CHAPTER 7

A bilingual electronic dictionary of idioms

Erla Hallsteinsdóttir

In this article the initial results of a research project on bilingual lexicography and idioms are presented. The results discussed here include (a) the evaluation of the potential and advantages of electronic dictionaries in comparison to printed dictionaries, and (b) how that potential is exploited to represent the lexicographical content in the German ↔ Icelandic dictionary of idioms.

1 Introduction

In this article I will present the initial results of a research project on bilingual lexicography and phraseology at the University of Iceland. The project was supported by a post-doctoral fellowship from The Icelandic Centre for Research. The aim of the project is to analyse linguistic and lexicographal theories and evaluate practical solutions in order to provide a basis for the conceptual work on and the compilation of an electronic German ↔ Icelandic dictionary of idioms. The results discussed here include (a) the evaluation of the potential of the computer as a medium for dictionaries in comparison with printed dictionaries and (b) how that potential is best exploited to represent lexicographical information in a dictionary of idioms.

Since the phraseological sources used in my research do not share a uniform definition of an idiom (see, for instance, Burger et al. (eds.) 2003 and Dobrovolskij 1997), the term ‘idiom’ in this article is to be understood in a broad sense covering a variety of phraseological units.

2 The lexicographical potential of the computer

Only few lexicographical theories deal with the development of electronic dictionaries. Many electronic dictionaries are by-products of printed dictionaries, and the conceptual lexicographical work is therefore based on the limits of the medium of printed books. The result is that many of those dictionaries, and especially the few
idiomatic dictionaries, do not exploit the potentials of the computer, but are merely printed dictionaries in electronic form (one exception being the Danish idiom dictionary *Den Danske Idiomordbog*: www.idiomordbogen.dk).

As mentioned above, one aim of this project is to describe the lexico graphical potential of electronic dictionaries, and how successfully they exploit the lexico graphical advantages of the electronic media over the print media. These advantages are:

1. Modern computer and database programs are not restricted by size and allow unlimited use of content such as text, graphics, audio and videos.
2. By using hyperlinks, the lexicographical content can be organized as a hypertextual network of information. The result is a multidimensional hypertext that extends beyond the boundaries of a two-dimensional printed page.
3. The software used should allow different user groups to adjust the search structures and the output of the database according to their needs.

In a very early phase of the conceptual work, I decided not to publish my own Icelandic-German dictionary as a printed book. This decision was made in order to be able to concentrate on the advantages of the computer without taking into account printing limitations.

3 Exploiting the lexicographical potential of the computer

3.1 Software

The database and publishing software *askSam* was used to compile the dictionary. The choice of *askSam* is based on the following features of the software:

1. *askSam* is very flexible as far as the organizing of information is concerned. It is a free-form text database for the storage of text as unstructured data. It also allows defined fields for structured data (as in a traditional database) with search and reporting functions. Cross-references and the organization of information with hypertext is possible.
2. *askSam* can import and export data in common formats, and it allows the insertion of graphics, pictures, sounds and video clips by Object Linking and Embedding (OLE).
3. The database size is only limited by the hardware used.
4. Professional publishing of the dictionary on CD-ROM and on the Internet is possible with *askSam Electronic Publisher* or *askSam Web-Publisher*. There is a free *askSam-Viewer* for users who do not have the *askSam* database software themselves.
3.2 Dictionary structure and database organization

The dictionary is structured as a modular database. There is a module for each language, with submodules that contain specific lexicographical information on different hypertext levels. Hyperlinks provide cross-references between the modules and/or types of information.

The idioms in each language are intended to represent a basic idiomatic corpus for German (or Icelandic) as a foreign language. This allows the possible compilation of multilingual dictionaries by adding idiom modules in other languages. However, each language pair must undergo individual contrastive analysis in order for language specific relations to be included.

3.3 Access to the dictionary

The dictionary's users access the lexicographical data through hypertext menus. These menus include an overview of the contents of the language modules, search options, word and idiom lists.

*askSam* offers a wide variety of search commands, including proximity, boolean, wildcard and fuzzy searches. The search functions can be combined into user-specific requests searching the whole database, a single module or specified fields.

In each language an alphabetically sorted idiom list provides access to the lexicographical data by hyperlinks. The list is sorted by the components of the idioms, whereby each component of an idiom occurs in the list.

The idioms are sorted onomasiologically by (a) the semantic fields of their components, and (b) the semantic concepts of the idiomatic meaning (see Hallsteinsdóttir 2005). Although the pre-arranged information in the hypertext menus should preferably be used to access the dictionary, it is possible to “leaf” through the lemmata in the database.

3.4 Dictionary components and document structure

Each idiom is represented in a main document in the database (traditionally a dictionary entry). Each main document has a visual size that can be viewed on two computer screens.

The first screen contains information about the idiom form (lemma, usage forms and usage restrictions, corpus examples, negation forms, formal relations to other idioms, search forms and frequency in corpus) and a paraphrase of the idiom’s meaning (see Figure 1).
Figure 1. Information in the database, screen one.

The second screen contains information on etymology, pragmatics, equivalence and data for the lexicographical work like basic metaphors, figurative aspects, concept(s) of the idiomatic meaning(s), onomasiological group of the literal meaning of the components, phraseological classification, comments on lexicographical work, etc. (see Figure 2).
Figure 2. Information in the database, screen two.

The underlined words in Figure 1 and Figure 2 represent hyperlinks. The areas with a grey background are predefined database fields. The layout is additionally structured by using colours. In the following sections I will describe the content of the different types of information in the dictionary.

3.4.1 The lemmata
Each main document begins with a lemma. I have chosen the simplest form of each idiom as the lemma. The main function of a lemma here is to differentiate and identify the idioms. In this connection the lemma (as a unique form of the idiom) is used as a bookmark when setting hyperlinks.
3.4.2 Usage forms and corpus examples

The usage forms of an idiom are listed in the main document in the field *Aussageformen* ('usage forms'). The external valency is shown formally here with the substitutes *jemand* ('someone') and *etwas* ('something'). The concrete form is given in the analysed corpus examples (see Burger 1998: 176 and Filipenko 2002: 45–46).

Complementary grammatical and lexical information is added where necessary, including information on formal variation (as NB!-Information in Figure 3).

Additionally, morphosyntactic and semantic restrictions found in the corpus examples are described. In the specific case where one idiom refers to a process and another idiom describes the result of this process (or a corresponding event, respectively) (see Filipenko 2002: 48), the relations between those idioms are shown with hyperlinks in the field *Aussageformen* ('usage forms'), (see the *Siehe auch* ('see also') information in Figure 3).

The analysed corpus examples are kept in a separate module in the database and linked to the main document by the hyperlink *Verwendungsbeispiele* ('corpus examples').

The corpus examples include examples of negation (linked by the hyperlink *Negation*) and the lexical modifications found in the corpus examples (see abridged version in Figure 4).
3.4.3 Search forms and frequency

The idioms in the database were selected by determining the frequency of over 5000 German idioms from dictionaries of idioms and dictionaries of German as a foreign language. The frequency test was carried out in the corpus Deutscher Wortschatz (www.wortschatz.uni-leipzig.de) at the University of Leipzig in April 2002 by using constructed search forms that correspond to possible usage forms of the idioms. The basis of the dictionary consists of the 300 idioms that had more than 1000 records in the text corpus (see Hallsteinsdóttir 2005). During the analysis of the corpus examples it was established that not only the ‘base’ idioms occur in the corpus examples but also ‘spin-off’ idioms. These idioms have been added to the dictionary as lemmata and linked to formally related idioms by hyperlinks in the field Weitere Phraseologismen mit der Form: “am Ende” (‘other idioms with the form …’). The search forms used occur in the field Suchform (‘search form’), with the total number of records in the text corpus in brackets. The frequency information enables users to see whether an idiom belongs to the core or the periphery of
the phraseology. The search forms (here: *am Ende*) represent hyperlinks to the total amount of corresponding corpus examples.

3.4.4 **Semantic and pragmatic information**

The meaning of the idiom is explained in the field *Bedeutung* (‘sense definition’) in simple terms, so that a non-native speaker would understand it (cf. Figure 1). If the idiom has more than one meaning, the different senses are numbered with letters.

The etymology of the idiom is explained in the field *Etymologische Erklärung* (‘etymological explanation’) (cf. Figure 2) as a “historischer Kommentar” (‘historical comment’) if it is relevant for the motivation behind the idiomatic meaning. The comment explains the opaque semantic structure by referring back to the origin of the idiom in order to make the meaning of the idiom transparent to the user (see Burger 1998: 180).

Synonymous and antonymous idioms are listed as hyperlinks in the field *Semantische Relationen* (‘semantic relations’).

The field *Pragmatische Angaben* (‘pragmatic information’) includes instructions for the use of the idiom. The information in this field is intended to be an explicit comment on how to use the idiom and – more importantly – it explains the usage restrictions (i.e. how not to use the idiom) in addition to the information in usage examples. Ideally this field should include all the information that a non-native speaker needs in order to be able to use an idiom (see Mudersbach 1998: 319 and Cheon 1998).

3.4.5 **Equivalence**

The bilingual part of the dictionary is limited to four fields. In my opinion, it is important to show the formal structure of the idiom components; each idiom is therefore translated word by word into Icelandic (or German, respectively) in the field *Wörtliche Übersetzung ins Isländische* (‘literal translation in Icelandic’). A translation of the meaning of the idiom from the field *Bedeutung* (‘sense definition’) is given in a separate field. The field *Äquivalente isländische Phraseologismen* (‘equivalent Icelandic idioms’) contains semantically (near-)equivalent idioms. The equivalence relations are explained in the field *Erklärung der Äquivalenzbeziehung* (‘explanation of equivalence relations’). The syntactic equivalence is also explained if the syntax differs. Idioms that are syntactically equivalent, but have different meanings, are highlighted in red and provided with an explicit comment, as a warning against ‘false friends’.

3.4.6 **Lexicographical data**

Some lexicographical data are important for the compilation of the dictionary, the design of the user interface and the lexicographical work, but irrelevant for the dic-
tionary users. These data are found at the end of each main document. The data include the explanation of the idiom metaphor on the basis of the relations between the literal meaning of the components of the idiom, the etymology and the idiomatic meaning. On the basis of this metaphorical explanation the idiom is marked semantically in specific fields.

These fields are used to establish the two kinds of onomasiological access in the user interface: (a) the classification of the semantic fields of the literal meaning of the idiom components is used to enable traditional onomasiological access to the dictionary, and (b) the classification of the idioms according to the semantic concept of the idiom enables conceptual access to the idiom meaning. The basis for the semantic concept of the idiom meaning is provided by Jón Hilmar Jónsson's conceptual classification of 33,000 Icelandic idioms in 840 concepts (see Jónsson 2002). The semantic marking of the idiom meaning can be used for further work on understanding and describing phraseological semantic relations. In the dictionary, the marking is used to group the synonymous and antonymous idioms in a simple and effective way.

4 Conclusion

The computer yields various new lexicographical options which have not yet been fully described. In this article I have shown some ways in which the lexicographical advantages of the computer can be exploited in a bilingual electronic idiomatic dictionary. This article contributes to a theory of computer lexicography still at an early stage, and I hope that my ideas will encourage further work in this field.

5 References

Deutscher Wortschatz: www.wortschatz.uni-leipzig.de [last accessed July 27, 2007]


Ordbogen over Faste Vendinger: www.idiomordbogen.dk [last accessed July 27, 2007]