

One, Two, Many: Customization and User Profiles in Internet Dictionaries

Lars Trap-Jensen

Society for Danish Language and Literature, Copenhagen

Recent textbooks in lexicography recommend the use of customization in e-dictionaries whereby users or dictionary-makers specify which information categories should be shown on the screen. In this paper I take a look at some online dictionaries and analyze how they solve the task. A few basic types are recognized, based on the answer to questions such as: are the user profiles specified by the user or by the lexicographer? Is the profile defined in relation to the look-up situation or to the user's general background and skills? Is the profile fixed or flexible? Must the profile be specified once and for all, before every look-up situation or can it be changed as the user navigates through the dictionary entry? For practical reasons, I confine myself primarily to English and Scandinavian dictionaries. The analysis formed part of the preparatory phase of the online version of The Danish Dictionary. Four months after the introduction we can now observe from the log files how users manage the various options they are given. The experience so far is that user profiles that require deliberate action from the user are rarely used. The same holds for other kinds of customization such as advanced search possibilities. For the dictionary-maker there is all the more reason to be careful about configuring the default setting.

1. Introduction

Towards the turn of the millenium, a characteristic trend among dictionary publishers was to have a growing stock of still more refined print dictionaries, accompanied by the recognition that dictionaries should be tuned to match the needs of specific user groups. As electronic dictionaries are gradually taking over and the sale of paper dictionaries is declining, it is likely that the trend will take a different course: one in which the electronic dictionary or dictionary site may encompass a single or a whole range of traditional dictionaries that can be adjusted in various ways to comply with the needs of particular user groups. A possible outcome is that the future will see fewer lexicographical products but those that are there will allow individual customization in the way they are presented on the screen.

In this paper, I explore the notion of user profiling and customization. A number of Internet dictionaries are equipped with tabs, buttons or other options that allow the user to specify different presentational modes. I take a look at some of them, analyze what types of solutions have been employed and discuss the problems and perspectives involved.

2. The past and the present

In the good old days – which in this context is no more than 15 years ago – a dictionary to most people was a book, i.e. a concrete physical object consisting of sheets of paper with text on them and bound in one or more volumes. The electronic dictionary (hereafter e-dictionary) is much less tangible and can perhaps best be characterized by its two-sided nature: it consists of a content component with data that is typically stored in a database, and a presentational component in the form of an interface through which the user can see a representation of the data on a screen, e.g. on a desktop computer, a pocket electronic dictionary, a smart phone, tablet computer or other mobile device. The two components are not entirely independent: the way data is organized in the database determines what is possible to query and present on the screen. But contrary to the printed dictionary, the e-dictionary screen allows the same data to be presented in many different ways, depending on the functional and aesthetic preferences of the developer as well as on the types of query permitted.

The notion of dictionary customization is still in its infancy. It has been recognized by some of the more recent textbooks in lexicography, but typically only by way of recommendation or tentative suggestion. Here are a few examples:

Build multiple user profiles, and let users customize their e-dictionary. Users have their own specific needs and skills (and these may change according to the task they are engaged in), so it is important to allow them to decide which information-categories should be displayed by default (and which can be accessed by an additional click)
Atkins & Rundell 2008: 245

We could introduce user profiles that would allow users to define the type of information categories they want to be displayed in the entries. These user profiles could be either user-specified or defined by the system on the basis of the information the users give about the task they want to perform with the help of the dictionary.
Varantola 2003:237

Today, we can aspire to move a step further by investigating what solutions have actually been implemented in e-dictionaries. I confine myself to online e-dictionaries, and for practical reasons I shall look primarily at English and Scandinavian dictionaries.

3. Types of customization

A user profile is nothing more than a convenient label for a particular subset of dictionary data that is presented to a user in a look-up situation. Viewed in this perspective it is important to note that 1) the number of information elements in the database may be – and often has to be – much larger than the user will ever see at a time. The selected user profile decides which elements will be concealed and which will be visible. And that 2) each information element should contain one and only one type of information. Even though the latter point is old news, it is nevertheless indispensable as both conditions should hold if the user profiles are to be functional. So, user profiling takes as a starting point the existing information structure in the database and presents a subset of this data to the user.

If we now turn to look at different dictionary solutions, we can list a few basic types. The types emerge from the answers to the following questions which are in turn closely connected to the strategy adopted:

- are the user profiles specified by the user or by the lexicographer?
- is the profile defined in relation to the look-up situation or to the user's general background and skills?
- is the profile fixed or flexible?
- must the profile be specified once and for all, before every look-up situation or can it be changed as the user navigates through the dictionary entry?

3.1. The flexible user-specified profile

A user-specified profile is one that is left to the user to configure: he or she selects which elements should be visible, either freely or from a selection chosen by the dictionary-maker. I am not aware of any e-dictionaries that allow completely free configuration but the Oxford English Dictionary Online allows the user to configure a number of information categories: pronunciation, spelling, etymology, quotations, date chart and, if available, additions. Figure 1 shows how this is realized as buttons at the top of the screen.

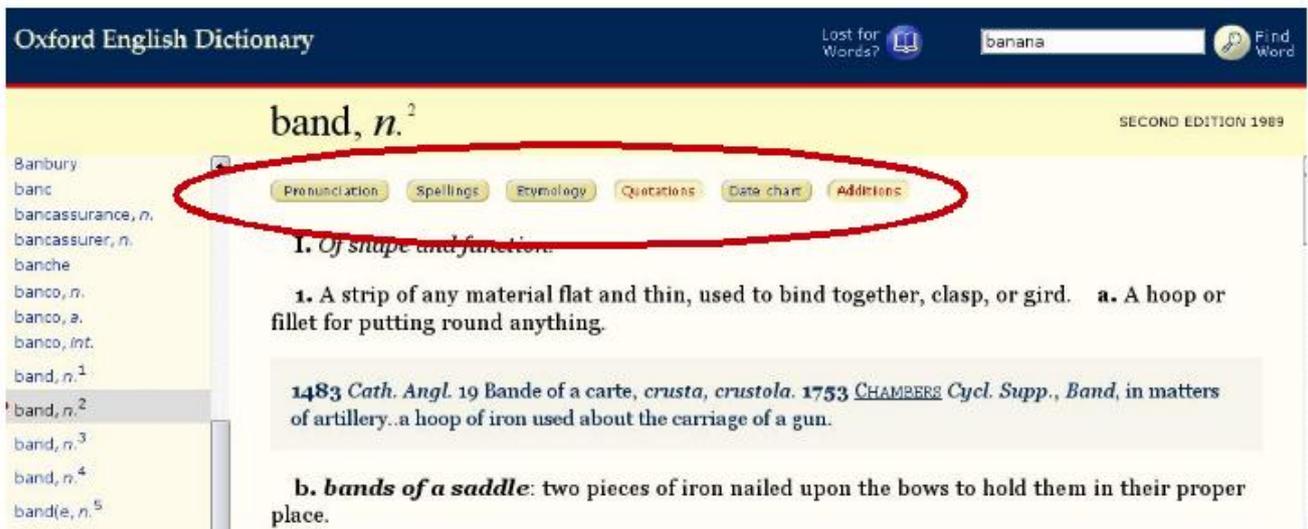


Figure 1. OED Online has buttons at the top of the screen that allow the user to show or hide certain information types

Red colour indicates that the element is visible, blue that it is suppressed and the selected setting is preserved in the session until the user changes it. In the default setting, quotations and additions are visible whereas the other categories are hidden. Customization may take place at any time, before or after the look-up, but the default setting is restored when a new session is opened.

3.2. The fixed user-specified profile

It would seem that OED's solution is a flexible one as the user is free to pick and choose information elements as he or she wishes. There are no pre-fabricated settings except the default, unlike the next type where the user can choose between a limited number of presentations, usually two or three. This is the solution adopted by Macmillan, where the user can choose between two settings, indicated by the buttons 'Show more' and 'Show less', as shown in Figure 2. As there are only two presentational modes, the button always shows the alternative mode.

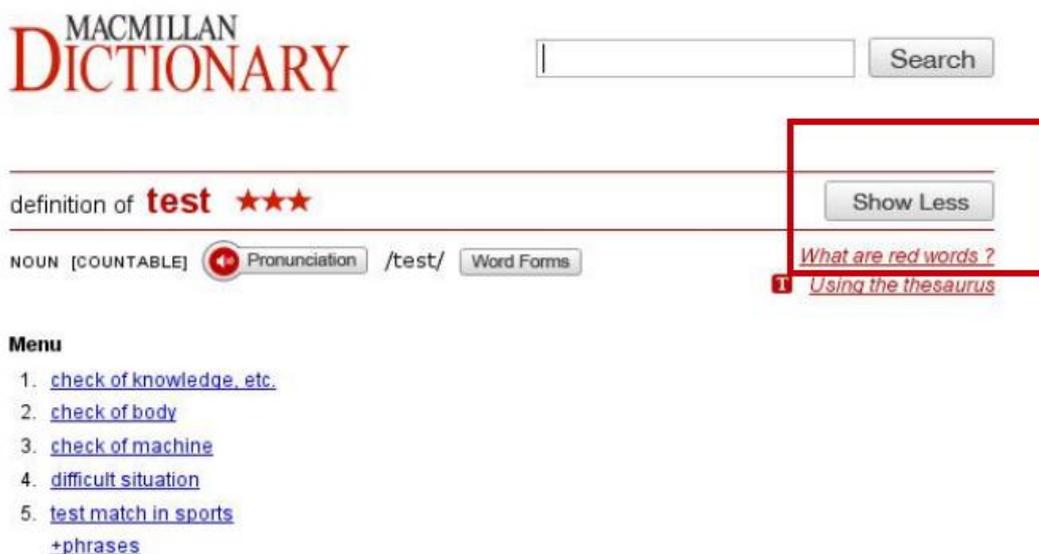


Figure 2. The button 'Show Less' allows users of Macmillan to hide parts of the entry

A similar solution was adopted in the project that I myself work with, The Danish Dictionary online. This dictionary also has two settings, both visible to the user in the form of tabs. In addition, The Danish Dictionary incorporates elements of the first type by allowing the user to open or hide the major information components of an entry at any time (meanings, phraseology and word formation). As opposed to the OED, however, the user changes the setting at the relevant point in the entry and not at the top.

It is also directed at different information elements, mainly because the option was introduced to promote ease of access and navigation rather than by considerations of user profiles. The possibility in The Danish Dictionary to see more than the maximum of two citations given in the default setting is closer to the OED solution, but again it should be set at the relevant point in the entry, not as a general setting for the whole entry.

The screenshot shows the interface of 'DEN DANSKE ORDBOG'. At the top, there are navigation links: 'English', 'Sideoversigt', 'Tilgængelighed', and 'Kontakt'. Below this is a header with 'DEN DANSKE ORDBOG' and 'MODERNE DANSK SPROG'. A breadcrumb trail reads 'du er her: forsider -> den danske ordbog -> ordbog'. The main content area features a search bar with 'frisisk' entered and a 'Søg' button. Below the search bar, there are two tabs: 'Kort visning' (selected) and 'Lang visning'. The entry for 'frisisk' is shown as an adjective with its grammatical information: 'BØJNING -r, -e' and 'UDTALE [ˈfriːsɪs]'. The 'Betydninger' section lists the primary meaning: '1. som vedrører friserne, deres kultur og de områder de bebor'. A sub-entry '1.a' provides a detailed definition: 'SOM SUBSTANTIV I NTETKØN vestgermansk sprog der tales i den hollandske provins Friesland, på visse af De Frisiske Øer og spredte steder i Nordtyskland omkring byerne Leer og Niebüll'. To the right, a 'Søgeresultat' section shows an alphabetical list of related terms, including 'frisé eller frise sb.', 'frise eller frisé sb.', 'frisée + frisé eller frise sb.', 'friser sb.', 'frisere vb.', 'frisering sb.', 'friserslæg sb.', 'frisésalat + frisé eller frise sb.', 'frisind sb.', 'frisindet adj.', 'frisisk adj.', 'frisk adj.', 'friskade sb.', and 'friskbagt adj.'. The left sidebar contains navigation options like 'Den Danske Ordbog', 'Søg i:', 'Om Den Danske Ordbog', 'Se også', and 'Relaterede søgninger'.

Figure 3. The Danish Dictionary has two tabs: 'Kort visning' ('short presentation') and 'Lang visning' ('long presentation')

3.3. The fixed profile defined by the situational needs

Finally, there are a few e-dictionaries that try to take into account the situational needs that users have, primarily inspired by the theoretical considerations developed by Bergenholtz and Tarp which they label 'functional'. In a Danish dictionary of phraseology (Bergenholtz et al.), the users can specify their current situation by clicking one of four radio buttons: 1) I am reading, but have a reception problem, 2) I am writing and want to use a particular expression, 3) I am writing and am looking for an expression with a particular meaning, and 4) I want to know everything about an expression. An example is shown in Figure 4.

Section 7. Dictionary Use

The screenshot shows the website 'Ordbøgerne over Faste Vendinger' from Aarhus School of Business. The search interface includes a dropdown menu for search criteria with 'indeholder' selected. A search result for 'Som man råber i skoven, får man svar' is displayed, showing its meaning and usage. A red box highlights the search settings options:

- Jeg søger information i ordbogen i forbindelse med, at jeg
 - læser en tekst, men forstår ikke betydningen af en fast vending
 - skriver en tekst med en bestemt fast vending
 - skriver en tekst og søger efter en fast vending med en bestemt betydning
 - ønsker at få mest muligt at vide om en eller flere faste vendinger

Figure 4. A Danish phraseological dictionary with four different settings which must be chosen before look-up. The default setting is 1) ‘reception problem’

The screenshot shows the front page of the 'Lexical Database for French (Base lexicale du français - BLF)'. The page is organized into several sections:

- Get information on:** Options for searching by word or word combination/expression.
- Get the translation of:** Options for translating words or word combinations.
- Verify:** Tools to check the use of words and the correctness of translations.
- Learn:** Resources for learning to express ideas, combine words, and avoid common errors.
- Do (a lot of) exercises:** Exercises on all aspects of the vocabulary.
- Interface:** Language selection (English, French, Dutch) and a feedback form.
- Accented letters and special characters:** A list of special characters for input.

Figure 5. The front page of the learner’s dictionary Base lexicale du français

A somewhat similar approach is taken by the Belgian project Base lexical du français, BLF (Verlinde, forthcoming), only here the users must specify their needs already on the front page. The BLF is a learner's dictionary and tries to make the user reflect on his or her need before accessing the database. There is a choice between six major categories with subcategories as can be seen from Figure 5's view of the front page. The rationale behind the BLF and the phraseological dictionary is in both cases to present the user with only the information needed for the particular task at hand and nothing else.

3.4. The two-in-one profile tailored to situational needs

Because most e-dictionaries available today were conceived as print dictionaries, the data structure is often less than optimal, with the consequence that the profiling potential is not fully exploited. Even if incorporated from the outset, the creation of an optimal data structure can be a costly affair: it takes more time and effort to present the same content in different versions no matter if it is done by creating separate versions or by marking up stepwise condensations of the full text. Nevertheless, it is an obvious and highly recommendable way of differentiating between user needs, and one which is also implemented in some reference works. To my knowledge, it is more widely used in encyclopedias and specialized dictionaries, probably because they often have very long and detailed explanations. The long explanation is useful for users that seek in-depth knowledge of a particular subject but they can be too long if the consultation is motivated by a simple reception problem. In the latter case the user is better helped with a short definition. Figure 6 shows an example from the Swedish National Encyclopedia where the user can choose between up to three different styles: 'lång' (long), 'kort' (short) and 'enkelt' (simple). Another example is the Danish Dictionary of Music which offers a long and a short explanation of the head-word.



Figure 6. The Swedish National Encyclopedia offers up to three different presentations of the same article

4. Discussion

There seem to be different theoretical analyses behind the user-specified and the need-dependent solutions. A recurring position in the functional approach of the Aarhus school is that monofunctional dictionaries are preferable to multifunctional (Bergenholtz & Vrang 2006). An argument against this view is the rather mundane observation that users simply do not buy several dictionaries for one language and use them alternately according to their

situational needs. But this argument is only valid for print dictionaries: with the possibilities offered by the e-dictionary it is now possible to get several monofunctional dictionaries in one. And this is just what has been implemented in the above-mentioned phraseological dictionary: the users can get just the data that match their needs by clicking on the relevant button. The approach is attractive as it aims at eliminating information that is not relevant in the look-up situation, thereby keeping things simple and clear. To be successful, however, it is required that the users are able to analyze their own needs in every look-up situation and pick the right button. According to my knowledge, there is not much evidence to support such a rational user behaviour (cf. Nesi, forthcoming) and our own experience is that the prospects are rather gloomy.

This is the reason why, in the planning of The Danish Dictionary, we decided on a simple model with only few possibilities. The flexible user-specified model (as exemplified by OED above) is attractive but was, at least temporarily, eliminated at the planning stage for the practical reason that categories like inflection, pronunciation and etymology usually do not take up much space on the screen and there would not be much to be gained by suppressing this kind of information individually. Instead, a simple solution with only two tabs was chosen: a short version with the basic information needed for reception (suppressing the following: pronunciation, etymology, grammatical information, related words and all examples except one citation for each sense), and a long version which included all available information.

Before launching the online version of The Danish Dictionary we carried out a small series of user tests in order to decide finally how many and how sophisticated user profiles we should allow. The test persons were simply asked to perform various look-up tasks and the results of their searches were registered as well as any comments they might have in connection with the task. The test results showed that the users had problems in finding rather basic information. For example, they could not find the pronunciation of a word because they did not realize they were in the short presentation mode, and they were generally unable to analyze their needs ('I don't care if it is reception or production, I just want to know what the word means'). This made us change the default setting from short to long presentation: it may be less clear and simple but at least everything is to be found there.

Although the functional approach, as expressed by Tarp (e.g. Tarp 2008) and Bergenholtz, has a point by directing attention to users' situationally defined needs, the theoretical analysis will not do on its own if the users are unable to understand and orient themselves among the choices offered. The Belgian BLF project seeks a different solution to the same underlying challenge: here the users have to choose between situations before they are allowed to perform a look-up. This approach looks promising but it also draws attention to a potential catch-22 situation: on the one hand, requiring too many options and clicks of users before they can get started may scare them away. And on the other hand, a model with immediate look-up and only few options may lead to inaccurate access and lack of clarity. Whatever the situation, we need more information about user behaviour to assess which solution works more effectively.

5. User behaviour as attested by log files

The online version of The Danish Dictionary was opened to the public in November 2009, and after the opening we have been able to log the actual use of the two presentational modes. Table 1 shows the distribution of the modes during a 2-week period in February 2010.

Total pageviews	Pageviews, short	Pageviews, long
294,284	2,519	291,765

Table 1. Pageviews with different presentation modes over a 14 day period

Not surprisingly, our suspicion from the user tests is confirmed: users very rarely make use of the possibility to view the short version of the entry, or rather: the non-default setting. Of the total pageviews more than 99 % are in the long mode and less than one percent (0.86 % to be exact) in the short. Although the overall picture is quite clear, it does of course not tell the whole story: As it is not possible to choose between the two modes on the front page and as many visit the dictionary only once, all pageviews resulting from a single front page search will be in the default – long – mode. And technically, as the change of tabs is carried out by using JavaScript, a change is only logged if the change in tab is followed by a search and a page reload. It is not possible for us to log, for example, if users shift tab position once or more when browsing a single entry. Nevertheless, it may also be interesting to view the changes between the two modes that are followed by page reload. This is shown in table 2.

long → short	short → long
2,576	1,447

Table 2. Changes between long and short presentation

The table shows that in slightly more than 4,000 pageviews the user has actively changed the presentation from the previous position. Out of the total number of pageviews this amounts to 1.37 % so the overall picture is not affected much. And although we have just seen that the log statistics underestimate the number of tab changes, as only changes followed by a page reload are registered, it should be safe to conclude that in the overwhelming majority of pageviews the default presentation is used.

The statistics alone do not, however, tell us *why* users do what they do. It may be that they find the long version appropriate in the look-up situation and feel no need to change to the short version. Alternatively, one may suspect that many users are unaware of the options and just use what is available to them. And a third position would be to maintain that the truth is a mixture of both these interpretations.

The choice of user profiles may be related to the different search options offered by some e-dictionaries. As with user profiles, the latter requires that the user must actively choose a different option than the one given by default, usually a more advanced search mode than the standard search. Such an advanced search option is for instance offered by OED Online, see Figure 7.

Advanced search

Search for **Entries** **Quotations** containing

in **full text**

AND (Operation A)

in **full text**

AND (Operation B)

in **full text**

Case-sensitive searching

Exact character searching

Options for NEAR/NOT NEAR:

1 word

before

after

before or after

Order of operations: **A then B**

Restrict search to results of

previous search

search number

Part of speech filter

<input type="checkbox"/> noun	<input type="checkbox"/> preposition
<input type="checkbox"/> pronoun	<input type="checkbox"/> interjection
<input type="checkbox"/> adjective	<input type="checkbox"/> prefix
<input type="checkbox"/> article	<input type="checkbox"/> suffix
<input type="checkbox"/> verb	<input type="checkbox"/> combining form
<input type="checkbox"/> adverb	<input type="checkbox"/> participle
<input type="checkbox"/> conjunction	<input type="checkbox"/> phrase

Figure 7. Advanced search option in OED Online

I have been informed (Penny Silver, personal communication) that the use of the advanced search option as the non-default choice is very similar to that of the non-default user profile found for The Danish Dictionary: it is scarcely used, less than 5 % of the searches are performed using the advanced search.

Teksteksempler: Standardsøgning

Find eksempler på brugen af et ord eller en vending. [Læs mere om søgning.](#)

Standardsøgning
Udvidet søgning
Formel søgning

Indtast søgeord:

Alle bøjningsformer Kun indtastede former

Figure 8. KorpusDK offers three different query modes

The same holds for a different part of our website. The dictionary is connected to a corpus of contemporary Danish, KorpusDK, at the same site, and here the user can query the corpus in three different ways, as shown in Figure 8: ‘Standardsøgning’ (standard query) finds a word and its inflected forms. If more words are entered, the query is for both words and their inflections with up to three intervening words. The ‘Udvidet søgning’ (extended query) allows the user to specify the part of speech, the precise inflected forms to look for and the number of intervening words allowed. Finally, the ‘Formel søgning’ (formal query) allows the advanced

user to adopt the full range of formalized queries offered by the underlying corpus query language used. The standard query is used by default. Table 3 summarizes how the three query modes were used in the same 2-week period that were used for tables 1 and 2.

Standard	Extended	Formal
5,427	92	208

Table 3. Use of query options in KorpusDK

Again, the same picture emerges: the default standard query was used 94.8 % of the times, with extended query accounting for 1.6 % and formal query for 3.6 % of the total.

6. Conclusions

Whether they adhere to one school of thought or another, most lexicographers welcome the possibility of showing exactly the relevant information categories in a particular lookup situation, no less and no more, tailored to the specific needs and skills of the user. For the lexicographer, this is a strong argument in favour of the e-dictionary over the printed dictionary: the electronic medium has solved some of the problems related to traditional dictionaries. For the same lexicographers, it may be disappointing that the users do not seem to take advantage of all these wonderful possibilities.

The lesson to learn is probably that both lexicographers and dictionary users must make an effort. Dictionary-makers cannot use the introduction of user profiles as a pretext for leaning back and do nothing but should be concerned with finding ways to improve presentation. On the other hand, people are not born with the skills to extract the wealth of data stored in dictionaries and other reference works efficiently and transform it into knowledge. It takes time to get accustomed to new ways of finding information, it may even require formal training.

Literature

Websites

Base lexicale du français, BLF: <http://ilt.kuleuven.be/blf>

Den Danske Ordbog (The Danish Dictionary): <http://ordnet.dk/ddo>

KorpusDK: <http://ordnet.dk/korpusdk>

Macmillan Dictionary: <http://www.macmillandictionary.com>

Musikordbogen (Dictionary of Music): <http://www.musikordbogen.dk/musik>

Ordbøgerne over Faste Vendinger (Dictionaries of Fixed Phrases): <http://www.idiomordbogen.dk>

Oxford English Dictionary: <http://dictionary.oed.com>

References

Atkins, B.T. Sue; Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford and New York: Oxford University Press.

Bergenholtz, H.; Vrang, V. (2006). 'Den Danske Ordbog: en ordbog for lingvister!'. In *LexicoNordica* 13. Oslo: Nordic Association for Lexicography. 185-196.

Nesi, H. (forthcoming). 'E-dictionaries and Language Learning: Uncovering Dark Practices'. In Granger, S.; Pacquot, M. (eds.). *Proceedings from eLEX 2009*. Louvain-la-Neuve (Cahiers du CENTAL series).

Tarp, S. (2008). *Lexicography in the Borderland between Knowledge and Non-Knowledge*. Tübingen: Niemeyer.

Varantola, K. (2003). 'Linguistic corpora (databases) and the compilation of dictionaries'. In: Sterkenburg, P. van (ed.). *A Practical Guide to Lexicography*. Amsterdam/Philadelphia: John Benjamins. Publishing Company. 228-239.

Verlinde, S. (forthcoming). 'From Paper-Based Electronic Dictionary to Leximats'. In Granger, S.; Pacquot, M. (eds.). *Proceedings from eLEX 2009*. Louvain-la-Neuve (Cahiers du CENTAL series).