communication Verbs), the *E-VALBU – Valenzwörterbuch deutscher Verben* (Valency Dictionary of German Verbs), two further discourse dictionaries – one about the "democracy" discourse around 1968, the other covering the keywords of the German reunification 1989/1990. Moreover, 300 entries from a corpus-based project on proverbs will be integrated into OWID. Thereby, OWID is a constantly growing resource for academic lexicographic work of the German language.

2. OWID as a dictionary net

In Engelberg/Müller-Spitzer (forthcoming), the term 'dictionary portal' is defined rather broadly: The only conditions are that the portal is accessed via a computer screen, that it provides access to more than one dictionary, and that the integrated dictionaries can still be accessed as stand-alone products. This includes everything from dictionary link lists to highly interlinked lexicographic nets. In Engelberg/Müller-Spitzer (forthcoming), we provide a number of criteria to distinguish between different types of dictionary portals. The main criteria for distinguishing the central types of dictionary portals pertain to (1) types of access provided, (2) the implemented cross-reference structures between dictionaries, (3) the ownership relation between the portal and the contributing dictionaries, and (4) the layout of the portal. With regard to distinctions in the implementation resulting from these criteria, dictionary portals are differentiated into dictionary nets, dictionary search engines, and dictionary collections (cf. Figure 1). OWID is a typical example of a dictionary net.

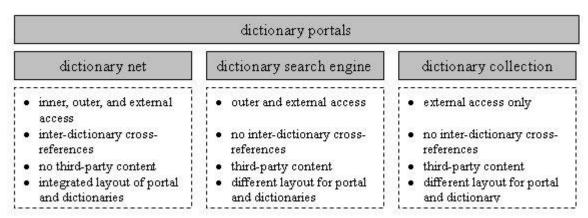| dictionary portals		
dictionary net	**dictionary search engine**	**dictionary collection**
• inner, outer, and external access • inter-dictionary cross-references • no third-party content • integrated layout of portal and dictionaries	• outer and external access • no inter-dictionary cross-references • third-party content • different layout for portal and dictionaries	• external access only • no inter-dictionary cross-references • third-party content • different layout for portal and dictionary

Figure 1. Typology of dictionary portals (cf. Engelberg/Müller-Spitzer forthcoming)

All resources in OWID are entirely compiled at the IDS and are integrated in a common layout; OWID does not contain any third-party content. Inter-dictionary cross-references exist, for example, from single word units from elexiko to corresponding multi word units from the *Feste Wortverbindungen*, i.e. from entry *blind* ('blind') to *Liebe macht blind* ('love is blind'). This interconnection is also visible by using the search option: if a user types in the word "blind", s/he gets a search result with the elexiko entry *blind* as well as all

corresponding entries from *Feste Wortverbindungen* with the explanation: "*blind*: base element of *Liebe macht blind*" etc.

Prerequisite for these common access- and navigation-possibilities across the different dictionaries is the same concept for the lexicographic data model. Data from all dictionaries in OWID are structured according to a tailor-made, fine-granular, XML-based data model. In this data model, similar content is modelled similarly, dictionary related differences are preserved.

3. Outer and Inner Access structures in OWID

Generally, OWID provides two main outer access possibilities: the main search box and an alphabetic register. Both outer access possibilities combine with a search option for including or excluding the individual dictionaries. The main search box works as follows: if a user types "frei" ('free') in the search box of OWID, he/she gets the results displayed in Figure 2.
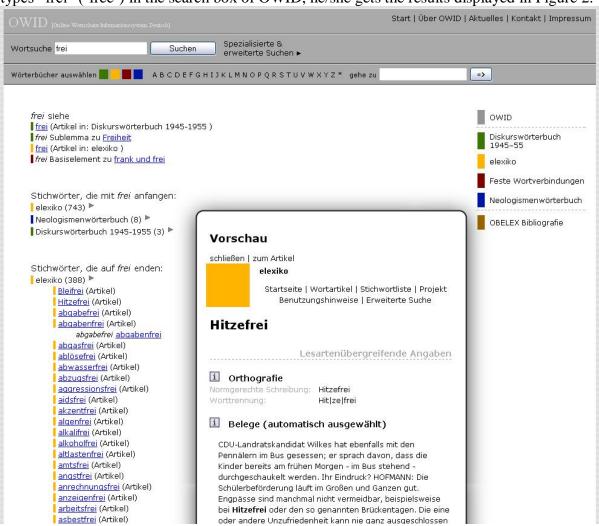

Figure 2. Results for the search term "frei" in OWID with a preview of the elexiko entry *hitzefrei* ('school's out today because of the heat').[1]

---

[1] The relaunch of OWID is in progress and is going to be online at the end of June 2010.

The search results in OWID are always divided into four groups: firstly, exactly fitting search results for the search string are presented; secondly, search results which begin with the search string such as *freiheitlich* ('liberal') or the neologism *Freisprechanlage* ('hands-free kit'); thirdly, search results which end in the search string like *bleifrei* ('unleaded') or *abgasfrei* ('zero-emission'), and lastly, search results which contain the search string such as *Befreiung* ('liberation'). In the online presentation, information from different dictionaries is rendered in different colours. Each dictionary is associated with a particular colour. The layout of the different dictionaries is embedded in the entire OWID layout.

If a user clicks on a search result, the entry is displayed in the context of the joined lemma-list of all OWID dictionaries (a rather special feature for dictionary portals). On the other hand, if a user chooses one individual dictionary via the option "choose dictionaries", only the headword list from the individual dictionary is displayed in the way as it is specified by the dictionary writers. In the discourse dictionary, for example, the dictionary-own headword list contains main headwords as well as sublemmas. The coloured little boxes on the right which stand for the different dictionaries signal to the user whether s/he had chosen a specific dictionary or whether s/he moves on the OWID level. If, for example, only one dictionary is chosen (like here the discourse dictionary), the corresponding box is displayed bigger than the other ones. With this differentiation – level of OWID in contrast to the level of a specific dictionary – we hope to meet two different user needs: firstly, searching for one word in no particular dictionary, or secondly, searching within one specific dictionary only and use this individual dictionary like a stand-alone product (cf. Figure 3).
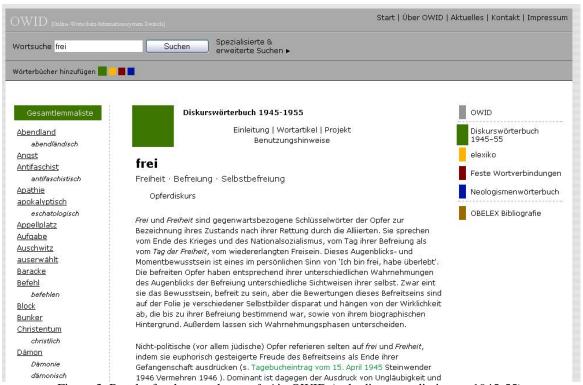

Figure 3. Results for the search term *frei* in OWID (only discourse dictionary 1945-55).

The second outer access structure besides the search box is the alphabetic register and the corresponding "go to"-box. Via this box, a user has a kind of fast outer access like leafing through a book and stopping at a certain place in the alphabet. In contrast to the search option explained above, entering "frei" in this search box, leads not to a search result, but starts a look-up in the OWID headword list and immediately displays the best matching entry with the corresponding part of the headword list (cf. Figure 4). Accordingly, if a user types only the first two letters of a word in this "go to"-box, the first matching headword with the entry is displayed, for example, typing in "fr" leads to the entry *Fracht* ('cargo'). Again, this alphabetic register combines with the search option for including or excluding the individual dictionaries. Therefore, it is possible to use the alphabetic register for navigating in one specific dictionary in OWID.
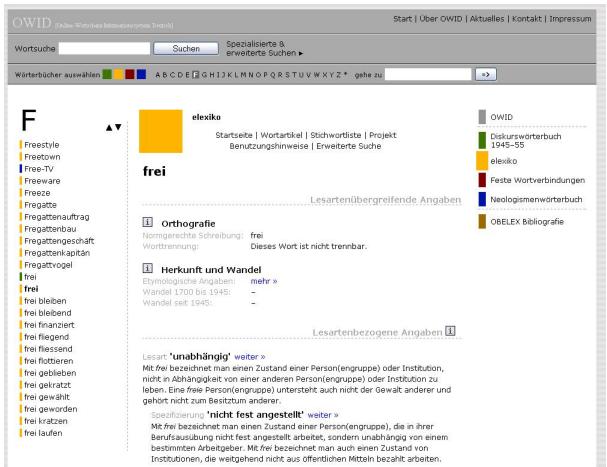

Figure 4. Accessing the entry *frei* via the alphabetic register

Inner access is provided through extended search options. To illustrate the granularity of the search options consider the following example: In the domain of word formation, it is possible to search for all explicit derivatives where the basis is the category "adjective". The search may be further limited to these lexemes which in the process of word formation changed their vowels, like *kürzlich* ('recently') derived from *kurz* ('short'). These search options in the field of word formation are provided across the elexiko and the neologism dictionary due to the consistent concept of data modelling (cf. Figure 5). Other extended search options are provided across other OWID dictionaries. However, extended search

options in this granularity are currently offered to lexicographers only and not to OWID users. It is one of the main tasks for the future in OWID, to bring the search options for OWID users into line with the internal ones for lexicographers (as far as it corresponds to user's needs) and therefore make full use of the potential which the granular data modelling offers.

Figure 5. Extended search options for lexicographers

To intensify the character of OWID as a dictionary net, new ways of grouping the different dictionaries together (especially the different discourse dictionaries) as well as new forms of visualisations will be realised in OWID in the near future. For example, we want to develop a synopsis of the different discourse dictionaries where one is able to see which words occur in the relevant discourses at various times.

4. Non-dictionary resources in OWID

Besides the main function as a lexicographic portal, OWID provides another service for researchers in the context of online lexicography, the "Online-Bibliography of Electronic Lexicography (OBELEX)" (cf. the poster of Müller-Spitzer/Möhrs in these proceedings). All publications recorded in OBELEX are labelled by keyword and language. Information on dictionaries is currently not included in OBELEX; the main focus is on metalexicography. However, we are working on a database with information on online dictionaries as a supplement to OBELEX. Another supplement to the dictionaries in OWID which will be presented online till summer 2010 is the "OWID-Korpussuche", a corpus query site to look up the IDS-corpora of written German which cover 3,75 billion of words (cf. http://www.ids-mannheim.de/kl/projekte/korpora/). In contrast to the corpus query system Cosmas 2 (http://www.ids-mannheim.de/cosmas2/) which provides a lot of extended search options and possibilities to analyse corpus data, e.g. in form of a cooccurrence analysis, but is otherwise not always easy to use, the corpus search via OWID has the intention to offer a simplified access to the IDS corpora which may be used without any previous experience or tutorial. With all these resources and functionalities, OWID serves as the first adress for all resources concerning Internet lexicography at the IDS.

5. Closing remark

Altogether, OWID is a special kind of dictionary portal not only owing to its content but also because of the way in which it brings together the different dictionaries and by virtue of its access possibilities as well as some special features of its mode of representation.

As said before, the main tasks for the future are to enhance OWID with further dictionary resources, to improve the inner access structures so that they exhaust the possibilities of the data model, and to customize the presentations of the dictionaries as well as the search options according to the user's needs. Therefore, we are currently running a project of user research on online dictionaries with the aim to learn more about the preferences of users in certain using situations and how a user-adapted view may be developed (cf. www.benutzungsforschung.de).

References

DICTIONARY.COM = Dictionary.com. An Ask.com Service. Published by Dictionary.com 2009. Online, May 4, 2009, http://dictionary.reference.com/.

OWID – Online-Wortschatz-Informationssystem Deutsch (2008ff.), ed. by Institut für Deutsche Sprache, Mannheim, www.owid.de.

Engelberg , S.; Müller-Spitzer, C. (forthcoming) 'Dictionary portal'. In R.H. Gouws; U. Heid; W. Schweickhard; H.E. Wiegand (eds) *Dictionaries. An international encyclopedia of lexicography. Supplementary volume: Recent developments with special focus on computational lexicography*, Berlin / New York: de Gruyter.

YOURDICTIONARY.COM = YourDictionary.com. Published by LoveToKnow, 1996-2009. Online, May 8, 2009, http://www.yourdictionary.com/.