

## A model for integrated dictionaries of fixed expressions

Henning Bergenholtz<sup>1,2,3</sup>, Theo Bothma<sup>2</sup> and Rufus Gouws<sup>3</sup>

<sup>1</sup> Aarhus University; <sup>2</sup> University of Pretoria; <sup>3</sup> Stellenbosch University

<sup>1</sup> Center for Lexicography

<sup>2</sup> Department of Information Science

<sup>3</sup> Department of Afrikaans and Dutch

Fuglesangs Allé 4

Private Bag X20

Private Bag X1

8210, Aarhus V

Hatfield, 0028

Matieland, 7602

Denmark

South Africa

South Africa

E-mail: hb@asb.dk, tbothma@up.ac.za, rhg@sun.ac.za

### Abstract

This paper discusses a project for the creation of a theoretical model for integrated e-dictionaries, illustrated by means of an e-information tool for the presentation and treatment of fixed expressions using Afrikaans as example language. To achieve this a database of fixed expressions is compiled wherein data are treated in such a way that access can be provided through a variety of dictionaries for specific situations, based on specific lexicographic functions, e.g. the cognitive function as well as the communicative functions of text reception and text production. From one database, the user will have access to six monofunctional dictionaries of fixed expressions. Each one of these dictionaries provides a view on selected fields of the database, i.e. a search is carried out on selected fields in the database and only the data in specific fields that are relevant for the specific dictionary are displayed. There are unique user needs that may not necessarily be satisfied by means of these six dictionaries. Individualised search facilities will therefore be provided to enable a user to retrieve data from a single data field or a user-specified selection of data fields. Phase two will provide the option of setting up a user profile, an extension of data fields and linking to external data sources. The result of the project will therefore be a comprehensive database of Afrikaans fixed expressions that may be accessed through six monofunctional dictionaries, as well as individualised search options, user profiling and the possibility to display additional data on demand.

**Keywords:** Fixed expressions; databases; integrated dictionaries; monofunctional and polyfunctional dictionaries

### 1. A purportedly user-friendly e-idiom dictionary

For many centuries lexicographers have proudly claimed that specifically their own dictionary was user-friendly and satisfied the needs of all users as well. This was, and still is, an immunising and self-serving assertion in most cases. It is based on real, factual research only to a limited extent, and at the same time it is an advertising measure to persuade potential dictionary buyers. However, one thing has changed. Up to some 30 years ago, dictionary user research was *de facto* nonexistent. This was formulated quite succinctly by Wiegand, who referred to the "known unknown" that needed to be researched (Wiegand, 1977: 59). Apparently, much research has indeed been done. Numerous surveys of all kinds have been conducted, but much of this assumed the form of memory-based questions such as: "How often do you use a dictionary? Daily? Weekly? Monthly? Rarely?"; "What kind of information do you look for? Grammatical? Orthographic? Semantic?"; "What kind of entries do you look for? Collocations? Examples? Items about style?" These days it is hardly possible any more to read and understand all contributions made to and by user studies. In our view this is not worthwhile either, as such surveys mostly ask questions which are constructed instead of being rooted in reality. The research should be conducted on real users with their real and specific needs and on their use of dictionaries, but in most cases it is not. A user with a cognition-related information need may be looking explicitly for certain types of data (examples, for instance). That is not what users with a need for communication-related information do. They have a problem

concerning reception, text production or translation and are hoping to get the necessary help in this regard. Ordinary users do not know exactly what lexicographers (and linguists in particular) call such items. We therefore hold the view that lexicographers should not act as dictionary philologists or interpreters of what users remember about their use of dictionaries, but should especially develop new concepts on the basis of theoretical considerations concerning the needs of certain types of users in foreseen situations. It is not the actual user that matters, but the potential user and his/her potential needs in situations anticipated by the lexicographer. For these needs the lexicographer develops a (new) tool of which he/she assumes that it can satisfy the needs he/she foresees. Within the function theory such dictionaries are typically monofunctional, i.e., they address a specific need of a specific user group in a specific situation (see, for example, Bergenholtz, 2010; Bergenholtz, 2011; Bergenholtz & Bergenholtz, 2011; Tarp, 2007; Tarp, 2008; Tarp, 2009a; Tarp, 2009b; Tarp, 2011). Yet a majority of practical and theoretical lexicographers assume that all dictionaries should always provide as much as possible data for the identified user groups, and therefore always were, and should continue to be, polyfunctional (see Bergenholtz 2010).

The above introduction reflects the arguments which have been frequently put forward in lexicographic discussions in recent years. Tarp (2002) proposed the following basic division into two types of lexicography: In **contemplative lexicography**, existing dictionaries are analysed and users are questioned about their use of existing dictionaries to date. In **transformative**

**lexicography**, theoretical analyses of the potential user situations, the respective user conditions and the user needs are used to develop new concepts for compiling new dictionaries, typically monofunctional dictionaries. On the basis of theoretical analyses the lexicographer therefore decides what the characteristics are of the monofunctional dictionaries that will satisfy specific user needs. In the case of the Centlex dictionaries developed at the University of Aarhus, no general surveys on the use of these dictionaries are undertaken, but feedback in the form of e-mails is analysed and taken into consideration. Moreover, log file analyses are done which, in selected cases, are linked to enquiries among a handful of users (see Bergenholtz & Johnsen 2005; Bergenholtz & Johnsen 2007).

Such log file analyses and feedback can lead to small changes, but also to a complete redesign of the dictionary, as was the case with the e-idiom dictionary by Vrang, Bergenholtz & Lund (2003-2005) (*Den danske Idiomordbog*). This was a dictionary of idioms containing the relatively large number of 8500 idioms. It had been designed especially as a reception dictionary, as it contained only meaning items. In the user guide and in the outer text on the structure of the dictionary, the meaning of 'idiom' was explained clearly. The authors received a fair number of e-mails from users with feedback on this dictionary. None of these mails asked what actually constitutes an idiom. They were, however, quite frequently asked why this or that combination of words could not be found in *Den danske Idiomordbog*. The typical answer to this question was that the expression in question is a proverb, not an idiom, and is therefore not in this dictionary. This happened regardless of the fact that the terms were clearly defined in the user guide.

During the period from mid-2003 until mid-2004 the number of unsuccessful dictionary searches was relatively high. (Misspelled search terms are included here, but amounted to fewer than 3% of searches; searches for unlemmatised idioms are also included, but these searches amounted to less than 1% of the searches.)

| Number of searches in <i>Den danske Idiomordbog</i> |       |
|-----------------------------------------------------|-------|
| With result                                         | 70.4% |
| Without result                                      | 29.6% |

Table 1: Percentage of successful and unsuccessful searches in *Den danske Idiomordbog*, 2003-2004

There are two sides to the bare figures for successful and unsuccessful uses of *Den danske Idiomordbog*. The positive side is that the users find the idiom they were looking for in more than 70% of all enquiries. The negative side is that in about 26% of all enquiries (cases with incorrect spelling and deficient lemmas have been deducted from the 29.6%) users were looking for 'idioms' which are not idioms but proverbs, sayings ('winged

words'), standard formulations, multiword expressions from other languages and many more. We do not believe that another definition of idioms would have improved this rate. On the other hand, for an internet dictionary of idioms there is an obvious solution: Don't make one at all; make one that contains all forms of fixed expressions. Moreover, the user with a reception problem does not even need to know what kind of fixed expression he/she is dealing with; he/she needs only the meaning. This insight led to a new concept for a new Danish database with fixed expressions from which several monofunctional dictionaries are offered to users (see Bergenholtz, 2011).

The preceding insight led to the decision to compile a database for fixed expressions in Afrikaans, rather than a database of idioms.

The concept for this database with several monofunctional dictionaries is the point of departure for the concept of the Afrikaans database presented here. The new database differs from the previous concept in some respects, however, especially as regards the number of fields for item types. The Danish database has 14 fields, the new database has 36. Also, this is a database for two languages, viz. Afrikaans and English, not for one. Nevertheless, the basic concept remains intact. A database and a dictionary are not the same thing. A single dictionary can be extracted from a database, and the result will normally be a polyfunctional dictionary. From a database, on the other hand, as many dictionaries could be extracted as are deemed relevant on the basis of theoretical considerations and experience with earlier databases, and these should be function-oriented monofunctional dictionaries (see Bergenholtz & Tarp 2002; Bergenholtz & Tarp 2003; Bergenholtz & Tarp 2005).

## 2. Afrikaans dictionaries with fixed expressions

Afrikaans dictionaries represent a wide-ranging typological variety, compiled to assist different users in finding assistance with regard to both language for general purposes and language for special purposes. Within the category of general dictionaries various monolingual and bilingual dictionaries offer an extensive presentation of fixed expressions. The category of restricted dictionaries also include a few dictionaries that focus on fixed expressions, cf. Malherbe (1924), De Villiers & Gouws (1988), Botha, Kroes & Winckler (1994), Prinsloo (1997) and Prinsloo (2009). The extent and nature of both the macrostructural coverage and the treatment in these dictionaries of fixed expressions differ considerably. They share one feature and that is that they have been produced in printed format. The only non-printed version of Afrikaans fixed expressions can be found in the presentation and treatment of this category of lexical items in those general monolingual and bilingual dictionaries that are available in CD ROM

format or online. Afrikaans has a real need for an e-dictionary of fixed expressions. The advantage of the fact that all dictionaries of fixed expressions are only available in printed format is that no bad e-dictionary exists, and a transformative approach to the planning and compilation of an e-dictionary of fixed expressions does not have to pay any attention to any electronic predecessor. In the following sections the plan for an innovative e-dictionary that deals with Afrikaans fixed expressions will be discussed.

### 3. Concept for a database system for Afrikaans fixed expressions

The database system consists of the database itself and the database management system. The database is developed in MySQL. It is integrated in a database management system that has been developed using open source software (HTML, XML, XSLT, Perl, CGI and related technologies). The database management system has a comprehensive administrative back-end which manages access, data security and integrity, including aspects such as version control and back-up. The system has two further interfaces, viz. an interface for the researchers contributing the data for the different fields in the database and an interface for end-users through which they obtain access to the dictionaries and other customization functionalities.

In principle, hundreds or even thousands of fields related to one or more phenomena could be provided in a database with one or more languages. For instance, there is a total of 84 fields in a Danish, English and Spanish accounting database, from which 23 different dictionaries are offered to users at present (see Bergenholtz 2011). In this case, only fields with different types of data – apart from the lexicographers' notes on the work in progress – which also finds its way into one or more dictionaries are provided for in the database. If one or more collaborating lexicographers have data at their disposal which are not intended to be presented in at least one dictionary, specific fields could be created for such data so that it could perhaps be accommodated in one or more additional dictionaries. The limit must be drawn where the number of fields becomes so large that the lexicographer loses sight of the big picture and the first presentation of the database takes too long.

The order of the fields in the "Field" column in Table 2 is a working order; the order in the individual dictionaries is determined for each respective dictionary.

| Field                                |
|--------------------------------------|
| 1. Core field                        |
| 2. Meaning in Afrikaans              |
| 3. Internet link to meaning          |
| 4. Further meaning item in Afrikaans |
| 5. Meaning in English                |
| 6. Grammar                           |
| 7. Comment on grammar                |

|                                                              |
|--------------------------------------------------------------|
| 8. Internet link to grammar                                  |
| 9. Background remark(s)                                      |
| 10. Comment on background remark(s)                          |
| 11. Internet link to background remark(s)                    |
| 12. Fixed expression(s) in Afrikaans                         |
| 13. Remarks on the fixed expression(s)                       |
| 14. References to fixed expression(s)                        |
| 15. Internet link to variants, e.g. statistical              |
| 16. Fixed expression(s) in English translated from Afrikaans |
| 17. Style                                                    |
| 18. Comment on style                                         |
| 19. Internet link to style                                   |
| 20. Classification of the fixed expression                   |
| 21. Comment on classification                                |
| 22. Collocation(s)                                           |
| 23. Comment on collocation(s)                                |
| 24. Internet link to collocation(s)                          |
| 25. Example(s)                                               |
| 26. Comment on example(s)                                    |
| 27. Internet link to example(s)                              |
| 28. Synonym(s)                                               |
| 29. Comment on synonym(s)                                    |
| 30. Internet link to synonym(s)                              |
| 31. Antonym(s)                                               |
| 32. Comment on antonym(s)                                    |
| 33. Internet link to antonym(s)                              |
| 34. Associated concept(s)                                    |
| 35. Key word(s)                                              |
| 36. Memo field                                               |

Table 2: Data fields for the database of fixed expressions in Afrikaans

There is not enough space here to justify all fields. However, some fields which are not self-explanatory do require some explanation. Field 12 contains only one expression in some cases, but in others there are more if variants exist, e.g. the same core of a fixed expression combined with different verbs. If one wants to, one can call this a lemma field. We do not; that would rather be field 1, called the core field, which is identical to Field 12 if there are no variants and contains only the words it has in common with the variants in Field 12 if there are variants. This field is used for automatic searches on the one hand, and on the other for items the user can use as links if search results are displayed as a list, or if synonyms or antonyms are provided. The field contains key words with all the lexical words, including irregularly conjugated forms which occur in the fixed expression(s) of a particular card. In Field 22, we use the term 'collocations' in the sense of combinations of words in which the fixed expression occurs. A collocation is never a complete sentence, unlike the data in Field 25, where 'example' refers to a full sentence. Field 9 contains a brief history behind the full expression; if there are two different histories and it is not clear which one is correct, both are given. In addition, in some cases reference is

made to background histories which are given in various textbooks and dictionaries (but are not necessarily correct). Lastly, field 34 contains associated concept(s). This refers to concepts which can be associated with the meaning and use of the fixed expression. Finding such concepts could be very time consuming if some sort of semantic system were applied. That is not how it is done here. In fact, the editor's lexicographic instruction is to write down up to five such associated concepts within 30 seconds (but never more).

#### 4. Six dictionaries with fixed expressions

At present the concept provides for six dictionaries – five monofunctional and one polyfunctional. It is a model that can be used not only for these languages and this language combination, but in principle also for at least all Indo-European languages and probably also for other language families, for example the Austronesian languages.

##### 4.1 MEANING OF FIXED EXPRESSIONS

Access to the first dictionary is gained by pressing the button "I am reading a text, but do not understand the meaning of a fixed expression". Here the user enters an expression or part of an expression in the search field and obtains the desired information, i.e. the meaning of the fixed expression. This dictionary is called MEANING OF FIXED EXPRESSIONS. When a search is done in this dictionary, the program looks in two of the fields in the database in the order indicated by figures<sup>1</sup> (see column 1 in Table 3 below). For this dictionary a maximising search is done. The user obtains one or several articles with the content of three of the fields in the database (see column 3). In other words, the user receives only a small part of the database as article, but it is exactly the part that is needed to solve a reception problem. If more than 10 articles are found they are displayed as a list where the content of the core field and the first line of the meaning are shown.

The following tables show only those fields which are used for a search and those fields from which data are presented in the dictionary article for the specific dictionary. Because of space limitations in the headings of the tables, "Search" is used as an indication of the fields that are searched and the numbers indicate the order in which the search is carried out. Similarly, "Entry" refers to the fields that are shown to the user and the numbers indicate the ordering sequence of the fields in the specific dictionary article. "List" is used as an indication of which data are displayed when a list is needed (i.e., when more than 10 articles are found).

<sup>1</sup> In a maximising search the order does not really matter, as all subresults for each individual search are added up in the overall result. In a minimising search this is different. In this case the search ends after searching one field if one or more results are found. Therefore the next fields are not searched.

| Search | Field                                | Entry | List                   |
|--------|--------------------------------------|-------|------------------------|
| 1      | 1. Core field                        |       | 1                      |
|        | 2. Meaning in Afrikaans              | 1     | 2 1 <sup>st</sup> line |
|        | 5. Further meaning item in Afrikaans | 2     |                        |
| 2      | 12. Fixed expression(s) in Afrikaans | 3     |                        |
| 3      | 35. Key word(s)                      |       |                        |

Table 3: Search and data fields in the dictionary  
MEANING OF FIXED EXPRESSIONS

##### 4.2 USE OF FIXED EXPRESSIONS

The second dictionary is activated by pressing the button "I am writing a text with a specific fixed expression". Here the user enters a fixed expression or part of it in the search field and obtains information about the use of the fixed expression, including its meaning, grammar, collocations, example sentences and synonymous or antonymous fixed expressions. In other words, the search is expression-specific. We call this dictionary USE OF FIXED EXPRESSIONS. When this dictionary is activated, four fields of the database are searched; however, this is a minimising search, where the search is terminated after one field type has been searched and other fields are therefore not searched. The items relevant to text production are reflected as figures in column 3; if there are more than 10 articles, a list is shown.

| Search | Field                                  | Entry | List                   |
|--------|----------------------------------------|-------|------------------------|
| 1      | 1. Core field                          |       | 1                      |
|        | 2. Meaning in Afrikaans                | 3     | 2 1 <sup>st</sup> line |
|        | 3. Internet link to meaning            | 5     |                        |
|        | 4. Further meaning item in Afrikaans   | 4     |                        |
|        | 6. Grammar                             | 8     |                        |
|        | 7. Comment on grammar                  | 9     |                        |
| 2      | 12. Fixed expression(s) in Afrikaans   | 1     |                        |
|        | 13. Remarks on the fixed expression(s) | 2     |                        |
|        | 17. Style                              | 6     |                        |
|        | 18. Comment on style                   | 7     |                        |
| 4      | 22. Collocation(s)                     | 10    |                        |
|        | 23. Comment on collocation             | 11    |                        |
| 5      | 25. Example(s)                         | 12    |                        |
|        | 26. Comment on examples                | 13    |                        |
|        | 28. Synonym(s)                         | 14    |                        |
|        | 29. Comment on synonyms                | 15    |                        |
|        | 31. Antonym(s)                         | 16    |                        |
|        | 32. Comment on antonyms                | 17    |                        |
| 3      | 35. Key word(s)                        |       |                        |

Table 4: Search and data fields in the dictionary USE OF  
FIXED EXPRESSIONS

### 4.3 FIXED EXPRESSIONS WITH A SPECIFIC MEANING

The third dictionary is activated by pressing the button "I am writing a text and am looking for a fixed expression with a specific meaning". Here the user can enter one or several words with a specific meaning and find expressions with this meaning or part of this meaning. The user then receives information about the use of the expression, including its meaning, grammar, collocations, example sentences and synonymous or antonymous fixed expressions. In other words, the point of departure is a meaning, which can be very wide and can therefore yield many hits. If a more restricted meaning is used as the search string, fewer hits may be found or even none at all. This dictionary is called FIXED EXPRESSIONS WITH A SPECIFIC MEANING. When a search is done in this dictionary, the program looks in three of the fields in the database, in the case of a maximising search. The data are presented as in the dictionary mentioned above (USE OF FIXED EXPRESSIONS), as the function is the same, i.e. assistance with text production problems.

| Search | Field                                  | Entry | List                   |
|--------|----------------------------------------|-------|------------------------|
|        | 1. Core field                          |       | 1                      |
| 1      | 2. Meaning in Afrikaans                | 3     | 2 1 <sup>st</sup> line |
|        | 3. Internet link to meaning            | 5     |                        |
| 2      | 4. Further meaning item in Afrikaans   | 4     |                        |
|        | 5. Grammar                             | 8     |                        |
|        | 6. Comment on grammar                  | 9     |                        |
|        | 12. Fixed expression(s) in Afrikaans   | 1     |                        |
|        | 13. Remarks on the fixed expression(s) | 2     |                        |
|        | 17. Style                              | 6     |                        |
|        | 18. Comment on style                   | 7     |                        |
|        | 22. Collocation(s)                     | 10    |                        |
|        | 23. Comment on collocation             | 11    |                        |
|        | 25. Example(s)                         | 12    |                        |
|        | 26. Comment on examples                | 13    |                        |
|        | 28. Synonym(s)                         | 14    |                        |
|        | 29. Comment on synonyms                | 15    |                        |
|        | 31. Antonym(s)                         | 16    |                        |
|        | 32. Comment on antonyms                | 17    |                        |
| 3      | 34. Associated concept(s)              |       |                        |

Table 5: Search and data fields in the dictionary FIXED EXPRESSIONS WITH A SPECIFIC MEANING

One can then click on the core expression to get to the dictionary article which gives a meaning that fits the context. An article will be displayed with a set of corresponding data, as was illustrated above in the USE OF FIXED EXPRESSIONS dictionary. Although the data presentation of the two dictionaries is identical, the

dictionaries are not. In the dictionary FIXED EXPRESSIONS WITH A SPECIFIC MEANING access is gained by means of a meaning-oriented search, as in a printed dictionary with a systematic macrostructure and with one or more registers, whereas the dictionary USE OF FIXED EXPRESSIONS corresponds to a dictionary with an alphabetic macrostructure without registers. But the information the user is looking for to assist him/her with the production of a text is the same for both dictionaries.

### 4.4 KNOWLEDGE ABOUT FIXED EXPRESSIONS

For the Danish dictionaries with fixed expressions mentioned above there are four dictionaries, the first three like those presented here and a fourth which shows all fields in the database. Here we found that the two text production dictionaries accounted for only about 9% of all user actions during the second period in Table 6. A comparison with the log file analysis from the earlier period (2007), when there was only one production dictionary, shows that this share is relatively stable. Compared with the polyfunctional dictionary, which shows everything, the reception dictionary showed a substantial shift in the user actions between the two periods, which are presented in Table 6 as absolute figures and as percentages.

| 27 February 2007 until 17 December 2007 |         |        |
|-----------------------------------------|---------|--------|
| Understanding a text                    | 51 242  | 60.33% |
| Writing a text                          | 5 294   | 6.23%  |
| All data                                | 28 405  | 33.44% |
| 17 December 2007 until 1 December 2008  |         |        |
| Understanding a text                    | 154 239 | 29.57% |
| Writing a text with a known expression  | 19 386  | 3.72%  |
| Writing a text with a known meaning     | 27 052  | 5.19%  |
| All data                                | 320 865 | 61.52% |

Table 6: Usage statistics for the Danish dictionaries of fixed expressions

Feedback from a random selection of users showed that the change is explained by the fact that many users are looking particularly for the historical (= generic) background to the fixed expression and selected the dictionary that displayed all data for this reason. In view of this experience, we therefore offer a separate dictionary that supplies such historical data as well as meaning items. It is therefore a cognitive dictionary in which a maximising search is performed (left column) and the items of the respective fields are shown in the third column in the order indicated. We call this dictionary KNOWLEDGE ABOUT FIXED EXPRESSIONS.

| Search | Field                                      | Entry | List                   |
|--------|--------------------------------------------|-------|------------------------|
| 1      | 1. Core field                              |       | 1                      |
|        | 2. Meaning in Afrikaans                    |       | 2 1 <sup>st</sup> line |
|        | 9. Background remark(s)                    | 6     |                        |
|        | 10. Comment on background remark(s)        | 7     |                        |
|        | 11. Internet link to background remark(s)  | 8     |                        |
| 2      | 12. Fixed expression(s) in Afrikaans       | 1     |                        |
|        | 13. Remark(s) on the fixed expression(s)   | 2     |                        |
|        | 14. References to fixed expression(s)      | 3     |                        |
|        | 20. Classification of the fixed expression | 4     |                        |
|        | 21. Comment on classification              | 5     |                        |
| 3      | 35. Key word(s)                            |       |                        |

Table 7: Search and data fields in the dictionary  
KNOWLEDGE ABOUT FIXED EXPRESSIONS

#### 4.5 AFRIKAANS-ENGLISH DICTIONARY OF FIXED EXPRESSIONS

We call the fifth dictionary the AFRIKAANS-ENGLISH DICTIONARY OF FIXED EXPRESSIONS. It is a communication dictionary with the function of translation. It is not an ideal translation dictionary, however, as no grammatical information on the English equivalents is presented and no translations of collocations or examples are supplied in Afrikaans.

| Search | Field                                                        | Entry | List |
|--------|--------------------------------------------------------------|-------|------|
| 1      | 1. Core field                                                |       | 1    |
|        | 2. Meaning in Afrikaans                                      | 2     |      |
|        | 3. Internet link to meaning                                  | 3     |      |
|        | 4. Further meaning item in Afrikaans                         | 4     |      |
|        | 5. Meaning in English                                        | 6     |      |
| 2      | 12. Fixed expression(s) in Afrikaans                         | 1     |      |
|        | 16. Fixed expression(s) in English translated from Afrikaans | 5     |      |
| 3      | 35. Key word(s)                                              |       |      |

Table 8: Search and data fields in the dictionary  
KNOWLEDGE ABOUT FIXED EXPRESSIONS

#### 4.6 COMPREHENSIVE KNOWLEDGE ABOUT FIXED EXPRESSIONS

The sixth dictionary is activated by pressing the button "I want to know as much as possible about fixed expressions". We call it COMPREHENSIVE KNOWLEDGE ABOUT FIXED EXPRESSIONS. It is a traditional polyfunctional dictionary that shows all fields (except for

the field for working notes). A minimising search is performed.

| Search | Field                                                        | Entry | List |
|--------|--------------------------------------------------------------|-------|------|
| 1      | 1. Core field                                                | 1     | 1    |
|        | 2. Meaning in Afrikaans                                      | 12    |      |
|        | 3. Internet link to meaning                                  | 13    |      |
|        | 4. Further meaning item in Afrikaans                         | 14    |      |
|        | 5. Meaning in English                                        | 15    |      |
|        | 6. Grammar                                                   | 19    |      |
|        | 7. Comment on grammar                                        | 20    |      |
|        | 8. Internet link to grammar                                  | 21    |      |
|        | 9. Background remark(s)                                      | 16    |      |
|        | 10. Comment on background remark(s)                          | 17    |      |
|        | 11. Internet link to background remark(s)                    | 18    |      |
| 2      | 12. Fixed expression(s) in Afrikaans                         | 7     |      |
|        | 13. Remark(s) on the fixed expression(s)                     | 8     |      |
|        | 14. References to fixed expression(s)                        | 9     |      |
|        | 15. Internet link to variants, e.g. statistical              | 10    |      |
|        | 16. Fixed expression(s) in English translated from Afrikaans | 11    |      |
|        | 17. Style                                                    | 2     |      |
|        | 18. Comment on style                                         | 3     |      |
|        | 19. Internet link to style                                   | 4     |      |
|        | 20. Classification of the fixed expression                   | 5     |      |
|        | 21. Comments on classification                               | 6     |      |
|        | 22. Collocation(s)                                           | 22    |      |
|        | 23. Comment on collocations                                  | 23    |      |
|        | 24. Internet link to collocations                            | 24    |      |
|        | 25. Example(s)                                               | 25    |      |
|        | 26. Comment on examples                                      | 26    |      |
|        | 27. Internet link to examples                                | 27    |      |
|        | 28. Synonym(s)                                               | 28    |      |
|        | 29. Comment on synonyms                                      | 29    |      |
|        | 30. Internet link to synonyms                                | 30    |      |
|        | 31. Antonym(s)                                               | 31    |      |
|        | 32. Comment on antonyms                                      | 32    |      |
|        | 33. Internet link to antonyms                                | 33    |      |
|        | 34. Associated concept(s)                                    | 34    |      |
| 3      | 35. Key word(s)                                              | 35    |      |
|        | 36. Memo field                                               |       |      |

Table 9: Search and data fields in the dictionary  
COMPREHENSIVE KNOWLEDGE ABOUT FIXED EXPRESSIONS

## 5. Forthcoming attractions

Ultimately, a database and the dictionaries extracted from it are never finished, as new cards can constantly be added and those that have already been made can also be expanded or corrected. Our aim is to build up a database of 10,000 to 15,000 cards. However, we will already offer the users the lexicographically recorded expressions when only 1,000 cards are ready. In the further course of the work we will, as explained with reference to the Danish dictionaries above, amend or add specific / additional data on the basis of log file analyses and user feedback, as well as on the basis of further research on and experimentation with different concepts and tools for manipulating data in the e-environment.

Provision has already been made for expansion. The intention is to give users the opportunity to define their profiles, to define their search criteria and to select fields and the order in which they are displayed. For some fields we intend providing the option of displaying more detailed information on request and access to advanced tools. We assume that only a small number of users will make use of these options. Nevertheless, when they do, even more dictionaries will be extracted from one and the same database. It may not be possible to give each of the new, user-defined dictionaries a functional description, as has been done here. However, such options will be "capable of meeting all the users' needs in specific types of situations" (Tarp 2009a: 292) by providing "dynamic articles [...] structured in different ways according to each type of search criteria", "articles that are especially adapted", resulting in "the 'individualization' of the lexical product, adapting to the concrete needs of a concrete user" (Tarp, 2009b: 57-61).

### 5.1 User profiling

We intend providing users with the possibility to define a user profile at the beginning of a consultation session; see Bothma, 2011 for details about user profiling technologies. Users will be able to set up a persistent profile that will remain active across multiple user sessions, but will be able to either reset or change this profile at any stage. Profiles will enable users to define the specific dictionary they intend consulting during a specific interaction session. For example, a user who is reading a text and regularly needs help only with the meaning of fixed expressions may set his/her profile to use the dictionary MEANING OF FIXED EXPRESSIONS as the default dictionary. A user will also be able to set personalised search options (as discussed below) as default.

### 5.2 Personalized search and display options

The six dictionaries discussed above are six different customised views on the database. Each of these dictionaries is defined in terms of a specific type of user need defined by the lexicographer. Each of the dictionaries is monofunctional in terms of a text reception, text production, text translation or a cognitive

information need (in addition to a "traditional" polyfunctional dictionary). It is possible to provide any further number of monofunctional dictionaries in terms of the lexicographer's analysis of perceived user needs. However, it is also possible to provide the user with the option to define his/her own search and therefore define his/her own personalised / customised dictionary. The principles are discussed in Bothma, 2011 and Bergenholtz & Bothma, 2011. We intend providing such customised advanced search facilities where the user can define exactly which data are to be displayed. The user will be able to display the data of only a single field or any combination of fields to satisfy unique information needs in a given situation.

### 5.3 Additional fields for more detailed information

Currently we assume that all users require the same amount of detail when accessing a dictionary article by means of any of the six dictionaries and / or the customisation options. However, this is not necessarily the case. Some users may require only a brief description whereas others may require a detailed exposition. This obviously does not apply to all fields, but could typically apply to, for example, background remarks (fields 9-11) and examples (fields 25-27).

A user may require only a few brief comments about the origin and/or history of a fixed expression, or, alternatively, could require a comprehensive exposition on the origins of an expression, alternative views about the origin, a discussion about erroneous or popularly held beliefs about the origins of the expression, etc. The database should make provision to satisfy these individualised user needs as well. The content required for these details can be provided by a member of the lexicographic team (probably a team member who has a background or interest in history, heritage and culture studies) or could be a link to external source(s) where the background of a fixed expression may have been discussed in detail. We intend providing such a facility for expansion. These data can be made accessible on demand, either by means of a "Read more" button when data of fields 9-11 are displayed or by adapting the user profile at the start of the consultation session.

The current database structure makes provision for examples with comments about and links to the original contexts of the examples. We provide a highly selective list of examples to illustrate meaning and use of a specific fixed expression. However, we foresee that in individual cases users may require either more examples or additional detail. For example, in a text production situation, a user writing a historical novel may require to know which of two current variants of a fixed expression was used (or was the more common variant) at the time the novel takes place. This requires access to data typically not within the database and tools for text manipulation that are not associated with a

lexicographical database. (One of a number of dictionaries that does incorporate such a facility is the *Base lexicale du français* (BLF) (<http://ilt.kuleuven.be/blf>) which provides the user with the option of linking to various corpora, including a set of documents of the European Parliament and Wikipedia. The selection of the examples does not require any input from the lexicographer as the BLF and the corpora are linked automatically. These examples are displayed by the BLF only when the user requires this and the possible information overload is displayed on demand.) In the above example a user may require to see the actual examples in context, i.e., a concordance of examples in a keyword in context (KWIC) format; alternatively, a user may require to see a table that provides only a statistical analysis of the occurrence of variants at a specific time. The two options require two different types of tool, namely a tool that can present “raw” corpus data in a KWIC format as well as a tool that can do statistical analysis of the “raw” corpus data and present the results in statistical tables. We hope to incorporate such facilities in due course. This will, however, require a considerable amount of both theoretical and empirical research and depends on the availability of suitable corpora. Research issues that need to be taken into account to incorporate such a facility are, *inter alia*:

- How should the data in the external database(s) be marked up to enable access to specific data at a fine level of granularity? In terms of the above example, granularity may include mark-up for different time periods, different genres, etc.
- How are word form variants such as inflections and conjugations to be handled? For example, does the database require detailed tagging of morphological forms beforehand, or would it be possible to link to the “raw” text of the corpora on the fly without prior tagging?
- What type of tools will be required to make this type of searching/linking possible?

#### 5.4 Multi-language databases

Currently, the database makes provision for Afrikaans and only a single field for English. It is feasible to use the concepts and database structures outlined here for other languages as well, as indicated above. It is therefore feasible to create multiple interlinked databases for fixed expressions in multiple languages. For translation purposes such multiple databases could be interlinked via a pivot language, for example English. Existing databases of fixed expressions could also be linked, even if the data fields in the different databases are not identical. The minimum requirement would be that there are at least a minimum set of corresponding fields, or that translation tables between different fields can be created.

### 6. Conclusion

Some of the envisaged expansions discussed above may not necessarily currently be commercially feasible since the time required to do the programming or to write / collate / select the data may simply be too much to

complete the dictionary in a reasonable time. In addition, some of these expansions may not be what individual users may require. However, if researchers do not experiment with concepts and technologies that currently do not seem commercially realistic or feasible, innovation in e-information tools will be stifled. Such “blue sky” research could eventually lead to e-information tools that are not only incrementally better than those that are currently available, but provide different tools through disruptive innovation. The current project therefore has two aims:

- To create a database of fixed expressions, as well as to develop the necessary database tools, administrative backend, user interface and search functions, that enable users to have access to a number of monofunctional and one polyfunctional dictionaries. To result in a useful product this database and set of tools has to be completed in a limited timeframe (even though further extensions and updates need to be added regularly).
- To provide a platform to experiment with disruptive technologies and see to what extent any of these technologies can add value for the user in providing access to information in terms of the user’s specific information need in a given user situation.

Such “blue sky” research is absolutely essential to ensure that not only better but different types of e-tools are developed. After all, the development of new cars is not left up to the drivers. One can ask drivers about which aspects of their cars they are not quite satisfied, and the designers and manufacturers of cars can then make the required improvements. However, drivers do not possess the know-how and the technical creativity that is necessary to design and develop cars that are totally new, much better and also manufactured quite differently. As Henry Ford allegedly said, “If I had asked people what they wanted, they would have said faster horses.” e-Dictionaries are no different. Users may help to improve e-dictionaries incrementally, but only fundamental research in metalexicography, user needs, database technologies and principles of information organisation, access and retrieval will result in different types of e-tools.

### 7. References

- Bergenholtz, H. (2010). Needs-Adapted Data access and data presentation. In *Doctorado Honoris Causa del Excmo. Sr. D. Henning Bergenholtz*. Valladolid, pp. 41-57.
- Bergenholtz, H. (2011). Access to and presentation of needs-adapted data in monofunctional internet dictionaries: In P.A. Fuertes-Olivera, H. Bergenholtz (eds.) *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London & New York: Continuum 2011, pp. 30-53.
- Bergenholtz, H., Bergenholtz, I. (2011). A dictionary is a tool, a good dictionary is a monofunctional tool. In P.A. Fuertes-Olivera, H. Bergenholtz (eds.)



- e-Lexicography: The Internet, Digital Initiatives and Lexicography*. 2011. London & New York: Continuum, pp. 188-207.
- Bergenholtz, H., Bothma, T.J.D. (2011). Needs-adapted data presentation in e-information tools. *Lexikos*, in press.
- Bergenholtz, H., Johnsen, M. (2005). Log files as a tool for improving Internet dictionaries. *Hermes*, 34, pp. 117-141.
- Bergenholtz, H., Johnsen, M. (2007). Log files can and should be prepared for a functionalistic approach. *Lexikos*, 17, pp. 1-20.
- Bergenholtz, H., Tarp, S. (2002). Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. *Lexicographica*, 18, pp. 253-263.
- Bergenholtz, H., Tarp, S. (2003). Two opposing theories: On H.E. Wiegand's recent discovery of lexicographic functions. *Hermes*, 31, pp. 171-196.
- Bergenholtz, H., Tarp, S. (2005). Wörterbuchfunktionen. In I. Barz, H. Bergenholtz & J. Korhonen (eds.) *Schreiben, Verstehen, Übersetzen und Lernen: Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*. 2005. Frankfurt a.M./Bern/New York/Paris: Peter Lang, pp. 11-25.
- Botha, R.P., Kroes, G. & Winckler, C.H. (1994). *Afrikaanse idioome en ander vaste uitdrukkings*. Halfweghuis: Southern.
- Bothma, T.J.D. (2011). Filtering and adapting data and information in the online environment in response to user needs. In P.A. Fuertes-Olivera, H. Bergenholtz (eds.) *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. 2011. London & New York: Continuum, pp. 71-102.
- De Villiers, M., Gouws, R.H. (1988). *Idiomewoordeboek*. Cape Town: Nasou.
- Malherbe, D.F. (1924). *Afrikaanse spreekwoorde en verwante vorme*. Bloemfontein: Nasionale Pers.
- Prinsloo, A.F. (1997). *Afrikaanse spreekwoorde en uitdrukkings*. Pretoria: J.L. van Schaik.
- Prinsloo, A.F. (2009). *Spreekwoorde en waar hulle vandaan kom*. Cape Town: Pharos.
- Tarp, S. (2002). Translation dictionaries and bilingual dictionaries. Two different concepts. *Journal of Translation Studies*, 7, pp. 59-84.
- Tarp, S. (2007). Lexicography in the Information Age. *Lexikos*, 17, pp. 170-179.
- Tarp, S. (2008). *Lexicography in the borderland between knowledge and non-knowledge. General lexicographical theory with particular focus on learners' lexicography*. (Lexicographica. Series Maior 134). Tübingen: Max Niemeyer.
- Tarp, S. (2009a). Reflections on lexicographical user research. *Lexikos*, 19, pp. 275-296.
- Tarp, S. (2009b). Reflections on data access in lexicographic works. In S. Nielsen, S. Tarp (eds.) *Lexicography in the 21st Century. In Honour of Henning Bergenholtz*. (Terminology and Lexicography Research and Practice, Volume 12). 2009. Amsterdam: John Benjamins, pp. 43-65.
- Tarp, S. (2011). Lexicographical and other e-tools for consultation purposes: Towards the individualization of needs satisfaction. In P.A. Fuertes-Olivera, H. Bergenholtz (eds.) *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. 2011. London & New York: Continuum, pp. 55-70.
- Vrang, V., Bergenholtz, H. & Lund, L. (2003-2005). *Den danske Idiomordbog*. Database and layout: Richard Almind. [www.idiomordbogen.dk](http://www.idiomordbogen.dk).
- Wiegand, H.E. (1977). Nachdenken über Wörterbücher. In *Nachdenken über Wörterbücher*. Mannheim/Wien/Zürich, pp. 51-102.