Online dictionaries – how do users find them and what do they do once they have?

Henrik Lorentzen & Liisa Theilgaard

Abstract

In general, user behaviour studies on online dictionaries have focused on user behaviour once the user is on the site. But before a potential user even reaches this stage, he or she must succeed in finding the dictionary on the web. In this paper we investigate users' linguistic search strategies before they enter our dictionary site, *ordnet.dk*. What kind of search engine queries are successful and why (not)? Similarly, we have studied the site search queries. Are the search strategies the same? Taking the no-match searches as a starting point, we have asked ourselves if our content and search functionality correspond to the search behaviour of the users, that is if we can give an answer to the users' queries and if data is organized and presented in an appropriate way. Given the results of these analyses, we decided to make several changes to the site in order to optimize user access and attract new users. These changes and their ensuing results are presented. Furthermore, we present and discuss the results of a user survey conducted in October-November 2011.

1. Background, tools, terminology

The website *ordnet.dk*, which is the object of this paper, consists of three parts: The Danish Dictionary (Den Danske Ordbog, DDO), Dictionary of the Danish Language (Ordbog over det danske Sprog, ODS) and KorpusDK. In the following, we will concentrate on DDO.

In our study of user behaviour we have primarily used two tools, Google Analytics and our log files. Google Analytics (cf. Clifton 2010) provides a wide variety of key metrics on user behaviour, for example the daily number of visitors, and on how the users access the site. At the same time we monitor the users through our own log files (cf. Almind 2008, Nygaard and Fjeld 2008, Bergenholtz and Johnsen 2007), in particular so-called no-match lists, that is lists of the items that the user does *not* find when searching in DDO.

The study was initiated in September 2010 and it led to a number of changes to the site. The effects of these changes were measured in January 2012.

When talking about 'visits' we refer to the number of individual sessions started by a user. If the visitor has been inactive for 30 minutes or more, new activity is counted as a new visit.

2. How do the users find the dictionary?

There are three different ways to access the dictionary:

(1) The user enters a query into a search engine and clicks a link to the dictionary in the search engine result page.
(2) The user arrives directly at the dictionary by clicking a bookmark or by typing the site URL directly into the browser.
(3) The user accesses the dictionary through a referring link on another website.

In September 2010, 49% accessed DDO with a search engine query (99% Google), 33% accessed DDO directly, and the remaining 18% accessed DDO from a reference in another website. Thus, users apparently prefer to use a search engine as their point of entry when

looking for linguistic information, which in turn makes it very important to appear as a search result in Google, preferably on the first result page, to ensure that a potential user becomes aware of the dictionary.

3. How do the users search in a search engine?

In order to study the users' behaviour in a search engine, we have examined all search terms entered by the users in September 2010 which resulted in a visit. We have no access to search engine search logs, so we have no data to study potential search terms not resulting in a visit.

Firstly, we looked at the number of words in every unique search term, a word being a string without white spaces. It appears that 60% of the users wrote one word, 23% two words and 17% three or more words, meaning that about 40% search terms contain more than one word. The average query length is 1.69 words, and the maximum number of words is 33.

Secondly, we have grouped all search terms into three categories: (1) meta words (e.g. *Danish dictionary*, *ordnet.dk*), (2) linguistic phrases (a dictionary word or expression, e.g. *insinuate*, *to be kicked upstairs*) or (3) a combination of the two (e.g. *nonchalant meaning*, i.e. the user is looking for the sense of the word *nonchalant*) (see Table 1; for ease of understanding, all examples are given in English despite the actual data being in Danish):

Table 1. Distribution of search terms from a search engine in September 2010.

	Number of search terms	Search terms in %	Number of visits	Visits in %	Pageviews per visit	Pageviews in %	Examples
Meta word	1,160	25%	28,355	64%	47,125	30%	*ordnet.dk, Danish dictionary, ddo*
Linguistic phrase	2,220	48%	13,290	30%	104,400	66%	*dissolve, today to day*
Combination	1,251	27%	2,780	6%	7,250	5%	*nonchalant meaning, cheat synonym*
Total	4,631	100%	44,425	100%	158,775	100%	

The 4,631 different search terms have resulted in 44,425 visits, that is an average of 10 visits per search term. However, the five most effective search terms *dictionary*, *Danish dictionary*, *The Danish Dictionary*, *ordnet* and *ordnet.dk* account for as much as 39% of all visits, and altogether, the most successful search terms are meta words, which in turn has to do with the type of pages the web crawlers were allowed to index. In contrast, linguistic search terms generate far more pageviews per visit (66%) than meta word search terms do (30%).

For fear of insufficient server capacity it was initially decided only to allow indexation of the front page, help pages and other material about the dictionary, that is about 25 pages from DDO and about 65 pages altogether in *ordnet.dk*. Accordingly, no dictionary entry pages were indexed automatically, except for a few in cases where other websites had linked to a particular dictionary entry page. But this is at odds with the users' actual search behaviour – where linguistic phrases account for 48% of the search terms and 66% of the pageviews per visit – so there is a huge user potential in allowing for linguistic phrases to turn up in a search engine result.

4. How does the user search in the search box?

When we compare the users' behaviour in a search engine to their behaviour once they have accessed DDO and use the site search box, the picture looks quite different.

For this part of the study, we have used data from March 2011 since the logging of search terms from the search box was not yet switched on in Google Analytics in September 2010. Here, it appears that 96% of users search for just one word while 2.4% search for two words and 1.6% for three or more words. Compared to the 60% one-word searches in a search engine, it is obvious that the user behaves differently when using the site search box.

The sheer number of different search terms (194,660) rules out a quantitative treatment and a systematic classification, but a quick glance through the search terms shows that the vast majority of users search for a linguistic phrase. Wild cards as well as meta searches and combined searches are hardly ever used. Clearly, the user conceives the site search box as a dictionary look-up function as it is generally used in electronic dictionaries as opposed to other websites where it is used for full text search on the whole site.

5. What does the user not find?

Using our log files, we have examined all registered site search terms in September 2010, 20.6% of which gave no result. This number is relatively low compared to two earlier studies of Scandinavian online dictionaries: 31% (Nygaard and Fjeld 2008: 59) and 41% (Hult 2008: 78). In their study of a dictionary between Sesotho sa Leboa and English, de Schryver and Joffe (2004: 191) reported success rates from 67 to 75%, allowing us to assume that 25 to 33% of searches gave no match.

Secondly, we have analyzed and categorized the 1,000 most frequent unsuccessful search terms into a number of types. The 1,000 search terms represent 12.6% of all unsuccessful search terms. A given search term may occur in more than one category, for example both as a spelling error and as an inflectional form. The most frequent type of unsuccessful search terms was spelling and typing errors (30%) and the second-most frequent was inflectional forms (20%). We were surprised that so many inflectional forms were searched for. One possible explanation is that we offer an open-search function which allows the user to look up any word form on a web page by highlighting it and then right-clicking it to look it up directly in the dictionary.

Other types of unsuccessful search terms are (in declining order) foreign words, proper nouns, compounds, new words, multi-word expressions, abbreviations and derivatives – types which indicate what users look for in vain and what we should consider providing a wider coverage of in DDO.

6. Changes

With these results in mind, we decided to make a number of changes to the website. Above all, we wanted to improve our visibility in search engine results, so we allowed web crawlers to index all individual dictionary entry pages, so that instead of fewer than 100 indexed pages, the total number of indexed pages amounts to approx. 100,000 pages for DDO and 350,000 for *ordnet.dk* as a whole.

Secondly, something had to be done about the large number of inflectional forms. Consequently, we have added the present participle of verbs, which is often used as an adjective, but in DDO it is generally treated in the entry for the verb. Similarly, the genitive

forms of all nouns have been added. Finally, the results of the latter function is now also displayed in the more prominent middle section of the page (and not only under the search box in the right-hand sidebar) in order to attract the user's attention to the suggestions presented when a search renders no match.

7. Results

The indexation of all dictionary entry pages has had a remarkable effect. The number of daily visits has increased from approx. 3,000 in September 2010 to approx. 19,600 in January 2012, and the curve shows no signs of flattening off. Obviously, this impact is also reflected in the distribution of the users' path to the dictionary. In September 2010, 49% of the visits came from a search engine whereas in January 2012 it was 84%. Through this indexation we are now visible where users look for information (even linguistic), that is in the search engines, mainly Google. Not only new visitors (84%) access our site through a search engine, even 84% of the returning users use a search engine as their point of entry. And quite often we now figure on the first result page (with ten results): The average position of the top 1,000 daily queries resulting in a click to our web site in January 2012 was 2.4. Add to this that the users are now referred directly to the answer in the form of a dictionary entry page rather than the front page.

The number of new users has also increased from 31,190 (34%) in September 2010 to 221,800 (36%) new visitors in January 2012. Thus, we have succeeded in attracting more new visitors as well as turning many of them into returning visitors.

An analysis of the search terms leading users to the dictionary through a search engine in January 2012 shows 'the long tail' distribution; eleven times as many users access the dictionary through a search result compared to September 2010, but at the same time the number of successful search terms is almost 27 times as high, that is 123,175 compared to 4,631 in September 2010. To illustrate this, 2,530 of the search terms in September 2010 or 55% resulted in one visit whereas in January 2012, 84,773 of the search terms or 69% resulted in one visit. This can be attributed to the indexation of all dictionary entry pages which increases the chance of a match with a much larger variety of search terms. A similar observation can be made in the distribution of the types of search terms: before it was primarily meta search terms like *dictionary* that were successful, but now users are much more likely to get a match with linguistic phrase search terms because the headword is part of the title of the dictionary entry pages. All in all, the indexation of the dictionary pages has been the single most decisive factor in terms of matching the users' search behaviour: they search for linguistic information through a search engine and they often use actual dictionary words in their search terms.

The addition of inflectional forms has also proven successful. Out of the 1,000 most frequent unsuccessful search terms from September 2010, all inflectional forms except two can now be dealt with, either thanks to the added inflectional forms or through the 'did you mean' function. And when it comes to spelling and typing errors, the search algorithm is often able to suggest the intended word. Furthermore, we refer to ODS (a large historical dictionary with 225,000 entries) if there is no match in DDO. This means that out of the 1,000 most frequent unsuccessful searches from September 2010, we can now handle 45%, 10% being referrals to ODS. Assuming that the result for the frequent searches can be transferred to all unsuccessful searches, we are able to deal with 45% or almost half of the 20.6% unsuccessful searches, meaning that only about 10% of all searches cannot be dealt with – and this we find to be a very satisfactory result.

8. User survey

In order to substantiate the previous study and to understand our users better, we decided to conduct an online user survey. The survey was posted on 20 October 2011 and was carried out until 12 November when we reached 1,082 completed questionnaires. It was a voluntary survey; users were asked to participate by clicking a banner link on the entry page ordnet.dk/ddo or on any search results page above the site search box.

The survey contains questions about the user (age, gender, education, knowledge of site), use (frequency, situation, information type and success rate), evaluation (quality, usability, intelligibility) and finally an option to give praise and complain in plain text. The survey was designed and carried out by an intern from the University of Copenhagen using SurveyXact developed by Rambøll Management Consulting (www.surveyxact.com).

The online survey was supplemented by a qualitative part in which oral protocols were made for ten informants. They were asked to perform six tasks while 'thinking aloud', thus allowing the conductor of the experiment to log what they were doing and saying. This part was followed by an interview where the informant and the interviewer discussed the tasks and DDO in general.

9. Results of the online survey

The data showed an even distribution of males (541) and females (541) at an average age of 37.4 years across a wide age span, which implies that our target group is very broad. The respondents were distributed among the following age groups: 0-19 yrs. 17.5%, 20-39 yrs. 39.6%, 40-59 yrs. 30.1% and finally 60+ yrs. 12.8% When it comes to educational level, respondents aged 25 and over belong to the well-educated part of the population: 52% with a university-level education, whereas almost every respondent aged under 25 was still at school or college.

Being voluntary, the survey is more likely to attract respondents who are frequent users, more interested, informed or concerned about our site and thus have stronger or more extreme viewpoints. And accordingly, our data revealed that only 8% of the respondents were new visitors compared to 36% new visitors in our Google Analytics data. The remaining 92% returning visitors turned out to be frequent users indeed: 44.3% use DDO several times a week, 30.9% daily, 13.3% several times a month and 6.0% once a month or less, and finally 5.5% once a week.

The respondents use our dictionary at work or in school, but astonishingly many use it at home as well. This holds especially for our first-time users (see Table 2; 'use this visit' means 'for what purpose are you using the dictionary now').

Table 2. Respondents' use.

	Frequent users use in general	Frequent users use this visit	New users use this visit
At work	37.0%	35.3%	19.8%
At home/spare time	30.2%	34.8%	35.8%
In school/education	27.6%	24.8%	25.9%
Other	5.3%	5.1%	18.5%

What do users look up, and why? It is no surprise that a need for knowledge about meaning and use as well as about spelling is reported as a frequent motive for using the dictionary. Etymology is also a popular information type for both frequent and new users whereas the existence of a word, fixed expressions, and inflection seem to appeal more to frequent than to

new users. The latter category reports a surprisingly great interest in 'other things' which may stem from the fact that they are new to the site (see Table 3).

Table 3. Information types looked up.

	Frequent users use in general	Frequent users use this visit	New users use this visit
Meaning and use	84.9%	69.1%	41.5%
Spelling	41.9%	26.5%	28.0%
Idiom or fixed expression	29.1%	15.8%	6.1%
Whether a word exists	27.3%	17.2%	9.8%
Inflection	26.1%	17.0%	7.3%
Etymology	24.9%	15.2%	14.6%
Part of speech	9.8%	6.5%	3.7%
Pronunciation	9.4%	7.3%	6.1%
Other	4.4%	8.2%	24.4%
Language tips	2.8%	1.8%	0.0%

The category 'other' as well as the final option to give praise or complain revealed valuable information regarding information types in demand. Many respondents were looking for synonyms and would like to find more synonyms, idioms and fixed expressions, usage examples, etymologies and a better coverage of LSP vocabulary and foreign words. As far as synonyms are concerned, we are currently working on a conceptual dictionary project (Lorentzen and Trap-Jensen 2011), the material of which will add a new dimension to our online dictionaries by making onomasiological searches possible.

10. Results of the qualitative survey

Whereas the online survey yields information about what the users report about their behaviour – an approach often regarded as methodologically uncertain (cf. Hatherall 1984, Tarp 2006) – the protocol and interview approach is more likely to shed light on actual user behaviour and shortcomings of the dictionary web site. Certain parts of the oral survey confirm results from the online survey, for instance the way users tend to find information on the web. When asked to find the DDO website, half of the informants started by using Google and another 30% used Google when the URL they typed did not lead to DDO. Mostly, the oral survey gives valuable information about the design and structure of the dictionary page. The informants were primarily faced with difficulties when asked to select the correct homograph (out of three) and to locate a fixed expression. Both results have encouraged us to facilitate the lookup process, partly by giving the list of fixed expressions a more prominent position in the search result, partly by adding a gloss to each homograph in order to make it easier for the user to select the right word.

11. Concluding remarks

We are thus able to draw several positive conclusions from our preliminary experiences with user studies and search engine optimization: Being a freely available non-profit dictionary site, we have experienced a remarkable success in allowing web crawlers to index all our pages so we can only encourage others to do so, provided you have the hardware capacity. The number of visitors to *ordnet.dk* has more than quintupled since we started this process in September 2010, and the number is still growing as of January 2012.

In this paper we have tried to combine the results of statistical analysis with a user survey

based on a questionnaire, and this combination seems to prove useful in the future development of our dictionary site. In order to improve our web site architecture it will be relevant to focus on a quantative study of the way users navigate through dictionary pages and entries for instance by tracking clicks on internal links; and when it comes to user interface design, another relevant path to explore is eye-tracking studies, cf. Tono (2011: 125): 'eye tracking can yield much more detailed moment-by-moment observations about how users interact with dictionary information'.

References

A. Dictionaries
Hjorth, E., Kristensen, K. et al. 2003–. *Den Danske Ordbog*. Copenhagen: Det Danske Sprog- og Litteraturselskab. (DDO). Accessible online at: http://ordnet.dk/ddo.
Dahlerup, V. et al. 1918–2005. *Ordbog over det danske Sprog* incl. supplement. Copenhagen: Det Danske Sprog- og Litteraturselskab. (ODS). Accessible online at: http://ordnet.dk/ods.

B. Other literature
Almind, R. 2008. 'Søgemønstre i logfiler.' *LexicoNordica* 15: 33–55.
Bergenholtz, H. and M. Johnsen 2007. 'Log Files Can and Should Be Prepared for a Functionalistic Approach.' *Lexikos* 17: 1–21.
Clifton, B. 2010. *Advanced Web Metrics with Google Analytics*. (2nd edition). Indianapolis: Wiley Publishing, Inc.
De Schryver, G.-M. and D. Joffe 2004. 'On How Electronic Dictionaries are Really Used.' In G. Williams and S. Vessier (eds.), *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6–10, 2004*. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud, 187–196.
Hatherall, G. 1984. 'Studying dictionary use: Some findings and proposals.' In R. R. K. Hartmann (ed.), *LEXeter '83 Proceedings: Papers from International Conference on Lexicography at Exeter, 9-12 Sept. 1983, Lexicographica Series Maior 1*, 183–189. Tübingen: Niemeyer.
Hult, A.-K. 2008. 'Användarna bakom loggfilerna – redovisning av en webbenkät i Lexin online Svenska ord.' *LexicoNordica* 15: 73–91.
Lorentzen, H. and L. Trap-Jensen 2011. 'There And Back Again – from Dictionary to Wordnet to Thesaurus and Vice Versa: How to Use and Reuse Dictionary Data in a Conceptual Dictionary.' In I. Kosem and K. Kosem (eds.), *Electronic lexicography in the 21st century: New applications for new users. Proceedings of eLex 2011*, 175–179.
Nygaard, L. and R. V. Fjeld 2008. 'Analyse av søkelogger for bedre søkmuligheter i elektroniske ordbøker.' *LexicoNordica* 15: 57–72.
Tarp, S. 2006. 'Kan brugerundersøgelser overhovedet afdække brugernes leksikografiske behov?'. *LexicoNordica* 15: 5–32.
Tono, Y. 2011. 'Application of Eye-tracking in EFL Learners' Dictionary Look-up Process Research.' *International Journal of Lexicography* 24.1: 124–153. doi: 10.1093/ijl/ecq043