1 Dictionary portals and other lexical systems

Electronic media, in particular the Internet, provide ample opportunity to store, combine, and cross-reference information of any sort. While experienced users of print dictionaries have always turned to dictionaries of different types in order to satisfy their needs, it took them some effort to identify and physically assemble all these dictionaries. With the potential to do away with this physical restriction and, in addition, to provide new search options and a broad net of cross-references beyond what is implementable in print dictionaries, electronic media appeal to dictionary users and producers (cf. Pulitano 2003, De Schryver 2003). Collections of electronic dictionaries, sometimes cross-referenced and provided with new access structures, are often referred to as dictionary portals. Within the Internet domain, the word “portal” is used to refer to Internet pages that are designed to serve as starting points for searches for information in particular thematic domains. They contain links to other webpages, search functions, and web applications (cf. Müller-Spitzer 2007; Klosa, Lemnitzer & Neumann 2008). Different kinds of portals can be distinguished. Internet service providers make general Internet portals available as link-based starting pages to the WWW; search portals like Google or Altavista provide a search-based entrance to the Internet, and, similarly, thematically more specialized portals provide access to information on certain domains, companies, scientific institutions, etc. We will use the term dictionary portal to refer to a data structure (i) that is presented as a page or set of interlinked pages on a computer screen and (ii) provides access to a set of electronic dictionaries, (iii) where these dictionaries can also be consulted as standalone products.
Dictionary access can be provided to the dictionary as such – e.g., its starting page – or to elements of its macrostructure and/or microstructure (cf. Section 2). Dictionary portals shall be distinguished here from two other kinds of lemma-based information portals, i.e., portals that provide access to reference works in which information is attributed to lemma signs (cf. Figure 1). On the level of data content, dictionary portals differ from encyclopedic portals; where dictionary portals connect reference works on linguistic information associated with lexical entries encyclopaedic portals connect reference works on world knowledge associated with lexical entries. On the level of data description dictionary portals differ from lexicological portals: Dictionary portals provide access to a set of reference works with information on lexical items that have been compiled in lexicographic processes; lexicological portals allow users to access lexical resources that are mainly of a non-lexicographic sort such as (i) lexical information automatically extracted from corpora, like example sentences or co-occurrence data, (ii) data from electronic grammars pertaining to lexical items, (iii) didactic material concerning lexical items or classes of lexical items.

![Figure 1: Dictionary portals and related lemma-based information portals.](image)

Dictionary portals are represented by search engines like ONELOOK as well as highly interlinked portals like OWID or WÖRTERBUCHNETZ. STOUT ENCYCLOPEDIAS and N-ZYKLOP constitute encyclopedic portals: The first one is a list of links to encyclopedias, the second one is a searchable database of encyclopedias. An example for a lexicological portal is DEUTSCHER WORTSCHATZ: It combines different kinds of data on lexical items which are automatically extracted from text corpora such as collocations, frequencies, morphological properties, and semantic relations. In contrast to dictionary-free lexicological portals such as DEUTSCHER WORTSCHATZ other lexicological portals are construed around a single dictionary. Based on a digitized version of the “Wörterbuch der deutschen Gegenwartssprache” and the “Diccionario crítico etimológico castellano e hispánico”, respectively, the DWDS and the PORTAL DE LÉXICO HISPÁNICO (cf. Claveria et al. 2008) constitute such dictionary-based lexicological portals.

The properties displayed in Figure 1 constitute prototypes of lemma-based information portals. Actual instances of portals can be blends of these types; for example, in some cases, dictionary and encyclopedic portals are combined. ENCYCLOPEDIA.COM allows searches for entries in general encyclopedias like “The Columbia Encyclopedia” as well as specialized encyclopedias and dictionaries like “Webster’s New World Dictionary”. Similarly, Bartleby’s
Internet portal BARTLEBY.COM provides simultaneous access to dictionaries like “Roget’s Thesaurus” and “The Columbia Encyclopedia” as well as reference books without a lemma-based access structure. In particular for portals dealing with specialized language (e.g., OVER 2,500 GLOSSARY LINKS), it is often difficult to decide about the type of the reference works integrated. Many of them show properties of dictionaries as well as encyclopedias. Dictionary portals and lexicological portals can also blend into each other. Sato’s (2003) REBLIN allows access to a dictionary portal consisting of nine dictionaries, retrieves concordances from a web-as-corpus application, and provides access to a movie database that links words with clips from movies where the word is uttered.

In the remainder of the article, the focus will be on dictionary portals with an emphasis on general monolingual and bilingual dictionaries.

2 Typology of dictionary portals

The term ‘dictionary portal’ has been 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 standalone products. This includes everything from dictionary link lists to highly interlinked lexicographic nets. In this section, we provide a number of criteria to distinguish between different types of dictionary portals (cf. also Engelberg & Lemnitzer 2009). 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.

1) The main function of dictionary portals is to provide access to a set of dictionaries in a way that the dictionary user obtains the kind and amount of information he needs. Three main types of access structures within dictionary portals can be distinguished (cf. Figure 2): (i) External access structures provide access to the dictionary surface: They consist of links to the starting pages of the dictionaries listed; all searches for particular lexicographic information will be carried out within the dictionaries themselves, i.e., outside the portal (e.g., OVER 2,500 GLOSSARY LINKS). (ii) Outer access structures serve to access dictionary entries by searching through all the lemmata of the integrated dictionaries (e.g., AKADEMIE-WÖRTERBUCHPORTAL; cf. Figure 4 below). (iii) Inner access structures allow to call up information from specific information positions within the microstructure of the entries of the participating dictionaries (e.g., CANOO.NET, OWID).
2) In some dictionary portals, entries in different integrated dictionaries are explicitly cross-referenced. For example, lemma \( f \) in dictionary 1 and lemma \( f \) in dictionary 2 can be connected by an explicit mutual link that shows up in both entries and allows to switch between these entries (cf. Figure 2). Besides the provision of inner access, explicit inter-dictionary cross-reference is evidence for highly integrated dictionary portals.

3) The possibility to implement certain kinds of access structures and cross-references is connected to questions of the right of ownership. It is easier to increase the interconnectedness between the dictionaries if the owner of the portal is identical to the owner of the integrated dictionaries. Unfortunately, information of this sort is not always easy to come by. Only some portals explicitly inform their users about this situation: “Note that Dictionary.com does not produce all the dictionaries that appear on our Web site; we simply make them available. In that respect we are more like a bookstore or library than a publisher.” (DICTIONARY.COM, April 23, 2009)

In some cases, the institution running the portal does not own the lexicographic material but nevertheless, has privileged access to the dictionary structures. This is often the case if the institution running the portal has carried out the digitalization of the dictionaries, for example, with the WÖRTERBUCHNETZ whose dictionaries have been digitized by the University of Trier’s competence center for processes of electronic access and publication. Thus, we can distinguish dictionary portals where portal and dictionaries have the same owner from those where they have different owners (i.e., integrate third-party content). This classification is partly but not completely extensionally identical with the distinction between portals that exert influence on the digital structure of the participating dictionaries and those which do not.
4) Some dictionary portals have an integrated layout for the portal and its dictionaries so that navigation and search facilities are similar across the dictionaries (e.g., WÖRTERBUCHNETZ, Leo, OWID). The layout of other dictionary portals is restricted to the portal page itself. After having accessed a dictionary, the user is confronted with the particular idiosyncratic surrounding of this particular dictionary.

The four distinctions presented above lead to the typology of dictionary portals represented in Figure 3, which distinguishes three types of dictionary portals: dictionary nets, dictionary search engines, and dictionary collections. The basic idea behind this typology is to distinguish dictionary portals by the degree to which they integrate the participating dictionaries. The typology is to be understood as rendering prototypes to which actual dictionary portals correspond to more or less. The prototypical dictionary net allows inner, outer, and external access to dictionaries and shows a high degree of inter-dictionary cross-reference and an integrated layout for the portal and the dictionaries. The owner of the portal has privileged access to the dictionaries and usually has the right of ownership to the dictionaries. A typical example for a dictionary net is OWID. Dictionary search engines usually provide outer and external access and do not exhibit inter-dictionary cross-reference. Portal and dictionary owners are not identical, and third-party content is represented in its own layout. ONELOOK is a typical example. Dictionary collections only provide external access to dictionaries that are usually not owned by the institution that runs the portal. An example for a dictionary collection is OVER 2,500 GLOSSARY LINKS.

<table>
<thead>
<tr>
<th>dictionary portals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>dictionary net</strong></td>
</tr>
<tr>
<td>• inner, outer, and external access</td>
</tr>
<tr>
<td>• inter-dictionary cross-references</td>
</tr>
<tr>
<td>• no third-party content</td>
</tr>
<tr>
<td>• integrated layout of portal and dictionaries</td>
</tr>
<tr>
<td><strong>dictionary search engine</strong></td>
</tr>
<tr>
<td>• outer and external access</td>
</tr>
<tr>
<td>• no inter-dictionary cross-references</td>
</tr>
<tr>
<td>• third-party content</td>
</tr>
<tr>
<td>• different layout for portal and dictionaries</td>
</tr>
<tr>
<td><strong>dictionary collection</strong></td>
</tr>
<tr>
<td>• external access only</td>
</tr>
<tr>
<td>• no inter-dictionary cross-references</td>
</tr>
<tr>
<td>• third-party content</td>
</tr>
<tr>
<td>• different layout for portal and dictionary</td>
</tr>
</tbody>
</table>

Figure 3: Typology of dictionary portals.

Besides the four central criteria the typology is based upon, there are other properties according to which dictionary portals can be described: (5) the presentation of the search results, (6) the languages described, (7) the number of dictionaries integrated, and (8) the medium.

5) Dictionary portals opt for various kinds of data presentation. The result of a lemma search based on outer access can be presented within the context of a joined lemma list as in the WÖRTERBUCHPORTAL (cf. Figure 4), or it can be displayed without this context as in ONELOOK. The information given about the search item can be accessible by a list of links that lead into the entries of the dictionaries that contain the lemma (e.g., ASL DICTIONARY PORTAL, WÖRTERBUCHNETZ), or the dictionary entries can be displayed in their entirety on the screen as with DICTIONARY.COM.
6) Dictionary portals can be distinguished as to whether they are devoted to one language (mono-language portals) or to several languages (multi-language portals). Portals centered on one language provide access to monolingual dictionaries and/or to bilingual dictionaries that relate the language the portal focuses on to other languages (e.g., CAMBRIDGE DICTIONARIES ONLINE, ASL DICTIONARY PORTAL). An example for multi-language portals are YOURDICTIONARY.COM and, focusing on bilingual dictionaries, the PONS PORTAL.

7) The number of dictionaries integrated into a portal varies considerably. Dictionary nets with their highly connected microstructures often only contain a handful of dictionaries while dictionary collections and search engines are often based on a large number of dictionaries. According to their web pages (April 27, 2009), the dictionary collection YOURDICTIONARY.COM lists “more than 2,500 dictionaries and grammars in over 300 languages”, and the dictionary search engine ONELOOK indexes 1009 dictionaries.

8) The article at hand focuses on Internet dictionary portals. However, dictionary portals, in particular dictionary nets and search engines, also come as handheld dictionary devices and CD-ROM-based PC portals. Producers of portable devices such as handheld dictionaries (e.g., SEIKO TR-700) began in the 1980s to offer sets of dictionaries provided with common outer or inner access. A little later reading pens (e.g., the WizCOM SUPERPEN PROFESSIONAL) also integrated several dictionaries. These portals were followed by CD-ROMs that offered similar functions under a more elaborate user interface, primarily produced by the large British dictionary publishing houses (e.g., BOOKCASE). Except for dictionary collections, Internet-based dictionary portals were the last step in this development (cf. Nesi 2009 on the development of English electronic dictionaries).
3 Dictionary collections
From a technical point of view, dictionary collections are the simplest examples of dictionary portals, because usually they are just websites containing links to other online dictionaries (e.g., YOURDICTIONARY.COM, OVER 2,500 GLOSSARY LINKS, and ERLANGER LISTE). While the dictionaries collected are usually third-party content, in some cases the portal and the listed dictionaries constitute a unit since the owner of the portal and the dictionaries are identical and all dictionaries are presented within the interface provided by the portal (e.g., LEO). Dictionary collections are usually realized as link lists. However, other forms of access and presentation are conceivable. Some collections provide not only links to dictionaries but also integrate the search masks of some of the dictionaries or other portals (e.g., YOURDICTIONARY.COM, ERLANGER LISTE, cf. Figure 5).

![Figure 5: Dictionary collection ERLANGER LISTE with a search mask for an integrated German-English dictionary and an integrated dictionary portal (ONELOOK).](image)

Furthermore, database searches instead of link lists are about to be realized within dictionary collections. The Online-Bibliography OBELEX, which currently references metalexicographic work on electronic lexicography, will be extended by the addition of a dictionary collection. The external access to the electronic dictionaries will not be provided via a link list but in form of a searchable online database (cf. Engelberg, Klosa & Müller-Spitzer 2009).
4 Dictionary search engines

Typically, the institution that runs a dictionary search engine indexes the integrated dictionaries. Most dictionary search engines only index the lemma lists (e.g., LEXILOGOS, WORDREFERENCE, ASL DICTIONARY PORTAL, AKADEMIE-WÖRTERBUCHPORTAL), some of them also index the full text or parts of it (ONELOOK). Most of them have in common not only that they are based on index building processes but also that they deal with third-party content. Dictionary search engines like ONELOOK often explicitly inform their users about this relation: “OneLook.com, like most Internet search services, provides a way to find and explore third-party content but does not incorporate it directly into the site. We’re committed to respecting the intellectual property of the dictionary and glossary providers we index.” (ONELOOK, April 23, 2009)

ONELOOK is an example of a dictionary search engine that not only provides outer access to the integrated dictionaries but also allows users to restrict the search to either definitions or translations. In addition, it offers a “Reverse Dictionary” option: “Here we search our references for words that have definitions conceptually similar to the words you search for. We do this using a motley assortment of statistical language processing hacks.” (ONELOOK, May 2, 2009). For example, a search for “ship cargo transport” renders a list of words containing – among others – expressions for transport ships like freighter, tanker, lighter, and container ship, sorted by relatedness (more than 1,000 results).

The method outer access is provided can differ from the prototypical case. In LEXILOGOS, a language has to be chosen before starting the look-up. In CAMBRIDGE DICTIONARIES ONLINE – not a prototypical dictionary portal because of the doubtful status of the integrated dictionaries as standalone products – a dictionary has to be chosen before a search can be performed. This amounts to a combination of external and outer access that does not allow access to the integrated dictionaries simultaneously.

Dictionary search engines usually show a hit list as result where each hit constitutes a link into the corresponding dictionary. By clicking on a search result, the user leaves the portal and enters an individual dictionary; thereby, dictionary search engines provide a metastructure for accessing many dictionaries as third-party content. Some index-based portals do not share this prototypical feature of dictionary search engines. As we have seen, DICTIONARY.COM does not present hit lists but displays entire entries from all dictionaries that contain the lemma. DICTIONARY.COM is not a typical search engine in another respect. It allows outer access to the integrated dictionaries but no external access to the dictionaries themselves. This makes it difficult to see whether the dictionaries listed on the “about” page are really standalone products.

A further example of a presentation of search results rather untypical for search engines is the ASL DICTIONARY PORTAL. It allows a view of the search results (“Alternate ASL Dictionary Portal”) where the videos appear on the same page as the hit list rather than on a separate page (cf. Figure 6). The integrative way of data presentation almost reminds one of a dictionary net.
Other portals, like bartleby.com, deviate from pure dictionary search engines with respect to their content in that they index encyclopedias and fictional texts as well. One of the general problems of large search engines indexing hundreds of sites is that the integrated dictionaries differ very much in quality. Although portals like OneLook claim to observe a certain standard for the integration of a dictionary ("Your dictionary should be free of spelling errors in its headwords, and its definitions must be accurate.", OneLook, May 2, 2009), de Schryver (2003: 157) is rather sceptical about search engines: "All the search data are shown in long lists, results from trustworthy sources and downright amateurish concoctions all mixed up."

5 Dictionary nets
Typically, with dictionary nets (e.g., Canoo.net, Wörterbuchnetz, OWID), the owner of the portal is also the owner of the digital content of the integrated dictionaries or at least has privileged access to it. This makes it possible to structure the dictionary content according to the needs of the portal and to install cross-references between the different dictionaries. Dictionary nets often share a common concept of data modelling and structuring. OWID, for example, influences the data structures of the integrated dictionaries in order to develop its potential as a dictionary net (cf. Müller-Spitzer 2008; Engelberg, Klosa & Müller-Spitzer 2009; for another example see Bird 2002 on a portal for dictionaries of American indigenous languages). This allows users to search for lemmata in different dictionaries that share certain properties, which are explicitly encoded in the corresponding information positions. For example, all lexemes that are classified as explicit derivations with an adjectival basis can be retrieved in OWID. Search results of this advanced search option are – among others – headwords like Ersti (coll. for ‘freshman’) or schwächeln (‘to be ailing / indisposed’) from a dictionary of neologisms and headwords like Aktivität (‘activity’) or ergänzen (‘to complement’) from a general monolingual dictionary. The prerequisite of such search
options is a fine-grained cross-dictionary data modelling. Since information on word formation is modelled identically in both XML-based dictionaries in OWID (cf. Figure 7), an explicit cross-dictionary search can be based on one common XPath-term (here: //expl-derivation/exd-basisA[@basistyp="adjektiv"]).

Creating and administrating connections between the dictionaries in a dictionary net is a demanding task. Another interesting example of the different ways to accomplish this task is the WÖRTERBUCHNETZ that contains different dictionaries of regional or temporal varieties of German. The printed versions of these dictionaries are only partially interlinked because of the historical order of their compilation (older ones cannot link to newer ones). Thus, the digital version is being interlinked more thoroughly using statistical methods and methods from information retrieval (cf. Burch 2008: 137f.). For example, based on similarities between the dictionary entries, these methods should allow to reveal the relatedness of butterblume (entry in the “Deutsches Wörterbuch”) and its dialectal equivalent Gackel(s)-blume (entry in the “Pfälzisches Wörterbuch”) and to install a cross-reference between the two entries. The online version of the WÖRTERBUCHNETZ demonstrates initial interesting results of this processing method.

Besides providing inner access in cross-dictionary searches, dictionary nets often also establish explicit cross-references between the dictionaries (OWID, VERBUND MITTELHOCHDEUTSCHER WÖRTERBÜCHER, cf. Figure 8) and a joined headword list that builds the framework of the portal (OWID).
The **VERBUND MITTELHOCHDEUTSCHER WÖRTERBÜCHER** (cf. Fournier 2001) deviates from typical dictionary nets in that it does not provide a common outer access structure. Its net-like properties come from the large number of explicit cross-references at the level of microstructures. However, a superordinate meta lemma list in which corresponding lemmata of different dictionaries are mapped onto a meta lemma, which subsumes its diachronic and dialectal variants, could strengthen the net-like character of this portal.

Dictionary nets differ in the extent to which the integrated dictionaries are meant to complement each other. In **OWID** and **WÖRTERBUCHNETZ**, the dictionaries have been compiled independently, notwithstanding the common data modelling. A somewhat different example of a dictionary net is **CANOO NET**, a network of complementary dictionaries, giving spelling or morphological information stemming from different dictionaries. In a sense, the information coming from the different dictionaries, which can also be used as standalone products, adds up to one virtual dictionary entry on the user surface.

This is to say, as far as the presentation of search results is concerned, dictionary nets offer the possibility to create virtual microstructures that combine data from information positions stemming from different dictionaries. However, this has only been realized in a very rudimentary manner so far. Figures 9 and 10 show how **CANOO NET** and **OWID** present information on a lemma, which has been retrieved from four, respectively, three integrated dictionaries, partly as explicit information, partly as links leading into the entries of the dictionaries.
Figure 9: Search result for the search term “bunt” in CANOONET; the information stemming from the defining dictionary (“Bedeutungswörterbuch”) is given explicitly; information from the dictionaries of orthography, inflection, and word formation is accessible via links.

Figure 10: Part of the search result for frei* in OWID; sublemmata and information about base elements of collocations and morphological properties like comparative formation and orthographic variants (“Schreibvariante”) are retrieved from different dictionaries and are presented partly explicitly and are partly accessible via links. (The links are distinguished by their color according to the dictionary the links lead to.)
6 Closing remark

Dictionary portals are based on different strategies. While dictionary nets come closest to the concept of a set of highly interlinked lexicographic resources accessible by advanced search options, they often integrate only very few dictionaries. Thus, they are probably still only attractive for a small but specialized community of users. Search engines with a large number of dictionaries indexed cater to those users who want a lot of dictionary data to choose from. The improvement of search algorithms will certainly increase their value for different kinds of user needs. However, users of search engines have to be experienced enough to assess the quality of the search results.

“Two decades ago, researchers like Arnold [...] or Kay [...] predicted that dictionaries would cease to be merely products, and would start to be offered as a user-oriented online service.” (De Schryver 2003: 157) Dictionary portals have certainly followed this course from the single lexicographic product to a general lexicographic information service. Lacking research on the use of dictionary portals, it is too early to say how far dictionary portals already meet the needs of their users. The high number of user requests mentioned on the starting pages of some portals (e.g., Dictionary.com) shows at least that there is a great demand for such online services.

References

Cited online resources


BOOK CASE = BOOKcase. [Pc-based dictionary software that allows simultaneous access to dictionaries from the Routledge, Cambridge, and Oxford series as well as to Internet search engines (cf. Nesi 2009: 468f.).]


RebLin = RebLin. [Software that integrates several dictionaires, a web-as-corpus application and a movie database; cf. Sato (2007).]


Other references


