TERMINOLOGY and LEXICOGRAPHY RESEARCH and PRACTICE 10

# Dictionary Visions, Research and Practice

EDITED BY

Henrik Gottlieb

Jens Erik Mogensen

John Benjamins Publishing Company

Dictionary Visions, Research and Practice

# *Terminology and Lexicography Research and Practice (TLRP)*

*Terminology and Lexicography Research and Practice* aims to provide in-depth studies and background information pertaining to Lexicography and Terminology. General works include philosophical, historical, theoretical, computational and cognitive approaches. Other works focus on structures for purpose- and domain-specific compilation (LSP), dictionary design, and training. The series includes monographs, state-of-the-art volumes and course books in the English language.



**Volume 10**

Dictionary Visions, Research and Practice. Selected papers from the 12th International Symposium on Lexicography, Copenhagen 2004
Edited by Henrik Gottlieb and Jens Erik Mogensen

# Dictionary Visions, Research and Practice

## Selected papers from the 12th International Symposium on Lexicography, Copenhagen 2004

*Edited by*

Henrik Gottlieb
University of Copenhagen

Jens Erik Mogensen
University of Copenhagen

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

**Library of Congress Cataloging-in-Publication Data**

International Symposium on Lexicography (12th : 2004 : Copenhagen, Denmark)
Dictionary visions, research, and practice : selected papers from the 12th International Symposium on Lexicography, Copenhagen, 2004 / edited by Henrik Gottlieb and Jens Erik Mogensen.
p. cm. (Terminology and Lexicography Research and Practice, ISSN 1388-8455 ; v. 10)
Includes bibliographical references and index.
1. Lexicography--Congresses. I. Gottlieb, Henrik. II. Mogensen, Jens Erik, 1964- III. Title.
P327.I55 2004
413'.028--dc22 2007037109
ISBN 978 90 272 2334 0 (Hb; alk. paper)



John Benjamins Publishing Co. · P.O. Box 36224 · 1020 ME Amsterdam · The Netherlands
John Benjamins North America · P.O. Box 27519 · Philadelphia PA 19118-0519 · USA

# Table of contents

**Part 4. LSP lexicography**



**Part 5. Dictionaries and the user**



**Part 6. Etymology, history and culture in lexicography**

# Preface

The Twelfth International Symposium on Lexicography at the University of Copenhagen, the select proceedings of which are hereby published, took place on April 29 – May 1, 2004. In this volume, we present 19 papers from lexicographers in 11 countries: Canada, Denmark, Germany, Great Britain, Iceland, Israel, Latvia, The Netherlands, Poland, South Africa and Spain.

Since 1994 the International Symposium on Lexicography has opened with the *Otto Jespersen Memorial Lecture*, and the 2004 speaker was the founding father of the symposium, Arne Zettersten (Copenhagen), whose paper is entitled *Glimpses of the future of English-based lexicography*. The two other plenary lectures of the symposium were *Politik und Sprachpolitik in der Lexikographie*, by Henning Bergenholtz and Sven Tarp, and *Neuere Aspekte einer Theorie und Typologie von Wörterbuchartikeln und ihre Praxisrelevanz,* by Herbert Ernst Wiegand. All 19 papers – of which 15 are in English – have been anonymously peer-reviewed and selected. The primary criteria for selection were: quality, relevance, and variety of approaches and topics. The fields covered include online lexicography, dictionary structure, phraseology in dictionaries, LSP lexicography, user-friendliness in dictionary making, plus etymology, history and culture in lexicography.

The volume is a collection of papers covering a wide field within current lexicography. It opens with three chapters on central topics concerning *Online Lexicography*: Stating that online dictionaries have to be treated as specific products, Cristina Gelpí investigates quality parameters of online bilingual dictionaries and proposes a selection of main features on the basis of examples from specialized English-Spanish Dictionaries. The contribution of Annette Klosa and Carolin Müller-Spitzer deals with the grammatical data in "elexiko", an extensive electronic dictionary of contemporary German. In his paper about graphical user interfaces, Włodzimierz Sobkowiak argues for more radical ways of representing pronunciation in electronic dictionaries.

In the section on *Dictionary Structure*, Rufus Gouws argues that lemmatization should not be the result of linguistic, but user-driven lexicographical criteria such as lexicographical functions and the needs of the users. Andrejs Veisbergs discusses some central problems connected with the reversal of English-Latvian dictionaries in order to create a basis for compiling a new Latvian-English Dictionary.

The section *Phraseology in Dictionaries* comprises two contributions: Ken Farø discusses the equivalence of idioms in German-Danish dictionaries and argues that pragmatic facts, rather than structural similarities, should be the deciding factor in establishing lexical equivalents, and Erla Hallsteinsdóttir describes the advantages of electronic dictionaries over printed dictionaries in representing idioms in a German↔Icelandic dictionary.

In the section *LSP Lexicography*, Patrick Leroyer presents a number of new dictionary functions relevant for bringing dictionary design into accord with corporate image, while Sandro Nielsen & Lise Mourier describe a new function-based internet accounting dictionary.

The section *Dictionaries and the User* comprises four chapters: In the first of these, Ilan J. Kernerman considers the qualities and drawbacks of monolingual dictionaries for advanced learners of English. In the second chapter, Renata Szczepaniak investigates advanced learners' use of receptive dictionaries. Next comes Geart van der Meer's analysis of the grammatical information in five major English learner's dictionaries, the so-called 'big five', which are all shown to have serious problems in incorporating 'advanced' grammatical information. Finally, Herbert Ernst Wiegand proposes a new definition of the term *dictionary article* and introduces several new lexicographical terms in the process.

The volume concludes with the section *Etymology, History and Culture in Lexicography.* Arleta Adamska-Sałaciak deals with the lack of interlingual equivalence, questioning whether borrowings from the source language can be used as translation equivalents in a bilingual dictionary. In their language-political contribution, Henning Bergenholtz & Sven Tarp discuss the relation between lexicography and linguistics, distinguishing between a general, a specific and a selecting language policy. Andreas Gröger demonstrates the progress made from the second to the third edition of the *Oxford English Dictionary* as regards its German etymologies, and Marcin Overgaard Ptaszyński discusses the peculiarities of usage information in 18th century English↔Danish lexicography. In her contribution, Roda P. Roberts argues that dictionaries are not only linguistic, but also cultural objects. She examines how national culture, transnational culture and subculture are all reflected in lexicography. Arne Zettersten concludes this volume with a look into the future of English-based lexicography, concentrating on a number of topical issues regarding the potential of electronic dictionaries.

The editors wish to thank the authors for placing their manuscripts at our disposal. Thanks also to all participants – old friends as well as newcomers – for joining the symposium. We are indebted for financial support to The Department of English, Germanic and Romance Studies at the University of Copenhagen. Finally,

we are grateful to our student assistants, Søren Dahl and Aske Kammer, for helping us organize the symposium and for the preparation of the papers for the printers.

Henrik Gottlieb & Jens Erik Mogensen, editors
Center for Translation Studies & Lexicography,
Department of English, Germanic and Romance Studies,
University of Copenhagen

Part 1

# Online lexicography

CHAPTER 1

# Reliability of online bilingual dictionaries

Cristina Gelpí

Online bilingual dictionaries are specific products, different from printed dictionaries. Coverage is not the only criterion to determine dictionary quality. Parameters as the origin of the web page, user-specificity, access control, security elements, degree of hypertextuality, permeability, interaction, adaptation to lexicographical functions, access system to data, etc. are also important for describing online dictionaries. This paper proposes a selection of main features of online bilingual dictionaries, according to lexicographical and digital parameters, and includes examples from online specialized bilingual English-Spanish dictionaries.

## Introduction

The basic purpose of a bilingual dictionary, according to Zgusta (1971: 294), is to coordinate with the lexical units of one language those lexical units of another language which are equivalent in their lexical meaning. The relation among users, lexicographic functions and structures is what determines the features of bilingual dictionaries.

The medium in which bilingual dictionaries are presented is one of the elements that determine the selection, organization and representation of information in dictionaries. The Internet, from this point of view, is a good context to present not only new kinds of lexicographical information, but also new kinds of lexicographical products.

The access to lexicographical online information is easy, quick and cheap. One consequence of these qualities is the popularization of online lexicographical products. Popularization, or dissemination, may have advantages, but may also imply major disadvantages, as it makes it easy to find low-quality lexicographical products alongside good products. This situation can produce a lack of confidence and sense of helplessness in users.

Users often do not have explicit criteria to measure the quality of the lexicographical products that they use. Even the expert user may have difficulties in choosing the best dictionary for a given communicational purpose. These inconveniences pose a problem for users, because they do not have points of reference to decide if a given dictionary is "good" or "bad". When facing lexicographical information, users tend to trust in dictionaries, but this is not always the best policy.

The ordinary user does not usually have the need to evaluate dictionaries as a lexicographer does. Even though some users do not think about lexicographical evaluation, dictionaries may still be evaluated – not least online dictionaries, due to their dissemination. The question is then: how can we know the reliability of online dictionaries? How can we know if a dictionary is up to date, complete and exhaustive? And how can a user measure whether a dictionary is conceived to resolve her or his communicative needs?

To answer these questions, a lexicographical evaluation system is needed. However, such a system needs to have some criteria, which have not yet been firmly established. Before an evaluation system can be established, a list of quality indicators of online dictionaries must be provided.

My purpose in this paper is to present a list of quality indicators, based on two kinds of parameters. On the one hand, main lexicographical features of online dictionaries, i.e. author, addressee and/or user, lexicographical function and lexicographical structures. On the other hand, online-specific lexicographical requirements: usability, permanence and updating, degree of digitization and hypertextuality, and permeability and interaction of online bilingual dictionaries.

## 2 Main lexicographical features of online dictionaries

### 2.1 The author of the dictionary

Every dictionary has an author who is the responsible for quality. The author of a dictionary may be identified by the dictionary user. It is always important that the user knows who is offering the data. The issuer of information is generally an individual or an institutional author. An institutional author is acknowledged via his or her institution, while individual authors are independent and, frequently, suspected of not being very good or authoritative.

For example, a good online English-Spanish dictionary, The Collins Concise Spanish Dictionary (http://wordreference.com/), is promoted by a large publishing house (HarperCollins Publishers, 2002). The reliability of a publishing house is a guarantee of the quality of its dictionaries. But we can also find other online bilingual dictionaries, such as English-Catalan dictionaries[1], that are not supported by any important publisher. In such cases, unambiguous information about the author is basic[2], as we can see at http://eubca1.eubca.edu.uy/diccionario/diccion.

htm), a dictionary of organization and knowledge representation, by Mario Barité (from Uruguay). Here, the author is presented with a lot of personal information: personal data, education, lexicographical experience, etc.

In printed dictionaries, the author usually appears on the cover and is thus identifiable as the person responsible for the work. In contrast, authorship of online dictionaries is not always clear. In most cases, a simple © with a name is the only information provided. Sometimes, implicit information is related to DNS (.com, .es, .net, .edu, are clues to take into account). But the desirable situation is to have a minimum amount of information about the author of a dictionary. At the same time, some information about the lexicographical background of the author is desirable.

Moreover, authorship of lexicographical data should be made clear when online dictionaries have been borrowed from existing information sources[3].

## 2.2 The addressee and/or user of the dictionary

Dictionaries must be able to resolve most of the communicative needs of their users. As a result, any lexicographical decision related to the quantity and the type of information included in dictionaries should take into account the needs of its users.

However, user-oriented specialization may not in all cases be the best option, because of the difficulty of establishing a user typology and because of the multiplicity of user types. The economic and functional inconveniences of producing dictionaries devoted to a restricted audience are likewise important.

To address these inconveniences, it is useful to establish a distinction between user and addressee: 'Addressee' refers to the person for whom the dictionary is conceived. 'User' is a whoever uses the dictionary, not necessarily the type of consumer the author or the publisher had in mind. Good dictionaries have at least one addressee and several users (a good example can be found at www.ati.es/novatica/glosario/glosario_internet.html, a basic Internet glossary where information is adapted to user needs: plain language for definitions, a limited amount of specialized information, acronyms in lemma position, etc.). In contrast, most online bilingual English-Spanish dictionaries, both general and specialized, can be considered as undefined from the user's point of view[4].

But in order to take user needs into account, dictionaries do not necessarily have to cater only for users. Bilingual dictionaries exist as textual products irrespective of users, as a catalog of information units. Even though it is technically possible, online lexicography has not developed many systems allowing us to select different kinds of information according to user profile. That is a crucial lexicographical task that remains to be done. Mechanisms to select language or path in dictionaries are not rare (see for example www.wordreference.com) but at present we do not have general or specialized dictionaries that allow us to select and obtain only one kind of information.

### 2.3 The lexicographical function and the objectives of the dictionary

To appropriately resolve the communicative needs of users and addressees, it is important for dictionaries to specialize in regard to the information they include. A relation between user needs and dictionary function may be established, because different users have different needs, so one large dictionary is not enough to resolve all kinds of communicative needs.

Several lexicographers have discussed this question and pointed out that it is not possible in the same dictionary to devote attention to both SL and TL users. This impossibility confirms the need to have different dictionaries according to lexicographical functions. Still, this specialization is only useful in certain contexts. For printed bilingual dictionaries operating with the established production-comprehension opposition, it works well. But in an online context, and with reference to specialized dictionaries, the distinction becomes less interesting.

In online bilingual dictionaries it is more helpful to talk about objectives or purposes if lexicographical functions are included. The field of specialized translation is a useful example: a good online bilingual dictionary conceived to help users in specialized translation contains information concerning comprehension as well as production. See, for example, www.ssa.gov/espanol/glossintro.html (an English-Spanish glossary of Social Security Terminology that contains everyday words and expressions as well as technical Social Security terms: from parts of the body to very specific concepts. It is intended to promote uniformity in language use and prevent misinterpretation of Spanish language materials issued by the Social Security Administration)[5].

Another specific characteristic of online bilingual dictionaries, related to dictionary objectives, is the diversity of subjects. Even though printed dictionaries may be quite diverse, printing is expensive, whereas the web is cheap. A simple search on the words "dictionary, glossary, vocabulary or terminology" in a search system results in a great amount of dictionaries[6].

### 2.4 Lexicographical structures

As printed dictionaries, online dictionaries may be described from the point of view of structures. Following Hausmann and Wiegand (1989: 328–360), and trying to adapt their proposal to online bilingual dictionaries, lexicographical structures may be specified as hyperstructure, access structure, macrostructure, microstructure and iconic structure.

#### 2.4.1 *Hyperstructure*

Hyperstructure includes general organization of the dictionary and is conceived as a textual word list structure. Online bilingual dictionaries must contain the front matter (which includes presentation text and orientation text, such as introduction, presentation and users' guide), nomenclatures or lemmata (which include the word list and other nomenclatures) and back matter (for specialized online bilingual dictionaries it is important to show conceptual information)[7]. According to dictionary type, addressee, lexicographical function and main objective, dictionaries must select certain types of information to be included in each part of the dictionary, and therefore exclude other types of information.

#### 2.4.2 *Access Structure*

Once a user has found a dictionary, navigation inside it starts. Online English-Spanish bilingual dictionaries usually start with an alphabetic menu to access data, or with a search option.

Menus in dictionaries, whether vertical or horizontal, are important, and it is usual for the first online dictionary screen to offer a menu with different components. The menu of online bilingual dictionaries only presents a first level of information[8], and it is not easy to find online bilingual dictionaries with a complete representation of contents[9].

Access structure leads to the article or to the information in the article. Online dictionaries may be polyaccessible, and may offer different ways to access data. A good example of access structure may be found at www.sice.oas.org/Glossary/GLOSSARY.ASP (glossary of foreign trade, where information is found via an alphabetical index, or by using the general search system).

#### 2.4.3 *Macrostructure*

The macrostructure of online bilingual dictionaries is open-ended and only closed superficially. Online dictionaries are always 'under construction', even though they appear as finished products. Being evolutionary, temporary and under construction is not a limitation for online dictionaries, but rather a characteristic. At the same time, online dictionaries should promote the access to external and complementary information.

The selection of lemmata is the result of using criteria that include or exclude words from the dictionary. The criteria used should be clear, explicit and coherent with the objectives of each dictionary. Completeness, exhaustiveness and coherence are specific attributes, because a dictionary should be more than an incomplete list of words, and more than a database of lexical units. The criteria for evaluating the quality of the macrostructure should be the same as those for printed dictionaries.

With regard to the presentation form of entries, the consumer usually accesses only the information needed. Consumers often do not get complementary information (what we call "cerca per aproximació", or fuzzy search). Excluding some commendable exceptions (as www.wordreference.com), most online English-Spanish bilingual dictionaries offer a very simple search system. However, a combined system has been proposed, cf. http://mx.education.yahoo.com/reference/dict_en_es/ (The American Heritage and Spanish Dictionary, including a two-step access to data).

#### 2.4.4 *Microstructure*

Lexicographical discourse in printed dictionaries is homogeneous and compressed, and compression is usually caused by lack of space. The absence of space restrictions is one of the major advantages of online dictionaries.

The various indications of each article must be conditioned by the general characteristics of the dictionary. The presence of definition, phonetic information, synonyms, grammatical information, examples, etc. is not gratuitous but justified by the addressee, lexicographical function and dictionary objectives. Therefore the absence of space restrictions must not be a reason to include indiscriminate information and, at the same time, it is important to make better use of space.

#### 2.4.5 *Iconic structure*

Online iconic structure is clearly different from analogue iconic structure. Differences in the morphology of information are the most important changes, because multimedia resources allow the introduction of image, sound and video in dictionary entries.

A good online iconic structure should include such multimedia resources, but unfortunately this is not very usual in current online bilingual dictionaries. Besides their ornamental value, multimedia resources should be considered appropriate information. At the moment, only images are used as a resource (cf. www.jilt33.com/dti/, a dictionary of computer science).

Thus far I have outlined the main lexicographical features of online bilingual dictionaries. But just as important are the specific online requirements of such dictionaries, precisely because they are products used in an Internet environment. Amongst them, I would like to emphasize usability, updating, degree of digitization, and permeability.

## 3 Specific online requirements

### 3.1 Usability

An online bilingual dictionary is usable only if it is easy to use. Usability is the quality of a system that makes it easy to find, easy to learn, and easy to remember. Lack of usability can cost time and effort, and can greatly determine the failure of a system. Given a choice, people will tend to buy user-friendly systems.

Online usable bilingual dictionaries must provide several alternatives to the users, and they must do so relying on visual content; they must use markup and style sheets, and do so properly; they must use natural language; they must provide contextual and orientational information; they must provide clear navigation mechanisms; and, of course, they must ensure that the lexicographical information is clear and simple[10].

### 3.2 Permanence and updating

One of the most important differences between digital dictionaries and printed ones is stability, i.e. permanence of data. If we consider stability, it is obvious that printed dictionaries exist; they may be found and purchased, and nobody questions their existence. In contrast, the existence of online dictionaries is ephemeral, and minor errors may render them impracticable.

Permanence may be one of the big inconveniences of online bilingual dictionaries, because of the potential daily changes of digital context. An online dictionary is active during an undetermined period of time, and this instability can be negative for users' confidence.

If an online dictionary is created as an open-ended product, some explicit marks of updating should be visible, because information is offered almost in real time. *Last updated*, or simply *updated,* are some of the most used references, but other formulations are also possible: year of creation – current year; indication of update frequency, etc.

Online bilingual dictionaries with a printed version are more stable than true online bilingual ones. For this reason, updating is more important in more unstable dictionaries. Even though updating is a factor that favors reliability, most online bilingual dictionaries do not display this information. Some commendable exceptions are www.doctorproaudio.com/doctor/diccionario.htm, with indications as "2001–2004", and "Última actualización: 5 de julio de 2001" (www.ati.es/novatica/glointv2.html).

At the same time, if an online bilingual dictionary is organized in several informational layers, which is desirable, it is important to measure the reliability and updating of links carefully.

### 3.3 Degree of digitization and hypertextuality

All dictionaries share a lot of aspects with hypertext, if hypertext is conceived as an alternative form of organizing lexicographical information. Online bilingual dictionaries are ideal products for "hypertextualization".

But unfortunately not all online bilingual dictionaries are really designed as online products. Most of them are digitized printed products, or just transformed Word documents, without any consideration for digital resources and possibilities. As a quality standard of online bilingual dictionaries, proper digitization is required.

### 3.4 Permeability and interaction

Some online bilingual dictionaries are defined as interactive products. Such dictionaries must allow interaction with consumers and provide some system to invite users to participate, but at the same time they have to establish filters to avoid indiscriminate addition of data (a good example of interaction is www.insflug.org/colaboracion/recursos/glosario/alpha.php3, a glossary of Internet terms).

## 4 Conclusion

In conclusion, coverage is not the only criterion used to determine (online) dictionary quality. Factors such as web page origin, user type, adaptation to lexicographical functions, data-accessing system, lexicographical structures, updating or interaction, are also important points to take into account in order to establish criteria for evaluation of online lexicographical products.

This paper has focused on online English-Spanish dictionaries and has presented some factors for determining the reliability of such products. The reliability of online bilingual dictionaries is seen to be based on various aspects. In particular, an online bilingual dictionary must have a real and public author; it must be oriented to the main addressee or user; it must be designed according to its lexicographical function and main objectives; it must be adequately organized with regard to hyperstructure, access structure, macrostructure, microstructure and iconic structure. At the same time, a good online bilingual dictionary should be usable, updated, hypertextual and should offer some degree of interaction.

The panorama of existing online bilingual English-Spanish dictionaries is large and, in some cases, very satisfactory. Other lexicographical products, however, are in need of modification. I hope I have suggested ways to improve these products.

## Notes

**1.** For example (http://dacco.sourceforge.net/engcat/index.php and www.muth.org/Robert/Spanish).

**2.** In Catalonia (Spain), according to article 2 of Law 34/2002, of July 11, each web page should identify its issuer by name, e-mail address and business address.

**3.** An explicit reference should be enough, as is the case with www.oftalmored.com/diccionario/ (a dictionary of ophthalmology) and www.ati.es/novatica/glosario/glosario_internet.html#electronic%20book (an Internet glossary).

**4.** Some examples can be found: www.freedict.com/onldict/spa.html and http://diccionarios.elmundo.es/diccionarios/cgi/lee_diccionario.html.

**5.** See also the www.gencat.net/justicia/llengua/termin/altres.htm dictionary of sewerage (Catalan, English and Spanish), and www.najit.org/proteus/back_issues/orrantia.htm, about legal immigration terminology.

**6.** See, for instance, www.yourdictionary.com, an index of dictionary indices, where lexicographical products are classified alphabetically, from 'abbreviations' to 'wine' (includes almost 2,000 dictionaries in more than 260 languages and more than 150 specialized fields).

**7.** Some examples (not excellent) may be found in www.leeds.ac.uk/ifcc/PD/dict/spandict.html (a glossary of clinical laboratory sciences).

**8.** See, for example, http://members.aol.com/translatrs/dictionaries/ins-ensp.html, a Spanish insurance & annuity glossary.

**9.** A good example of a search option system is www.wordreference.com, where the search option includes all entries containing a selected word.

**10.** Cf. the discussion in Gelpí 2003 and 2004, and the Web Content Accessibility Guidelines at www.w3.org/TR/WAI-WEBCONTENT.

## References

*American Heritage Spanish Dictionary:* http://mx.education.yahoo.com/reference/dict_en_es/

*Diccionario bilingüe de Refuerzo de Sonido:* www.doctorproaudio.com/doctor/diccionario.htm

*Diccionario de Organización y Representación del Conocimiento: Clasificación, Indización, Terminología*: http://eubca1.eubca.edu.uy/diccionario/diccion.htm

*Diccionario de Términos Informaticos:* www.jilt33.com/dti

*Diccionario bilingüe de Refuerzo de Sonido Profesional:* www.abcdatos.com/tutoriales/tutorial/l11092.htm

*English-Spanish Glossary of Social Security Administration Terminology:* www.ssa.gov/espanol/glossintro.html
Gelpí, Cristina. 2003. "El estado actual de la lexicografía: los nuevos diccionarios". In *Lexicografía española,* A. M. Medina Guerra (ed.), 307–328. Barcelona: Ariel.
Gelpí, Cristina. 2004. "Diccionaris digitals especialitzats per la temàtica: estat actual i perspectives" *Scripta Nova. Revista Electrónica de Geografía y Ciencias Sociales* 170 (69). www.ub.es/geocrit/sn/sn-170–69.htm
*Glosario básico inglés-español para usuarios de Internet:* www.ati.es/novatica/glointv2.html
*Glosario de Términos:* www.insflug.org/colaboracion/recursos/glosario/alpha.php3
Hausmann, Franz Josef and Herbert Ernst Wiegand. 1989. "Component Parts and Structures of General Monolingual Dictionaries: A Survey" In *Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography /Encyclopédie internationale de lexicographie,* 1, Hausmann, Franz Josef; Oskar Reichmann; Ernst Wiegand and Ladislav Zgusta (eds.), 328–360. Berlin-New York: De Gruyter.
*SICE Dictionary of Trade Terms:* www.sice.oas.org/Dictionary/GT_e.asp
*Web Content Accessibility Guidelines 1.0:* www.w3.org/TR/WAI-WEBCONTENT
*WordReference. Online French, Italian and Spanish Dictionary*: http://wordreference.com
Zgusta, Ladislav. 1971. *Manual of Lexicography*. Praha: Academia

[All Web-based references last accessed July 25, 2007]

CHAPTER 2

# Grammatische Angaben in *elexiko* und ihre Modellierung

Annette Klosa und Carolin Müller-Spitzer

The project *elexiko* compiles an extensive, monolingual dictionary of contemporary German. This contribution deals with the grammatical data in this dictionary; it is not only described how these are arranged content-wise depending on corpus data, but also how they were modelled.

Das Projekt *elexiko* erarbeitet ein umfangreiches, einsprachiges Wörterbuch des Gegenwartsdeutschen. In diesem Beitrag geht es um die grammatischen Angaben in diesem Wörterbuch; es wird nicht nur erläutert, wie diese inhaltlich in Abhängigkeit vom Prinzip der Korpusbasiertheit gestaltet sind, sondern auch, wie sie modelliert wurden.

In diesem Beitrag soll es um die grammatischen Angaben und Prinzipien ihrer Modellierung im Projekt *elexiko* des Instituts für Deutsche Sprache (IDS) in Mannheim gehen. Zugleich erörtert diese Darstellung Chancen und Probleme des zu Grunde liegenden lexikografischen Prinzips der Korpusbasiertheit. Das Projekt *elexiko* erarbeitet ein einsprachiges Wörterbuch des Gegenwartsdeutschen von perspektivisch großem Umfang. Dieses soll im Internet publiziert werden, wobei schon in der Lemmaliste online recherchiert werden kann, die 300.000 Einträge umfasst (www.elexiko.de). Außerdem sind rund 700 ausgearbeitete Wortartikel online. Die weitere Erarbeitung und Publikation erfolgt zukünftig innerhalb von Arbeitsmodulen.

## 1 Grammatische Angaben in *elexiko*

Im Bereich der grammatischen Angaben eines *elexiko*-Artikels wird die Grammatik des Stichwortes immer in seinen einzelnen Lesarten beschrieben. Die grammatischen Angaben sind dabei so konzipiert, dass der Lexikograf zunächst die Wortart des untersuchten Stichwortes bestimmen muss. In Abhängigkeit von dieser

Entscheidung sind dann einzelne Angaben zu machen. Diese sind im Vergleich zur traditionellen Lexikografie umfangreicher. So werden alle vorgesehenen grammatischen Informationen zu einem Stichwort immer angegeben, egal ob sie erwartbar und vom Nutzer selbst bildbar wären, weil sie z. B. regelmäßig sind, oder umgekehrt. Damit kommt das Projekt einer Forderung der Metalexikografie nach, die u. a. Wiegand (1985: 96) formuliert hat, nämlich "dass zur Gewinnung einer grammatischen Information möglichst wenig Zweitnachschlagegänge nötig werden" (vgl. auch Mugdan 1989: 519).

### 1.1 Chancen des Prinzips der Korpusbasiertheit

Eines der zentralen Prinzipien bei der lexikografischen Arbeit an *elexiko* ist das Prinzip der Korpusbasiertheit. Zentraler Ansatz für das Projekt ist, alle Angaben, egal ob zur Grammatik, zur Bedeutung, zur Verwendung usw., konsequent aus dem *elexiko*-Textkorpus zu gewinnen. Dies hat seinen Grund darin, dass sich das Projekt für den aktuellen und öffentlichen Sprachgebrauch interessiert, d. h. dafür, wie die Sprecher und Sprecherinnen der deutschen Sprachgemeinschaft üblicherweise miteinander kommunizieren.

Das *elexiko*-Korpus ist ein dynamisches, also ein so genanntes Monitorkorpus, welches regelmäßig erweitert und aktualisiert wird, um die jeweils neuesten Entwicklungen verfolgen und damit aktuelle Beschreibungen liefern zu können. Für die Analyse des öffentlichen Sprachgebrauchs bietet sich die Auswertung von Zeitungstexten besonders an, da diese sprachlich den Fokus der öffentlichen Diskussionen widerspiegeln. Das *elexiko*-Korpus besteht daher hauptsächlich aus Zeitungs- und Zeitschriftentexten. Diese Textsorte setzt sich nicht nur aus einer Vielzahl an verschiedenen Texttypen (Reiseberichte, Leserbriefe, Anzeigen, Auszüge aus Romanen, Amtstexte usw.) zusammen, sondern deckt darüber hinaus auch zahlreiche Themen aus Politik, Kultur, Wissenschaft, Sport, Tourismus u. a. ab und ist somit als Datengrundlage für die Zielsetzung von *elexiko* besonders gut geeignet.

Das *elexiko*-Korpus ist Teil des IDS-Korpus, das weltweit größte Textkorpus zur deutschen Schriftsprache mit bald 2 Milliarden laufenden Textwörtern. Das Korpus wird mithilfe verschiedener intelligenter automatischer Recherche- und Analysewerkzeuge, der so genannten COSMAS-Plattform, erschlossen. Dabei handelt es sich um ein komplexes Werkzeug zur Strukturierung von Massendaten, das zu vielen verschiedenen Zwecken eingesetzt werden kann, z. B. für einfache Wortsuchen oder Volltextsuchen, aber auch für chronologische Sortierungen oder zum Erkennen von Gesetzmäßigkeiten bei Wortzusammensetzungen und vielem mehr.

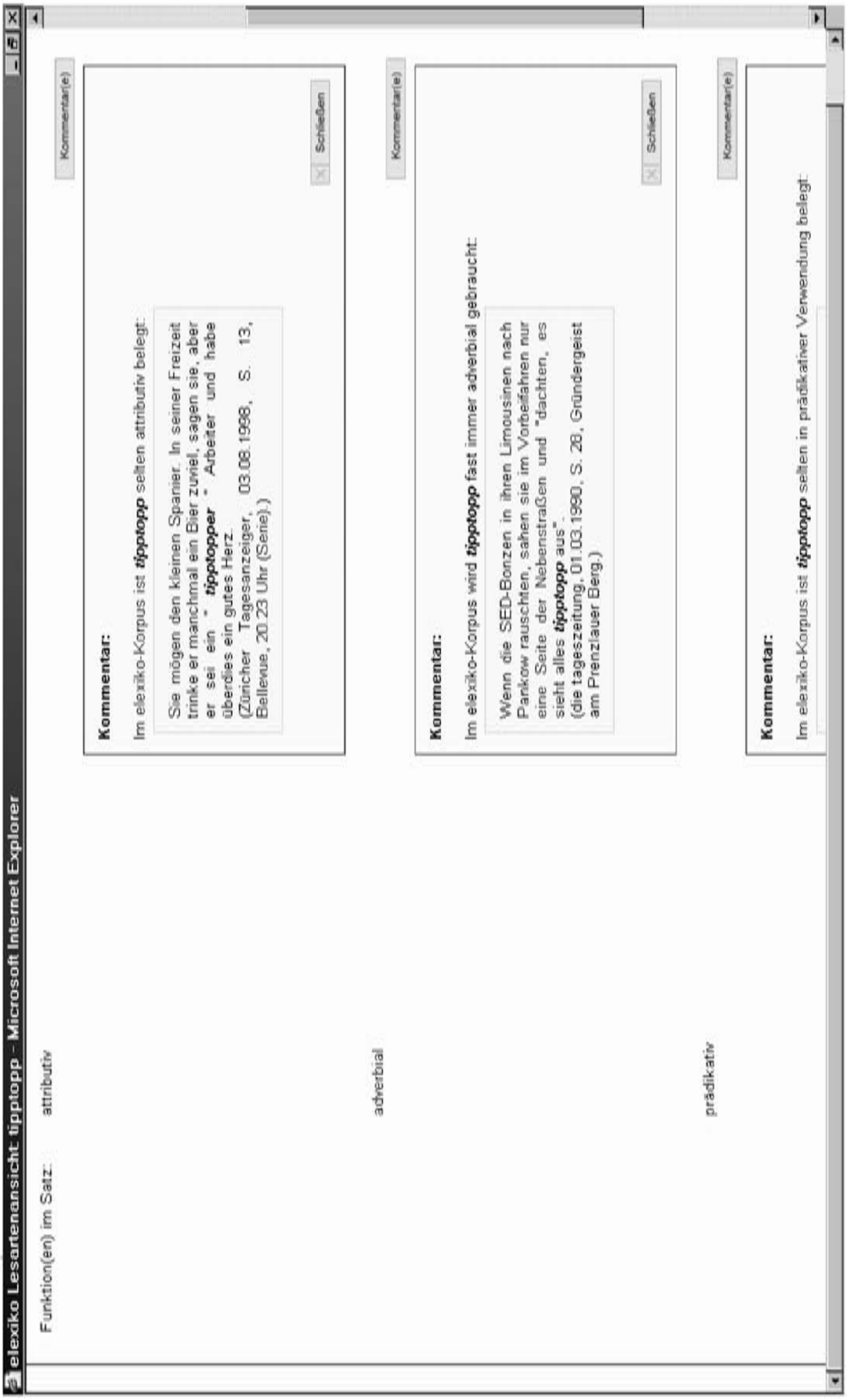


**Abbildung 1.** Artikel "tipptopp" in *elexiko*

Aus dem *elexiko*-Textkorpus sind nicht nur frequenzbasierte Angaben dazu zu gewinnen, welche Wörter überhaupt in die Lemmaliste aufgenommen werden, sondern es werden natürlich auch Information zu einem Suchwort, die in den entsprechenden Artikeln in schon publizierten Wörterbüchern noch nicht enthalten sind, ermittelt. Daneben gibt es Fälle, in denen der Korpusbefund eine Präzisierung der bereits in gedruckten Wörterbüchern enthaltenen Informationen ermöglicht. Innerhalb der grammatischen Beschreibung sind beispielsweise eine Reihe von Verwendungshäufigkeitsangaben vorgesehen, die nur mithilfe von Korpusrecherchen erarbeitet werden können. Als Beispiel sollen hier die Funktionen des Adjektivs *tipptopp* im Satz untersucht werden. Im *elexiko*-Korpus sind 26 Belege für attributives und flektiertes *tipptopp* enthalten und 376 Belege für unflektiertes und hauptsächlich adverbial gebrauchtes *tipptopp*. Im Wortartikel zu *tipptopp* wird dieses auffällige Verhältnis von etwa 1:14 zwischen attributiver und (überwiegend) adverbialer Verwendung unter den grammatischen Angaben deshalb in Form von entsprechenden Kommentaren gesondert vermerkt. Dabei werden die unterschiedlichen Verwendungshäufigkeiten mithilfe eines so genannten "standardisierten Häufigkeitswortes" nach einem Vorschlag von Bergenholtz und Mugdan (1984b: 63) markiert: "fast immer" für 90 % der Belege und mehr, "meist" für 70 % bis unter 90 % der Belege, "oft" für 50 % bis unter 70 % der Belege, "auch" für 30 % bis unter 50 % der Belege, "manchmal" für 10 % bis unter 30 % der Belege, "selten" für unter 10 % der Belege, "genauso" für genau 50 % der Belege. Entsprechend lauten die Einträge in der Online-Fassung des Artikels (vgl. Abbildung 1).

Die *elexiko*-Artikelstruktur ermöglicht es schließlich, im Bereich der Grammatik Angaben, die in traditionellen Wörterbüchern durchaus enthalten, jedoch nicht recherchierbar sind, gezielt zugreifbar zu machen. Systematisch wird beispielsweise im *elexiko*-Artikel nicht nur die Wortart angegeben, sondern es werden auch das syntaktische Verhalten bestimmter Wortarten, Valenzeigenschaften bestimmter Wortarten oder morphologische Charakteristika erfasst.

Die *elexiko*-Artikelstruktur ist in einer XML-Dokumenttypdefinition (kurz: DTD) festgehalten. Durch die Anwendung der DTD werden die Daten inhaltlich ausgezeichnet (vgl. hierzu Abschnitt 2. zur Modellierung der grammatischen Angaben). Die DTD wurde z. T. schon anhand von Musterartikeln entwickelt, damit sie den vorgesehenen lexikografischen Inhalten wirklich angemessen ist. Im Bereich der Grammatik basieren die in der DTD verankerten einzelnen Angaben daneben grundsätzlich auf (Zifonun et al. 1997), einer dreibändigen "Grammatik der deutschen Sprache", die am Institut für Deutsche Sprache erarbeitet wurde. Die *elexiko*-DTD versucht in diesem Bereich, die in dieser Grammatik enthaltenen Kategorisierungen und Beschreibungen so explizit wie möglich zu machen. So wird etwa für alle Adjektive in den Wortartikeln jeweils auf die einzelne Lesart bezogen vermerkt, ob sie steigerbar sind oder nicht, wie das Beispiel *international* zeigt (vgl. Abbildung 2).

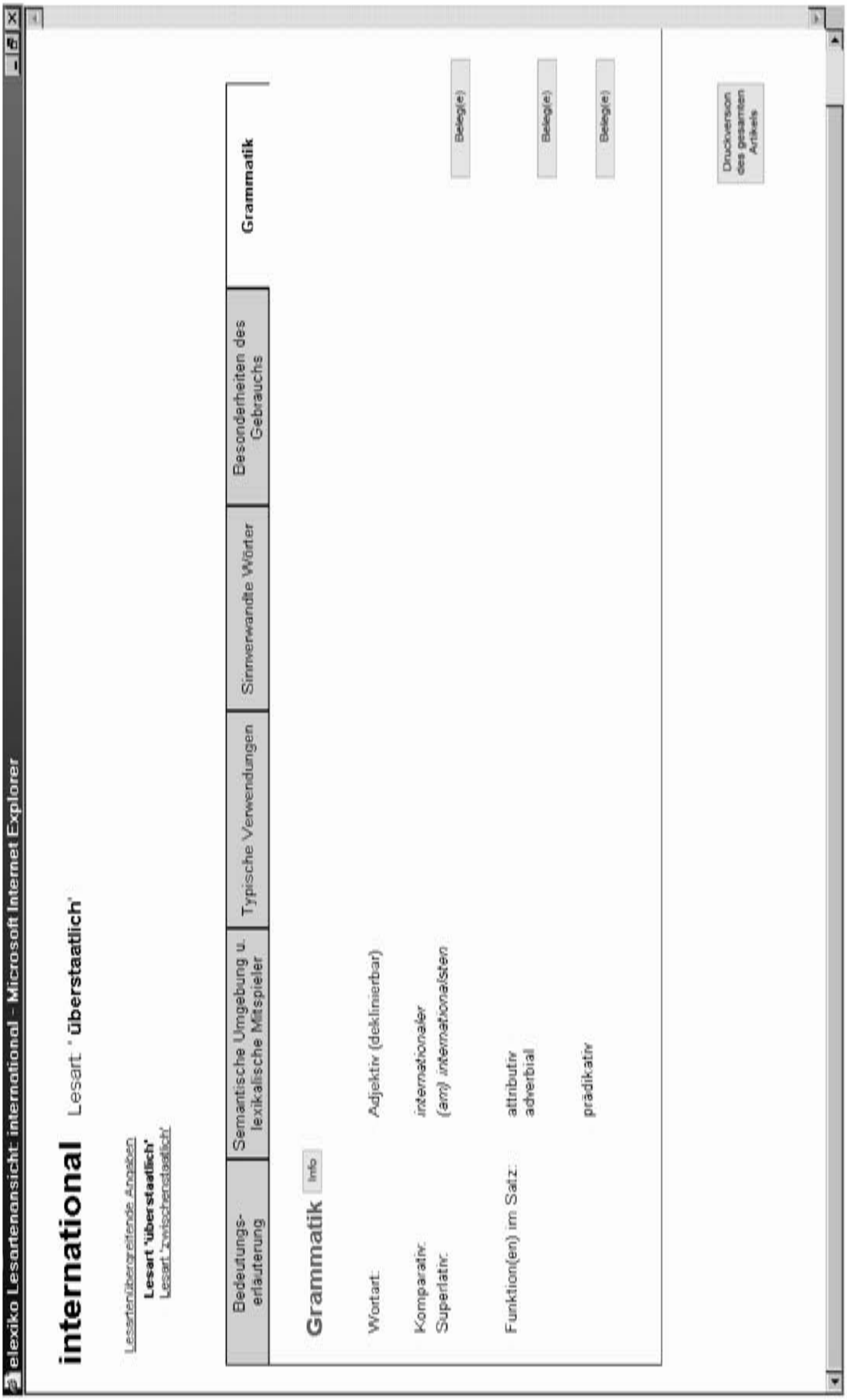


**Abbildung 2.** Artikel "international", Lesart 'überstaatlich' in *elexiko*

Der Rückbezug auf das Grammatikmodell, das der grammatischen Beschreibung im Projekt *elexiko* zu Grunde liegt, ist dadurch gewährleistet, dass die Onlinenutzer immer wieder auf die Möglichkeit hingewiesen werden, in Grammis, dem Grammatischen Informationssystem des Instituts für Deutsche Sprache, das ebenfalls über das Internet recherchierbar ist (http://hypermedia.ids-mannheim.de/grammis), nachzuschlagen. So beziehen sich unsere grammatischen Angaben explizit auf ein bestimmtes Grammatikmodell mit bestimmten Beschreibungskategorien (vgl. Bergenholtz/Mugdan 1984a: 16).

### 1.2 Probleme, die sich aus der Korpusbasiertheit ergeben

Neben den Vorteilen, die eine korpusbasierte Erarbeitung der grammatischen Angaben bietet, entstehen natürlich auch Probleme, wenn man ausschließlich mit Korpusdaten arbeitet. So muss auch das Projekt *elexiko* damit umgehen, dass sich Empirie, Norm und lexikografische Ansprüche nicht immer decken. Besonders zwei Konfliktgebiete (vgl. Farø 2005) haben sich dabei während der Erarbeitung der Wortartikel für *elexiko* als problematisch erwiesen: a) Das Korpus bietet einen Befund, welcher die aus der traditionellen Grammatik in die grundsätzliche Artikelstruktur bzw. DTD übernommene Kategorisierung oder die vorgesehenen Informationen nicht bestätigt; b) Das Korpus hat Lücken.

Als Beispiel für das Phänomen, dass Korpusbefund und angenommene grammatische Kategorisierungen nicht übereinstimmen, sollen im Folgenden bestimmte flexivische Eigenschaften der Adjektive dienen. So wird traditionell, und auch in Zifonun et al. (1997), zwischen deklinablen und indeklinablen Adjektiven unterschieden, wobei als indeklinabel gemeinhin Farbadjektive wie *rosa* gelten. Diese Unterscheidung soll nach der *elexiko*-Artikelkonzeption für jedes beschriebene Adjektiv getroffen und im Artikel festgehalten werden. Dies hat den Zweck, dass der Gesamtbestand an Artikeln später einmal nach allen deklinablen oder allen indeklinablen Adjektiven durchsucht werden kann.

Für das Adjektiv *rosa* ergibt die Recherche im *elexiko*-Korpus fast 5.000 Treffer für unflektiertes Vorkommen in attributiver Stellung *(rosa Rosen)* und knapp 20 Treffer für flektiertes Vorkommen in der gleichen Stellung *(rosane Rosen)*. Das Adjektiv *rosa* ist damit nicht wirklich indeklinabel; im Artikel wird also eingetragen "deklinabel". Darüber hinaus gibt es aber zusätzlich einen Hinweis darauf, dass *rosa* gemeinhin als indeklinabel gilt, und es sich hier also wohl um ein Normproblem handelt (vgl. Abbildung 3).

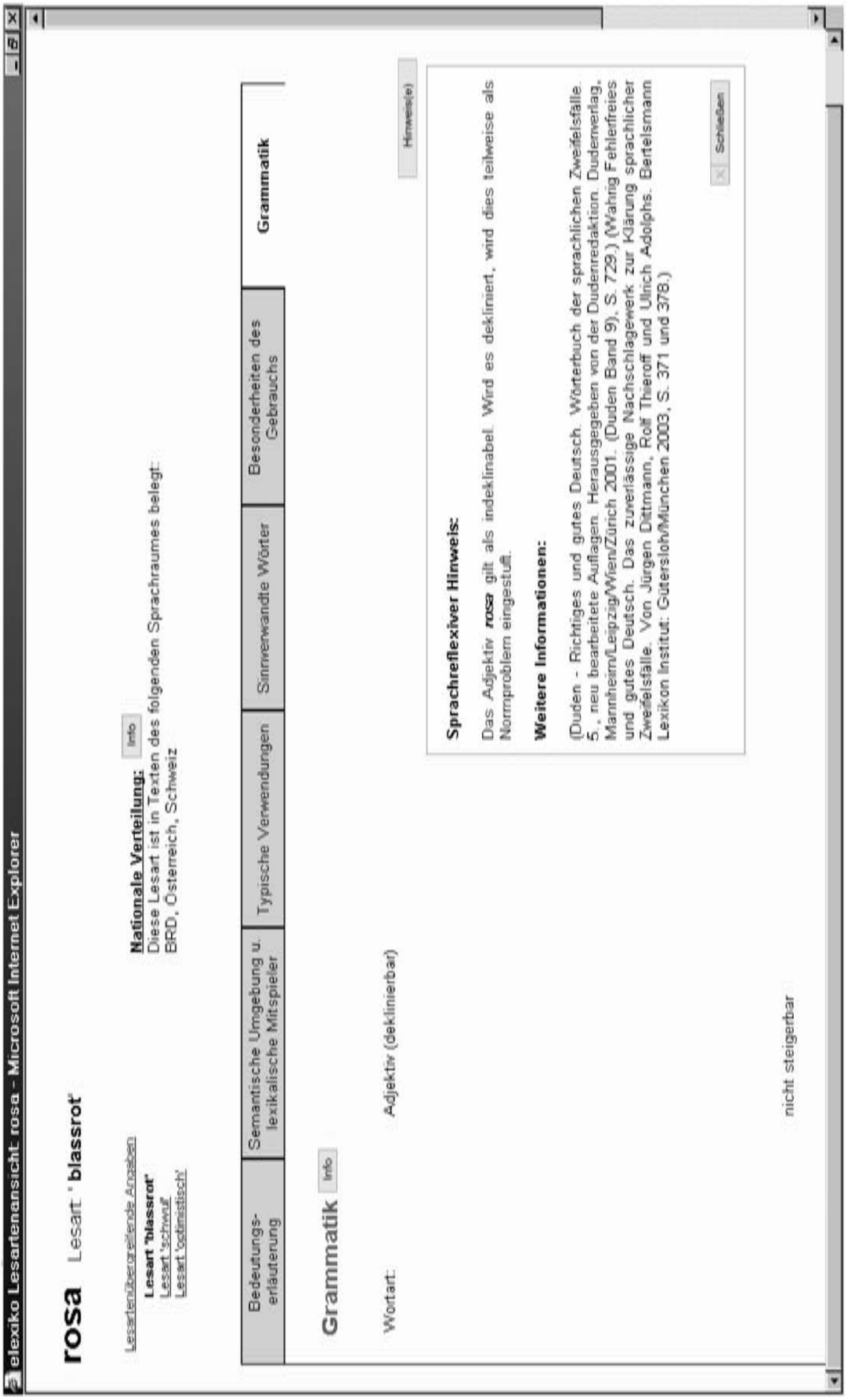


**Abbildung 3.** Artikel "rosa", Lesart 'blassrot' in *elexiko*

Werden durch die Korpusanalyse Abweichungen zu den angenommenen Kategorien oder Eigenschaften festgestellt, müssen sie in *elexiko* also erfasst werden. Dies geschieht sowohl durch eine insgesamt umfänglichere und ausführlichere Beschreibung wie auch durch die Möglichkeit, durch ergänzende Belege, Hinweise und Kommentare auf den abweichenden Korpusbefund hinzuweisen. Mit dieser korpusbasierten Beschreibung ist gegenüber der traditionellen Lexikografie der Fortschritt zu erzielen, dass traditionelle Kategorisierungen anhand eines Korpus hinterfragt werden und der tatsächliche Sprachgebrauch (im *elexiko*-Korpus) dokumentiert wird. Das Projekt *elexiko* stellt so im Idealfall im Lauf der Bearbeitungs- und Publikationszeit Material bereit, das Grammatiker später einmal benutzen können, wenn sie eine wirklich synchrone, "empirische" Grammatikschreibung anstreben. Auf der anderen Seite kann nicht innerhalb der lexikografischen Beschreibung von *elexiko* die Grammatik komplett neu geschrieben werden – im Einzelfall muss man also, nicht zuletzt aus Gründen der Praktikabilität, sicherlich abwägen, welche Befunde man im Artikel festhalten will und welche nicht, denn Massendaten so detailliert auszuwerten, wie hier vorgeführt, kostet viel Zeit.

Ein weiteres Problem für die Erfassung der grammatischen Angaben in *elexiko* ist vor folgendem Hintergrund zu sehen:

> When you analyze corpus data, you constantly find that inflectional forms given in the dictionary are in fact not attested in actual text productions, or the opposite situation of inflectional forms occurring that are not part of the official paradigm. The problem is primarily one of interpretation: is the absence of an inflectional form an indication that it does not exist, or is it an indication that the corpus is simply not large enough? (Trap-Jensen 2002)

Wie geht *elexiko* damit um, wenn das zu beschreibende grammatische Phänomen oder die gesuchte Flexionsform zu einem Stichwort gar nicht belegt ist, das *elexiko*-Korpus also Lücken hat?

Zu starken Verben werden etwa traditionellerweise in Wörterbuchartikeln bestimmte finite Formen, die Stammformen, angegeben, und das ist auch in *elexiko* so. Als Beispiel soll nun das Verb *weben* dienen; lässt man es mithilfe von COSMAS im *elexiko*-Korpus suchen, generiert der Wortformenoperator folgende Liste an Zeichenfolgen, die der Grundform "weben" zugeordnet werden (vgl. Abbildung 4).


**Abbildung 4.** Ergebnisfenster Wortformensuche zu “&weben” in COSMAS

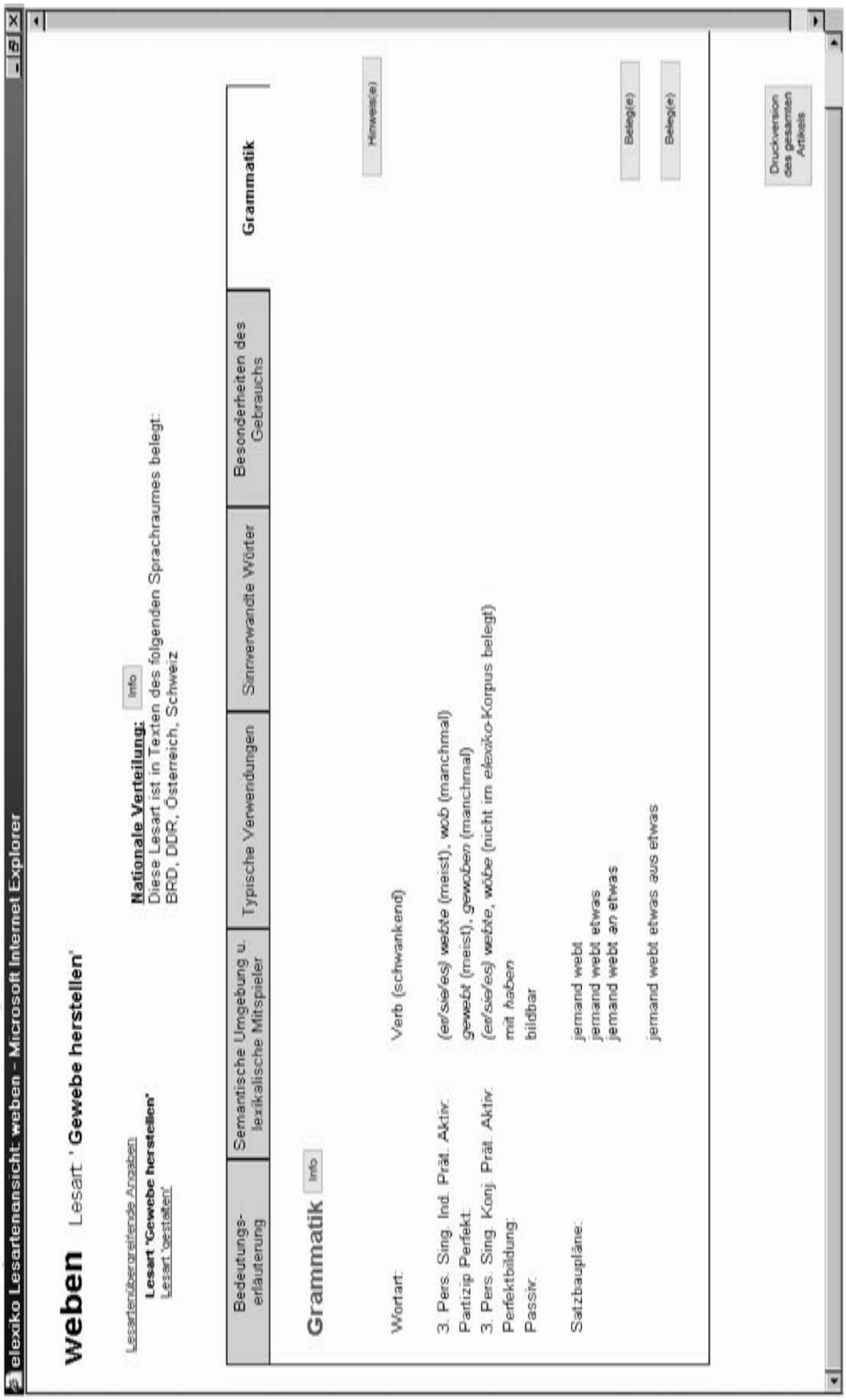


**Abbildung 5.** Artikel “weben”, Lesart ‘Gewebe herstellen’ in *elexiko*

Eine Überprüfung aller Belege zur Zeichenfolge *wöb* ergibt beispielsweise, dass es sich hierbei nicht um die Verbform der 3. Person Singular Konjunktiv Präteritum zu *weben* handelt (verkürzt um das auslautende -*e*), sondern um das Kürzel eines Journalisten der Berliner Zeitung. Somit ist im *elexiko*-Korpus diese finite Form des Flexionsparadigmas von *weben* nicht belegt. Die Struktur von Verbartikeln in *elexiko* sieht aber vor, dass mehrere finite Formen verzeichnet werden, und zwar immer die 3. Person Singular Indikativ Präteritum und das Partizip II; daneben die 3. Person Singular Konjunktiv Präteritum, falls der Stammvokal im Konjunktiv Präteritum nicht mit dem im Indikativ Präteritum übereinstimmt, und schließlich bei Verben mit Stammwechsel im Präsens die entsprechenden Formen. Strenge Korpusbasiertheit würde nun bedeuten, dass nur die Formen im Artikel vermerkt werden, die im Korpus belegt sind. In letzter Konsequenz bedeutet das, dass man gar nicht von Lücken im Korpus sprechen dürfte. Dieses Vorgehen hat aber den Nachteil, dass eine für viele potenzielle Nutzer des Wörterbuches wichtige Information nicht bereitgestellt wird. Das Projekt *elexiko* hat sich deshalb dafür entschieden, u. a. eine sozusagen "fehlende" Form in Verbparadigmen kompetenzgestützt nachzutragen, aber zugleich zu vermerken, dass sie im *elexiko*-Korpus nicht belegt ist (vgl. Abbildung 5).

Auf diese Weise soll sowohl dem Anspruch der Korpusbasiertheit wie den Interessen der Nutzer Rechnung getragen werden, ohne dass *elexiko* zugleich präskriptiv wird.

## 2 Modellierung in *elexiko*

Im ersten Teil dieses Aufsatzes wurden bestimmte Aspekte aus dem Bereich der grammatischen Angaben aus *elexiko* erläutert. Dabei wurden auch Ausschnitte aus fertigen Artikeln gezeigt, wie sie allen Benutzern im Internet zugänglich sind. In diesem Zusammenhang war die Rede davon, dass alle Angaben in unseren Artikeln explizit markiert sind und den potenziellen Benutzern so für eine gezielte Recherche zur Verfügung stehen sollen. Im zweiten Teil dieses Aufsatzes soll es nun darum gehen, den Projektbereich genauer zu erläutern, in dem die wesentlichen Voraussetzungen für diese späteren Möglichkeiten der Darstellung und der Recherche der lexikografischen Daten geschaffen werden. Dieser Bereich ist für die Wörterbuchbenutzer nicht sichtbar, ist aber wesentlich für die Lexikografen beim Erarbeiten der Artikel und schafft eben die Voraussetzungen für die Möglichkeiten der Publikation der Artikel. Es geht demnach hier mehr um die texttechnologischen Aspekte von *elexiko*.

### 2.1 Anforderungen an die Modellierung lexikografischer Daten

Jedes Projekt, bei dem es um die Erfassung und Verarbeitung von Textdaten geht, steht heutzutage vor Fragen wie:

- In welcher Weise sollen die Textdaten erfasst werden?
- In welcher Form werden die Textdaten gespeichert?
- Mit welchen Hilfsmitteln wird das Textmaterial für die weitere Arbeit oder Veröffentlichung erschlossen? (Vgl. Büchel/Schröder 2001: 8)

In lexikografischen Projekten ist der dritte Punkt – der Zugriff auf die Daten – besonders zentral, da die Realisierung der Zugriffsmöglichkeiten ein wichtiges Unterscheidungskriterium für gedruckte vs. elektronische Wörterbücher darstellt. "From a user's perspective, the most innovative aspect of EDs is probably the retrieval system" (Schryver 2003: 146). Dabei geht es nicht allein um die Formulierung von bestimmten Suchanfragen, sondern auch um die flexible Präsentation des Suchergebnisses. Daher wird gefordert, abhängig von Benutzungssituationen aus einem Wörterbuchartikel nur die gesuchten Angaben anzeigen zu können. "It is one thing to be able to store ever more data, but another thing entirely to present just the data users want to response to a particular look-up." (Schryver 2003: 178, ähnlich Klosa 2001: 97) "Die Benutzerschnittstelle von digitalen Wörterbüchern" sollte daher "an Typen von Benutzungssituationen adaptierbar sein." (Storrer 2001: 64) Die Voraussetzungen für diese flexiblen Zugriffsmöglichkeiten müssen bei der Modellierung geschaffen werden. Denn die Art der Modellierung und der Datenaufbereitung entscheidet darüber, wie gezielt und genau auf diese Daten später zugegriffen werden kann. "What you mark is what you get" (Gloning/Welter 2001: 128), so formulieren es Gloning/Welter bezogen auf eine XML-basierte Modellierung. Neben den gezielten Zugriffsmöglichkeiten sind folgende Anforderungen an die Modellierung lexikografischer Daten noch zusammenfassend zu nennen: Daten aus unterschiedlichen Medien müssen in eine Datenbasis integriert werden können und lexikografische Daten müssen an externe Ressourcen wie z. B. Textkorpora angebunden werden können.

Die bis jetzt genannten Anforderungen an ein Modellierungskonzept beziehen sich vor allem auf die Funktionalität elektronischer Wörterbücher. Ein weiterer wichtiger Bereich ist die Computerunterstützung im lexikografischen Prozess. Auch für die Lexikografen ist dabei ein gezielter Zugriff auf die lexikografischen Daten wichtig, um z. B. leichter und verlässlicher Konsistenzprüfungen der Beschreibungssprache einzelner Angaben vornehmen zu können. Daneben sollten die Lexikografen in der formalen Einhaltung der Artikelstruktur unterstützt werden. So fordert Wiegand, dass die Verantwortung für die Einhaltung des Artikelformats an ein Computerprogramm abgetreten werden könne, da dies bei anspruchsvollen Wörterbüchern eine Entlastung für den Lexikografen bedeute (Wiegand 1998: 232). Traditionell werden Schreibanweisungen für die Artikelarbeit in Instruktionsbüchern oder Manualen festgehalten. Das Problem in der lexikografischen Praxis ist jedoch oft, dass besonders eine komplexe Artikelstruktur ohne umfangreiche technische Unterstützung kaum einzuhalten ist.

> Konventionen, die die standardisierten Wörterbuchartikel erfüllen müssen, werden normalerweise in Instruktionsbüchern festgehalten. Das Einhalten dieser Konventionen wird manuell und oft nur stichprobenartig überprüft. In Anbetracht der hochkomplexen Textform 'Wörterbuchartikel' ist eine redaktionelle Betreuung eines Wörterbuchs mit mehr als 60000 Artikeln allerdings sehr schwierig und letztendlich sind inhaltliche und formale Mängel kaum auszuschließen. (Heyn 1992: 187)

Außerdem sollten lexikografische Angaben ohne Strukturanzeiger, d. h. ohne Kommata, typografische Auszeichnungen etc. eingegeben werden können. Diese Unterstützungen können eine wesentliche Erleichterung für die lexikografische Arbeit darstellen.

Im Zusammenhang mit der Modellierung lexikografischer Daten ist auch eine mögliche Schnittstelle zur automatischen Sprachverarbeitung zu beachten. Denn das Zusammenwirken von Lexikografie und automatischer Sprachverarbeitung kann sehr konstruktiv sein.

> NLP needs dictionaries, und dictionary-makers can use NLP to make better dictionaries, so there is a great potential for synergy between the two activities. [...] To date, NLP research has used dictionaries und dictionaries have used NLP, but the two processes have not been brought together. (Kilgarriff 2000: 105)

Auf das genaue Zusammenwirken kann in diesem Zusammenhang nicht weiter eingegangen werden. Jedoch ist im Sinne einer konstruktiven Schnittstelle für das Modellierungskonzept festzuhalten, dass es hinreichend streng sein sollte, sodass die lexikografischen Daten gut automatisch weiterverarbeitet werden können und dass der inhaltliche Gehalt der Daten so maschinenlesbar transparent gemacht werden sollte, dass er für andere Anwendungen ausgenutzt werden kann.

Wie bereits ausgeführt, sollen die Daten ohne Strukturanzeiger, also ohne typografische Auszeichnungen etc. eingegeben werden können; gleichzeitig soll der inhaltliche Gehalt der Daten kodiert werden. Eine Trennung von Inhalt und Layout ist also eine der zentralen Anforderungen an die Modellierung.

## 2.2 Das Modellierungskonzept und seine Anwendung in *elexiko*

Die Modellierung erfolgt in *elexiko* XML-basiert. XML (eXtensible Markup Language) ist eine international festgelegte Syntax, eine Metasprache, die die

Entwicklung von Auszeichnungssprachen zur Beschreibung von Daten hinsichtlich ihrer hierarchischen Struktur und ihrer inhaltlichen Einheiten ermöglicht. (XML Standard [Third Edition]) Die Wahl von XML und verwandten Standards als Modellierungssprachen ist hier in verschiedener Hinsicht sinnvoll: Eines der wichtigsten Ziele von XML bzw. des Vorläufers SGML (Standard Generalized Markup Language) ist gerade die Trennung von Inhalt und Lay-out der Daten. Deshalb ist XML bzw. SGML schon früh für Verlage interessant gewesen:

> The use of a single, internationally agreed syntax was not the only reason why a generic markup language was the appropriate solution für publishers that wanted to publish to multiple media. A fundamental concept in the use of generic markup languages is a move from formatting markup to content markup. (Gennusa 1999: 28)

Außerdem bietet XML als weltweit anerkannter Industriestandard eine bestmögliche Garantie dafür, dass die Daten langfristig weiterverwendet werden können. Eine XML-basierte Modellierung erfolgt in Form einer DTD (Document Type Definition) oder einem XML-Schema (das Für und Wider dieser beiden Formen ist hier nicht zu diskutieren). In einer solchen XML-DTD oder einem XML-Schema werden die Regeln festgelegt, der alle dazugehörigen Instanzen, d. h. in lexikografischen Projekten meist alle Artikel, zu gehorchen haben. Dies kann über Parsing-Prozeduren in XML-basierter Software geprüft werden. Neben vielen anderen Punkten, die für die Wahl von XML sprechen, wie z. B. die Verfügbarkeit von Software, ist daher v. a. wichtig, dass der Einsatz einer XML-basierten Modellierung eine gute Strukturführung der Lexikografen beim Erarbeiten der Artikel ermöglicht.

Allein mit dem Einsatz von XML ist das Modellierungskonzept jedoch noch kaum beschrieben, da XML auf viele Weisen eingesetzt werden kann. Daher muss genauer spezifiziert werden, wie die Modellierung erfolgen soll. Modellierungskonzepte unterscheiden sich von ihrer Intention her u. a. darin, ob sie allgemeine Richtlinien für eine Modellierung festlegen, die dann maßgeschneidert für ein Projekt umgesetzt werden, oder einen konkreten Modellierungsvorschlag entwickeln, der für möglichst viele Projekte anzuwenden ist, also eine maßgeschneiderte vs. eine Standard-Modellierung. Der bekannteste Vorschlag für eine Standard-Modellierung ist die Wörterbuchstruktur der TEI (Text Encoding Initiative), einen neuen Vorschlag dazu gibt Franziskus Geeb mit leXeML (Geeb 2001). Ein Vorteil dieser Initiativen oder Vorschläge besteht darin, dass keine eigene Modellierung im Projekt entwickelt werden muss. Nachteilig ist jedoch, dass entweder projektspezifisch starke Anpassungen vorgenommen werden müssen oder die Struktur sehr allgemein und – im Fall der TEI – sehr weich ist (s. u. a. Schmidt/Müller 2001: 37ff.). Eine solche Struktur bietet daher meist nicht in dem Maße eine Unterstützung im lexikografischen Prozess, wie das eine maßgeschneiderte Modellierung kann.

In *elexiko* wurde angestrebt, dass das Inhaltsstrukturenprogramm so genau wie möglich in der Modellierung abgebildet werden sollte, um den Lexikografen eine bestmögliche Strukturführung zu bieten. Ein Grund ist, dass voraussichtlich über einen längeren Zeitraum hinweg in verschiedenen personellen Zusammensetzungen an Wörterbuchartikeln gearbeitet wird, sodass eine möglichst umfängliche automatische Konsistenzkontrolle der Daten sehr wichtig ist. Für *elexiko* wurde daher eine maßgeschneiderte Modellierung entwickelt. Erklärtes Ziel dieses Modellierungskonzeptes ist es, den inhaltlichen Gehalt der lexikografischen Daten und damit auch den genuinen Zweck (i. S. v. Wiegand 1989: 426), weshalb sie von den Lexikografen für die potenziellen Benutzer angesetzt wurden, so genau wie möglich zu kodieren und transparent zu machen. Unsere Hypothese ist dabei, dass potenziell jede angesetzte Angabe für einen gezielten Zugriff durch potenzielle Benutzer interessant sein kann. Dieser genuine Zweck liefert im Rahmen dieses Modellierungskonzeptes auch die Benennung für die jeweilige Klasse von Daten. Die Auszeichnung der lexikografischen Daten ist damit sehr granular, wodurch sie flexibel und gezielt zugreifbar werden. Die Modellierung ist darüber hinaus möglichst streng, d. h. in ihr wird so genau wie möglich festgelegt, welche Angaben in welcher Reihenfolge in den *elexiko*-Artikeln zu erarbeiten sind. Diese Reihenfolge bezieht sich allerdings nur auf die Erarbeitung der Daten und noch nicht auf ihre Präsentation. Diese strenge Modellierung bietet die geforderte Unterstützung im lexikografischen Prozess, da formal geprüft werden kann, ob die Artikelstruktur eingehalten wurde oder nicht.

## 2.3 Erarbeiten von Artikeln

Dies soll kurz demonstriert werden am Beispiel des Artikels *international*, der schon im ersten Teil des Aufsatzes erwähnt wurde. Abbildung 6 zeigt einen Screenshot des Artikels *international* im XMetaL-Editor, einem sehr weit verbreiteten Editor zum Erstellen von XML-Dateien. In einem solchen Editor wird beim Erstellen eines Artikels die zugehörige DTD geladen, in der die Regeln für die Artikelerstellung festgelegt sind.

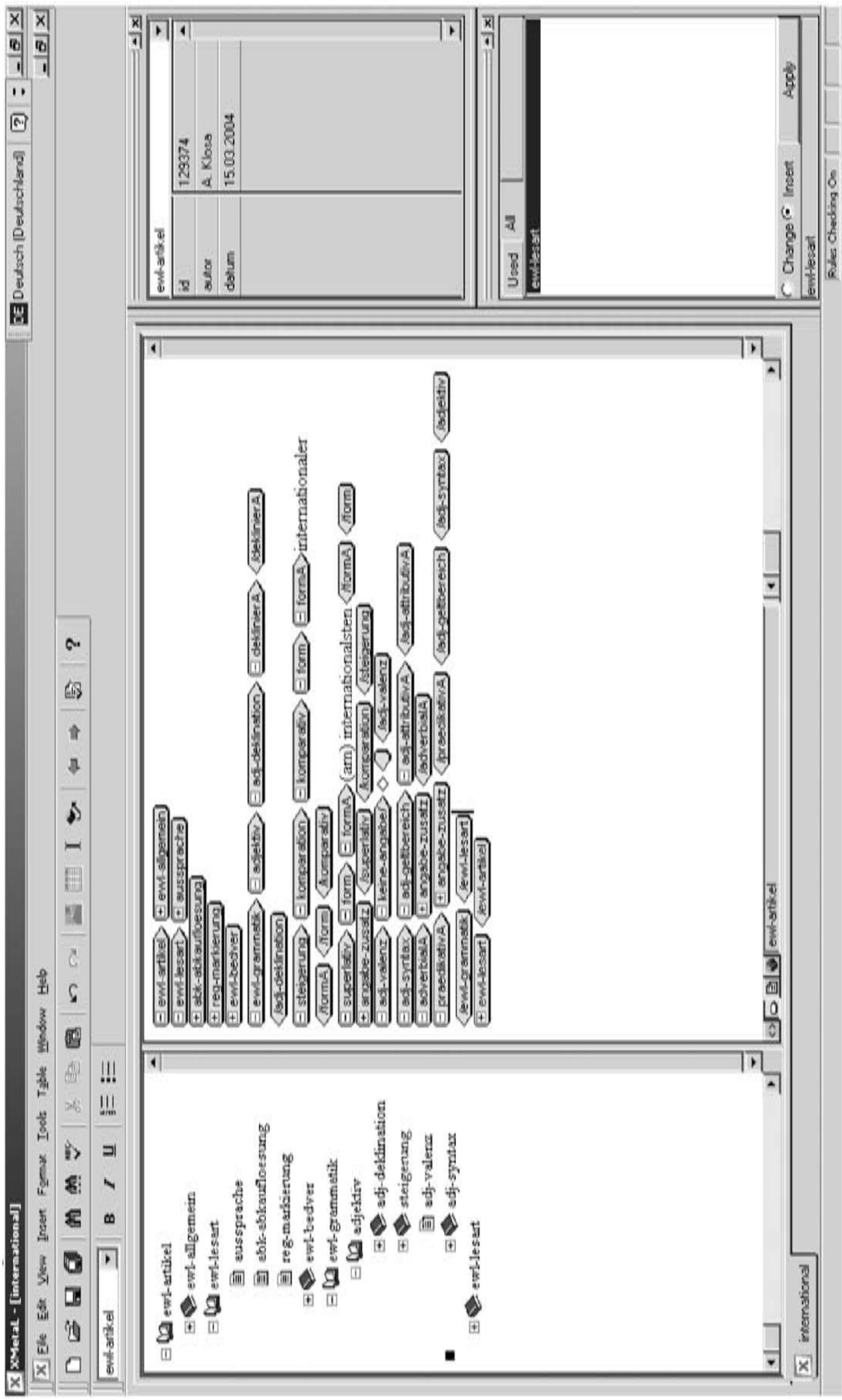

**Abbildung 6.** Bildschirmansicht des Artikels "international" im XMetaL-Editor

Auf der linken Seite des Bildschirms sieht man einen Strukturüberblick über die Teile des Artikels. Es führt in diesem Zusammenhang zu weit, diese zu Grunde liegende DTD-Struktur detailliert zu erläutern. Für einen kurzen Überblick nur so viel: Es handelt sich hier um die Struktur für ein Einwortlemma, daneben gibt es in *elexiko* Mehrwortlemmata wie "toter Hund" und Wortelementlemmata wie "-lich" oder "-heit". Die Angaben zu einem ewl-artikel, d. h. einem Einwortlemma-Artikel, sind in zwei große Gruppen aufgeteilt: in lesartenübergreifende Angaben, die unter dem Element `ewl-allgemein` gefasst werden, und lesartenbezogene Angaben, die mit dem Element `ewl-lesart` ausgezeichnet werden. In *elexiko* wird auch die Grammatik lesartenbezogen angegeben; deshalb ist das Element `ewl-grammatik` unter den lesartenbezogenen Angaben angeordnet. Innerhalb von `ewl-grammatik` sieht man hier die einzelnen Angabegruppen, die zu Adjektiven gegeben werden und die zum Teil schon im ersten Teil dieses Artikels erläutert wurden: Angaben zur Deklinierbarkeit, zur Steigerung, zur Valenz und zur Syntax. In der Mitte sieht man den Artikel selbst, auch hier kann man Teile des Artikels ein- und ausblenden. Rechts oben sieht man Attribute, die Elementen zugeordnet werden, und rechts unten nach den Regeln der DTD auszuwählende Elemente.

Nun wurde schon mehrfach herausgestellt, dass die Lexikografen beim Eingeben der Artikel durch die Modellierung in dem Sinne unterstützt würden, dass die formale Einhaltung der Artikelstruktur geprüft wird. Dies funktioniert in der *elexiko*-Umgebung folgendermaßen: Die Lexikografen bekommen an der jeweiligen Stelle im Artikel, an dem sie gerade arbeiten, nur die Angaben im Editor angeboten, die durch die DTD an der jeweiligen Stelle erlaubt sind. Will ein Lexikograf beispielsweise die Grammatik zu *international* in der Lesart 'überstaatlich' im Artikel ausfüllen, muss er zunächst die Wortart bestimmen, hier also `adjektiv`. Abhängig von der Wortart öffnen sich dann jeweils unterschiedliche Angabeklassen. Bei Adjektiven sind es Angaben zur Deklinierbarkeit, zur Steigerung, zur Valenz und zur Syntax. Unter der `deklinierA` muss beispielsweise über ein Attribut ausgewählt werden, ob *international* in der Lesart 'überstaatlich' deklinierbar ist oder nicht. Als Demonstration nehmen wir an, dass der bearbeitende Lexikograf vergisst, den passenden Attributwert auszuwählen. Daraufhin muss er etwas zur Steigerung aussagen: entweder er kann `keine-angabe` geben oder er wählt das Element aus, falls das Adjektiv steigerbar ist. Wählt er das Element `komparation` aus, muss er den Komparativ und den Superlativ angeben. Zur Valenz gibt es im Fall von *international* `keine-angabe`. Unter Syntax werden die im ersten Teil beschriebenen Angaben zum Geltungsbereich gegeben. Man wählt also das Element `adj-geltbereich` aus und sieht rechts unten die Auswahl von attributiv, adverbial und prädikativ. Im Falle von *international* sind alle drei Möglichkeiten des Geltungsbereiches nacheinander auszuwählen. Mit dieser Auswahl sind die Angaben zum Geltungsbereich gemacht.

An diesem Beispiel kann man auch erkennen, dass in der Modellierung unterschieden wird zwischen absolut obligatorischen, relativ obligatorischen und fakultativen Angaben. Diese Unterscheidung geht zurück auf Wiegand (1989). Absolut obligatorische Angaben sind dabei solche Angaben, die immer gegeben werden sollen und auch gegeben werden können, wie z. B. eine Lemmazeichengestaltangabe, oder hier bei Adjektiven die Angabe zur Deklinierbarkeit. Relativ obligatorische Angaben sind dagegen Angaben, die nicht zu jedem Lemmazeichen gegeben werden können, wie z. B. Komplemente bei Adjektiven, die aber stets angegeben werden sollen, wenn es möglich ist. In der Modellierung werden diese relativ obligatorischen Angaben von rein fakultativen Angaben folgendermaßen unterschieden: Die Angabe selbst wird in der DTD, d. h. im ‘Regelwerk’ für die Erstellung der Artikel, – wie die absolut obligatorischen Angaben – als obligatorisch definiert. Innerhalb dieser Angabe oder Angabegruppe gibt es allerdings ein Ausweichelement `keine-angabe`. Damit wird deutlich gemacht, dass diese Angabe oder Angabegruppe immer bearbeitet werden soll und Lexikografen Stellung dazu nehmen müssen; anders als dies bei fakultativen Angaben der Fall ist. Mit dieser Modellierung soll auch gewährleistet werden, dass diese Angaben nicht versehentlich vergessen werden.

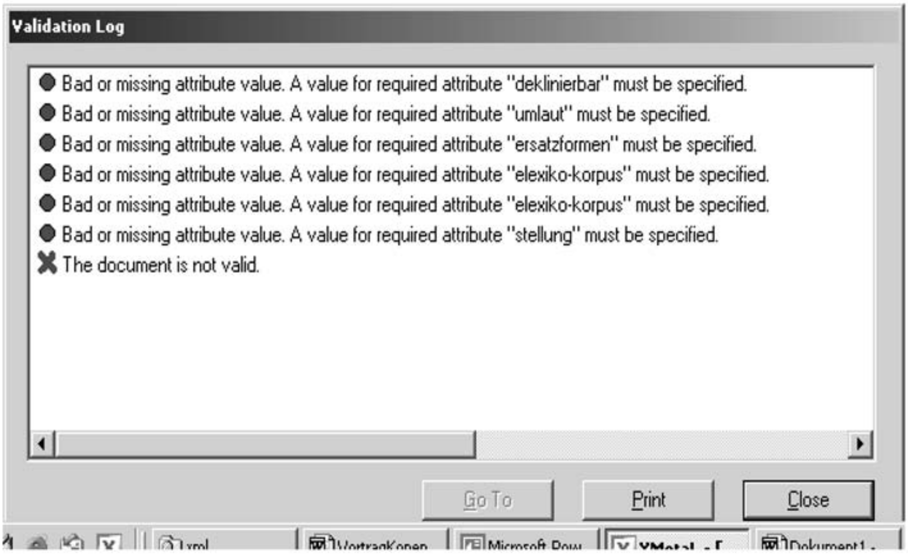


**Abbildung 7.** Fehlermeldung beim Validieren des Artikels

Ist der Lexikograf mit seinen Angaben zur Grammatik fertig und will den Artikel abschließen, muss er den Artikel zunächst validieren. Das bedeutet, dass der eingegebene Artikel gegenüber der in der DTD festgelegten Struktur geprüft wird. Da die Modellierung in *elexiko* sehr genau und streng ist, kann an dieser Stelle geprüft werden, ob der Lexikograf die festgelegte Artikelstruktur eingehalten hat oder nicht. Führt er diese Validierung im eben erläuterten Artikel durch, bekommt er die in Abbildung 7 gezeigten Fehlermeldungen.

Zunächst muss er also den Attributwert zur Angabe der Deklinierbarkeit ausfüllen, die er vorhin vergessen hatte. Dann wurde übersehen, dass zum Element `komparation` Attribute zu `umlaut` und `ersatzformen` auszufüllen sind. Diese sind im Falle von *international* beide auf "nein" zu setzen, weil die Steigerungsformen regelmäßig sind, also ohne Umlaut und Ersatzformen gebildet werden. Außerdem wurde bei den Steigerungsformen noch nicht vermerkt, ob sie im *elexiko*-Korpus belegt sind oder nicht. Dies muss aber in *elexiko* – wie vorhin ausgeführt – zu allen Formen gesagt werden. Die Steigerungsformen von *international* sind in unserem Korpus belegt, also werden die Attributwerte auf "ja" gesetzt. Die letzte Fehlermeldung besagt schließlich, dass für den Geltungsbereich `attributiv` etwas zur Stellung ausgesagt werden muss: für *international* ist dabei `praenominal` auszuwählen. Validiert der Lexikograf den Artikel nun erneut, ist die Prüfung erfolgreich. So werden die Lexikografen unserer Erfahrung nach sehr gut in der formalen Einhaltung der Artikelstruktur unterstützt.

## 2.4 Flexibilität der Darstellung

Es wurde immer wieder herausgestellt, dass diese granulare Modellierung dazu dienen soll, die Daten der lexikografischen Datenbasis auf verschiedene Weise im elektronischen oder gedruckten Wörterbuch präsentieren zu können. Zur Darstellung der in XML vorliegenden lexikografischen Daten wird die Extensible Styling Language, kurz XSL, eingesetzt. Mit ihr lassen sich so genannte Stylesheets erstellen, in denen die Darstellung der Daten festgelegt wird. Im Prinzip kann dabei für jedes Element und jedes Attribut spezifiziert werden, wie es den potenziellen Benutzern dargestellt werden soll. Der Vorteil ist dabei, dass separat zu den eigentlichen Daten die Darstellung der Daten spezifiziert wird. Je granularer dabei die Datenauszeichnung ist, desto spezifischer kann demgemäß die Präsentation definiert werden. Außerdem können mehrere Stylesheets für die gleiche Datenbasis erstellt werden, sodass die gleiche Datenbasis in verschiedener Weise präsentiert werden kann. Zum Beispiel könnte ein Stylesheet für ein elektronisches Wörterbuch und eines für ein gedrucktes Wörterbuch erstellt werden. Darüber hinaus ist es möglich und auch sinnvoll, dass für ein elektronisches Wörterbuch mehrere Stylesheets erstellt werden, die adaptiv zu Benutzungssituationen eingesetzt werden.

Dieses Prinzip soll hier kurz an einem Beispiel erläutert werden, und zwar am Artikel *international* anhand der grammatischen Angaben in der Lesart 'überstaatlich'. Im Moment wird in *elexiko* nur ein Stylesheet verwendet, in dem die Darstellung aller Angaben im Artikel festgelegt wird. Die Darstellung der grammatischen Angaben im Artikel *international* in der Lesart 'überstaatlich' ist in Abbildung 2 gezeigt. Die gleichen Daten könnten aber auch durch eine Veränderung des Stylesheets in anderer Weise präsentiert werden. Für DaF-Nutzer könnte man z. B. festlegen, dass die Angaben wie in Abbildung 8 präsentiert werden. Hier wird die Angabe zur Steigerung verständlicher ausformuliert, und zu den Angaben zum Geltungsbereich sind die entsprechenden Belege standardmäßig geöffnet, um die Angaben besser zu verdeutlichen.

Für Experten hingegen könnte man ein weiteres Stylesheet einsetzen, in dem die Präsentation wie in Abbildung 9 gezeigt spezifiziert wird. Hier werden nur kurz die grundlegenden Eigenschaften in anderer Terminologie ("deklinabel, komparabel") genannt und die beobachteten syntaktischen Funktionen aufgezählt. Belege werden hier nicht mit angeboten. Diese Darstellungsmöglichkeiten sind für die genannten Benutzergruppen nicht zu Ende gedacht; sie sollen hier mehr der Demonstration dienen.

So können also für unterschiedliche Benutzergruppen in unterschiedlichen Benutzungssituationen unterschiedliche Darstellungen festgelegt werden, die alle auf die gleiche lexikografische Datenbasis zurückgreifen. Diese unterschiedlichen Darstellungen können ohne Veränderung der zu Grunde liegenden Datenbasis erreicht werden. Die Kombination von granular inhaltlich ausgezeichneten Daten und deren separat spezifizierte Darstellung ermöglicht daher eine sehr weit reichende Flexibilität der Darstellung. Denn so ist es möglich, abhängig von Benutzungssituationen und abhängig von bestimmten Suchanfragen nur eine Teilmenge von Daten anzeigen zu lassen, die Daten unter terminologisch anderen Überschriften zu präsentieren etc. Auch ist es möglich, für die Darstellung im gedruckten Wörterbuch die typografische Realisierung und das Generieren von Strukturanzeigern auf diese Weise festzulegen.

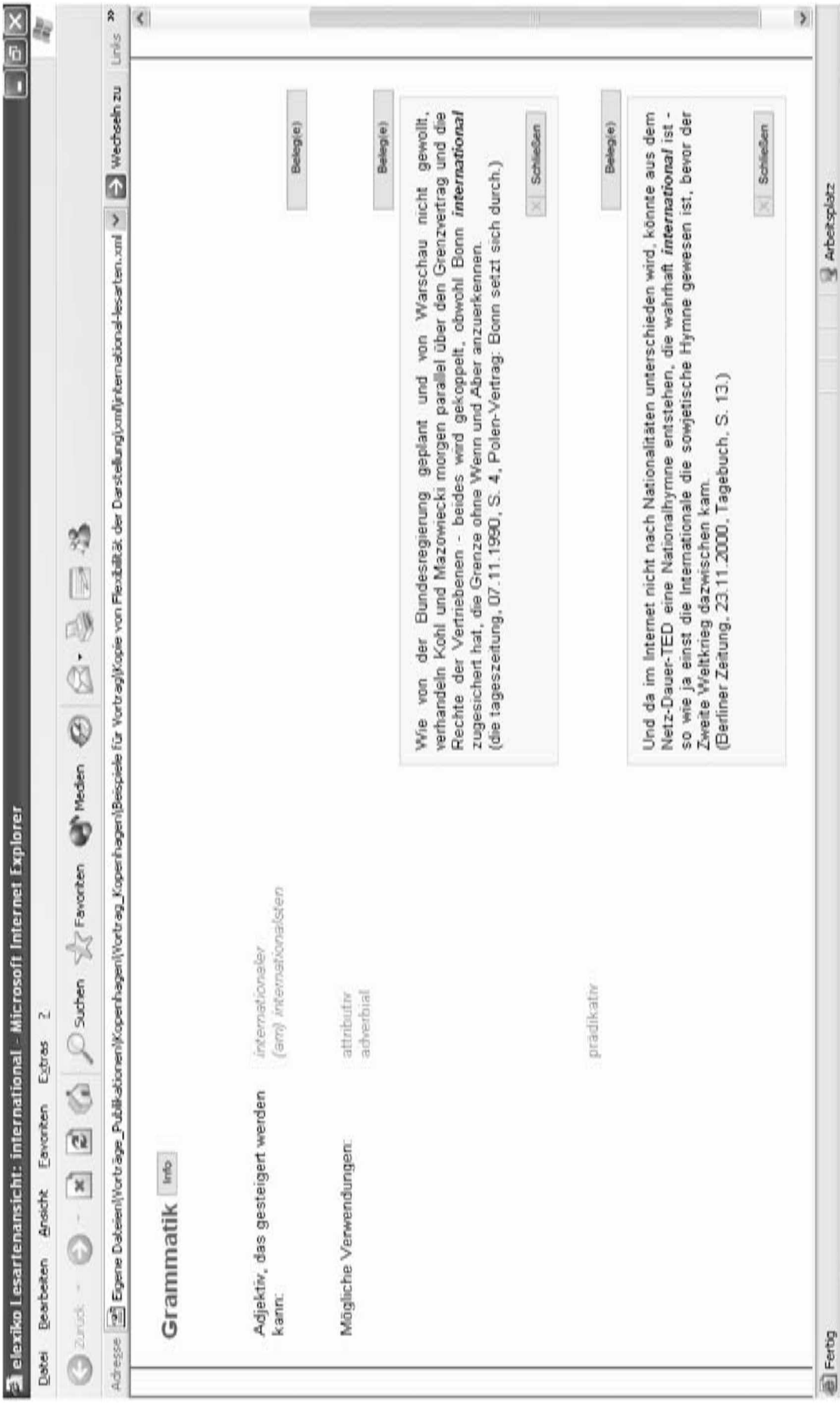


**Abbildung 8.** Mögliche veränderte Darstellung der grammatischen Angaben im Artikel "international" für DaF-Nutzer

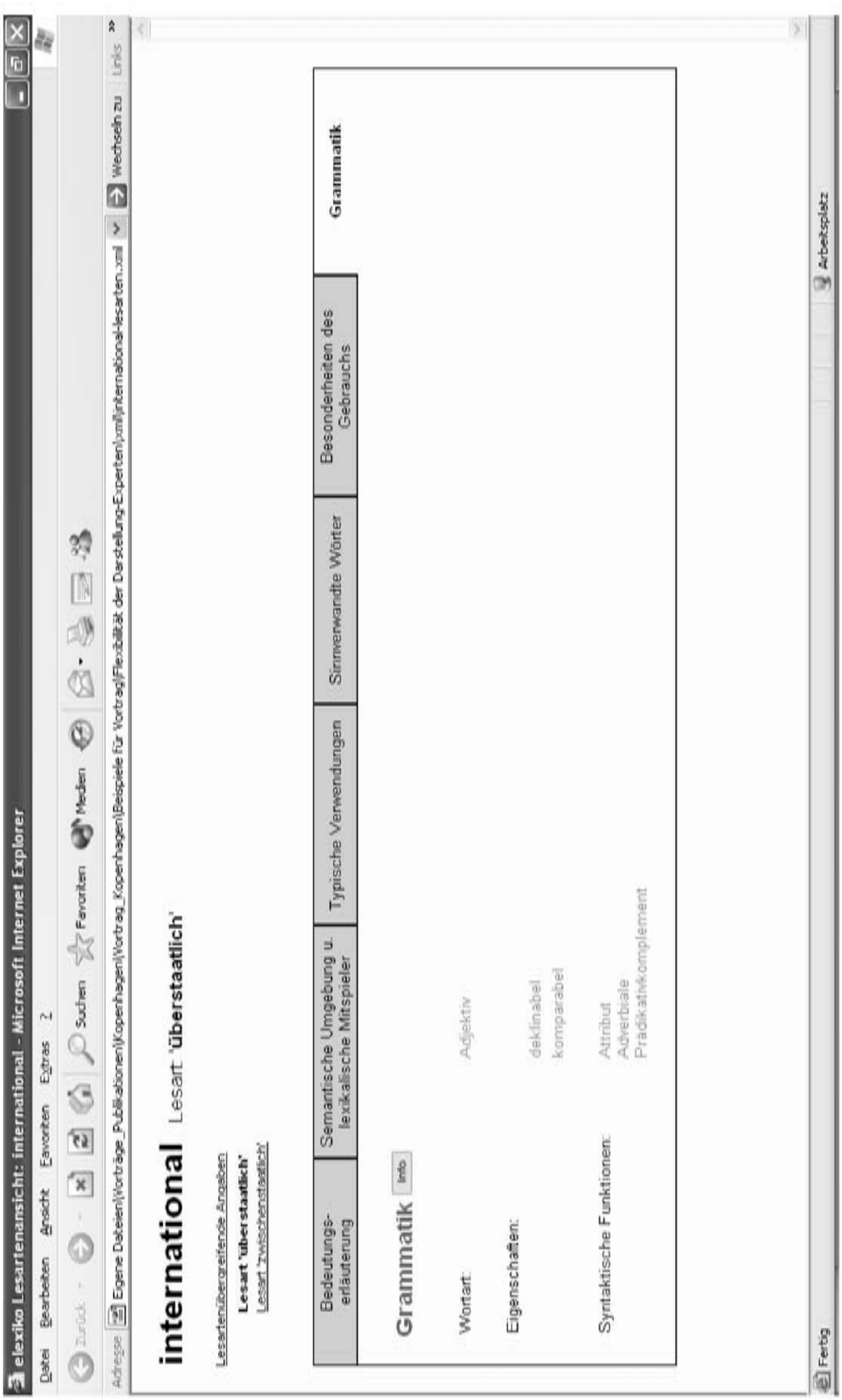


**Abbildung 9.** Mögliche veränderte Darstellung der grammatischen Angaben im Artikel "international" für Experten

## 2.5 Was noch zu tun ist

Auch in diesem Bereich des Projektes gibt es aber natürlich noch viel zu tun. Anfangs wurden die Anforderungen an ein Modellierungskonzept vorgestellt, die für die Entwicklung der Modellierung relevant waren. Mit dieser jetzt entwickelten und angewandten Modellierung ist auf Ebene der lexikografischen Datenbasis die Voraussetzung für einen sehr flexiblen Umgang mit den lexikografischen Daten auf Ebene der Präsentation geschaffen. Allerdings werden diese Möglichkeiten noch nicht in vollem Umfang in wirkliche Funktionalität für die Wörterbuchbenutzer umgesetzt.

Was jetzt schon als Vorteil dieser Modellierung zum Tragen kommt ist die – auch immer wieder geforderte – Unterstützung der Lexikografen bei der Einhaltung der formalen Artikelstruktur. Dies hat sich auch u. a. bei der Einarbeitung einer neuen Kollegin sehr gut bewährt. Doch das volle Potenzial der gezeigten Modellierung kann erst dann wirklich gezeigt werden, wenn man

1. adaptiv zu Benutzungssituationen flexible Darstellungsmöglichkeiten entwickelt und
2. sehr umfangreiche Recherchemöglichkeiten anbietet.

Recherchemöglichkeiten sind zwar ansatzweise angeboten, doch sie sind bei weitem noch nicht so umfassend, wie sie von der Aufbereitung der Daten her sein könnten. Das Problem ist also nicht (wie sonst oft in lexikografischen Projekten), dass die Daten nicht richtig aufbereitet sind, und von daher solche Möglichkeiten überhaupt nicht zur Verfügung stehen. Vielmehr ist auf dem Weg von der lexikografischen Datenbasis hin zur Präsentation der Daten im elektronischen Wörterbuch noch vieles zu tun.

## Literatur


Bergenholtz, Henning und Joachim Mugdan. 1984a. "Vorwort". In *Studien zur neuhochdeutschen Lexikographie V*, H. E. Wiegand (Hg.) [Germanistische Linguistik 3–6, 1984], 7–19. Hildesheim, Zürich und New York: Georg Olms.

Bergenholtz, Henning und Joachim Mugdan. 1984b. "Grammatik im Wörterbuch: Von *Ja* bis *Jux*". In *Studien zur neuhochdeutschen Lexikographie V*, H. E. Wiegand (Hg.) [Germanistische Linguistik 3–6, 1984], 47–75. Hildesheim, Zürich und New York: Georg Olms.

Büchel, Gregor und Bernhard Schröder. 2001. "Verfahren und Techniken in der computergestützten Lexikographie". In *Chancen und Perspektiven computergestützter Lexikographie*, I. Lemberg/B. Schröder/A. Storrer (Hgg.) [Lexicographica. Series Maior 107], 7–28. Tübingen: Niemeyer.

Farø, Ken. 2005. "Korpuslexikographie und Flexionsmorphologie". In *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography. May, 2–4,* *2002 at the University of Copenhagen,* H. Gottlieb/J. E. Mogensen/A. Zettersten (Hgg.) [Lexicographica. Series Maior 115], 217–226. Tübingen: Niemeyer.

Geeb, Franziskus. 2001. "leXeML – Vorschlag und Diskussion einer (meta-)lexikographischen Auszeichnungssprache". *Sprache und Datenverarbeitung* 2 (2001): 27–61.

Gennusa, Pamela L. 1999. "Evolution and use of generic markup languages". In *SGML und XML. Anwendungen und Perspektiven*, W. Möhr/I. Schmidt (Hgg.), 27–50. Berlin/Heidelberg/New York: Springer.

Gloning, Thomas und Rüdiger Welter. 2001. "Wortschatzarchitektur und elektronische Wörterbücher: Goethes Wortschatz und das Goethe Wörterbuch". In *Chancen und Perspektiven computergestützter Lexikographie*, I. Lemberg/B. Schröder/A. Storrer (Hgg.) [Lexicographica. Series Maior 107], 117–132. Tübingen: Niemeyer.

Heyn, Matthias. 1992. *Zur Wiederverwendung maschinenlesbarer Wörterbücher. Eine computergestützte metalexikographische Studie am Beispiel der elektronischen Edition des "Oxford Advanced Learner's Dictionary of Current English"* [Lexicographica. Series Maior 45]. Tübingen: Niemeyer.

Kilgarriff, Adam. 2000. "Business Models for Dictionaries und NLP". *International Journal of Lexicography* 13 (2/2000): 107–118.

Klosa, Annette. 2001. "Qualitätskriterien der CD-ROM-Publikation von Wörterbüchern". In *Chancen und Perspektiven computergestützter Lexikographie*, I. Lemberg/B. Schröder/A. Storrer (Hgg.) [Lexicographica. Series Maior 107], 93–101. Tübingen: Niemeyer.

Mugdan, Joachim. 1989. "Information on Inflectional Morphology in the General Monolingual Dictionary". In *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. 1. Teilband [Handbücher zur Sprach- und Kommunikationswissenschaft 5.1], F. J. Hausmann/O. Reichmann/H. E. Wiegand/L. Zgusta (Hgg.), 518–525. Berlin und New York: De Gruyter.

Schmidt, Ingrid und Carolin Müller. 2001. "Entwicklung eines lexikographischen Modells: Ein neuer Ansatz". In *Chancen und Perspektiven computergestützter Lexikographie*, I. Lemberg/B. Schröder/A. Storrer (Hgg.) [Lexicographica. Series Maior 107], 29–52. Tübingen: Niemeyer.

Schryver, Gilles-Maurice de. 2003. "Lexicographer's Dreams in the Electronic-Dictionary Age". *International Journal of Lexicography* 16 (2/2003): 143–199.

Storrer, Angelika. 2001. "Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie", In *Chancen und Perspektiven computergestützter Lexikographie*, I. Lemberg/B. Schröder/A. Storrer (Hgg.) [Lexicographica. Series Maior 107], 53–69. Tübingen: Niemeyer.

Trap-Jensen, Lars. 2002. "Descriptive and Normative Aspects of Lexicographic Decision-Making: The Borderline Cases", In *Proceedings of the Tenth EURALEX International Congress. Vol. II,* A. Braasch/C. Povlsen (Hgg.), 503–508. Kopenhagen: CST.

Wiegand, Herbert Ernst. 1985. "Fragen zur Grammatik in Wörterbuchbenutzungsprotokollen. Ein Beitrag zur empirischen Erforschung der Benutzer einsprachiger Wörterbücher", In *Lexikographie und Grammatik, Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.-30.6.1984*, H. Bergenholtz/J. Mugdan (Hgg.) [Lexicographica. Series Maior 3], 20–98. Tübingen: Niemeyer.

Wiegand, Herbert Ernst. 1989. "Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven". In *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. 1. Teilband [Handbücher zur Sprach- und Kommunikationswissenschaft 5.1], F. J. Hausmann/O. Reichmann/H. E. Wiegand/L. Zgusta (Hgg.), 409–462. Berlin und New York: De Gruyter.

Wiegand, Herbert Ernst. 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie.* 1. Teilband. Berlin und New York: De Gruyter.

XML Standard [Fourth Edition]: www.w3.org/TR/REC-xml. (27.7.2007)

Zifonun, Gisela; Ludger Hoffman und Bruno Strecker. 1997. *Grammatik der deutschen Sprache*. Berlin und New York: De Gruyter.

CHAPTER 3

# Innovative phonetic interfaces for electronic dictionaries

Włodzimierz Sobkowiak

The graphical user interfaces (GUIs) of contemporary electronic dictionaries keep to the old tradition of two windows: (1) a dynamically searchable macrostructural wordlist on the left, and (2) a static microstructural entry panel on the right. Other boxes, menus, frames and flashcards are just embellishments of this fundamental standard interface scheme. Pronunciation in particular is graphically represented in but one century-old method: as a phonetic transcription field appearing (or not) right after the entry headword. In this paper I argue for the application to the representation of EFL electronic dictionary pronunciation of the new Java- and Flash-enabled animated GUIs based on concept-mapping techniques, recently proposed for database querying by various authors. The benefits to learners include: (a) improved customizability, (b) explicit mapping of phonetic processes, relations and groupings (homophony, allophony, similarity, minimal-pairs, assimilations, deletions, L1 substitutions, etc.), (c) direct intuitive manipulability enhancing retention, (d) potential for incidental data-driven learning and serendipity, and others.

## Introduction

Seen from the internet perspective, with all its vibrant cornucopia of lavishly colourful banners, dynamic animation, high-quality streaming video and multi-channel MP3 stereo sound, electronic dictionaries must appear as a niche of primordial tranquillity and stability. One would be hard-pressed to find a dictionary, either in the stationary CD-ROM-mounted format, or indeed in a web-based form, which would venture beyond the standard tradition-sanctioned graphical user interface (GUI) of two or three tiled panels with (a) the scrolling word-list, (b) the chosen entry, and (c) something extra, depending on the dictionary (illustration, hit-list, exercise module, or the like). While some exceptions exist, which will be briefly discussed further on, the standard view of an electronic dictionary seen by the user is most likely to be like in Figure 1, which is a screen dump of the

CD-ROM-based *Cambridge English Pronouncing Dictionary* (CEPD), 16th edition published in 2003.

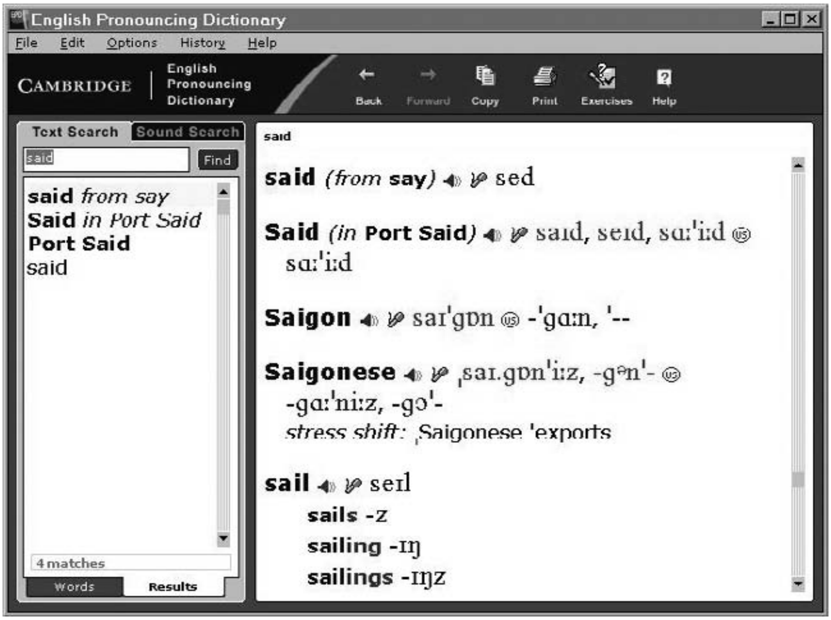


**Figure 1. The GUI of CEPD on CD-ROM**

Notice in particular that the graphical representation of pronunciation in electronic dictionaries (where it appears at all, that is) is hardly different from its canonical printed form as seen in reference books spanning the whole of the last century: just a piece of phonetic transcription duly graphically differentiated by colour or brackets/slashes from the rest of the entry's microstructure. True enough, there is now sound of the recorded audio rendition of the looked-up word, to be retrieved by clicking on the word itself or the loudspeaker icon, as the case may be, but this lexico-phonetic mode of pronunciation representation in electronic dictionaries is outside of the scope of this paper, and will not be discussed further. The interested reader is referred to http://ifa.amu.edu.pl/~swlodek/public.html for my constantly updated bibliography on this very issue.

The static, essentially nineteenth century, view of dictionary contents, pronunciation in particular, presented to the user may surprise in the field of research and commerce which – out of all domains of widely conceived linguistics – is among the most innovative. Additionally, electronic dictionaries and word-lists are the very hub of many Natural Language Processing (NLP) applications, which are now at the very cutting edge of conceptual and technological applications. Mind- and concept-mapping techniques, intelligent database and corpus querying, user-adaptive multimedia, distributed neural nets, 3D visualization modes, human-like animated interactive agents, automatic speech recognition (ASR) and text-to-speech synthesis (TTS) – these are but selected areas of potential relevance for designing electronic dictionary GUIs. Rather than speculating on why this potential has so far been mostly unnoticed in lexicography, in what follows I will take a much narrower look: I will argue for the application of some of the existing innovative visualization techniques to the representation of pronunciation in learners' electronic dictionaries of English as a foreign language (EFL). The main advantages to the learner will be presented as well as some of the potential problems.

## 1 Some recent innovative data visualization GUIs

Dynamic data visualization is not an entirely new concept, of course, but it was hard to implement before the advent of computers. Even afterwards data was normally displayed statically on the screen of the video display unit (VDU) of early computers. Text could be scrolled in four directions, and some simple graphics could be obtained to render tables and graphs, for example (see, e.g. Wong and Kuo 1982), but truly dynamic animation only came with the demise of DOS and the advent of Windows. One of the best-known dynamic database visualization schemes originated at that time (see Fairchild, Poltrock and Furnas 1999 and http://www.si.umich.edu/~furnas/Papers/SEMNET.txt). A number of ingenious and widely different database visualization techniques followed, including two- and three-dimensional, hierarchical (e.g. Grokker http://www.groxis.com/service/grok or Websom http://websom.hut.fi/websom/), radiating (e.g. TouchGraph http://www.touchgraph.com or TopicMap http://highwire.stanford.edu/help/hbt/index.dtl), web-like (e.g. Concept Space http://conceptspace.london.edu/) and hyperbolic (e.g. H3Viewer http://graphics.stanford.edu/papers/h3cga/). With the growing need to visualize the content of the exploding internet about a decade ago the main thrust of research moved in this direction, with visualization engines now catering specifically for the fuzzy and noisy web data (e.g. KartOO http://www.kartoo.com). Andrews 2002, McKiernan 2001 and 2003 have many more examples and some discussion.

All these developments have not gone completely unnoticed in the lexicographic circles. The makers of KirrKirr, one of the best known innovative-GUI dictionaries have this to say on their web-page: "Clearly the greatest potential for electronic dictionaries is to take advances in storage and visualisation technology to those who could truly benefit from a better interface to dictionary information, such as language learners" (http://www-nlp.stanford.edu/kirrkirr/ausweb99/html/@e-dictionaries@.html). This quote seems rather uncontroversial, so it is disappointing to discover very few lexicographic projects which would pay heed:

- Constellation (http://graphics.stanford.edu/papers/const/const.pdf)
- Visual Thesaurus by ThinkMap (http://thesaurus.plumbdesign.com/index.jsp)
- KirrKirr (http://www-nlp.stanford.edu/kirrkirr/ausweb99/present.html)

With the commonly observed disregard for matters phonetic in lexicography (cf. Sobkowiak 1999 for an extended discussion) it is only to be expected that none of these projects applies the new visualization techniques to the representation of pronunciation. Rather, they are all semantically oriented, with links, clusters and movement signaling meaning relations. Ignoring the rich structure of the underlying *phonetic* relations in a lexical database unfortunately has a long tradition, which I will not discuss here. Let this summarizing quote from 2002 suffice:

> The lexicographic perspective has thus traditionally been (and to quite an extent continues to be) idiographic: concerned with a description and analysis of *cases*, *tokens* and *idiosyncrasies*, conveniently categorized in terms of 'words' (Sobkowiak 2005: 513).

In the following sections of the paper I argue that applying innovative GUI techniques to electronic dictionaries, and especially to representing pronunciation, can bring substantial advantages to learners.

## 2 Advantages and problems of innovative GUIs in learner dictionaries

### 2.1 Adaptivity/customizability

The limit of thinkable customization in electronic dictionary GUI appears to be – to dictionary makers, reviewers and users alike – like in this review: "you can also adjust the layout of the dictionary interface, so that the screen splits either vertically or horizontally, allowing records to be displayed left to right or top-down, depending on your preferred consultation mode" (Herrmann 1999). On a similarly approving tone, Oppentocht and Schutz notice in 2003 (!) that: "in most electronic dictionaries the size and the colour of the letters can be adjusted to the convenience of the individual user" (Oppentocht and Schutz 2003: 221). Yet, if it is true that "learners need a system that can accommodate their own, idiosyncratic, and probably frequently changing ideas of vocabulary organization" (O'Rourke 1998), the extent of customization so praised by Herrmann and Oppentocht and Schutz is by far not sufficient. As has been rather forcefully argued in contemporary foreign language teaching (FLT) methodology, learners differ widely along a number of dimensions: intelligence, cognitive style, proficiency, linguistic background, personality, etc. Pedagogy should ideally accommodate all these different variables in designing methods, techniques, syllabuses and resources to suit each particular learner. Translated into requirements of a suitable foreign language dictionary GUI, this modern methodology will reduce to something like these two pertinent quotes: "interface allows learners to visualize relations between items and categories, without predefining or implicitly favouring any particular taxonomy" (O'Rourke 1998), and: "This means that the dictionary can deal with a broader range of intentions and a greater range of language competency than is possible with printed copy" (Jansz et al. 1999).

### 2.2 Foregrounding rules/regularities/patterns/clusters

As mentioned above, current dictionary interfaces, both traditional (paper) and electronic, fail miserably in highlighting lexical patterns inherent in the enormous and highly interrelated lexical database. An ordinary dictionary is a representation of alphabetically listed word-chunk lexical idiosyncrasies. The recent additions of study panels of all sorts (false friends, homonyms/graphs, grapho-phonemic regularities, etc.) appear to be just stop-gap measures of mostly commercial significance, rather than breakthroughs in dictionary design. Even the revolutionary invention of hypertext has only expedited the look-up technique rather than changed the underlying lexicographic philosophy of one long word-list.

On the other hand, innovative GUIs based on concept-mapping theories and techniques foreground structure, make relations explicit, visualize links and clusters, help in noticing and formulating regularities. In the words of one authority in the area of knowledge representation: "Besides readability, graphic notations often have heuristic value in helping human readers (either students or researchers) to discover patterns that would be difficult or impossible to see in the linear form" (Sowa 2002).

### 2.3 Iconicity, simulation of physicality/tangibility, conceptualizing by doing

Due to our animal ancestry going back hundreds of millennia, humans learn best by seeing and manipulating: *Tell me and I'll forget; show me and I may remember; involve me and I'll understand*. This phenomenon has been extensively studied by both psychologists and pedagogues: "Our understanding of the world is fundamentally linked to visual stimulation and the tactile experience of manipulating objects in our environment" (Lynch 1994: 23, quoting Piaget 1954, Bruner 1966, Kay 1988, Kay 1990). Innovative database GUIs capitalize on this genetic endowment by creating a simulated environment where users can actually manipulate conceptual objects, both abstract and concrete, represented as icons. For example, "the *Spider* attempts to layout a graph of information by using a combination of "Informotions," which are modeled on physical properties such as magnetism, elasticity and viscosity" (http://www.plumbdesign.com/products/spider.html). "The *Hierarchy* configuration builds on the Spider by adding one Informotion, "Tree Motion", which can be configured to act only on those entities that contain hierarchical information. Tree motion acts like a constant "wind" that blows lower entities down, and higher entities up" (http://www.plumbdesign.com/products/hierarchy.html). Other simulated dimensions of physicality used in innovative GUIs are, for example, node distance or link salience/colour signifying the strength of connection.

### 2.4 Exploration, interactivity, navigation, incidental/opportunistic learning, heuristics

In the truly customizable, iconic, quasi-physical environment of the lexical database users can engage in authentic interaction with the vocabulary content, going much beyond the customary lexical lookup. The contents of the dictionary can be actively explored, regularities serendipitously investigated, concepts metaphorically 'opened' and 'closed', relations restructured and invented, clusters discovered and deleted. To what extent these new potentials are used will of course depend on the circumstances of the learning situation, but even in the most traditional glotto-pedagogical paradigm of grammar translation there might then be place for some learner-initiated data-driven vocabulary work (or fun). This will only be possible, however, with dictionary GUIs which, at least to some extent, reflect the structure of the human mental lexicon, where

> the various lexical networks or fields are closely connected and form a kind of multi-dimensional word-web. Within the human word-web, each word is linked to other words or groups of words in a variety of ways and can be retrieved from different starting points (Abel and Weber 2000: 808).

Also in a perspective much wider than strictly lexicographic, extensive research shows (see Cañas et al. 2003 for an excellent overview) that computer-assisted concept-mapping fosters retention and understanding of the studied material.

### 2.5 Some GUI problems and their solutions

Quite apart from strictly computational difficulties, the new database GUIs present certain characteristic conceptual problems as well (see e.g. Munzner 1998, Munzner et al. 1999, Fairchild et al. 1999 for some discussion). For example, a multiply customizable, three-dimensional, fully multimedialized and animated interface carries with it the cost of complexity. One facet of this is screen clutter, a problem well-researched in the field of computer science and human-computer interaction, if less so in electronic metalexicography. There are many solutions to screen clutter, of varying age and effectiveness. With respect to database GUIs in particular, they include, among others: zooming, varied sampling density, clustering, 3D perspective, node pruning/deletion, mouse-over hovering tips, distinctive vs. desaturated colour, graph self-organization, fisheye views, and others.

Other potential and actual problems include excessive jiggle, especially in Java-animated displays (solution: freezing/anchoring, amplitude reduction) or the once much-discussed side-effect of hypertext: the lost-in-3D-space syndrome (solution: maps, indexes, trails, marker-dropping, backtracking, landmarks). New search techniques must be elaborated once the familiar linear arrangements of data are abandoned (I owe this note to the anonymous reviewer of this paper).

## 3 The need for innovative *phonetic* GUIs in electronic dictionaries

Prima facie, the functionalities of the innovative electronic dictionary GUIs, both actual and potential, should make them a standard tool for the presentation of phonetic information in all lexical resources. How pronunciation is accessed in the dictionary should best be left to the user, to decide through a complex but friendly interface. This interface should allow, or better still: stimulate, the learner to discover phonetic regularities as well as idiosyncrasies. After all, "The methodological tradition of phonetics, [... as opposed to lexicography], is that of nomology: preoccupation with *rules*, *trends* and *laws*", and "the rules are inevitably there, hidden under the surface of tokens, and the learner is bound to assimilate them inductively in direct proportion to his/her use of the dictionary" (Sobkowiak 2005: 513 and 514). In a similar context, designers of one innovative electronic dictionary summarize an important lesson which lexicographic phoneticians can learn from psycholinguists: "Words which present phonological assonances or similar rhythmic patterns are clustered together" (Abel and Weber 2000: 808).

Of course, it is not only assonances and rhythm that are at stake here. After all, "[p]honetic lexicography is at the very start of the long road to fully account for such prosodic features of words in context as the mentioned stress shift, liaison, rhythm, elision or juncture" (Sobkowiak 2005: 514-515). As an example of one such phonetic phenomenon which is poorly treated in standard EFL dictionaries, and to which innovative GUIs could be applied with some advantage, take phonostylistic reduction:

> In my dialect the words 'been' and 'Ben' are pronounced the same in isolation. But 'been' shortens to [bn] in 'I've been wanting to'. 'Ben' never shortens in this way. So has anyone developed a policy for indicating such pronunciation variation in their dictionary? (Ron Moe on <lexicographylist>, 12th April 2004)

The answer to Moe's question is of course: yes, many dictionaries, especially for learners, indicate fast-speech pronunciation of function words. What matters, however, in the context of this discussion, is that this variation is represented in the static traditional manner of isolated lexical representations, with no indication of graduality of the process, no hyperlink to other affected lexical items, no pertinent word-list, no minimal pair information or practice of the kind mentioned by Moe, indeed hardly anything beyond the listing of two phonetic transcription strings: one for 'strong', and the other for 'weak'. In pronouncing dictionaries, such as the CEPD exemplified in Figure 1 above, one can at least expect a short note on usage, which is however as static as the rest of the entry.

That there is an increasingly felt need for some more dynamic representation of pronouncing information in electronic dictionaries can be seen in yet another <lexicographylist> message, which has been brought to my attention by Gilles-Maurice de Schryver:

> From: "Louis Janus" <janus005@umn.edu>
> To: <lexicographylist@yahoogroups.com>
> Sent: Tuesday, February 24, 2004 9:28 AM
> Subject: [Lexicog] graphic representation of sounds
> [...]
> \> I'm searching for ideas to give a good graphic representation of the
> \> vowel length, syllable stress, and tone (Norwegian has two tonemes)
> \> -- all those three are related, and I'm thinking here that perhaps
> \> animation (with the long or short part of the word stretching), and
> \> also showing up and down tone patterns. I definitely think this type
> \> of graphic would help learners (while they listen to the word, for
> \> example), but it has to be simple and clear.

It is perhaps not surprising that the aspect of FL pronunciation most in need of a dynamic, animated, innovative display is the broadly conceived prosody: segmental length, stress and rhythm, tone, intonation, fast speech reductions. This is where the traditional phonetic representation, carried over from paper dictionaries into the electronic realm, fails most miserably. The techniques of innovative GUIs discussed in this paper could help, with their animation, colour, hyperlinks, 3D views, drags-and-drops, and more, but are sadly unexplored and under-applied.

## 4 An example of an innovative *lexi-phonetic* GUI with Semantica®

In this final section of the paper I will exemplify how one particular type of GUI, implemented as part of the mind-mapping package Semantica (http://www.semanticresearch.com/) can be used for graphically representing lexico-phonetic information in an electronic dictionary or other kinds of computer-assisted EFL resources. As a matter of fact, Semantica is not particularly well suited for the purpose; like many similar programs, it is heavily biased semantically, with no provision for font control, for example, which immediately excludes easy manipulation of phonetic transcription. The aim of this exercise, however, is not to advertise this particular package, but rather to show what can be achieved phono-lexicographically with a simple off-the-shelf application. The actual implementation of these ideas in an electronic dictionary GUI is another matter altogether, of course, and should best be left to professional programmers.

To fully demonstrate the functionalities of an innovative animated GUI would of course require an interactive presentation, which is not possible in this traditional medium of print (one of the increasingly unbearable paradoxes of contemporary writing on multimedia). The best that can be expected here, as opposed to the actual presentation in Copenhagen, is the demonstration and discussion of some screen shots. As the lexical hub (what is called 'central concept' in Semantica) of the little network of phono-lexical relations to be discussed below I selected the word *said*. This was done for two reasons: (a) it is one of the commonest English wordforms to be introduced at the very beginning of EFL teaching and learning, (b) it presents many phonetic difficulties (at least to Polish learners of English) in a relatively short form.

Let me begin by observing that the GUI of Semantica offers all the advantages itemized in section 2 of this paper: it is to quite an extent customizable, it foregrounds patterns, it simulates physicality, and allows interactive exploration. To be fair, one should add that it also falls prey to some of the classic GUI problems, such as excessive jiggle, for example.

In Figure 2 "a visualization known as the *graphic frame* is used. A graphic frame represents a two-dimensional neighbourhood slice of a much larger and more complex three-dimensional knowledge structure" (http://www.semanticresearch.com/semantic/visualization.php), and as such can be easily accommodated on a printed page. "Navigating a semantic network is done in a straightforward hypertextual manner, much like a traditional web browser. Clicking on any of the related concepts will navigate to that concept as the central concept, displaying a new graphic frame" (ibid.). Some of the concept nodes have additional information attached to them. In figure 2 this is the case of /sed/, /seit/, /seid/ and /saed/ (the <ae> standing for the phonetic ash character, i.e. /æ/). This additional information is called *knowledge object* in Semantica; it carries the paper-clip icon, can be any multimedia element, and can also be attached to relations, not only to concepts.

So what are the advantages of this kind of phono-lexical GUI to the EFL learner? First: representation. The learner, like in a standard dictionary, will be told that <said> 'is correctly pronounced as' /sed/, and will be able to click to listen to a model recording. This is where electronic dictionary phonetic information normally stops. With an innovative Semantica-like GUI the learner will be able to learn much more (clock-wise in Figure 2):

– Common, L1-sensitive, mispronunciations are provided, some of which may turn out to be homophonous with other words (e.g. <sad>), which is easily discovered simply by clicking on them, and thus revealing "a new graphic frame".
– Troublesome grapho-phonemic mismatches are highlighted, such as <said>/<laid>, which tend to lead learners up a garden path to create hard-to-eradicate habits.
– Each of the constituent sounds of the word has a life of its own, so to say, with its own phonetic problems for learners, characteristic interference from their mother tongue, phono-graphemic (ir)regularities, etc. And each can be clicked to reveal this information, leading further and further into the network.
– Idiosyncratic phono-lexical problems can be listed, such as the confusing <Port Said>, for example.
– Rhyme is an important phonetic phenomenon to most speakers generally, and to learners in particular. Rhyme can be used as a vehicle for highlighting different spellings of the same sounds, a notorious problem for EFL learners.
– <Said> is, of course, an irregular past-tense form of say, something which ordinary dictionaries are quick to mention. Clicking on <say> in Semantica will reveal further phono-lexical facts, this time radiating from this basic form of the lexeme.
– Finally, minimal pairs are still popularly used in teaching foreign language pronunciation, even in the otherwise communicative paradigms of FLT. The Semantica GUI provides an easy way to append a few minimal pair items with each word. Each of these will be a fully-fledged node in its own right, of course.

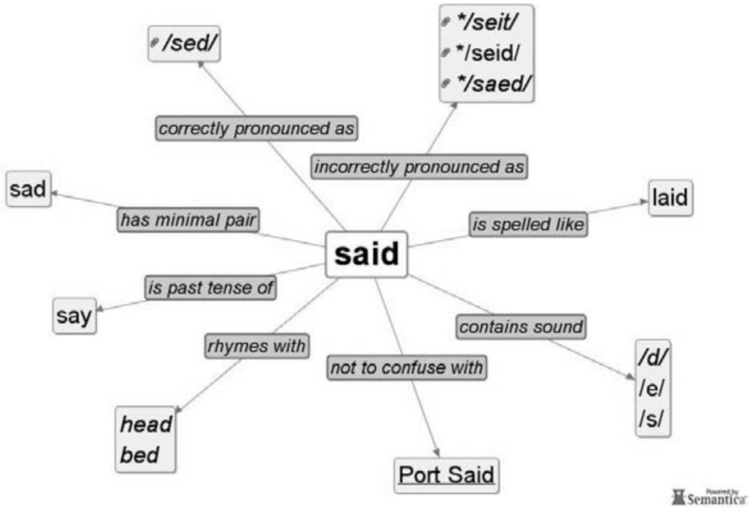


**Figure 2. An example of lexi-phonetic representation with Semantica®: a graphic frame view**

All this information in the form of concepts, relations and knowledge objects can be provided by the lexicographer, but also by the user, i.e. the learner or the teacher, as the case may be. Everything on screen is freely manipulable and deletable. Every element can be temporarily hidden from view to reduce clutter or to accommodate learners of particular characteristics.

The *graphic frame* visualization is not the only way to view and manipulate data in Semantica's GUI. To fully appreciate the (quasi) three-dimensional nature of the display, one should go into the so-called *global view*, where *all* concepts and relations can be shown simultaneously in an animated 3D lattice gracefully wiggling in the middle of the screen. It is not possible to reproduce this effect here, of course, but its frozen 2D rendering is shown in Figure 3. It is immediately obvious from this view that clutter can be a serious obstacle to legibility. At the same time, however, this view forcefully demonstrates the complex network of relations, which is conceptually by far more true a representation of lexico-phonetic information in a dictionary than the standard listing of transcription.

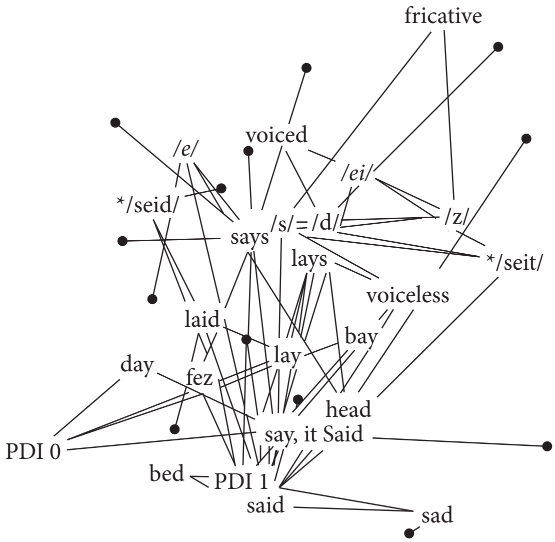


**Figure 3.** ***Said* in Semantica's *global view***

## 5 Conclusions

The growing popularity of ThinkMap's (formerly Plumb Design) *Visual Thesaurus* and Christopher Manning's *KirrKirr* (Manning, Jansz and Indurkhya 2001) attest to the budding realization among electronic dictionary makers and users that innovative interfaces are the right answer to new dictionary lookup habits of the highly computer-savvy generation. Increasingly, dictionaries are used as fully fledged learning resources, at least in the exploding foreign language arena in the rapidly uniting Europe (see, e.g. Zock 2002 or Zock and Quint 2004). In this context, patching up the standard (and boring) interface with a few additional pull-down menus, hyperlinks or hotspots will not do, either pedagogically or commercially. It is perhaps not surprising that the two technologies preferentially used for the design of the new interfaces discussed in this paper, namely Java and Flash, are the recent offspring of the internet, to which they have added in the last decade unprecedented degrees of interactivity and animation. Both *KirrKirr* and *Visual Thesaurus* are available on-line, for free download and restricted lookup, respectively. Neither, however, is a standard dictionary. We should now appeal for, and expect to see very shortly, the application of such innovative GUIs to 'ordinary' semasiological electronic dictionaries. In the current situation on the dictionary market odds are that the first such application will be a learner's dictionary of English as a foreign language. Another relatively safe prediction is (unfortunately to a phonolexicographer) that pronunciation will be the last to benefit from this breakthrough in GUI design.

## References


Abel, A. and V. Weber. 2000. "ELDIT – a prototype of an innovative dictionary". In *Euralex 2000 proceedings,* U. Heid et al. (eds), 807–18. Stuttgart: IMS.

Andrews, K. 2002. *Information visualisation. Tutorial notes*. Graz: Graz University of Technology. [www2.iicm.edu/ivis/ivis.pdf]

Bruner, J. 1966. *Towards a theory of instruction*. New York: W. W. Norton.

Cañas, A. J. et al. (eds.). 2003. *A summary of literature pertaining to the use of concept mapping techniques and technologies for education and performance support*. Pensacola, FL: The University of West Florida. [www.ihmc.us/users/acanas/Publications/ConceptMapLitReview/IHMC%20Literature%20Review%20on%20Concept%20Mapping.pdf]

Corris, M. et al. 2000. "Bilingual dictionaries for Australian languages: user studies on the place of paper and electronic dictionaries". In *Euralex 2000 proceedings,* U. Heid et al. (eds), 169–181. Stuttgart: IMS.

Fairchild, K. M.; S. E. Poltrock and G. W. Furnas. 1999. "SemNet: three-dimensional graphic representations of large knowledge bases". In *Readings in information visualization: using vision to think,* S. K. Card, J. D. Mackinlay and B. Schneiderman (eds.), 190–206. San Francisco, CA: Morgan Kaufmann Publishers Inc.

Herrmann, F. 1999. "Review of Larousse Chambers CD-ROM Advanced Dictionary English-French / French-English". [www.fhphd.org/LarousseChambersAD.html]

Manning, C. D.; K. Jansz and N. Indurkhya. 2001. "Kirrkirr: software for browsing and visual exploration of a structured Warlpiri dictionary". *Literary and Lingustic Computing* 16 (1): 123–39.

Kay, A. 1988. "Doing with pictures makes symbols: communicating with computers", Videotape. Stanford, CA: University Video Communications.

Kay, A. 1990. "User interface: A personal view". In *The art of human-computer interface design,* B. Laurel (ed.), 191–207. Reading, MA: Addison-Wesley.

Lynch, P. J. 1994. "Visual design for the user interface. Part one: Design fundamentals". *Journal of Biocommunications* 21 (1): 22–30.

McKiernan, G. 2001. "New age E-journals, indexes, and services". *Science & Technology Libraries* 21 (3–4): 223–33.

McKiernan, G. 2003. "New age navigation: innovative information interfaces for electronic journals". *The Serials Librarian* 45 (2): 87–123.

Munzner, T. 1998. "Exploring large graphs in 3D hyperbolic space". *IEEE Computer Graphics and Applications* 18 (4): 18–23.

Munzner, T. et al. 1999. "Constellation: a visualization tool for linguistic queries from MindNet". *Proceedings of the IEEE Symposium on Information Visualization*, October 24–29 1999, San Francisco. [http://graphics.stanford.edu/papers/const/const.pdf]

Oppentocht, L. and R. Schutz. 2003. „Developments in electronic dictionary design". In *A practical guide to lexicography,* P. van Sterkenburg (ed.), 215–27. Amsterdam: John Benjamins.

O'Rourke, B. 1998. "A network-based tool for organizing second-language vocabulary". In *Proceedings of ED-MEDIA/ED-TELECOM 98,* T. Ottmann and I. Tomek (eds), 1780–1. Charlottesville, VA: AACE.

O'Rourke, B. and K. Schwienhorst. 2000. "Learner databases and virtual worlds: Using computers to create collaborative learning environments", In *Integrating theory and practice in LSP and LAP: Papers from the ALC/IRAAL conference*, M. Ruane and D. P. Ó. Baoill (eds.), 123–132. Dublin: ALC & IRAAL.

Piaget, J. 1954. *The construction of reality in the child*. New York: Basic Books.

Sobkowiak, W. 1999. *Pronunciation in EFL machine-readable dictionaries*. Poznań: Motivex.

Sobkowiak, W. 2005. "Lexicographic phonetics or phonetic lexicography?". In *Symposium on Lexicography XI. Proceedings of the 11th International Symposium on Lexicography, May 2–4, 2002, at the University of* Copenhagen, H. Gottlieb, J. E. Mogensen and A. Zettersten (eds.), 511–519. [Lexicographica. Series Maior 115]. Tübingen: Niemeyer.

Sowa, J. F. 2002. „Semantic Networks". [www.jfsowa.com/pubs/semnet.htm]

Wong, K. T. and I. Kuo. 1982. "GUIDE: graphical user interface for database exploration". In *Proceedings of the 8th International Conference on Very Large Data Bases*, M. Kaufman (ed.), 22–32.

Zock, M. 2002. "Electronic dictionaries for men, machines or for both?". *3rd Papillon seminar, NII*, 35–41. Tokyo.

Zock, M. and J. Quint. 2004. "Why have them work for peanuts, when it is so easy to provide reward? One of the many possibilites of a dictionary converted into a drill tutor". *PAPILLON-2004 Workshop on Multilingual Lexical Databases*, Grenoble, 30 August – 1 September 2004.

[All web-based sources last accessed July 27, 2007]

Part 2

# Dictionary structure

CHAPTER 4

# Sublemmata or main lemmata

## A critical look at the presentation of some macrostructural elements

Rufus H. Gouws

The macrostructural presentation of lexical items as either main or sublemmata is too often done in an uncritical way – employing a system which is part of an established lexicographic tradition. This paper focuses on the need for a critical approach to the macrostructural positioning of lexical items, dominated by the needs and reference skills of the intended target user and the lexicographic functions prevailing in the dictionary. It is shown that the degree of textual condensation in the presentation of sublemmata can impede successful access to the lemma signs and an optimal retrieval of information. It is argued that the access structure of a dictionary is primarily a lexicographic device and not a linguistic instrument. The macrostructural positioning of lexical items should not be done in accordance with linguistic principles but should rather be the result of the implementation of user-driven lexicographic criteria.

### 1 Introduction

Once the selection of lemma candidates for a given dictionary has been completed, the data distribution structure of that dictionary should determine the way in which the lemmata should be presented and ordered in the macrostructure as well as the nature and extent of the lexicographic treatment allocated to each lemma. In current printed dictionaries adhering to a system of alphabetical ordering lemmatisation procedures lead to lemma candidates being accommodated as either main or sublemmata. In dictionaries with a strict alphabetical ordering all the lemmata will be presented as main lemmata in a vertical ordering, cf. example 1, taken from an English-Afrikaans dictionary:

(1) **balalaik´a,** balalaika,.
**bal´ance,** saldo, balans (in bank);..
**bal´ancebob,** skietlood;

**bal´ance bridge,** wipbrug;
**bal´anced,** ewewigtig, gebalanseer(d);
**bal´anced diet,** gebalanseerde dieet;
**bal´anced picture,** gebalanseerde (ewewigtige) voorstelling;
**bal´ancer,** koorddanser; balanseerder; stabiliseerder, stabilisator;
**bal´ance sheet,** balansstaat;
**bal´ance spring,** balansveer;
**bal´ance weight,** balanseergewig;
**bal´ance wheel,** skakelrat; onrus (in horlosie).
**bal´ancing,** balansering;
**bal´ancing-pole,** balanseerstok.

Where provision is made for the presentation of lemmata in a sinuous lemma file additional procedures of textual condensation come into play and the lemmata display a horizontal ordering. Textual condensation can have a further effect on these horizontally-ordered lemmata where the omission of the mutual component, usually the first stem, results in a clustering of partial lemmata presented as sublemmata, cf. example 2, taken from *Groot woordeboek/Major Dictionary* (hereafter abbreviated as GW) which is the condensed presentation of the same article stretch as given in example 1:

(2) **balalaik´a,** balalaika,
**bal´ance,** saldo, balans (in bank);..; die balans opmaak; afsluit (boeke); ~ **bob,** skietlood; ~ **bridge,** wipbrug; ~**d,** ewewigtig, gebalanseer(d); ~**d diet,** gebalanseerde dieet; ~**d picture,** gebalanseerde (ewewigtige) voorstelling; ~**r,** koorddanser; balanseerder; stabiliseerder, stabilisator; ~ **sheet,** balansstaat; ~ **spring,** balansveer; ~ **weight,** balanseergewig; ~ **wheel,** skakelrat; onrus (in horlosie).
**bal´ancing,** balansering; ~**-pole,** balanseerstok. (GW)

The use of main and sublemmata as two major types of lemmatization is seen in both monolingual and bilingual dictionaries. The publication of many new dictionaries sees a continuation of this well-established macrostructural system of main and sublemmata – often because it is an established lexicographic tradition and not because the lexicographers have decided on this procedure as a result of the identification of the needs and reference skills of their target users or to coincide with the functions of the specific dictionary. The argument is that what has been good so far will be good in future. This is a problem related to an approach where lexicographers plan their dictionaries only in terms of models on offer in existing dictionaries without negotiating the motivation for certain aspects of an existing model or the implications of the use of such a model for the new dictionary.

This does not necessarily lead to the publication of bad dictionaries. To quote a nineteenth century South African statesman: “one should take from the past what is good and use it to build the future.” However, practical lexicographers should not be content with merely following the patterns set forth in existing dictionaries. They need to take a critical look at existing patterns and models and they should expect theoretical lexicographers to do the same in their continued research efforts. This also implies that theoretical lexicographers should not restrict the scope of their research to the models on offer in existing dictionaries. In stead they should rather constantly question the continued value and applicability of these models. This may lead to the formulation of new strategies that would lead to new models, better dictionaries and more successful dictionary consultation procedures. However, it may also lead to a situation where theoretical lexicographers advise practical lexicographers to take a step backwards in order to re-employ some outdated methods in order to enhance the success that users will experience in their various situations of dictionary usage. As a part of the user-driven approach in lexicography the focus should constantly be on ways and means to improve dictionaries and when planning a dictionary or even when revising an existing dictionary lexicographers need to take a critical look at established procedures in order to determine whether they can still be regarded as functional and efficient for present-day lexicography.

## 2 Research regarding main and sublemmata and the existing practice

The use of main and sublemmata is part of a tradition in the lexicographic practice and one of the most commonly used space-saving procedures. However, this form of textual condensation is too often used in an uncritical way without giving any thought to its influence on dictionary usage situations.

Theoretical research has focused on main and sublemmata as different types of lemmata, cf. Wolski (1989) and Wiegand (1989), with an additional distinction in the category of sublemmata between nested and niched lemmata. In the category of nested lemmata a further distinction has been made between first level nesting and second level nesting, cf. Gouws (2001). Niching merely represents a change in the direction of ordering, i.e. from a vertical to a horizontal ordering, and is primarily introduced for space-saving reasons, cf. example 3, taken from the *Verklarende Handwoordeboek van die Afrikaanse Taal* (Odendal & Gouws (eds.) 2000; hereafter abbreviated as HAT):

(3) **mar´mer** s.nw. Fynkorrelige kalkgesteente met mosaïektekstuur; wat in die bou- en beeldhoukuns as grondstof gebruik word: *Kerksuile van*

*marmer. 'n Standbeeld van marmer. 'n Brokstuk marmer, glad, voorwêrelds, wit* (N.P. van Wyk Louw). ww. (gemarmer) Soos marmer verf, kleur: *Die muur is gemarmer.* [L. *marmor*] **marmer: ~aar, ~agtig, ~beeld, ~blad, ~blok, ~ing, ~papier, ~plaat, ~standbeeld, ~werker.**

**mar·mot´** (-te) *(< F.)* Dikwels in die verkleiningsvorm *marmotjie.* Knaagdiertjie, sterker gebou as 'n eekhoring en met reguit stert; fam. *Sciuridae.* [L. *mus* muis + *mons* berg] (HAT)

Nesting also leads to a change in the strict alphabetical ordering. With first level nesting the cluster-internal alphabetical order is maintained, but the external order, i.e. the alphabetical order between the last member of the nest and the following main lemma, deviates from a strict alphabetical order, e.g.

(4) **di·a·mant´** (-e) harde, kosbare edelsteen, bestaande uit kristallyne koolstof: *(On)geslypte diamante.* [G. *adamas* staal; L. *diamas* onbedwingbaar; G. *damao* mak maak] **diamant: ~agtig, ~bedryf, ~boor, ~delwer, ~gruis, ~houdend, ~kenner, ~ontginning, ~ring, ~vorming, ~winning.**

**di·a·mant´brui·lof** Herdenking van die 60ste verjaarsdag van 'n huwelik.

**di·a·mant´klo·wer** Ambagsman wat die klowersvlak van 'n diamant deur 'n fyn lyntjie aandui en dit daarlangs kloof.

**di·a·mant´ko·per** Iemand wat gemagtig is om ongeslypte diamante te koop en verkoop. (HAT)

Second level nesting introduces a cluster internal deviation from the alphabetical order with the internal ordering of lemmata determined by morpho-semantic motivations, cf. example 5, taken from *Nasionale woordeboek* (De Villiers et al. 1987, hereafter abbreviated as NW):

(5) **regering … 1.** bestuur, bewind.. **2.** bepaalde kabinet … **3.** owerheid …**regeringloos; regeringsamp, -amptenaar, -gebou, -koste, -pos, -vorm; regeringsaak, -stelsel** (by 1); **regeringsbeleid, -besluit, -blad, -hoof, -kringe, -man, -party, -tyd** (by 2) (NW)

From a linguistic point of view the use of second level nesting can be defended because the ordering is done according to strict linguistic principles and assists the user to retrieve the information from the partial article stretch along linguistic lines.

Recent research regarding main and sublemmata focused on the different ordering strategies but also on the classification. The critical criterion for the distinction between main and sublemmata is not whether their ordering is done in a vertical or a horizontal way. The nature of their presentation should determine whether they are main or sublemmata. A macrostructural element qualifies as sublemma when it can only be accessed via an entrance lemma because textual condensation has resulted in the horizontally-ordered macrostructural element being a partial lemma with a place-keeping symbol substituting the first stem, cf Wiegand (2002a; 2002b). In the following example (from Eksteen et al. 1997, in the following abbreviated as GW), the macrostructural element *dairy-farming*, presented as a partial lemma *~-farming* can only be found via the entrance lemma *dairy* and consequently qualifies as a sublemma.

(6) **dair´y, (dairies),** melkkamer; melkery; ~ **beef breed,** melkvleisras; ~ **breed,** melkras; ~ **cattle,** melkbeeste; ~ **cow,** melkkoei, ~ **factory,** suiwelfabriek; ~ **farm,** melkboerdery, melkery; ~ **farmer,** melk boer, suiwelboer; ~-**farmers' union,** suiwelbond; ~-**farming,** melkboerdery, suiwelboerdery; ~ **herd,** melkkudde; ~ **industry,** suiwelnywerheid, suiwelbedryf; ~**ing,** melkboerdery, suiwelboerdery; ~**maid,** melkmeisie; ~**man,** melkboer; melkverkoper; ~ **meal,** suiwelmeel; ~ **nut,** kragvoerkoekie; ~ **produce,** suiwelprodukte; ~ **product,** suiwelproduk; ~ **ranching,** veldmelkboerdery; ~ **woman,** melkboerin; melkverkoopster. (GW)

According to this approach horizontally-ordered lemmata that can be retrieved without the co-occurrence of an entrance lemma will also be regarded as main lemmata. This is illustrated by the horizontally-ordered lemmata in example 7, taken from the Labuschagne & Eksteen (eds.) (1993): *Verklarende Afrikaanse Woordeboek* (hereafter abbreviated as VAW), which function as full lemmata, in their alphabetical position, albeit in a horizontal ordering:

(7) **baga´sie. 1.** Reisgoed (koffers, handsakke, ens.). **2.** Voorrade en uitrusting van 'n leër (verouderd); *oortollige bagasie dra*, te vet wees; **bagasiebewys; bagasieburo; bagasiedraer; bagasiekaartjie; bagasiekantoor; bagasieruim; bagasiewa.** (VAW)

## 3 The user perspective

When planning a new dictionary the lexicographer has to be well-aware of the target user, the needs and the reference skills of this target user. This also implies being familiar with the typical situations of dictionary use relevant to the planned dictionary. The less developed the dictionary using skills of the target users, the more explicit and less complicated should the presentation and treatment be. In the recent planning of a monolingual dictionary compiled for primary school learners with Afrikaans as third or a fourth language the lexicographers agreed that the users of this dictionary should not be confused by the introduction of more than one macrostructural ordering system. Consequently the dictionary only displays a vertical strict alphabetical ordering without any niched or nested clusters.

To the well-trained dictionary user who is familiar with the macrostructural ordering system of a given dictionary it is usually not too difficult to access the desired lexical item, whether presented as a main or a sublemma, and in a dictionary where the intended target user group consists of experienced dictionary users the lexicographer has every right to utilize procedures of textual condensation, like partial lemmata presented in lemma niches and lemma nests, in the data distribution programme of the dictionary. However, as is the case in many learners' dictionaries, cf. COBUILD and the *Longman Dictionary of Contemporary English* (Sinclair (ed.) 1987 and Summers et al. (eds.) 1987), lexicographers refrain from using sublemmata due to the problems such a macrostructural procedure would cause for the typical target user of the dictionary.

## 4 Criteria for presenting macrostructural elements as either main or sublemmata

### 4.1 Simplex and complex words

A traditional approach in the macrostructural presentation in monolingual and bilingual dictionaries has been that simplex lexical items are included as main lemmata whereas complex items can be included as either main or sublemmata. However, the implementation of this criterion is complemented by a sub criterion applying to monolingual dictionaries only. In monolingual dictionaries the subsequent criterion that plays an important role follows from the semantic transparency of the complex item. Transparent items are typically included as horizontally-ordered lemmata, presented as either partial lemmata or full lemmata, cf. examples 4 and 7 respectively. These lemmata receive a restricted treatment whereas complex items which the lexicographer regards as semantically opaque are included as main lemmata with a full treatment. In example 4, the partial article stretch containing the partial lemmata *~agtig, ~bedryf, ~boor, ~delwer, … ~winning* implies that the complex lexical items *diamantagtig, diamantbedryf, diamantboor, diamantdelwer,… diamantwinning* are semantically transparent whereas the compounds *diamantbruilof, diamantklower* and *diamantkoper* are semanticaly opaque and therefore require a full treatment in the vertical ordering of main lemmata.

(4) **di·a·mant´** (-e) Harde, kosbare edelsteen, bestaande uit kristallyne koolstof: *(On)geslypte diamante.* [G. *adamas* staal; L. *diamas* onbedwingbaar;

G. *damao* mak maak] **diamant: ~agtig, ~bedryf, ~boor, ~delwer, ~gruis, ~houdend, ~kenner, ~ontginning, ~ring, ~vorming, ~winning.**

**di·a·mant´brui·lof** Herdenking van die 60ste verjaarsdag van ’n huwelik.

**di·a·mant´klo·wer** Ambagsman wat die klowersvlak van ’n diamant deur ’n fyn lyntjie aandui en dit daarlangs kloof.

**di·a·mant´ko·per** Iemand wat gemagtig is om ongeslypte diamante te koop en verkoop. (HAT)

Although this may be an established system used in many monolingual dictionaries its implementation implies two diverse grounds for the presentation and treatment of lexical items. Simplex items are included on account of their occurrence in the lexicon of the language and their frequency of usage qualifying them for inclusion in the lemma candidate list of the dictionary. Complex items also qualify on account of their usage frequency for inclusion in the lemma candidate list but their eventual inclusion and presentation in the macrostructure of the dictionary is determined by their morphological nature and certain aspects of their meaning. Different criteria of presentation and treatment apply to these different types of lexical items. For the lexicographer and the trained and knowledgeable dictionary user this does not pose a problem. The average dictionary user, however, has trouble grasping these different standards for lemmatization and will find it difficult to know exactly when the linguistic aspects, the morpho-semantic criteria, should override the lexicological criteria, i.e. the mere occurrence of an item as part of the lexicon of the treated language. The dictionary user has no objective criteria to ascertain where a given compound will be found: in the vertical or in the horizontal ordering of lemmata

Monolingual dictionaries often have clusters of unexplained complex items as niched or nested lemmata, cf. examples 3, 4 and 5. This scant treatment follows from the implied semantic transparency of these complex items. Semantic transparency is a lexical item specific feature, as the lexicalisation of items as either simplex or complex items is a language specific feature. The relation between two languages in terms of simplex and complex items is arbitrary. A simplex item in one language may have an equivalent that is also a simplex or its equivalent could be a complex item in a given target language. Consequently the presentation of “unexplained” complex items in monolingual dictionaries cannot have a parallel of “untranslated” complex items in bilingual dictionaries. The sublemmata in a bilingual dictionary often receive more or less the same treatment as the main lemmata in the vertical ordering, cf. example 6. Unfortunately a lack of consistency and an exaggerated space-saving approach too often lead to an inferior treatment of the lexical items included as sublemmata in a bilingual dictionary, cf. example 8:

(8) **earl´y,** vroeg, tydig; spoedig; vroegryp, vroegtydig; ~ *ARRIVAL,* vroeë (vroegtydige, voortydige) aankoms; *an HOUR ~*, 'n uur te vroeg; *KEEP ~ hours,* vroeg opstaan en vroeg na bed gaan; *(both) ~ and LATE,* vroeg en laat; *as ~ as OCTOBER,* al in Oktober; *the ~ PART,* die begin; *~ PEACHES,* vroeë perskes; *Early ROSE,* ellierous (spel); *in ~ TIMES,* in die vroegste tye; ~-**bearing,** vroeg(draend) (plant, boom); ~ **bird,** vroegopstaner; ~-**flowering,** vroeg(bloeiend); ~**ish,** vroeërig; ~ **mass,** vroegmis; ~ **retirement,** vroeë aftrede (uittrede); ~-**warning system,** vroegwaarskustelsel, vinnige alarmstelsel. (GW)

The treatment of the main lemma (*early*) includes among others the provision of translation equivalents and collocations. The treatment of the sublemmata attached to this article is restricted to the provision of translation equivalents. The morphological structure of these lexical items, i.e. their being complex words, influences the way in which they are presented macrostructurally but it also impedes the microstructural treatment they receive. Once again the lexicographic presentation is done on linguistic grounds with little or no attention to the user and his/her abilities to access the presented data. The lexicographic functions and the typical situation of dictionary usage, so important in a user-driven approach to lexicography, should have a much more compelling influence on the structure and presentation of data in a dictionary. Lexicographic work should not be characterised by a rigid adherence to linguistic conventions which are to the detriment of successful dictionary usage.

## 4.2 Idioms

Idioms constitute another type of lexical item often presented as sublemmata but not always recognized as such. One of the traditional ways of treating idioms is to include them in the article of a lemma representing one of the keywords of the idiom. Although such a presentation in fact eschews the status of idioms as independent lexical items, practical reasons make it difficult to lemmatise idioms as vertically ordered main lemmata, cf. Gouws (1989;1990). By including idioms in a separate text block, attached to or included in the article of a lemma representing a keyword taken from the idiom, the idiom does get the status of a macrostructural treatment unit, i.e. a horizontally ordered sublemma. Unfortunately the treatment allocated to idioms is usually of a lesser extent than the treatment allocated to other lemmata. Idioms are sublemmata because they can only be accessed via the entrance lemma, cf. the following examples:

(9) **egg,** (n) eier; ei *(vero.);* vrugbeginsel, ovarium; *BAD ~*, vrot eier; niksnuts, persoon met skadelike invloed, skobbejak; *put all one's ~s in one BASKET,*

al jou eiers in een mandjie sit; alles in een onderneming waag; alles op een kaart sit; *FRIED* ~, gebakte eier; *as FULL as an ~ is of meat,* propvol; *better an ~ today than a HEN tomorrow,* een voël in die hand is beter as tien in die lug; *NEWLAID ~(s),* pasgelegde eier(s); *POACHED* ~, geposjeerde eier, kalfsoog; *what has that to do with the PRICE of ~s?* hoe raak dit die saak? *SCRAMBLED ~(s),* roereier(s), struifeier(s); *as SURE as ~s are ~s,* so waar as vet; *TREAD on ~s,* op eiers loop; (v) met eiers gooi; met eiers bedek; eiers versamel; ~ *on,* aanhits, aanspoor, aanpor;.. (GW)

(10) **baad´jie** (-s, -tjie) Kledingstuk vir die bolyf, meestal vir mans, oor 'n hemp, trui of onderbaadjie maar onder 'n jas gedra. UITDR.: *Prosedeer oor 'n baadjie en hou jou* ***broek*** *klaar vir die onkoste,* as jy hof toe gaan, kan dit jou bankrot maak. *Iets met* ***broek*** *en baadjie kook* – sien onder BROEK. *Die* ***geel*** *baadjie (Jannie se baadjie) aanhê,* jaloers wees. *Op jou baadjie* ***kry,*** pak kry, verloor. *Iemand op sy baadjie* ***takseer,*** hom oppervlakkig, na die uiterlike beoordeel. *Baadjie* ***uittrek,*** jou regmaak om op te tree, veral om te baklei. **baad´jietjie. baadjie: ~knoop, ~kraag, ~sak.** (HAT)

In the bilingual dictionary, example 9, the idioms have been placed in the same text block as usage examples with no possibility for the user to distinguish between idiom and usage example, cf. Gouws (1996) for a more comprehensive discussion of this presentation. To illustrate the problem with which users are confronted in their dictionary consultation the idioms in example 9 have been underlined for the sake of this paper. The treatment of these sublemmata consists of one or more translation equivalents with no indication whether these translation equivalents are also idioms in the target language. From a text production perspective the users have very little guidance to assist them in their endeavour to produce texts that belong to the same style and register as the source language text.

Users experience fewer problems with the monolingual dictionary (example 10) because the idioms are all grouped into a single text block preceded by the structural indicator *UITDR.* (abbreviation for *uitdrukking* = expression). But still the treatment is restricted to an explanation of meaning without any attention to cotext or context guidance. Although the dictionary presents a consistent application of the system the system is not based on the needs prompting the typical usage situations. Users consult dictionaries to find out how to use idioms and this need is not addressed in this treatment. Yet again an outdated system impedes the type of knowledge transfer so desperately needed by the dictionary users. Idioms and other lexical items presented as sublemmata may not be treated in an inferior way just because of their macrostructural positioning.

## 5 Presenting lemmata in alphabetically ordered dictionaries

Looking at the way in which lemma candidates are presented in alphabetically ordered dictionaries, the following classification could be made:

- A strict vertically ordered alphabetical presentation of lemmata.
- A predominantly strict vertically-ordered presentation of lemmata with the presentation of idioms constituting the single deviation from the vertical ordering.
- A strict alphabetically ordered presentation of lemmata which allows a horizontal ordering to complement the vertical ordering, i.e. niching.
- An ordering where vertically-ordered lemmata follow a strict alphabetical presentation but where horizontally-ordered lemmata deviate from the strict alphabetical ordering, i.e. first and second level nesting. The deviation from the strict alphabetical ordering is done on morphological or semantic grounds or on morpho-semantic grounds.

Following from this it is quite clear that, for printed dictionaries, a traditional ordering system like the alphabetical procedure often is distorted, by a system determined by linguistic criteria, e.g. semantic transparency, morphological status, etc. In some dictionaries second level nesting also presents a kind of a thematic ordering where words that represent concepts belonging to the same semantic field are grouped together in a cluster. The application of linguistic criteria to determine the ordering or the presentation of lemmata is not restricted to sublemmata. In the presentation of homonyms a similar criterion is put to use. This, however, will not be discussed in the present paper.

## 6 Practical instruments

Dictionaries need to be regarded as practical tools in the hand of the specific user. The data presented in general monolingual and bilingual dictionaries primarily falls within the domain of linguistic data, and the typical dictionary user very often consults a dictionary for linguistic information, e.g. the meaning of a word, a target language translation equivalent, orthography, pronunciation, etc. Although a dictionary can be regarded as a container of primarily linguistic knowledge and although the user typically regards a dictionary as such, the lexicographer should refrain from employing an approach motivated on linguistic grounds for the access to lexical items included as lemmata in a dictionary. The access structure is a lexicographic and not a linguistic device. By utilizing both the alphabetical system along with a sinuous ordering lexicographers are hybridizing the access to data and this can so easily be to the detriment of the intended target user.

In the planning of a new dictionary the lexicographers should rather try to determine what the dictionary using skills of the intended target users will be and what way of presenting the lemmata would ensure successful access and an optimal retrieval of information. The typical user of a monolingual dictionary is not familiar with the notion of semantic transparency and would prefer similar search routes to find compounds like *honeybee* and *honeymoon.* The present system in many dictionaries actually compel the user to realize beforehand that *honeybee* is a transparent compound and would therefore be ordered according to different criteria than the opaque item *honeymoon*. This represents a form of macrostructural schizophrenia.

The macrostructural ordering problems in some bilingual dictionaries are increased by the use of procedures of multiple niching, cf. Gouws (2001). Multiple niching favours a strong adherence to the alphabetical ordering but allows the formation of niches. The lemmata that are candidates for a niche, attached to a single main lemma, are often separated and put into different niches due to their being alphabetically interrupted by either an unrelated main lemma or a complex lemma, belonging to the cluster of niche lemma candidates but singled out by the lexicographer for a more comprehensive treatment and an elevation to main lemma status, cf.:

(11) **bad,** (s) **(-dens),** bath, bathtub; swimming-bath; **(baaie),** mineral bath, spa, hot spring, mineral spring; *die baaie GEBRUIK,* take the waters; *die KIND 'n ~ gee,* bath the child; *'n ~ NEEM,* have a bath; *'n WARM ~*, a hot bath; (hot) mineral spring; (w) **(ge-),** take a bath, give a bath; ~**bande,** bath straps; ~**benodig(d)hede,** bathing requisites (requirements); ~**broek,** bathing drawers; ~**dery,** bathing; ~**gas,** visitor to a watering place; ~**greep,** grab bar; ~**handdoek,** bath towel; ~**handskoene,** bath gloves; ~**hokkie,** bathing-cubicle(-box); ~**huis(ie),** bathing booth; bath; bagnio.

**badineer´, (ge-),** banter.

**bad:** ~**inrigting,** bathing-establishment; ~**jie,** little bath; ~**kaartjie,** bathing-ticket; ~**kamer,** bathroom; ~**kamerspa,** bathroom spa; ~**kamervloer** bathroom floor; ~**ketel,** geyser; hot-water installation; ~**klere,** bathing-wear; ~**kom,** bath basin; ~**kostuum,** bathing-costume; ~**kraal,** bathing-enclosure; ~**kuip,** bathtub; ~**kussinkie, (-s),** bath pad(s); ~**kuur,** bathing-cure; ~**mat,** bath mat.

**badmin´ton,** badminton.

**bad´mus,** bathing-cap.

> **bad: ~olie**, bath oil; **~pak** = **badkostuum; ~plaas, ~plek,** bath; bathing-place; watering-place; health resort; seaside resort; **~seep,** bath soap; **~seisoen,** bathing-season; ~**(s)huis,** bathhouse; bathing booth; **bads´kamer** = **badkamer; ~skoen,** bathing-shoe; **~sout,** bath-salts; **~spons,** bath sponge; **~water,** bath water. (GW)

Such a system confuses the user even more because all the horizontally-ordered lemmata that should be expected to be candidates for one niche are spread over various niches, interrupted by related and unrelated main lemmata and presented in niches introduced not by a main lemma but by a lemma part in niche entrance position. In these dictionaries the alphabetical bias dominates the macrostructural ordering. Yet the lexicographer still feels the need to use some traditional space-saving mechanisms but the niching is done on a randomly selective basis with the lexicographer selecting certain niche lemma candidates to be promoted to main lemmata with a position in the vertical ordering of the dictionary. Judged from a user-perspective such a procedure is extremely unfriendly and one can imagine that the typical dictionary usage situation could so easily lead to a failed dictionary consultation procedure.

## 7 Lexicographic functions

One of the important more recent contributions in the field of theoretical lexicography is the emphasis on lexicographic functions, cf. Bergenholtz & Tarp (1995; 2002; 2003) and Tarp (2000). The compilation of any dictionary needs to be preceded by a clear identification of the functions that should prevail in the dictionary. Every component and aspect of a dictionary should fall within the scope of lexicographic functions. This does not only apply to the central list but also to the outer texts of a dictionary. In this regard Gouws (2004) and Gouws & Steyn (2005) argue in favour of a transtextual approach to lexicographic functions. Lexicographic functions should also prevail in the macrostructural selection, cf. Tarp & Gouws (2004), as well as in the presentation of these entries.

The inclusion of sublemmata as unexplained lemmata in a monolingual dictionary can be justified from a text reception point of view. If the dictionary has been compiled for mother-tongue speakers of the specific language then a mere listing of complex items in a cluster attached to the article of a main lemma could guide the users in their comprehension of the language. Where the particular main lemma represents a polysemous lexical item a system of nesting could be more beneficial than the presentation of a lemma niche because in the nest the sublemmata could be grouped according to that sense of the mutual item prevailing in the specific complex form. From a text production perspective the mere listing of sublemmata has to be regarded as insufficient guidance to achieve the purpose of the desired lexicographic function.

In bilingual dictionaries the lexicographer should be even more careful with the use of sublemmata. Where text production is the prevailing lexicographic function care should be taken to ensure a similar treatment of sublemmata as allocated to main lemmata. The space-saving procedure should be restricted to the macrostructural positioning of the lemmata and should not target their microstructural treatment. Yet again, the lexicographer should not rely on an often overestimated linguistic intuition of the target user to ensure a successful retrieval of information from the dictionary. A diminished treatment of sublemmata in bilingual dictionaries which is restricted to a translation equivalent without any context or cotext entries adds to the frustrations of the user consulting the dictionary for text production purposes.

## 8 Challenges for practical and theoretical lexicography

Theoretical lexicography and research in this field may never be isolated from the lexicographic practice and its needs and demands, cf. Wiegand (1996). The emphasis on the position of the dictionary user, the user situation and the usage situation so prevalent in modern-day lexicography should be noticeable in both theoretical and practical lexicography. Metalexicographers need to make a decision whether they are continuing their theoretical work by doing more of the same thing or whether they take a critical look at their current theoretical approach in order to determine whether the implementation of their theories will still benefit the user in the best possible way. This also applies to decisions regarding main and sublemmata. Metalexicographers need to work on a model that allows a clear and unambiguous identification and description of the different structural components of a dictionary, the different types of entries and the various data categories. This, however, should not be the ultimate aim of the theoretical work but only the means to an end to achieve the ultimate purpose, i.e. to devise a model that can assist the identified target user of a given dictionary with the kind of dictionary usage situation he/she needs and, consequently, a successful dictionary consultation procedure that leads to a successful retrieval of the kind of information that prompted the search.

When working on a theoretical level it is important to distinguish between main and sublemmata and to have a clear description of these concepts. This distinction should be of such a nature that practical lexicographers can interpret and apply it. In the lexicographic practice the selection of candidates to be entered as sublemmata should be done in an uncomplicated way, motivated by a single criterion and rather not in terms of criteria coming from diverse grounds. When a dictionary opts for a sinuous lemma file dictionary users should know, e.g., that all compounds will be included in a horizontal ordering, irrespective of their degree of semantic transparency, and that these sublemmata will be treated in a way that will assist the users sufficiently in the dictionary usage situations envisaged for the specific user group. In terms of the allocated treatment the distinction between main and sublemmata should not be seen as a distinction between first and second class treatment units. The lexicographer may still decide that some complex items do not qualify for a treatment due to their semantic transparency. Such items may be included as entries in an article slot set aside for results of word-formation processes with the lexical item represented by the lemma as first component. These entries are not presented as macrostructural elements.

## 9 In conclusion

Practical and theoretical lexicographers should take another close look at the way in which sublemmata are selected, presented and treated. Although it may be done in accordance with established methods, these methods often rely on a purely linguistic or even an arbitrary basis which no longer exhausts the needs of the target users. Every aspect of the lexicographic process, including the decisions regarding main and sublemmata, should be done to ensure the best possible results when real users really use the dictionary. This may prompt lexicographers to take a much stronger pragmatic approach with less attention to purely linguistic ordering principles in their sublemmatic endeavours.

## References

Bergenholtz, Henning and Sven Tarp (eds.) 1995. *Manual of Specialised Lexicography.* Amsterdam: John Benjamins.

Bergenholtz, Henning and Sven Tarp. 2002. "Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen." *Lexicographica* 18: 253–263

Bergenholtz, Henning and Sven Tarp. 2003. "Two opposing theories: On Wiegand's recent discovery of lexicographic functions." *Hermes* 31: 171–196.

De Villiers, Meyer et al. (eds.) 1987. *Nasionale woordeboek.* Cape Town: Nasou.

Eksteen, Louis C. et al. (eds.) 1997[14]. *Groot Woordeboek/Major Dictionary.* Cape Town: Pharos.

Gouws, Rufus H. 1989. *Leksikografie*. Cape Town: Academica.

Gouws, Rufus H. 1990. "Vaste uitdrukkings as multileksikale lemmas in verklarende Afrikaanse woordeboeke." *Tydskrif vir Geesteswetenskappe* 30 (4): 265–283.

Gouws, Rufus H. 1996. "Idioms and Collocations in Bilingual Dictionaries and their Afrikaans Translation Equivalents." *Lexicographica* 12: 54–88.

Gouws, Rufus H. 2001. "The use of an improved access structure in dictionaries." *Lexikos* 11: 101–111.

Gouws, Rufus H. 2004. "Outer texts in bilingual dictionaries." *Lexikos* 14.

Gouws, Rufus H. and Mariza Steyn. 2005. "Integrated outer texts: a transtextual approach to lexicographic functions." In *Schreiben, Verstehen, Übersetzen, Lernen: Zu ein- und zweisprachigen Wörterbüchern mit Deutsch,* I. Barz, H. Bergenholtz and J. Korhonen (eds.).

Labuschagne, F.J. and Louis C. Eksteen (eds.). 1993[8]. *Verklarende Afrikaanse Woordeboek.* Pretoria: J. L. van Schaik.

Odendal, Francois F. and Rufus H. Gouws. 2000[4]. *Verklarende Handwoordeboek van die Afrikaanse Taal.* Midrand: Perskor.

Sinclair, John (ed.) 1987. *Collins COBUILD English Language Dictionary.* London: HarperCollins.

Summers, Della et al. (eds.) 1987. *Longman Dictionary of Contemporary English.* Essex: Longman Group.

Tarp, Sven and Rufus H. Gouws. 2004. "Wie leer wat uit Afrikaanse (aan)leerderwoordeboeke?" *Tydskrif vir Geesteswetenskappe* 44(4): 276–298.

Tarp, Sven. 2000. "Theoretical challenges to LSP lexicography." *Lexikos* 10: 189–208.

Wiegand, Herbert E. 1989. "Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: Alphabetische Anordnungsformen und ihre Probleme." In *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography,* F. J. Hausmann et al. (eds.) 371–409. Berlin: De Gruyter.

Wiegand, Herbert E. 1996. "Zur Einführung." In *Wörterbücher in der Diskussion II*, H. E. Wiegand (ed.), VII-XV. Tübingen: Max Niemeyer Verlag.

Wiegand, Herbert E. 2002a. "Zur Makrostruktur und zu den äußeren Zugriffsstrukturen im de Gruyter Wörterbuch Deutsch als Fremdsprache." In *Perspektiven der pädagogischen Lexikographie des Deutschen II,* Herbert E. Wiegand (ed.) 413–442. Tübingen: Max Niemeyer.

Wiegand, Herbert E. 2002b. "Über textuelle Strukturen der Wörterbuchartikel und Artikelnischen im de Gruyter Wörterbuch Deutsch als Fremdsprache". In *Perspektiven der pädagogischen Lexikographie des Deutschen II,* Herbert E. Wiegand (ed.), 497–595. Tübingen: Max Niemeyer.

Wolski, Werner 1989. "Das Lemma und die verschiedenen Lemmatypen." In *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography,* F. J. Hausmann et al. (eds.) 360–371. Berlin: De Gruyter.

CHAPTER 5

# Semantic aspects of reversal of a set of bilingual dictionaries

Andrejs Veisbergs

New opportunities offered by information technology make it possible to apply new methods to lexicography. This paper describes a project that will result in compiling a new Latvian-English dictionary of unprecedented size and quality by reversing the definition entries of the many English-Latvian dictionaries currently in print. The paper discusses:

1. the current lexicographic situation, which forms the backdrop for the reversal project;
2. the process of reversal itself, conclusions about the reversal project, its pros and cons, problems and findings; and
3. corollary conclusions about the contents and structure of dictionaries used for reversal. The last area is in some ways the most interesting, as such projects might have a future in Latvia and elsewhere.

## 1 The current situation

Current Latvian lexicography lacks a comprehensive language corpus. This means that bilingual dictionaries with Latvian as the source language (the A part) endemically suffer from a smaller number of entries compared to the second language (the B part). The absence of a corpus gives ground to various speculation about the current state of the lexicon and also feeds the purist tendencies still rife in the country. Moreover, it entails an absence of information about the lexis as such – about frequencies of usage, meanings, stylistic registers, diachronic changes, e.g. the most frequent word of greeting in Latvian for the last 50 years, *čau,* was not recorded in a dictionary until 1996. This information is essential for lexicographers working on bilingual dictionaries with Latvian as the source language.

Limited lexicographical presence of new and not-so-new vocabulary in turn leads to multiple variants of many frequently used terms and notions, because they are not "fixed" in dictionaries. This deepens the chasm between the real language and the one reflected in dictionaries. The possible use of Latvian-language internet material as a corpus (Grefenstette 2002) seems to be less advantageous than reversal of bilingual dictionaries with Latvian as the B language, because of the variety mentioned above and the small proportion of terminological texts in the corpus. Latvian bilingual dictionaries are in high demand at present, as new language contacts are being established and huge amounts of translations (including EU texts) are being produced.

Consequently, all possibilities of expanding the volume of the lexicon in dictionaries should be pursued, and one of these is the reversal of existing English-Latvian dictionaries. For historical reasons, it so happens that these dictionaries have always tended to be better and broader in scope compared to Latvian monolingual ones, which were always affected by purism (Latvian as well as Soviet) and various restrictions of colloquial, borrowed and non-standard lexis. The use of reversing techniques can lead to substantial volume growth of dictionary entries as well as updating and finding "missed" words and translations.

However, it must be mentioned that reversal has a relatively negative reputation in Latvia, because several companies offer sometimes flawed bi-directional English and Latvian online dictionaries (see www.tilde.lv/dictionary and www.dictionary.site.lv) based on reversed, monodirectional (i.e. for Latvian users) dictionaries. Unedited reversals into Latvian often lead to serious, and anecdotal, mistakes such as impossible back-translations, strange Latvian entries and long lists of undifferentiated idiographic and stylistic synonyms. Unclear, unmarked and semi-alphabetic enumeration of English definitions when looking for a translation of a Latvian word are most confusing for the uninitiated.

The English translations of, for example, *televizors* ('television set') are provided in the following order: *goggle box, idiot box, television receiver, television set, televisor, telly, television, box, tube box, boob box, teevee*; and the entry for *skaists* ('beautiful') offers the following enumeration: *beauteous, pulchritudinous, beautiful, lovely, gallant, seraphic, likely, fair, sheen*.

Non-experts and people used to printed dictionaries have a tendency to use the first variant in the list. In addition, the amount of lexis offered (rarely labelled and with no extra information) is too big for a normal user. As a result, lay users (unaware of the dictionary reversal business) consider the electronic dictionaries dubious, while experienced users find reversal dangerous per se.

## 2 Reversal

There are different views on the efficiency of dictionary reversal. A study of the reversal of an English-Swedish dictionary suggests that the economy of time and money is not particularly great, because only about 30% of the reversed material could be retained (Geisler 1999). On the other hand, Estonian experience (Tamm 2002) reports successful solutions in reversing a Dutch-Estonian dictionary. It is inevitable that serious editing is required, which may be so time-and-effort consuming that it cannot be considered worthwhile. Yet one also has to consider the fact that, for well-known reasons and as alluded to above, smaller languages (like Latvian) normally have better bilingual dictionaries *from* than *to* other languages. These are normally based on the use of several large, high-quality, monolingual source-language dictionaries (such as Oxford, Duden, Hachette) worked upon by local language experts with expert technical assistance. As a result the A part is generally *comprehensive* and the B part fairly *accurate and acceptable*. It seems reversal is especially good for large dictionaries which contain vast terminological material that is predominantly monosemous – 35–45% of the total number of entries according to Corda et al. (1998: 442) – and consequently undergoes reversal more easily.

The project with TILDE, the largest Latvian software company, was carried out in two stages. The first stage included reversing an English-Latvian dictionary (the notorious internet version) as well as many English-Latvian terminological dictionaries (most Latvian terminological dictionaries exist in this form) and imposing this material on *the existing entries* of the Latvian-English dictionary (Veisbergs 2001), thus enlarging the material. Some simplification was done in the reversal process, like removing some bracketed grammatical information, and reversed phrases were often not found under the entry where one would expect them. In the first stage of editing, the imposed part was easily recognisable (in a different font) bearing the tags from the source dictionaries, including field labels. This was followed by thorough editing (performed by linguists), consisting mainly of deleting endless equivalent meanings from various thematic fields, expunging redundancies, reordering senses and removing unnecessary working labels and tags, e.g. *dog* (customs), which signifies a dictionary of customs terms as the source. The result was a 25–30% increase in the volume of material. The second stage (now underway) consists of *adding new entries* from the reversed, mainly terminological sources, where one deals mainly with monosemous language pairs and the editors' main task lies in deciding whether the term is sufficiently frequently used to be included in a dictionary for the general user.

Below is a sample from letter S of the combined product (the material of the existing Latvian-English Dictionary in Times New Roman, the added material in Courier):

**saturīgs**
1. pithy;
2. `informative`;
3. `sapid`;
4. `thoughtful`

**savienojums**

1. joining; connection; junction;
   telefona savienojums – connection; put through;
   dabūt savienojumu – to get* through;
   elektrības savienojums – contact;
   savienojuma vieta – *formāl* juncture;
   savienojuma zīme – (Morzes ābecē) dash; hyphen;
2. *ķīm mat* combination; compound;
3. *tehn* coupling; union;
4. *filoz* connection;
5. combination;
6. tehn.joint;
7. bracing; brace;
8. *ķīm tehn el* junction;
9. *ķīm* linkage;
10. *tehn* meeting;
11. *ķīm tehn el* tie;

**savijums**

1. curlicue;
2. kink;
3. (tauvu) splice;
4. tangle;
5. contexture;
6. twist;

**savstarpējs a**

1. mutual a; reciprocal a;
   savstarpēja atkarība – interdependence; interconnectedness;
   savstarpējās attiecības – interrelation;
   savstarpēja saistība – correlation; interconnectedness; interrelation;
   **savstarpēja apmaiņa** – interchange;
   **savstarpēja iedarbība** – interaction; interplay;
   **savstarpēja piekāpšanās** – give-and-take;
   **savstarpēja sakarība** – interconnection;
   **savstarpēja saskaņa** – agreement;
   **savstarpēja slavināšana –** log-rolling *amer*;
   **savstarpējs sakars** – interconnection;
2. mutual relations;
   savstarpēji saistīts – correlated; interconnected;

savstarpēji sakari – intercommunication;
savstarpējās atzīšanas līgumi – *pol.* mutual recognition agreements;
iestāžu savstarpējie maksājumi – payments among institutions;

## 3 Conclusions about the reversed dictionary

Apart from purely technical problems, some of which took a lot of time to solve, such as difficulties matching the various systems of the dictionaries, and differing methods of variant inclusion (slashes versus brackets), the editing work offered interesting insights not only into the reversal process, but also concerning the original entry systems and lexical material (see below). Some of the problem issues interfering with smooth reversal and editing were comprehensive enough to be considered endemic in reversal cases. These fall into three main groups:

1. A multitude of synonyms, much repetition, redundancies (Newmark 1998) that have to be automatically or manually deleted.
2. The reversed material contained numerous definition-style translations (useless for productive reversal) instead of equivalents.
3. Connotative inaccuracies (lack of equivalence), the result of doubtful strategies in the source dictionaries.

One of the main problems in reversal is the long lists of entries with endless synonyms, only a few of which are adequate. These long lists come as a result of:

1. Overly generalised translations in source dictionaries;
2. Overly neutralised translations (loss of connotation);
3. Imprecise translations; or
4. The situation when Latvian has no precise counterpart, or the compilers have not been aware of it and have used a near-synonym.

What often emerge are huge groups of entries containing unimportant auxiliary components – again useless for reversal: *tāds, kas..; neliels...; cilvēks, kas...; bez....; būt....; izdarīt...;ar...;* ('that which…'; 'small…'; 'a person who…'; 'without…'; 'to be…'; 'to do…'; 'with…'):

– *tāds, kas izdala mitrumu* ('that which produces moisture');
– *tāds, kas iznāk divreiz mēnesī* ('that which is published twice a month')
– *tāds, kas lauž starus* ('such that breaks rays');
– *būt par partneri* ('to be a partner');
– *būt par pierādījumu* ('to serve as proof');
– *būt par rakstnieķeli* ('to be a poor writer').

This suggests that bilingual dictionaries, instead of translation equivalents, often operate with the definition types characteristic of monolingual explanatory, or encyclopedic, dictionaries. One of the reasons for this may be the use of monolingual dictionaries in the process of compiling bilingual dictionaries. Similarly, some idioms and set expressions tend to have explanatory translations, some of which are useless for reversal:

– *vientulības (izolācijas) trūkums* ('lack of solitude, isolation') – *goldfish bowl;*
– *Vinčesteras universitātes studentu izteiciens, paraža vai tradīcija* (an expression, saying or tradition of Winchester University students) – *notion.*

A similar, yet somewhat different problem is presented by slang – for two reasons:

1. slang is ephemeral and subject to rapid diachronic changes in both languages and consequently difficult to match; and
2. in many bilingual (including English-Latvian) dictionaries, English slang words tend to have rather neutral or mildly colloquial Latvian equivalents. As a result, for the reversed dictionary entry *ļoti* ('very'), 95% of the translations were English slang words of the intensifier type. The Latvian entry *nauda* ('money') was followed by about 50 slang counterparts, and so on, not to mention the enormous volume and scope of the usual derogatory terms.

The problem of stylistic non-equivalence can also be seen on a broader scale –numerous translations distort the connotative meaning of the entry. Most bilingual dictionaries tend to move from connotatively strong items in the source language towards a more neutral/standard register in the target language. When reversed, connotatively neutral words tend to acquire numerous bright and expressive translations (the English back-translation is provided by the author in brackets):

– *šūšanas rūpniecība* ('textile industry') – *rag trade;*
– *entomologs* ('entomologist') – *beetle sticker;*
– *izdarīt* plaģiātu ('to plagiarise') – *to crib;*
– *nieres* ('kidneys') – *waterworks* [not a very precise denotative equivalent];
– *čiks* ('nothing') – *fuck all.*

The connotative difference between denotative equivalents requires very thorough editing. While in the original English-Latvian dictionaries the neutralising approach is not disastrous and in some cases even acceptable, in a reversed dictionary it becomes a serious problem, distorting any equivalence. For example, *rag trade* in a serious business context, *waterworks* at a medical conference, or *fuck all* at the end of a fairy tale all speak for themselves.

Similar deviation can also be observed on the denotative level – in the case of non-equivalence, the translated meaning of the entry usually errs by becoming generalised. In reversal, accordingly, semantically broad entries are given narrower, more concrete counterparts:

– *platforma* ('platform') – *catwalk;*
– *iemīļots dzīvnieks* ('favourite animal') – *pet;*
– *transporta tunelis zem jūras šauruma* ('transport tunnel under a sea straight') – *Chunnel.*

One has also to consider the time frame when reversing dictionaries. As language is in a state of flux, some words and terms may have dated variants and definition-type equivalents which have been replaced by new terms today, as found in these entries:

– *kapāti dārzeņi ar majonēzi* ('cut vegetables with mayonnaise') – *tartar sauce* [today *tatāru mērce*];
– *remonta punkts autosacīkstēs* ('repairs point in car races') – *pit* [today *pitbokss*]

Finally, an interesting problem occurs when simultaneously reversing material of many dictionaries. When creating a cluster of (terminological) dictionaries, compilers obviously make use of existing bilingual dictionaries without avoiding the obvious errors which are repeated again and again in the reversed material. The editors seem to have lured by the fact that even if all their sources point to one and the same equivalent, this equivalent is sometimes wrong.

## 4 Corollary conclusions about the source dictionaries

Regarding input material, one of course has to consider that the dictionaries used for reversal were not planned or designed for such a future operation, and that the inadequacies should not be viewed as a reproach against the original compilers and editors. However, reversal makes it possible to see clearly some tendencies and flaws.

The most evident problem is connotative neutralisation. One can state that the Latvian part of the English-Latvian dictionaries tends to be simplistic and bland, avoiding nuanced equivalents. This is not obvious when looking at separate words, but it is evident in a larger reversed corpus: Here are some examples from the English-Latvian dictionary and (to the right), suggested, more exact, equivalents:

– ***weird*** *dīvains, savāds* *ērmīgs, ērmots, spocīgs*
– ***quaint*** *dīvains, savāds* *īpatns, vecmodīgs*
– ***odd*** *savāds, dīvains* *jocīgs, "īpatnējs"*

One is left with the impression that English is so much more varied and that Latvian has a rather primitive and connotatively limited set of words, which is not true. And this seems typical of most bilingual dictionaries (Veldi 2004: 533). I would argue that bilingual dictionaries should not simply provide denotative equivalents but also stress the connotative component. This refers both to the lower and higher layers of lexis and could be done by first giving the neutral counterpart, then adding a connotationally brighter counterpart. This would to some extent be a strategy for dealing with slang – the neutral counterpart would render the denotational meaning even if the target language slang equivalent were to change or become incomprehensible to the user, taking into account the fast change of slang terms.

Similar deviation can be also observed on the denotative level. With nouns, for instance, the meaning of the entry is usually generalised (discussed above), often with awkward definitions that focus not on the essential functions, but rather on a visual description:

- *disks uz kārts* ('disk on a pole') – *lollipop;*
- *dzeltenbrūngans ābols* ('yellowish brown apple') – *russet apple;*
- *liels apelsīns* ('a large orange') – *jaffa.*

There are also some non-existent English neologisms created on the basis of Latvian words, mostly found in one dictionary (ELDO 1995):

- *konkurētspēja – competitive capacity* (correct: *competitiveness*);
- *kravnesība – load carrying capacity* (correct: *cargo (carrying) capacity*).

In general we conclude that this way of augmenting existing bilingual dictionaries may be worth the effort. A serious increase has been observed in the number of entries and in the number of meanings and collocations (missed, new, terminological). For smaller languages whose lexicographical projects are often not commercially interesting, dictionary reversal allows a fast and considerable "beefing up" of existing resources and augmenting the list of existing words. The latter can also involve dictionaries of other language combinations (even if it is no secret that the source language tends to affect the Latvian part in a dictionary). On the other hand, there are inherent limitations, most of which were discussed above, which with many language combinations (having reliable and large monolingual and bilingual corpora) might render reversal impracticable and too time-consuming.

If the method of reversal is used in the future, hopefully one of the consequences of the present project would be that compilers avoid explanatory translations in bilingual dictionaries. Whenever this is not possible, the target component should give the nearest synonym and then explain the differences. Such an approach would provide semantically more coordinated material. Also more attention should be paid to the stylistic register and its retention in the translated part. Finally, the new dictionary should be available in both printed and electronic versions.

## References

Corda, A. et al. 1998. "Automatic Reversal of a Bilingual Dictionary: Implications for Lexicographic Work". In *Proceedings of the Eighth EURALEX International Congress.* Th. Fontanelle et al. (eds), 433–443. Liege: University of Liege.

*ELDO* 1995. Riga: LVI

Geisler, C. 1999. "Reversing a Swedish-English dictionary for the Internet". In *Language and Computers. Parallel corpora, parallel worlds. Selected papers from a symposium on parallel and comparable corpora at Uppsala University, Sweden, 22–23 April, 1999.* Lars Borin (ed.), 123–133. Amsterdam: Rodopi.

Grefenstette, G. 2002. "The WWW as a resource for lexicography". *In Lexicography and Natural Language Processing. A Festschrift in Honour of B.T.S. Atkins.* M. H. Corréard (ed.), 199–215. Grenoble: Euralex.

Newmark, L. 1998. "Reversing a One-Way Bilingual Dictionary". In *Proceedings of the Eighth EURALEX International Congress.* Th. Fontanelle et al. (eds), 405–410. Liege: University of Liege.

Tamm, A. 2002. "Reversing the Dutch-Estonian dictionary to Estonian-Dutch". In *Proceedings of the Tenth EURALEX International Congress,* A. Braasch and C. Povlsen (eds.), 389–400. Copenhagen: CST, University of Copenhagen.

Veisbergs, A. 2001. *The New Latvian English Dictionary.* Riga: EAI.

Veldi, E. 2004. "Phonosemantic Accuracy in Bilingual Dictionaries: English and Estonian Derogatory Words". In *Proceedings of the Eleventh EURALEX International Congress,* G. Williams and S. Vessier (eds.), 531–535. Lorient: UBS.

Part 3

# Phraseology in dictionaries

CHAPTER 6

# Idiomatische Äquivalenzprobleme: Ein ikonoklastischer Zugang

Ken Farø

In bilingual lexicography, the issue of idiom equivalence plays a remarkably small role within theoretical discussions. Nevertheless, it is a matter deserving much more consideration from bilingual lexicographers: On exactly what grounds are interlingual idiom "pairs" taken to be equivalent? It seems that this decision is very often taken on the basis of two factors that are of little use in this case, namely 1) *component quality and structure* of the idioms, and 2) *intuition* of the lexicographer. It should be considerered whether or not other – more suitable – methods exist that allow for functional decisions about idiom equivalence. A number of German-Danish dictionaries are used as a means of discussing this problem and pointing out possible solutions.

Die Frage nach der Äquivalenz spielt eine erstaunlich geringe Rolle in der Diskussion über Idiome in der bilingualen Lexikographie. Dabei ist es im Grunde genommen eine Frage, die alle bilingualen Lexikographen beschäftigen sollte. Denn auf welcher Grundlage wird eigentlich entschieden, ob ein interlinguales Idiom-"Paar" äquivalent ist? Meine Annahme ist, dass dies in sehr vielen Fällen aufgrund von zwei Faktoren geschieht, die zu diesem Zweck gar nicht geeignet sind, nämlich 1) *Komponentenqualität und -struktur* der Idiome, und 2) *Intuition* des Lexikographen. Dabei wäre zu überlegen, ob es nicht andere – zuverlässigere – Methoden gäbe, idiomatische Äquivalenzrelationen zu bestimmen? Anhand einer breiten Auswahl an Wörterbüchern mit Dänisch und Deutsch wird diesem Problem nachgegangen und gleichzeitig auf mögliche Lösungen hingewiesen.

## 1 Einleitung

In der metalexikographischen Literatur zur Phraseologie sind u. a. folgende Defizite festzustellen: Erstens steht meistens der monolinguale Aspekt im Zentrum der Bemühungen (vgl. u. a. Schemann 1989; Farø 2004c). Zweitens kann beobachtet werden, dass das Problem der *Äquivalenz* in den wenigen Arbeiten zu bilingualer Lexikographie entweder kaum eine Rolle spielt oder zumindest in einer eher postulierenden, theoretisch kaum fundierten Weise behandelt wird (so z. B. bei Schemann 1991; Farø 2000b). Auch Dobrovol'skij (1997: 46) stellt fest, dass der spezifisch lexikographische Anteil interlingualer Arbeiten zur kontrastiven Phraseologie sehr gering ist. Dieses Problem bildet den Ausgangspunkt des vorliegenden Beitrags. Für umfassendere theoretische Ausführungen und empirische Analysen sei auf die Dissertation Farø (2006) hingewiesen.

„Äquivalenz" heißt etymologisch bekanntlich so viel wie ‚gleicher Wert' (vgl. Kluge 2002). Nun sind Werte etwas Relatives, nämlich das, was ihnen das System an Funktionen beimisst. Ein klassisches Beispiel für relative Werte ist das Geld, das Saussure (1916: 94f.) mit der Sprache vergleicht. Der Wert des Geldes ist eine völlig konventionelle Größe: Obwohl jemand, der in einer Gemeinschaft aufwächst, in der z. B. mit Zahlen bedruckte Metallstücke und Papierscheine irgendwie „wertvoll" sind, sich schwer vorstellen kann, dass die Geldsubstanz an sich keinen absoluten Wert hat, ist die Substanz in der Tat gar nicht wesentlich – Muscheln täten es z. B. auch. Hauptsache diese sind von der Gemeinschaft als Werte akzeptiert. Die Tatsache, dass es sich um Metall oder Meeresprodukte handelt, ist für ihre Funktion im Grunde genommen irrelevant. Mit den Sprachzeichen verhält es sich genau so: Ihre Werte können nicht ohne Kodebezug ermittelt werden, ihre aktuellen Bedeutungen oder Funktionen sind weder logisch noch natürlich, wie der Sprachwandel und -vergleich zur Genüge zeigen.

In der Sprache existiert aber ein Phänomen, das oft so behandelt wird, als würde es sich dabei um absolute Werte handeln, und zwar die Idiomatik. Idiome lösen bei Linguisten, darunter auch Lexikographen, häufig eine Art „Bilderfetischismus" aus. In der bilingualen Lexikographie wird nicht selten mehr Wert auf Bilder als auf den Sprachgebrauch gelegt. Und es wird oft davon ausgegangen, dass das Bild an sich die Basis einer Äquivalenzrelation ausmachen kann. Dies ist methodisch und theoretisch problematisch und für den Wert eines Wörterbuchs als Kommunikationswerkzeug häufig direkt schädlich.

Ich möchte in diesem Beitrag zeigen, wie die Idiomatik in der bilingualen Lexikographie ein Äquivalenzproblem ersten Ranges darstellt. Dabei möchte ich als Ausgangspunkt für einen teilweise „ikonoklastischen" Umgang mit Idiomen im Wörterbuch plädieren. Ein solcher Umgang würde den kommunikativen Bedürfnissen des Benutzers m. E. eher entgegenkommen als der Drang des Lexikographen nach einer Art „Bilderharmonie". Es handelt sich nämlich bei der so genannten „Bildhaftigkeit" (vgl. dazu Burger 1989 und 2003) in den allermeisten Fällen um ein Epiphänomen der Sprachverwendung, und keineswegs um ein zentrales kommunikatives Element an den Idiomen. Gerade weil die Idiome im zweisprachigen Wörterbuch auf anderen Ebenen ein immenses Äquivalenzproblem ausmachen, wie etwa in Semantik, Syntax, Valenz oder Frequenz, würde sich ein alternativer Fokus lohnen. Zumindest wenn man davon ausgeht, dass es sich bei einem bilingualen Wörterbuch um ein Werkzeug handelt, das primär als Stütze bei Kommunikationsproblemen dient.

Nach einer kurzen Reflexion über die Relevanz des Begriffes „Pseudoäquivalenz" im Wörterbuch bespreche ich eine kleine Auswahl an zentralen Äquivalenzproblemen, die erfahrungsgemäß eine wichtige Rolle spielen. Dann wird der Begriff „idiomatische Äquivalenznormen" der Lexikographie eingeführt und mit den festgestellten lexikologischen Äquivalenzproblemen konfrontiert. Als nächster Schritt wird die Frage gestellt, inwiefern diese beiden Phänomene für die bilinguale Lexikographie relevant sind. Zum Schluss des Beitrags gebe ich eine Antwort auf die Frage „Wann ist ein lexikologisches ein lexikographisches Äquivalent?", indem ich diskutiere, wann ein idiomatischer Übersetzungsvorschlag auch ein lexikographisches Äquivalent ist.

## 2 „Pseudoäquivalente"? Beispiele für lexikologische Äquivalenzprobleme

Gibt es so etwas wie idiomatische „Pseudoäquivalente" im zweisprachigen Wörterbuch (vgl. Farø 2000a, 2000b)? Ja und nein. Denn um in einem bilingualen Wörterbuch von „Pseudoäquivalenz" sprechen zu können, müsste zunächst geklärt werden, welche lexikographischen *Funktionen* das Wörterbuch zu erfüllen beansprucht. „Pseudoäquivalenz" an sich gibt es nicht, sondern es gibt lexikologische, translatorische und lexikographische Pseudoäquivalenz (vgl. Farø 2006). In den beiden letzteren Fällen muss sogar jeweils relativ zur Übersetzungsstrategie und zu den lexikographischen Funktionen entschieden werden, ob „Pseudoäquivalenz" vorliegt oder nicht. Versuchen wir zunächst, uns ein Bild darüber zu machen, worin idiomatische „Pseudoäquivalenz" im *lexikologischen* Sinn überhaupt bestehen kann, und wie umfassend das Phänomen in einem Sprachenpaar wie Deutsch und Dänisch ist. Als Beispiele werden Idiome aus neueren Wörterbüchern herangezogen.

*Semantische* Äquivalenzprobleme sind verbreitet bei Idiomen. Die Ursache liegt häufig darin, dass sich in den Sprachgemeinschaften die Bedeutung des einen oder beider ursprünglich semantisch äquivalenter Idiome geändert hat. Dies trifft z. B. auf das konvergente Idiom-"Paar" *die Ohren steif halten* vs. *holde ørene stive* zu. Wenn man von den Informationen der Wörterbücher ausgeht, könnte man den Eindruck gewinnen, es handele sich um absolute lexikologische Äquivalente (vgl. u. a. POL, GY, LA, GA2, MU). Dies ist aber nicht der Fall, denn das dänische Idiom wird heute kaum mehr in der Bedeutung ‚den Mut nicht verlieren' (DUW) verwendet. Somit kann nicht von semantischer Äquivalenz gesprochen werden, zumindest nicht im lexikologischen Sinn. Andere Fälle solcher semantischen Äquivalenzprobleme sind etwa:

(1) sætte ngn stolen for døren ≠ jm den Stuhl vor die Tür stellen
(2) holde ngn stangen ≠ jm die Stange halten
(3) ved det grønne bord ≠ am grünen Tisch
(4) over stok og sten ≠ über Stock und Stein
(vgl. Farø 2004b)

Auch die *Bedeutungsstruktur* spielt bei Idiomen eine wichtige Rolle als lexikologisches Äquivalenzproblem. In einigen Fällen entspricht der semantische Wert eines ausgangssprachlichen Idioms gleich zwei verschiedenen zielsprachlichen Idiomen. Dies hat sich z. B. bei einer Untersuchung des von dänischen Lexikographen viel diskutierten Idioms *gå over åen efter vand* gezeigt (wörtlich etwa: „den Fluss/Wasserlauf überqueren, um Wasser zu holen"). Bei diesem Idiom kommen offensichtlich jeweils zwei deutsche lexikologische Äquivalente in Frage. Nämlich einerseits *mit der Kirche ums Dorf laufen* für das „Umständlichkeitssem", und andererseits *Wasser ins Meer tragen* für das „Überflüssigkeitssem". Beide Seme kommen im deutschen Idiom zum Tragen und können erst durch feine semantische Analysen diskriminiert werden (vgl. Farø 2004a). Auch *Schmetterlinge im Bauch haben* vs. *have sommerfugle i maven* verursachen Probleme, weil das „Verliebtheitssem" des deutschen Idioms im Dänischen nicht kodiert ist. Nur bezüglich des allgemeineren „Nervositätssems" sind die beiden Idiome lexikologisch äquivalent.

Es ist außerdem das Problem der *Selektion von Kontextelementen* zu erwähnen. Eine Untersuchung (Farø 2004a) hat ergeben, dass die im Allgemeinen als Äquivalente aufgefassten Idiome *etw an den Nagel hängen* und *lægge ngt på hylden* („etw aufs Regalbrett legen") keineswegs eine eindeutige lexikologische Äquivalenzbeziehung eingehen. Die Analyse einer größeren Menge von Korpusbelegen zeigt, dass die Objekte, die das deutsche Idiom selektiert, inhaltlich nicht mit den vom dänischen Idiom selektierten Objekten übereinstimmen. Während die häufig metonymisch ausgedrückten Objekte des deutschen Idioms immer, eine permanente Aktivität, die oft ausgeübt wird' bezeichnen, kommen beim dänischen zwei weitere Objekttypen vor, die man einerseits als eine ‚schlechte Gewohnheit', andererseits als ein ‚noch nicht realisiertes Projekt' beschreiben kann. Die Korpusdaten ergeben folgende Systematisierung[1]:

(5) *etw an den Nagel hängen:*
den Job (5x), den Beruf (3x), die Karriere (2x), die Gesangskarriere, die Arbeit, das Christ-Sein, den weißen Kittel, Tennis und Hockey, die Fußballschuhe, die Dienstwaffen, die [Priester-] Kutte, den Rennfahrerhelm, das Fliegen

– alle: 'eine Aktivität, die häufig ausgeübt wird [oft metonymisch ausgedrückt]'

(6) *lægge ngt på hylden:*

1 fodboldstøvlerne (,die Fußballschuhe'), handskerne (,die Handschuhe'), ketsjeren (,den Schläger'), bandet (,die Band'), hvepsetoxinerne (,die Wespentoxinen') – = etw an den Nagel hängen

2 fløjlshandsken ('den Samthandschuh'), cigaretten (,die Zigarette'), cigaretterne (,die Zigaretten'), salaten (,den Salat'), skvatmageriet og cykelhjelmen (,die Warmduscherei und den Fahrradhelm'), deres lesbiske markedsføringsforhold (,ihre lesbische Vermarktungsbeziehung'), stofmisbruget (,den Drogenmissbrauch'), sin hang til Elton John og blød 70'er-rock (,seinen Hang zu Elton John und softem Siebzigerjahre-Rock'), benzodiazepinerne (,die Benzodiazepine'), antihistaminerne (,die Antihistamine') – 'etw was als eine schlechte Gewohnheit aufgefasst wird'

3 eksklusivforslag ('Exklusivvorschläge'), filmen (,den Film'), kopimad (,Copy-Food'), Lov NT4 (,das Gesetz NT4'), Fun-sektionen (,die Fun-Sektion') – 'ein noch nicht realisiertes Projekt'

Von einer absoluten lexikologischen Äquivalenz kann hier also nicht gesprochen werden, obwohl diese Annahme aufgrund einer rein formalen, isolierten und intuitiven Analyse des Idiompaars nahe liegen würde.

Ich habe bisher einige Beispiele lexikologischer Äquivalenzprobleme in der dänisch-deutschen Idiomatik dargestellt. Mein Anliegen damit ist zu zeigen, auf wie vielen Ebenen Äquivalenzprobleme vorkommen können: Mindestens 20 unterschiedliche Äquivalenz*aspekte* können für die interlinguale Idiomatik registriert werden (vgl. Farø 2006). Darunter sind u. a. Diachronie (*jm vor den Kopf stoßen* vs. *støde ngn. for hovedet* (archaisch), Register (*tage ngt med et gran salt* („etw. mit einem [gran] Salz nehmen") vs. *etw cum grano salis nehmen*), lexikalische Kategorie (*have ål i sokkerne* („Aale in den Socken haben") vs. *heruntergerutschte Socken haben*) und (Semanto-)Frequenz (*ngt. er en by i Rusland* („etw. ist eine Stadt in Russland") vs. *etw sind (für jn) böhmische Dörfer*). Möglicherweise ist die Zahl solcher lexikologischer „Pseudoidiompaare" im Deutschen und Dänischen sehr viel höher, als im Allgemeinen angenommen wird. Die weit verbreitete Annahme, die lexikalischen Inventare zweier Sprachen würden relativ problemlos zusammenpassen, wird bei Idiomen als eine Illusion entlarvt.

Mit diesen Feststellungen ist aber noch keine Aussage über die Äquivalenz von Idiomen in der bilingualen Lexikographie gemacht worden. Der Frage, inwiefern solche idiomatischen Äquivalenzprobleme auch eine Crux für zweisprachige Wörterbücher darstellen, möchte ich mich jetzt zuwenden.

## 3 „Lexikographische Äquivalenznormen". Versuch einer Erklärung

Angesichts der großen Zahl an möglichen und vielfach auch empirisch feststellbaren lexikologischen Äquivalenzproblemen bei Idiomen möchte ich auf ein bemerkenswertes Phänomen aufmerksam machen. Es handelt sich um „lexikographische Äquivalenznormen". Diese definiere ich hier als eine signifikante Tendenz der bilingualen Wörterbücher, bei gewissen Idiomen immer die gleichen Äquivalentvorschläge zu geben. Dass solche Äquivalenznormen existieren, mag zunächst selbstverständlich erscheinen, kann es doch logischerweise nur *eine* Beziehung zwischen Zeichen zweier Sprachen geben. Folglich muss es auch *eine* optimale lexikographische Beschreibung dieser Beziehung geben. Nun wurde aber soeben festgestellt, dass es zwischen Idiomen eine erstaunlich große Zahl an rein lexikologischen Äquivalenzproblemen gibt. Unter den lexikographischen Normäquivalenten sind vielfach auch solche vertreten, die lexikologisch keineswegs unproblematische Äquivalente sind, wie z. B. das Idiompaar *etw an den Nagel hängen – lægge ngt på hylden*, das in 10 von 10 untersuchten Wörterbüchern als äquivalent aufgeführt wird (POL, GY1, GY2, PO, HØ, MU, GA1, GA2, LA, TA). Dass die Wörterbücher somit fehlerhaft wären, möchte ich an dieser Stelle noch nicht behaupten. Dennoch ist diese eiserne Homogenität angesichts der rein lexikologischen Voraussetzungen zunächst einmal recht erstaunlich.

Für solche lexikographischen Äquivalenznormen muss es eine Erklärung geben. Dass der Hintergrund nicht empirisch-lexikologische Untersuchungen sein können, ist klar, denn aus solchen wäre ja hervorgegangen, dass absolute Äquivalenz nicht gegeben ist. Meine vorläufige These ist, dass die Erklärung für solche lexikographischen Normäquivalente bei gleichzeitiger lexikologischer Heterovalenz ein psychologischer Mechanismus ist. Dieser bewirkt beim Lexikographen, dass er Idiome als tendenziell funktional gleichwertig auffasst, wenn sie ikonographisch oder strukturell (oder beides auf einmal) parallel sind. Es muss sich dabei nicht unbedingt um völlig parallele Komponentenstrukturen handeln. Eine gewisse Motivgemeinschaft (Farø 2000a; 2004d) reicht offensichtlich aus, um von Äquivalenz auszugehen (wie bei *etw an den Nagel hängen* vs. *lægge ngt på hylden*). Häufig scheinen Oberflächenentsprechungen entscheidend zu sein; eine sprachgebrauchsbasierte Feststellung von funktionaler Äquivalenz wird nicht vorgenommen.

Dies ist aber zunächst bloß als eine Hypothese anzusehen, denn welchen Status diese Äquivalenzprobleme in der Lexikographie haben, ist noch gar nicht klar. Bisher war die Rede ja von rein *lexikologischen* Problemen. Was weiterhin aussteht, ist der Versuch, die festgestellten lexikologischen Äquivalenzprobleme zu spezifisch *lexikographischen* Fragestellungen in Beziehung zu setzen. Dem werde ich mich jetzt zuwenden.

## 4 Zur Frage nach der Beziehung zwischen interlingualen lexikologischen Daten und dem bilingualem Wörterbuch

Welche Relevanz haben die besprochenen Probleme also für die *bilinguale*[2] *Lexikographie*? Hiermit wird ein klassisches Problem angedeutet, nämlich die Beziehung zwischen Lexikographie und Lexikologie (vgl. Wiegand 1999: 23ff.). Ich möchte diese Diskussion nicht erneut aufnehmen, sondern nur darauf verweisen, dass sie in der Regel monolingual geführt wird. Ist der Kontext aber bilingual, wird sie ungleich komplexer. Und sind Idiome der konkret zu behandelnde Lemmatyp, dann verkompliziert sich die Diskussion noch mehr.

Theoretisch kann man in einem bilingualen Wörterbuch zwischen einer Menge lexikographischer *Funktionen* unterscheiden. Nicht zuletzt Bergenholtz (u. a. 1997) und Tarp (u. a. 1995) haben sich um die Entwicklung der so genannten „Lexikographische Funktionslehre" verdient gemacht. Die Hauptidee hinter diesem Zugang zur Lexikographie ist pragmatisch, nämlich dass ein Wörterbuch an sich keinen Wert zu besitzen braucht, sondern ein Wörterbuch sollte in erster Linie anhand der *Funktionen* beurteilt werden, die es für den Benutzer zu erfüllen imstande ist. Zwischen dem Wörterbuch als Substanz und den Funktionen, die es zu lösen hat, besteht also keine 1:1-Beziehung.

Wenn man das Problem der Wörterbuchfunktionen zur Lemmakategorie Idiom in Beziehung setzt, dann ist eine Reihe von Fragen zu beantworten, die in Bezug auf Idiome in einem bilingualen Wörterbuch relevant sind. Wenn wir das Problem der Muttersprache des Benutzers zunächst ignorieren, sind bei der Kategorie Idiomatik v. a. die Funktionen *Rezeption*, *Übersetzung*, *Produktion* und *Metawissen* zu berücksichtigen.

Wenn ich anhand eines bilingualen Wörterbuchs ein L2-Idiom in einem Text *rezipieren* möchte, so besteht z. B. die Möglichkeit, dass ich ein äquivalentes Idiom des L1 als Bedeutungserklärung bekomme: Als reines Rezeptionsproblem stellt somit das lexikologisch problematische Idiompaar *etw an den Nagel hängen – lægge ngt på hylden* für einen dänischsprachigen Wörterbuchbenutzer kein *lexikographisches* Problem dar: Denn die Bedeutung des deutschen Idioms wird so mit dem dänischen Idiom wiedergegeben, d. h. das dänische Idiom kann immer in den Kontexten verwendet werden, in denen das deutsche Idiom gebraucht wird; somit kann es auch als Bedeutungsschlüssel zum deutschen Idiom benutzt werden. Gleichzeitig wird damit auch die Funktion *Übersetzung* L2-L1 erfüllt.

Betrachtet man die *Produktion* von L2-Texten von einem L1-Idiom ausgehend, so steht man hier vor einer schwierigen Aufgabe, denn Produktionswissen gehört doch schon bei anderen Wortschatzeinheiten zu den problematischen Funktionen. Bei Idiomen gestaltet sich diese Funktion als besonders schwierig. Beim eben genannten Idiompaar hat man das Problem, dass das deutsche und das dänische Idiom offensichtlich nicht in den gleichen Kontexten auftreten. Somit ist es wohl problematisch, wenn in den dänisch-deutschen Wörterbüchern *lægge ngt på hylden* mit *etw an den Nagel hängen* „übersetzt“ wird, weil dies ja den Benutzer zur Produktion von Textmustern animiert, die zumindest im von mir untersuchten Material von den muttersprachlichen Sprachbenutzern nicht sanktioniert werden. So lassen die Angaben der untersuchten Wörterbücher Sätze wie *?er hat die Zigaretten an den Nagel gehängt* und *?die Pläne zur Errichtung eines neuen Stadions sind an den Nagel gehängt worden* zu, was offensichtlich mit den typischen Sprachgewohnheiten im Deutschen nicht übereinstimmt. Stattdessen müsste konsequenterweise eine Bedeutungsdifferenzierung unternommen werden. Möglicherweise müsste sogar auf die Kategorie Idiom völlig verzichtet und nicht-phraseologische Äquivalente wie *aufgeben* müssten gewählt werden, sodass der zentrale semantische Gehalt vom Äquivalent wiedergegeben wird, obwohl damit das Register nicht wie beim Ausgangsidiom markiert ist.

Die beiden erwähnten Funktionen, *Rezeption* und *Produktion*, die auch die Funktion *Übersetzung* einschließen können, sind die kommunikativen Funktionstypen eines bilingualen Wörterbuchs. Es existieren aber auch nicht-kommunikative lexikographische Funktionen wie z. B. *metakommunikatives Wissen*. So liefert das phraseologische Wörterbuch TA zum dänischen Idiom *gå over åen efter vand* („den Fluss/Wasserlauf überqueren, um Wasser zu holen“) das französische Äquivalent *porter de l'eau à la mer*, gibt aber gleichzeitig an, dass dieses Äquivalent „veraltet“ ist. Hier werden metasprachliche Bedürfnisse berücksichtigt, z. B. von Benutzern, die an Nominationsmotiven interessiert sind. Der Lexikograph zeigt, dass es in anderen Sprachen ähnliche Entstehungsmotive gibt, warnt aber gleichzeitig indirekt vor dem Gebrauch des frz. Idioms als *kommunikatives* Äquivalent.

In der folgenden Übersicht wird dargestellt, welche lexikographischen Funktionen welche Äquivalenzrelationen erfordern, wobei von Benutzern mit Dänisch als Muttersprache ausgegangen wird:

(7) *Rezeption* (für Dänen): *etw an den Nagel hängen – lægge ngt på hylden*

(8) *Produktion* (für Dänen): *lægge ngt på hylden*
   1 (Tätigkeit als Beruf oder Hobby[3]): *etw an den Nagel hängen*
   2 (Laster od. Projekt) *aufgeben* (z.B. er hat die Zigaretten (od. das Rauchen) aufgegeben; die Hochzeit wurde aufgegeben)

(9) *Metawissen* (für Dänen): *gå over åen efter vand – porter de l'eau à la mer* (veraltet)

Bei Benutzern mit Deutsch als Muttersprache wären bei *Schmetterlinge im Bauch haben* folgende Äquivalenzrelationen anzugeben:

(10) *Produktion*: *Schmetterlinge im Bauch haben*
 1 (verliebt sein) *være forelsket*
 2 (nervös, gespannt sein) *have sommerfugle i maven*

(11) *Rezeption*: *have sommerfugle i maven – ein Kribbeln im Bauch spüren*

(12) *Metawissen*: *Schmetterlinge im Bauch haben – have sommerfugle i maven*

Diese Aufteilung ist natürlich formalisiert; in der lexikographischen Praxis werden diese Funktionstypen integriert – im nächsten Abschnitt wird auf diesen Aspekt eingegangen. Trotzdem lohnt sich zunächst, die Funktionen zu isolieren um das Bewusstsein für einen funktionalen Zugang zur Idiomatik zu stärken.

## 5 Wann ist also ein lexikologisches ein lexikographisches „Äquivalent"?

Wenn wir nochmals die Frage stellen, wann man bei Idiomen *lexikographisch* von Äquivalenz ausgehen kann, ist also zunächst festzustellen, dass „Wörterbuchäquivalenz" bei Idiomen ein relatives Phänomen ist. Ob eine echte Äquivalenzbeziehung oder Pseudoäquivalenz vorliegt, kann nur relativ zur jeweiligen Wörterbuchfunktion beurteilt werden. Um dies zu illustrieren, folgen einige Beispiele für beide Äquivalenzarten. Zunächst lexikographische Pseudoäquivalenz:

(13) *Produktion* (für Dänen): *lægge ngt på hylden – etw an den Nagel hängen*

(14) *Rezeption* (für Deutsche): *have sommerfugle i maven – Schmetterlinge im Bauch haben*

(15) *Metawissen* (für Dänen): *have sommerfugle i maven – ein Kribbeln im Bauch spüren*

Im ersten Fall führt die Angabe den Benutzer womöglich auf den Holzweg: Das deutsche Idiom stimmt nur zu einem Viertel mit dem dänischen bezüglich Anwendungskontexte überein. D. h. in 3/4 der Gebrauchssituationen wird der Benutzer anhand dieser Angaben unnatürliche Texte produzieren können. Auch im zweiten Beispiel muss von lexikographischer Pseudoäquivalenz gesprochen werden, weil eine nicht-kommentierte Angabe des deutschen Idioms den deutschen Rezipienten zu falschen Schlüssen verleiten kann: Das Verliebtsein würde in Kontexten mitschwingen, in denen es gar nicht intendiert ist. Und im letzten Fall ist, wenn wir an lexikographischem Metawissen interessiert sind, auch nicht von lexikographischer Äquivalenz die Rede. Denn hier wird verschwiegen, dass es im Deutschen ein konvergentes Idiom *Schmetterlinge im Bauch haben* gibt. Wer allein an der Sonderfunktion *Metawissen* interessiert ist, den stört es nicht, dass die beiden Idiome lexikologisch-funktional nicht völlig äquivalent sind.

Wenn man die genannten Funktionstypen konsequent in Relation zu den „lexikologischen Tatsachen“ berücksichtigt, könnten folgende Musterartikel vorgeschlagen werden:

Für Dänen:

(16) dä-de.: *have sommerfugle i maven – Schmetterlinge im Bauch haben*

(17) de-dä: *Schmetterlinge im Bauch haben*
  1 (gespannt, nervös sein): *have sommerfugle i maven*
  2 (verliebt sein): *være forelsket* („verliebt sein“)

(18) dä-de: *gå over åen efter vand*
  1 (etw zu umständlich machen) *mit der Kirche ums Dorf laufen*
  2 (jm etw geben, was er bereits hat) *Eulen nach Athen tragen*; *Wasser ins Meer schütten*

(19) de-dä: *mit der Kirche ums Dorf laufen – gå over åen efter vand*

(20) dä-de: *lægge ngt på hylden*
  1 (permanente Aktivität) *etw an den Nagel hängen*
  2 (Projekt; Laster) *etw aufgeben*

(21) de-dä: *etw an den Nagel hängen – lægge ngt på hylden*

Zu überlegen wäre außerdem, ob der Vorschlag von Bergenholtz/Mogensen (1993), Idiome im Wörterbuch *literal* zu übersetzen, falls kein ikonographisch konvergentes L2-Idiom vorhanden ist, in irgendeiner Form in den Wörterbuchartikel aufzunehmen wäre. Dies würde die Funktion Metawissen über das Nominationsmotiv des deutschen Idioms erfüllen, die im Beispiel (19) fehlt. Man könnte sich z. B. eine Literalisierung in Anführungszeichen vorstellen, wie sie sich in der Phraseologieforschung eingebürgert hat (vgl. Burger 1989). Dieser Aspekt erfordert aber eine besondere Markierung, damit die Literalisierung nicht als ein in allen Kontexten verfügbares *kommunikatives* Äquivalent aufgefasst wird. Und als einziges Äquivalent, wie es Bergenholtz/Mogensen (ibid.) implizit vorschlagen, würde sie sicherlich als solches interpretiert werden, was ich für problematisch halte.

## 6 Schlussfolgerungen

Was hier angeschnitten worden ist, ist das Problem der Etablierung von Wörterbuchäquivalenz bei Idiomen. Idiomäquivalenz im Wörterbuch wird sehr häufig nicht auf der Basis von kontrastiven lexikologischen Untersuchungen, sondern sozusagen „aus dem Bauch heraus“ bestimmt. Bei Idiomen *scheint* dieses Verfahren zu funktionieren, weil man anscheinend anhand der Komponenten die Bedeutung der Idiome erschließen kann: zumindest wenn ihre Nominationsstruktur transparent ist. Gerade bei dieser lexikalischen Kategorie treten aber viele Probleme auf, weil idiosynkratische Restriktionen verschiedener Art, Bedeutungswandel, sowie nicht zuletzt das allgemeine Prinzip der Arbitrarität (vgl. Farø 2006) methodisch keineswegs eine interlinguale Zuweisung von Äquivalenten ausschließlich anhand der Oberflächenstruktur zulassen.

Andererseits – und das ist der spezifisch lexikographische Gehalt dieses Beitrags – lässt sich aus den sehr vielen lexikologischen Äquivalenzproblemen, die in der Idiomatik beobachtet werden können, noch keineswegs etwas darüber ableiten, wie Idiome in bilingualen Wörterbüchern dargestellt werden sollen. Die Darstellung hängt hauptsächlich von den lexikographischen Funktionen ab, die das Wörterbuch erfüllen soll. Spezifische Funktionen neutralisieren häufig lexikologische Äquivalenzprobleme der Idiomatik. Je nach Konzeption des Wörterbuchs kann z. B. der Aspekt der lexikalischen Kategorie eine Rolle spielen oder nicht. Das gleiche gilt für die Äquivalenzaspekte *Syntax*, *Valenz*, *Register* usw. Selbst der Aspekt der *Semantik* kann nicht von vornherein als entscheidend betrachtet werden, denn vorstellbar wäre z. B. ein Wörterbuch, das die Beziehung von interlingualen Nominationsmotiven beschreibt, ein seit jeher reges Forschungsfeld[4] der Phraseologie, für das sich auch Laien häufig interessieren.

Im Allgemeinen kann aber festgestellt werden, dass die Anzahl an relevanten Äquivalenzaspekten umso größer wird – und damit umso schwieriger die Zuweisung von Idiomäquivalenten –, je mehr ein bilinguales Wörterbuch auf den Aspekt der Produktion von L2-Idiomen fokussiert.

Eine bilinguale Lexikographie, die dazu beitragen möchte, kommunikative Probleme zu lösen, kann keineswegs von den lexikologischen Tatsachen absehen; dies ist unumstritten. Beides, möglichst radikale lexikographische Reflexion und profunde lexikologische Forschung, sind unabdingbar. Leider sind, zumindest in der rein kommunikativ orientierten Ausrichtung der kontrastiven Lexikologie, noch große Forschungsdefizite zu beklagen.

Als Motto der hier skizzierten Haltung kann, *mutatis mutandis*, Wittgenstein zitiert werden: „Don't just look for the picture of an idiom, look for its *use*"![5]

## Noten

1. Die genaue semantische Bestimmung dieser Objekte kann natürlich nicht immer auf der Grundlage der isolierten Wörter erfolgen, sondern sie hat selbstverständlich aufgrund der textuellen Kontexte stattgefunden.

2. Obwohl ich Tarps (2004) Kritik am Terminus „bilinguales Wörterbuch" für relevant halte, weil sie auf das wichtige Problem aufmerksam macht, dass schon der Begriff fixe Vorstellungen

und Erwartungen bei Lexikographen sowie Wörterbuchbenutzern auslöst, halte ich ihn dennoch nicht für unbrauchbar. Denn man braucht eine Bezeichnung für sprachliche Werkzeuge, deren Aufgabe es ist, dem Benutzer beim Navigieren zwischen den Wortschätzen zweier Sprachen zu helfen – ohne dass damit gleich bestimmte Inhalte präsupponiert werden müssen. Ich halte es somit für berechtigt, vom „bilingualen Wörterbuch“ als einer lexikographischen Basiskategorie auszugehen, ohne dass ich sie genauer definieren möchte, als oben getan wurde.

3. Ich schreibe hierum der Lesbarkeit willen die dänischen Glossen auf Deutsch.

4. Auf der EUROPHRAS-Tagung in Basel 2004 beschäftigte sich etwa ein Drittel der Beiträge mit diesem Problem.

5. Das ursprüngliche Zitat Wittgensteins lautete folgendermaßen: “Do not look for the meaning of a word, look for its use” (Wittgenstein 1953: 21).

## Literatur


Andersen, Stig Toftgaard. 2001/1998. *Talemåder i dansk*. Ordbog over idiomer. Kopenhagen: Gyldendal (= TA).

Andresen, Jutta; Jette Mez und Eckhard Bodenstein. 2001. Langenscheidt Taschenwörterbuch Dänisch. München/Berlin: Langenscheidt (= LA).

Bergenholtz, Henning. 1997. ”Polyfunktionelle ordbøger”. *LexicoNordica* 4: 15–29.

Bergenholtz, Henning und Jens Erik Mogensen. 1993. ”Wörterbuchkritik in Dänemark“. *Lexicographica* 9: 8–35.

Bergstrøm-Nielsen, Henrik; Henrik Lange, Henry Verner Larsen et al. 1996/1991. *Dansk-tysk ordbog*. Kopenhagen: Munksgaard (= MU).

Bork, Egon; Holm Fleischer, Jens Erik Mogensen, Helmut Molly og Elisabeth Møller. 1999. *Dansk-tysk Ordbog.* 11. Ausgabe. Kopenhagen: Gyldendal (= GY1).

Brink, Karl-Erich. 1993. *Hals- und Beinbruch og 2017 andre tyske udtryk og talemåder*. Kopenhagen: Høst (= HØ).

– 2002. *Pons Praxiswörterbuch Dänisch*. Barcelona etc.: Klett (= PO).

Burger, Harald. 1989. ”„Bildhaft, übertragen, metaphorisch“. Zur Konfusion um die semantischen Merkmale von Phraseologismen“. In: Gréciano (Hg.): 17–29.

– 2003. *Phraseologie. Eine Einführung am Beispiel des Deutschen*. 2., überarbeitete Auflage. Berlin: Erich Schmidt Verlag.

Dahl-Blumenberg, Constanze og Michael Dahl-Blumenberg. 2001. Tysk-dansk/dansk-tysk ordbog. Kopenhagen: Gad (= GA2).

Dobrovol’skij, Dmitrij. 1997. „Kontrastive Idiomatik Deutsch-Russisch: Zur lexikographischen Konzeption“. In: Gréciano/Rothkegel (Hg.): 45–59.

Farø, Ken. 2000a. „Idiomatik i moderne dansk-tysk leksikografi“. *Hermes* 25: 176–202.

– 2000b. *Dansk-tysk idiomatik. En empirisk baseret konfrontation af det moderne danske og tyske idiominventar.* København: Københavns Universitet. (Magisterarbeit).

– 2004a. ”Hvornår går man over åen efter vand? Idiomatiske ækvivalensproblemer og bilingval leksikografi. Med dansk og tysk som eksempel”. *LexicoNordica* 11, 235–244.

– 2004b. ”Omkring det grønne bord. Dansk-tysk idiomatik og sprogforandring”. *Studier i nordisk* 2002–2003. Kopenhagen: Selskab for Nordisk Filologi, 51–75.

– 2004c. "Idiomer på Nettet: Den Danske Idiomordbog og fraseografien" (Review article). *Hermes* 32: 201–235.
– 2004d. "Vom geschossenen Papagai und anderen Schwierigkeiten. Phraseologische falsche Freunde im Deutschen und Dänischen". *Deutsch als Fremdsprache* 3, 152–157.
– 2006. *Idiomatizität – Ikonizität – Arbitrarität. Beitrag zu einer funktionalistischen Theorie der Idiomäquivalenz*. Kopenhagen: Universität Kopenhagen. (unpubliz. Dissertation).
– /Voller, Inge 2002. *Dansk-tysk/tysk-dansk ordbog*. Kopenhagen: Gyldendal (= GY 2).
Gréciano, Gertrud (Hg.). 1989. *Europhras 88. Phraséologie Contrastive. Actes du Colloque International Klingenthal – Strasbourg 12–16 mai 1988.* Strasbourg: Université des Sciences Humaines, Département d'Études Allemandes.
– /Annely Rothkegel (Hg.) 1997. *Phraseme in Kontext und Kontrast*. Bochum: Universitätsverlag Dr. N. Brockmeyer.
Hausmann, Franz-Josef; Oskar Reichmann, Herbert Ernst Wiegand und Ladislav Zgusta (Hgg.). 1989–1991. *Wörterbücher. Ein internationales Handbuch zur Lexikographie*. Berlin: de Gruyter.
Jensen, Erik; Sven-Aage Outzen og Karin Bienz. 2003. Politikens Tysk-Dansk/Dansk-Tysk Ordbog. Kopenhagen: Politikens Forlag (= POL).
Klosa, Annette; Kathrin Kunkel-Ratzum, Werner Scholze-Stubenrecht und Matthias Wermke. 2001. *Duden Deutsches Universalwörterbuch*. Mannheim: Dudenverlag (= DUW).
Kluge, Friedrich. 2002. *Kluge. Etymologisches Wörterbuch der deutschen Sprache.* 24. Auflage, bearbeitet von E. Seebold. Berlin/New York: de Gruyter.
Saussure, Ferdinand de. 1916/2001. *Grundfragen der allgemeinen Sprachwissenschaft*. 3. Auflage. Berlin/New York: de Gruyter.
Schemann, Hans. 1989. „Das phraseologische Wörterbuch". In: Hausmann et al.: 1019–1032.
– 1991. „Die Phraseologie im zweisprachigen Wörterbuch". In: Hausmann et al.: 2789–2794.
Tarp, Sven. 1995. „Wörterbuchfunktionen. Realistische und utopische Vorschläge für die bilinguale Lexikographie". *Germanistische Linguistik* 127–128: 17–62.
– 2004. "Hvad er en bilingval ordbog?". *LexicoNordica* 11: 5–22.
Thomsen, Knud Troels. 2000. *Tysk kombi*. Kopenhagen: Gad (= GA1).
Wiegand, Herbert Ernst. 1998. *Wörterbuchforschung*. 1. Teilband. Berlin/New York: de Gruyter.
Wittgenstein, Ludwig. 1945/1953. *Philosophical Investigations*, translated by G. E. M. Anscombe. Oxford:

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 Dobrovol'skij 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 *Ordbogen over Faste Vendinger*: www.idiomordbogen.dk).

As mentioned above, one aim of this project is to describe the lexicographical potential of electronic dictionaries, and how successfully. they exploit the lexicographical 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).

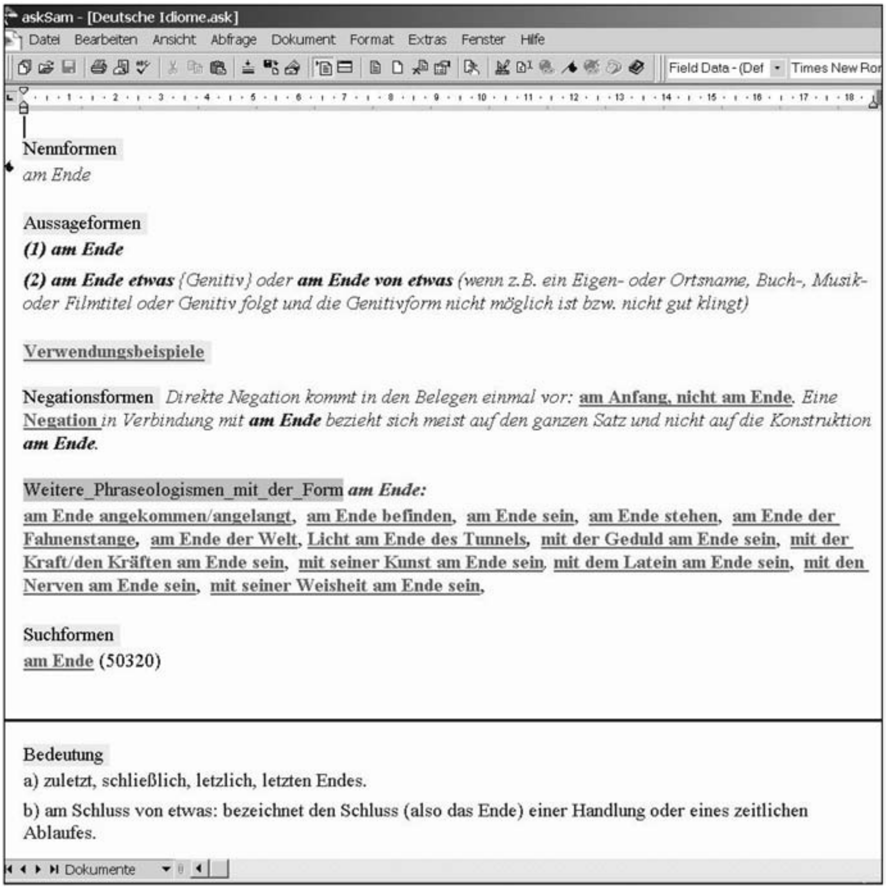


**Figure 1.** Information in the database, screen one.

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).

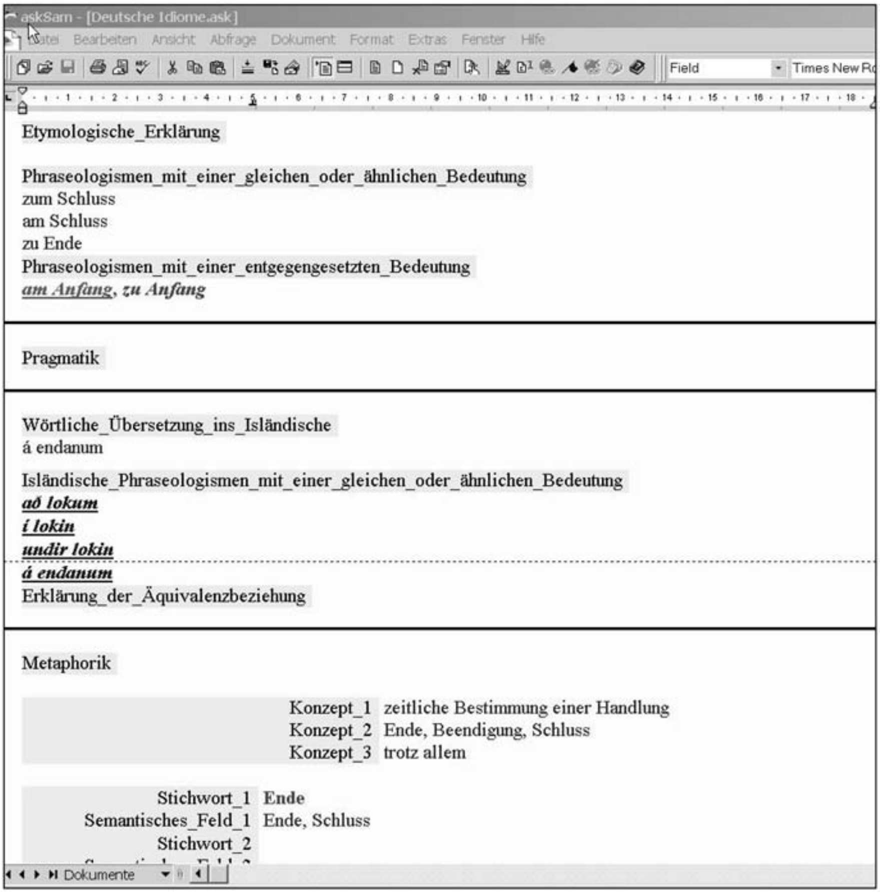


**Figure 2.** Information in the database, screen two.

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).

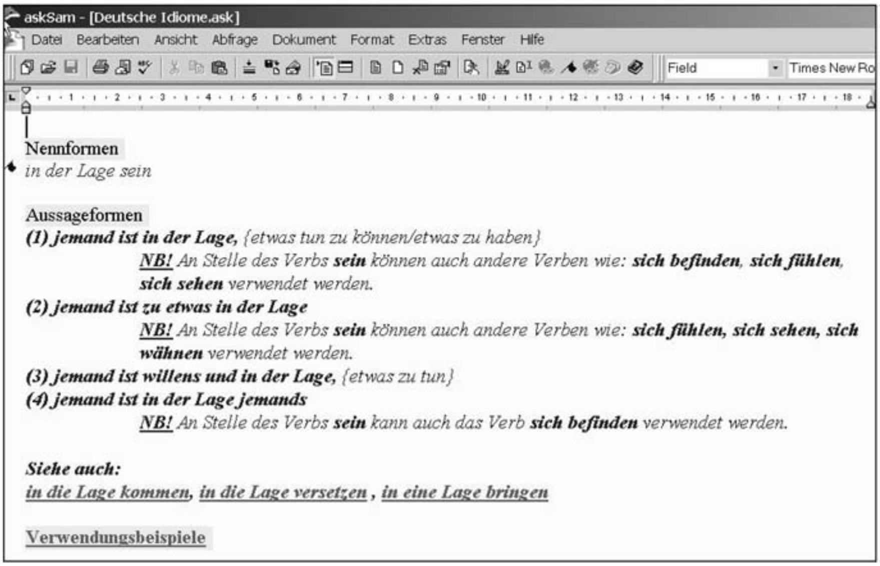


**Figure 3.** Usage forms with comments.

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).

**Nennform** in der Lage sein

**Aussageformen**

*(1)* jemand/etwas befindet sich in der Lage
jemand/etwas fühlt sich in der Lage
jemand/etwas ist in der Lage
jemand sieht sich in der Lage
jemand wähnt sich (durch etwas) in der Lage
*(2)* jemand/etwas fühlt sich zu etwas in der Lage
jemand/etwas ist zu etwas in der Lage
*(3)* jemand ist willens und in der Lage
*(4)* jemand/etwas befindet sich in der Lage jemands/etwas
jemand ist in der Lage jemands

NEGATION (...)
MODIFIKATIONEN
Erweiterung durch Adjektiv(e)

jemand/etwas befindet sich in der Lage
In der Begründung wurde darauf hingewiesen, daß sich Ungarn gar nicht **in der Lage** befinde, über die Mitgliedschaft zu entscheiden.

jemand/etwas fühlt sich (bestens/fachlich/gesundheitlich/körperlich/schwer/thematisch/... ) in der Lage
"Ich hätte damals von den Älteren erwartet', sagt Lafontaine, daß jemand aufsteht und sagt: 'Ich übernehme den Parteivorsitz' - so lange bis er, Lafontaine, sich gesundheitlich **in der Lage** gefühlt hätte, selbst die Parteispitze zu übernehmen.

jemand/etwas ist in der Lage
Ministerpräsident Professor Biedenkopf meint offensichtlich, daß der Staat eher **in der Lage** ist, einen derartigen Betrieb wirtschaftlich (und ökologisch) sinnvoll zu betreiben, als dies ein Privater vermag.

Fraglich ist für Kerber, ob die Kommission wirklich **in der Lage** ist, innerhalb der Ein-Monats-Frist zuverlässig zu entscheiden, ob sie die Fusion gleich freigeben kann oder ein förmliches Verfahren einleiten muß.

**Figure 4.** Analysed corpus examples.

### 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 dictionary 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

*Ask Sam – Making information useful:* www.asksam.com [last accessed July 27, 2007]

Burger, Harald. 1998. *Phraseologie. Eine Einführung am Beispiel des Deutschen* [Grundlagen der Germanistik 36]. Berlin: Erich Schmidt Verlag.

Burger, Harald; Annelies Häcki Buhofer and Gertrud Gréciano (eds.). 2003. *Flut von Texten – Vielfalt der Kulturen. Ascona 2001 zur Methodologie und Kulturspezifik der Phraseologie* [Phraseologie und Parömiologie 14]. Baltmannsweiler: Schneider Verlag Hohengehren.

Cheon, Mi-Ae. 1998. *Zur Konzeption eines phraseologischen Wörterbuchs für den Fremdsprachler. Am Beispiel Deutsch-Koreanisch* [Lexicographica. Series Maior 89]. Tübingen: Max Niemeyer Verlag.

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

Dobrovol'skij, Dimitri. 1997. *Idiome im mentalen Lexikon: Ziele und Methoden der kognitivbasierten Phraseologieforschung* [Fokus 18]. Trier: WVT Wissenschaftlicher Verlag.

Filipenko, Tat'jana V. 2002. "Beschreibung der Idiome in einem zweisprachigen Idiomatik-Wörterbuch (Deutsch-Russisch)". *Das Wort. Germanistisches Jahrbuch der GUS*: 43–62.

Hallsteinsdóttir, Erla. 2005. „Vom Wörterbuch zum Text zum Lexikon." In *Zwischen Lexikon und Text – lexikalische, stilistische und textlinguistische Aspekte*. Ulla Fix, Gotthard Lerchner, Marianne Schröder und Hans Wellmann (Hgg.), 325–337. Leipzig: Verlag der Sächsischen Akademie der Wissenschaften zu Leipzig.

Jónsson, Jón H. 2002. *Orðaheimur. Íslensk hugtakaorðabók með orða- og orðasambandaskrá*. Reykjavík: JPV Útgáfa.

Mudersbach, Klaus. 1998. "Ein Vorschlag zur Beschreibung von Phrasemen auf der Basis eines universalen pragmatischen Modells". In *Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium* [Lexicographica. Series Maior 84], Herbert E. Wiegand (Hg.), 319–347. Tübingen: Max Niemeyer Verlag.

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

Part 4

# LSP lexicography

CHAPTER 8

# Bringing corporate dictionary design into accord with corporate image

## From words to messages and back again

Patrick Leroyer

Corporate dictionaries are usually designed as terminological dictionaries according to a positivist approach, the selection of LSP lemmas and related information being guided by a systematic delimitation and representation of the subject field (Pedersen 1996). Their direct function is to facilitate knowledge transfer, whereas their indirect function is to facilitate LSP communication while ensuring a consistent use of terminology. Although well established, the terminological method seems transfer-focused, and should not stand alone. The concept of subject field should be extended to business area, and that of LSP communication to corporate communication and image. User needs should also be extended to the needs of the corporate organization as a whole. In this article, I will present new dictionary functions and planned methods for selecting, structuring and presenting the lexicographic information that should bring corporate dictionary design into accord with corporate image, and provide coherent lexicographic help at message level.

### 1 Introduction

The purpose of this article is to present briefly a number of new theoretical developments and planned practical solutions in the field of *corporate lexicography*, which should be of great interest to functional metalexicographic work in general. *Corporate lexicography* can be defined as the scientific study and development of *corporate dictionaries* for knowledge and communication purposes – internally as well as externally – that are genuinely made to fulfil the needs of the users as well as of the corporation itself. In this regard, corporate lexicography must necessarily be based on functional lexicographic theory, but should also include corporate dictionary functions that are specific to those needs. From a metalexicographic point of view, this definition raises a number of important questions. First: Is there any such thing as corporate lexicography? Is it not some kind of LSP lexicography in a specific environment? Is a corporate dictionary any different from any other kind of LSP dictionary? Can a corporation as such have any kind of lexicographic needs? And second, and even more important: Is corporate lexicography a lexicographic issue that is capable of contributing seriously to the development of a general lexicograhic theory? Is it desirable and possible to identify and define specific corporate dictionary functions? My answer to these questions is yes. I am convinced that corporate lexicography can contribute to the development of a general lexicographic theory in an original way. This is why I have chosen to include the expression "*into accord with corporate image*" in the title of this article. Although it might at first seem totally irrelevant to serious metalexicographic work, it is not a fashionable management expression, but lies at the *prescriptive* core of the new lexicographic functional framework.

## 2 Methodology – from words to messages and back again

Corporate image mapping is the dynamic rule of corporate communication. All corporations develop, adapt, promote and defend their image according to their strategic objectives, to their internal and external strengths and weaknesses, and to their competitors, in order to meet the expectations from their stakeholders. But image unfortunately is not directly accessible to the lexicographer. Image is transcribed into words, which are not ordinary content words, but *message components*, a message being a purposeful complex of coherent and distinctive discourse ideas, which can be transcribed into texts and genres according to the communication setting. The methodology used in this work has therefore been to apply corporate communication theory according to Libaert 2003 in order to analyse and classify messages, to perform genre analysis according to Bhatia 2004 in order to map their content schemata, and to perform lexicographic selection, structuring and presentation of message-dependent lexicographic data types belonging to the text example category according to Leroyer 2002.

This is why I call it "*from words to messages and back again*". The data presented here are extracted from the corporate website of NEG Micon, a major Danish manufacturer of wind turbine technology, and have been selected to illustrate major features of the new corporate dictionary design. Let us now go back to metalexicography and take a brief look at the functional paradigm. As lexicography is an independent science which welcomes help from other sciences (Tarp 2001: 23–24) I have not hesitated to include corporate communication theory as well as genre and discourse analysis. Still, the central object remains the same: It is the development of dictionaries determined by user needs and made to serve *communication and knowledge-oriented functions* in particular *user situations*. It is also worth noticing in this context that modern dictionary function theory (Tarp and Bergenholtz 2003) originated, at the beginning of the nineties, in a corporate environment framework. The theory however did not include any reference to corporate communication as such, communication-oriented functions being of textual essence only, in order to solve problems connected with text reception, translation or production encountered by the users.

Is a corporate dictionary then some kind of unspecified LSP dictionary? My answer to that question is no. LSP dictionaries are based on the idea of subject field and focus on subject field knowledge and subject field language knowledge according to user profiles. LSP dictionaries apply terminological methods, although their theory has also been modernized: The best of them should also contain the necessary information for helping users in connection with text-dependant functions. In this context, it is then assumed that terminology and terminography belong to specialized lexicography. Is a corporate dictionary an LSP dictionary merely located in a corporation? My answer to this question is also no, in as far as the terminology based approach still dominates, and no specific reference is made to corporate communication as such. This does not mean that terminology must be abandoned. It means that indispensable corporate communication aspects are absent, and that the scope of usability is therefore limited. It also means that the terminological work should be revised and that new priorities should be allocated to the dictionary design process. I will claim that a corporate dictionary does not represent a subject field as such: It represents a business area and its communication according to corporate image mapping. Such a dictionary includes a great number of subject fields, which are parts of the same business. Such a dictionary also replaces to a large extent the user as both user and author, as Humbley 2002 argues, and the lexicographer becomes a mediator. Still, I will go further and argue that the corporate dictionary should also contain lexicographic data types that represent message information, in order to help the user at the message level.

Major work in the field of corporate lexicography has been undertaken by Simonsen 2004, who rightly argues that a corporate dictionary is not simply located in the corporation. It is embedded, and should follow organizational principles. This brings us back to the definition level: The dictionary certainly exists because of user needs, but it can only exist properly in relation to corporate needs. Being a knowledge and communication-oriented tool, it can only exist properly if it also represents corporate communication and brings about solutions to solve its specific problems. The corporate dictionary necessarily becomes a *prescriptive tool,* planned and designed in accordance with corporate strategy and image. This of course has vast consequences for practical dictionary making and ongoing developments that are easy to see: choice of main language and interrelated languages, choice of represented business areas, choice of communication policy and language policy, choice of localization strategy, choice of selection criteria and adaptation of lexicographic data types, etc. etc. The very nature of the *toolification* pro-cess (Wiegand 1994: 104) is affected, as the dictionary should now include lexicographic data types representing the corporate communication at work itself. As a result, one of the prerequisites of all corporate lexicography is to ensure that dictionary planning should take place at strategic management and communication department level. In this respect, the lexicographer, as a central mediator, should be responsible for the overall dictionary coherence, its ongoing developments, data selection criteria and adaptation of data presentation to user and corporate needs. The data should also include message data from competitors in order to ensure the central distinctive quality of message conception and transcription, and of its lexicographic representation.

## 3 Two new message-dependant dictionary functions

What does this mean to the design process? It means that dictionary functions should be adapted to corporate communication management in order to help corporate users identify the communication problem, access knowledge about generic message conception, and consult examples of specific message transcription. I am now in a position to reveal two new corporate dictionary functions.

The first function – the *message planning dictionary function* – is knowledge-oriented. Its purpose is to help the user acquire, check or share knowledge about message planning policy at work in texts, in conformity with corporate communication plans. This is achieved by structuring lexicographic information on *message planning keywords* that are used as query words throughout all levels of corporate communication planning, and by presenting the information as matching text examples and documents from a dynamic text example and document retrieval system.

The second function – the *message transcription dictionary function* – is communication-oriented. Its purpose is to help the user in the process of transcribing and coining a specific message into text. It is realised by structuring lexicographic information on *message component keywords* and presenting the information as matching text examples linked to the entries of the central lemma list.

I will stress the fact that the two new dictionary functions fundamentally conform to the central epistemology of the functional paradigm. They are fully integrated in its typology, in what one might call the two *dictionary metafunctions* – knowledge on the one hand and communication on the other. Knowledge about message planning is obtained in much the same user situation as applies to the traditional knowledge functions: knowledge about language or language policy, or knowledge about the business area. Help with communication by way of information on message transcription is also obtained in much the same user situation as applies to text reception, production and translation. The purpose of the message transcription function is to assist message reception, translation, and production. I have not at present finished developing and refining the message transcription functions to message reception and translation, although these are in preparation. These functions will deal on the one hand with reading, understanding and interpreting messages from the corporation (and from competitors as well), and with translating and producing messages according to adapted rhetorical strategies on the other. This part of the work requires further development to fully integrate the traditional text-dependant and the new message-dependant functions.

## 4 "Powerful dictionary" – some practical solutions

I will now illustrate the theoretical points that have been made above by an example from real world corporate communication. In accordance with NEG Micon's generic message – "*Powerful future*" – I will call it "*Powerful Dictionary*", a planned dictionary designed to go far beyond terminological descriptions of wind turbine technology. The empirical base includes message sensitive lexicographic data types that are selected in accordance with the new dictionary functions, the methodology being, I repeat, to establish a Micon's message typology on the basis of Micon's communication planning – *from words to messages and back again*. The components below are the main online features of such a powerful dictionary design:

- Corporate corpus built on corporate communication and image mapping
- Indexing of text examples and documents according to text genre settings
- Indexing of texts and documents according to communication settings
- Extension of the lemma list with communication-based selection criteria
- Query driven access to information on message planning at text example and text genre level through functional message sensitive keywords
- Alphabetic access to information on message transcription at word level through presentation of matching text examples and related networks

Naturally, these new components should be fully integrated into already existing in house dictionaries, provided that the database has a modular architecture. The data base and server technology should be based on the Intranet model developed by Simonsen 2004, with an extranet extension.

### 4.1 Help in connection with message planning

Online access to information on message planning is obtained by message queries. The user can combine message sensitive keywords from the following search list: *objective, field and strategy*. In example 1 below, the combined query is "*secure top of mind awareness*", "*wind turbine technology*" and "*justify benefits of the technology*". In the result of the query, all matching text examples display the message strategy at work in text examples: acquiring and securing top of mind awareness in the business of wind turbine technology through the reasserting of the benefits of the technology itself.

**Example 1.** Searching for message planning information – "secure top of mind awareness"

→ **Search for message planning in text examples**
Objective = secure top of mind awareness
Field = wind turbine technology
Strategy = justify benefits of the technology
→ Display matching text examples
*[Designed for easy service] [Developed through experience] [NM110 designed for off-shore operations] [NM110 direct extension from NM80 and NM92] [NM110 – The next generation of offshore wind turbines]*

The communication problem in example 2 below is of a similar kind, the objective being "*gain recognition*" and the message planning strategy being the formulation and justification of a promise – here reliability through the technology itself.

**Example 2.** Searching for message planning information – "gain recognition"

→ **Search for message planning in text examples**
Objective = gain recognition
Field = technology
Strategy = promise and justify reliability
→ Display matching text examples
*[Enables preventive and proactive maintenance] [Ensures high reliability] [Flexible and safe internal crane system] [One year service interval] [Optimal lightning protection – Faraday cage]*

Access to message planning information should also be obtained through the combination of message planning keywords and the retrieval of the best matching text genre document. The data are then accessed through communication structuring query words at surface level and matching documents below. The system is based on a thematic classification of the communication data according to modern corporate communication theory. The lexicographic knowledge representation about message planning is determined by the main issues of the communication problem and communication function:

– What is the nature of the communication problem to be solved?
– Which communication function, i.e. type of communication field and theme such as environment, social responsibility, B2B, recruiting, influential, financial, product, sensitive?

In example 3 below, message planning query words available in the rolling menu of the search window are: *problem, function, issue, type, objective, strategy, topic, and genre* – the best matching documents being then retrieved from the document data base. The same procedure can also be applied for multilingual purposes, in particular for localization of the message planning. In each case, the hierarchy of message sensitive keywords is based on true whole/part relations refining the multiple search options available to the user:

**Example 3.** Searching for message planning information – bad news announcement

→ **Search for message planning in genres**

| | |
|---|---|
| problem => | bad news announcement |
| function => | sensitive communication |
| issue => | crisis |
| type => | technology |
| objective => | image restoration |
| strategy => | denial |
| topic => | main shaft rupture |
| genre => | press release |

→ Display matching genre

*On Monday 9 September 2002* [*a rupture of the main shaft*] *of a 500kW turbine with a rotor diameter of 43 meters was reported from a turbine in Germany installed in 1996. The event caused no personnel harm or injury. Investigations on site have revealed that* [tag/denial] *the rupture was clearly induced by a repair procedure causing weakening of the main shaft material and inadequate quality of the repair itself… The event is clearly singular/Investigations on site have revealed that the rupture was clearly induced by a repair procedure causing weakening of the main shaft material and inadequate quality of the repair itself* […].

### 4.2 Help in connection with message transcription

Access to message transcription information is obtained through the lemmatization of *message component keywords* and on their hyper linking to the text example and document database. It is not a simple semasiologic system since lemmas are not selected and marked because of their semantic or encyclopaedic meaning representation *per se*, but because of their functional ability to open networks of message transcription options that can be used during text-dependant tasks, these words being truly argumentation *topoï*. The message transcription data are recycled from the message planning data of the example and document retrieval system. In example 4 below, searching on *low* will perform a full text search and display all text examples of B2B product messages in which promises and justification are based on the benefits of the *low* distinctive *argumentation* network:

**Example 4.** Searching for message transcription information – "low", full text examples

| | |
|---|---|
| **installation** | *[Low weight – low installation costs]* |
| **low** | *[Low weight – low transportation costs]* |
| | *[Low weight – low installation costs]* |
| | *[Low weight of nacelle and rotor only 250 ton]* |
| **transportation** | *[Low weight – low transportation costs]* |
| **weight** | *[Low weight – low installation costs]* |

## 5 Conclusion

Lexicographic help at message level constitutes the genuine form and content expression of the two new corporate dictionary functions, as outer user access is facilitated through flexible, combined search options. Along with hyper linking of message data types to external dictionary components such as an introduction to the official corporate communication planning policy, communication actions and activities, guidelines for message creation according to the tonality of the language policy, and message authoring functionalities, these components should bring *corporate dictionary design into accord with corporate image*. It should be clear now that modern dictionary function theory constitutes the true theoretical and practical framework of corporate lexicography, and that the future of corporate dictionaries belongs to the ongoing development of the already well established dictionary functions. Knowledge-oriented functions must now include knowledge of corporate communication planning, while communication-oriented functions must now also include message-dependent functions. In this regard, the two new dictionary functions shed, albeit modestly, new light on the true nature of knowledge and communication-oriented dictionary functions. These are truly metafunctions that have paved the way for the continuous development of functional metalexicography in general, and of functional, corporate metalexicograpy in particular. Corporate dictionaries offer unique lexicographic opportunities and challenges to modern lexicography by truly bringing dictionary auditing, planning and making into the real world of corporations of today, so as to address human needs along with corporate needs. This will help to continuously register, transfer and share internal knowledge on the corporate communication planning policy on the one hand, and help users transcribe it properly when actually working on the multi dimensions of corporate texts for either internal or external purposes – for the benefit of corporate messages – on the other. From words to messages and back again…

## References

Bergenholtz, Henning and Sven Tarp. 2003. "Two opposing theories: On H. E. Wiegand's Recent Discovery of Lexicographic Functions". *Hermes. Journal of Linguistics* 31: 171–196.

Bhatia, Vijay K. 2004. *Worlds of Written Discourse*. London and New York: Continuum.

Humbley, John. 2002. "Nouveaux dictionnaires, nouveaux rapports avec les utilisateurs". *Meta* XLVII, 1: 95–104.

Leroyer, Patrick. 2002. *Les Exemples Textuels Unificateurs. Fondements et Applications pour le Dictionnaire de Traduction Technique*. Aarhus: The Aarhus School of Business.

Libaert, Thierry. 2003. *Le Plan de Communication. Définir et Organiser votre Stratégie de Communication*. Paris: Dunod.

Pedersen, Jette. 1996. *Koncernspecifik brancheordbog med særligt henblik på teknisk engelsk*. Aarhus: The Aarhus School of Business.

Simonsen, Henrik Køhler. 2004. "Nine Key Principles on Corporate LSP Intranet Lexicography". In *Proceedings of the Eleventh Euralex International Congress*, G. Williams and S. Vessier (eds), 603–613. Lorient: Université de Bretagne-sud.

Tarp, Sven. 2001. "Lexicography and the linguistic concepts of homonymy and polysemy". *Lexicographica* 17: 22–39.

Wiegand, Herbert Ernst. 1994. "Zur Unterscheidung von semantischen und enzyklopädischen Daten in Fachwörterbuchern". In *Fachlexikograhie, Fachwissen und seine Repräsentation in Wörterbüchern*, B. Schaeder and H. Bergenholtz (eds.), 103–159. Tübingen: Gunter Narr Verlag.

CHAPTER 9

# Design of a function-based internet accounting dictionary

Sandro Nielsen and Lise Mourier

The traditional definition of a dictionary needs to be replaced by one that defines the dictionary in terms of lexicographic functions, data and structures. These must be linked to the intended user groups, the users' linguistic and factual competences and their needs in the relevant situations of use. The functions of the dictionary include translation, production, reception and revision of texts, and knowledge acquisition, and the user groups may be divided into experts, semi-experts and interested laypeople. The dictionary discussed in this paper takes advantage of the space available in an open-ended electronic dictionary and is unique because it combines several data types relevant to the need of the users in function-related situations, including: definitions, grammatical information, collocations, phrases, cross-references, synonyms and antonyms. Both the factual and linguistic data are particularly helpful to the user in communication-oriented situations within a register-specific context such as accounting.

## 1 Introduction

In every societal and cultural context you reach a point where people are looking for something new rather than the tried and true. And this is where Internet dictionaries come in with their potential for presenting considerably more information than printed dictionaries. In August 2003 the Centre for Lexicography, Aarhus School of Business, in cooperation with Copenhagen Business School, published the Danish Accounting Dictionary and the Danish-English Accounting Dictionary for free use on the Internet at: www.regnskabsordbogen.dk. These dictionaries are genuine Internet dictionaries and not remakes of existing printed dictionaries. They therefore take advantage of the new technological possibilities and combine in one Internet dictionary information items that are normally found in two or more printed dictionaries. An easy-to-follow search structure and the use of easily identifiable search fields in the articles give the user quick access to the data.

Accounting is a challenging subject requiring much specialist background knowledge and with distinct terminological characteristics. Many Danish companies produce translations of their financial data, primarily in UK English, but today many companies address the US and therefore need to communicate in US English. By 2005, listed companies in the EU must apply the International Financial Reporting Standards (IAS/IFRS) and follow changed accounting rules and concepts. The new Danish Financial Statements Act 2001 as amended is widely based on IAS/IFRS and make statutory the paradigm shift in financial reporting that has spread lately from the US over IAS/IFRS to Europe. The main reasons for designing and making the dictionaries are the above-mentioned internationalisation and new legislation; the fact that accounting language is LSP code language and terminology and difficult to grasp for semi-experts, translators and laypeople: not just distinct terminological characteristics, but indeed background knowledge is a prerequisite (therefore definitions and contexts are included); because mapping of varieties of English versus Danish is imperative; and because no other Danish-English electronic Internet accounting dictionary and no Danish-English updated special accounting-only dictionary are available. This paper will introduce and discuss some of the major lexicographic principles and considerations underlying the preparation of the Danish-English Accounting Dictionary, in particular dictionary functions, data types and user needs in various types of use situation.

## 2 A new approach to defining the dictionary

From a practical as well as a theoretical point of view, it is imperative to describe the subject-matter of the project, i.e. the dictionary, at a general level of abstraction. A dictionary has traditionally been defined as a reference work that presents the vocabulary of a language, usually in alphabetical order, with explanations of meanings in the same language or with equivalents in another language (see e.g. Hartmann/James 2001: 41; and Sterkenburg 2003: 396). This type of definition may be acceptable if you are unfamiliar with the concept. Rather than approaching the dictionary from the traditional and linguistically inspired point of view, it may be more appropriate to introduce what may be called a lexicographic approach. This focuses on the significant features of a dictionary, as they reflect its raison d'être. In this context, the adjective 'significant' is used to refer to a feature of a dictionary that is important and meaningful for the purpose of describing the concept of dictionary; i.e. not trivial and arbitrary for the purposes of classification. A significant feature thus helps to shed light on the existence of a dictionary as an object of investigation, description and analysis, as well as a "workpiece". A dictionary may, and usually does, have other features than those described here as being significant.

A dictionary has at least three significant features, which together describe and explain the concept. First of all, the overriding feature is the general idea of fulfilling one or more functions. It refers to the dictionary itself without taking into account the external dictionary environment. In terms of textual analysis, this feature does not distinguish text (i.e. the electronic or printed dictionary) from context (i.e. the dictionary environment) but sees them as one. The feature applies to only one specific dictionary and focuses on its lexicographic potential as an aid in particular types of use situation.

Secondly, the dictionary contains the actual or physical realisation of lexicographic data that support its function(s). This feature makes it possible to compare dictionaries, their data, information types etc. by considering the external context of each realisation; this also allows us to find common properties among dictionaries. The second significant feature is dependent on the first feature, as it focuses on the lexicographic potential and the dictionary as a material object, which may be compared with other dictionaries at the data level. The interaction between the first and the second significant feature is important, because the potential of the dictionary is not realised until it is materialised in some form through the data presented.

Finally, the dictionary itself may be regarded as the third significant feature. The dictionary as a whole marshals its parts of lexicographic data into the task of fulfilling its function(s). This is the feature that makes it possible to unite the lexicographic data and the function(s) to produce a whole, and allows us to see the relation and interaction between individual realisations of data based on specific lexicographic principles. This is where the lexicographic structures come into play, whether they are order structures such as the macro- and microstructures or network structures such as the mediostructure (see Nielsen 2003a, 111–112; and Nielsen 2003b, 168–169). The important point here is that the intention behind the dictionary is not found in either the first, the second or the third significant feature, but in the aggregate of features, i.e. the electronic (or printed) dictionary.

Consequently, a dictionary may be defined as a lexicographic reference work that has been designed to fulfil one or more functions, contains lexicographic data supporting the function(s), and contains lexicographic structures that combine and link the data in order to fulfil the function(s). This definition applies to printed, electronic and Internet dictionaries, and it applies to existing, planned and imaginary dictionaries alike.

## 3 The overriding nature of functions

The first step in studying and describing a dictionary is to determine its function(s) at a lower level of abstraction. Firstly, the function(s) chosen by the lexicographers provide the basis for all other lexicographic decisions, from the selection of lemmata, over the selection of information types, to the selection of lexicographic structures. Secondly, the lexicographers must determine the basic needs of the users on the basis of the dictionary function(s) relative to the intended user group. This emphasises the overriding nature of the first significant feature.

It is possible to distinguish between two main types of function: communication-oriented and knowledge-oriented functions. Communication-oriented functions include reception of texts in the user's native language (L1) or in a foreign language (L2), the translation of texts into L1 or L2, the production of texts in L1 or L2, and the revision and editing of texts in L1 or L2. These functions, which focus on the transfer of a message from a sender to a receiver, are characterised by being text-dependent. Knowledge-oriented functions, which focus on deriving and verifying (propositional) knowledge and the acquisition of information, include the acquisition of factual and/or linguistic knowledge generally about the L1 or L2 culture, or in connection with a specific issue in the L1 or L2 culture, for instance in a learning context (such as a course on accounting): What is the difference between the Danish accounting terms *goodwill* and *badwill*? These functions, which focus on knowledge, i.e. data that – through a mental process – have become information that has a use or purpose, are characterised by being text-independent. The distinction between these two main types of function allows lexicographers to make multifunctional dictionaries (see e.g. Bergenholtz/Kaufmann 1997: 98–99; and Bergenholtz/Nielsen 2002: 5–6).

The lexicographers of a multifunctional dictionary will invariably have to give different priorities to the functions, as – for practical reasons – its data will support some functions more or better than others. The Danish-English Accounting Dictionary was designed to help the users in connection with one primary function, three secondary functions and one tertiary function:

1. Primary function:
1.1. Translation of Danish texts into English
2. Secondary functions:
2.1. Production of English and Danish texts
2.2. Reception of Danish and English texts
2.3. Revision and editing of Danish and English texts
3. Tertiary function:
3.1. Knowledge acquisition concerning Danish and English accounting issues

These functions must be linked to the intended user group, the users' linguistic and factual competences, and their needs in the relevant situations of use, to determine which types of data have to be included in the dictionary.

## 4 Determining the needs of the users

The second step in the function-based approach involves the identification of user competences. The factual competence, *in casu* accounting competence in a Danish and an English culture, and the linguistic competence, i.e. Danish and English specialised language, determine how the user will use a dictionary in a specific type of situation. The lexicographic literature proposes a distinction between three general groups of users: experts, semi-experts and interested laypeople (see e.g. Bergenholtz/Kaufmann 1997: 98–99). This triad is an appropriate basis for any specialised dictionary, but it requires supplementation in order to form a practical basis.

The lexicographers need to draw up a profile of the intended user group to relate the dictionary's functions to their needs and competences. The purpose of the user profile is to identify the major characteristics and lexicographic needs of the users, taking into account the factual and linguistic competences in relation to L1 and L2 and the relevant cultures. The Danish-English Accounting Dictionary has (1) a primary user group, made up of translators and secretaries, (2) a secondary user group, accounting experts, and (3) a tertiary user group, which consists of student accountants and translators, journalists as well as laypeople interested in Danish financial reporting information.

The primary user group consists of persons with considerable linguistic competence and small to medium factual competence. They are generally semi-experts in relation to both types of competence, taking into account that translators are experts within the field of language, but not within the specialised language of a specific subject field; the persons who make up the secondary group have a considerable factual competence and a small to medium linguistic competence. Accordingly, the Danish-English Accounting Dictionary contains information about linguistic and factual aspects relating to the foreign language to fulfil its functions as a production and translation dictionary. Linguistic and factual information is relevant to both user groups. The primary user group needs more factual information, such as definitions of the Danish terms, than the secondary group. Both groups need the same factual information specifying the difference between English equivalents in cases where it is necessary to distinguish between English variants, depending on meaning and context. Failure to include such information would have seriously impaired the dictionary's potential for fulfilling the user needs.

The best solution is for the lexicographers to compile a dictionary that contains data supporting the functions translation and production in the relevant situations of use. A traditional bilingual LSP dictionary focuses exclusively on terms, but in any LSP text the terms account for less than fifty percent of the words; this kind of dictionary does not assist the user with all the words surrounding the terms. The Danish-English Accounting Dictionary gives priority to the functions translation and production of accounting texts and contains a variety of data types in addition to terms and their definitions. It contains the following types of data supporting the primary, secondary and tertiary functions on the L1 side: orthography, grammatical irregularity, definition, usage note, collocation, phrase, synonymy and antonymy; and on the L2 side: translation equivalent, orthography, grammatical irregularity, usage note, collocation, phrase, synonymy and antonymy.

The satisfaction of user needs in types of use situations contemplated by the dictionary's function(s) requires a dictionary concept that differs from the traditional bilingual LSP dictionary. One focus point is the inclusion of syntagmatic information, which deals with the combination of words and their relationship such as collocations, phrases and sentences in L1 and their L2 translations. In containing all these different types of data, the Danish-English Accounting Dictionary offers considerably more help to the user than the traditional bilingual LSP dictionary in that it supplements the factual, defining data; and this is largely due to the space available in a genuine Internet dictionary relative to a printed dictionary. It is not possible here to comment on all the various data types in the dictionary, so the following discussion will centre on definitions, equivalents, collocations and phrases and how they support the primary and secondary functions with particular reference to the primary and secondary user groups.

## 5 The dictionary as a utility product

The lexicographic function(s) of a given dictionary is to provide assistance to a defined user group to cover the complex of needs that arise in a specific situation, and no data should be included if not on the basis of the dictionary's function(s) (Bergenholtz/Tarp 2003: 176–177). Hence, language dictionaries must help users in the processes of understanding language (text reception), producing language (text production) and/or translating from one language into another (text translation). LSP dictionaries are of course intended to help users solve language problems. However, for language within a specific technical subject field, i.e. a register-specific context, the dictionary would e.g. not be able to define particular terms without drawing on extra-linguistic knowledge. Monolingual LSP dictionaries are expected to contain factual information to help users learn the meaning of a particular term, but also to give them more knowledge. For bilingual LSP dictionaries the combination of language knowledge and background knowledge is exactly what make such dictionaries useful and unique. To achieve quality when translating accounting texts, the translator needs both language and special accounting knowledge (Mourier 2004: 146–147). Therefore, in the compilation of the Danish-English Accounting Dictionary, our mission has been to guide and give the user solutions, both for language and for extra-linguistic purposes. In this way, users are helped find the appropriate equivalent and at the same time provided with more knowledge. Therefore, to create an efficient LSP dictionary, it is important that definitions are included, not just to define the terms, but to create knowledge for the users. The importance of the knowledge function in our accounting dictionary is also reflected by the use of special information fields (see below).

The concept of our accounting dictionary as a utility product involves

- delimitation of subject-matter
- in-depth treatment rather than getting-it-all-in
- usability: focus on easy use and efficient solutions, fulfilling user requirements, and providing problem solutions.

The delimitation of the subject field to accounting – again further delimited to primarily external financial reporting – also delimits and decides the selection of corpus material and lemmata. A corpus consisting of texts relevant to the subject field – in this case chosen from accounting rules and regulation as well as empirical material – will ensure coverage of all, or at least most, relevant lemmata.

Usability involves making user considerations. Before specifying user needs, the user group must be defined – as discussed in section 4 above the accounting dictionary addresses somewhat heterogeneous users: accounting experts, semi-experts such as translators or journalists and interested laypeople. Next follows the mapping of user requirements that must be met, and above all this means that the user must be referred to the version of the lemma and the equivalent that the lexicographers judge proper and that no confusing alternative choices should be offered. Equivalent options should only be added when clearly distinguished from one another for various, specified uses, such as for polysemy and collocations (see section 5.5 below).

This involves informing users of the English equivalents in the relevant varieties of English, since accounting terminology is a subject field where the terminology differs widely when comparing GB, US and international English[1] (Mourier 2004: 149–150). As a result, the accounting dictionary is both descriptive and normative: descriptive as it includes various spelling versions – to help the user, even frequently misspelt lemmata are included with reference to the properly spelt lemmata. Normative as the dictionary advises the user of correctness, adds extra information to the lemmata in the form of extra-linguistic, encyclopaedic definitions and, in some cases, relevant information relating to e.g. synonyms, antonyms, cross-references or spelling of equivalents in the varieties of English (marked 'GB', 'US' or 'IAS/IFRS' or, when referring to an explanatory translation into English of the Danish lemma: 'DK'). Also in the overall structure of the dictionary, usability has been highlighted: the dictionary has been compiled from the outset as an Internet dictionary: in addition to wider space, this offers special electronic handling opportunities, such as links to sources and relevant websites etc. In the following, the data fields supporting the lemmata will be explained, illustrated by examples from the dictionary.

### 5.1 Definitions supporting a plurality of functions

Definitions first of all delimit the lemmata in particular contexts, enabling one equivalent only to be offered as the correct choice, besides including semi-technical terms and general terms as seen in their technical 'accounting' meaning. Second, definitions provide users with problem-solving and additional background knowledge as they explain the lemmata and often include additional relevant information (see Example 1 below). This is particularly helpful to translators who are not accounting experts. Specialists in accounting, on the other hand, need updated information about changed rules and new terminology[2]. Definitions must be true and faithful to the subject field – if not, experts' needs would not be met. However, some technical definitions in their original wording may be difficult to grasp by the broad user group; therefore, complex wording has been rephrased and simplified to meet the requirements of the user group at large. This has been done with careful faithfulness to sources to ensure quality.

**Example 1.** Definition written for heterogeneous user group

**kvartalsrapport**
Synonym
delårsrapport
quarterly report
Synonym
quarterly interim report
Betydning *[definition]*
En kvartalsrapport er en årsrapport, der kun omfatter et enkelt kvartal, og skal ligesom årsrapporten overholde reglerne i årsregnskabsloven for at være en kvartalsrapport. Kvartalsrapporten er ikke et krav for de børsnoterede virksomheder ligesom halvårsrapporten, men Fondsbørsen anbefaler udarbejdelse heraf.
*[A quarterly report is an annual report prepared for one quarter only, and like the annual report, it must comply with the provisions of the Danish Financial Statements Act for quarterly reports. Unlike the half-year report, the preparation of a quarterly report is not a requirement for listed enterprises, but the Copenhagen Stock Exchange recommends that it be prepared.]*
2. Henvisning *[2nd Reference]*
årsrapport
Kollokationer *[collocations]* …………

## 5.2 Equivalents solving translation and production problems

The English equivalents to the Danish lemmata must be true to the Danish language context and meaning, and translation is provided in the case of lexical gaps in any of the original English varieties. Hence, '*ledelse*' has the English equivalent 'supervisory board and executive board' (DK)[3] to correctly transfer the Danish term that is legally defined as the two-tier management body of a public limited company, consisting of both '*bestyrelse*' (supervisory board, i.e. board of directors) and '*direktion*' (executive board). Likewise, Danish acts, names of Danish institutions etc. have been provided with English translations, for acts those generally used in official English translations and for names those applied for English purposes by the institutions themselves.

To follow our mission statement we seek to guide and not confuse the user by providing one equivalent only. This means one British-English, one American-English and one IAS/IFRS-English where the equivalents are different terms. Where the terminology overlaps, only one equivalent is provided, and the lack of English-specific reference denotes the general use of this equivalent in English generally.

For polysemous Danish lemmata, the definitions are extremely important, as they will provide the specific meaning of the lemma, in this way combining it with the correct equivalent. Technically, the polysemous lemma is listed as one headword that leads to two definitions each with the proper equivalent:

**Example 2.** The three cards of the polysemous lemma 'afskrive'

**afskrive** *[headword]*

**afskrive**
write off
Betydning 1 *[definition 1]*
At afskrive er at omkostningsføre hele beløbet *[To 'afskrive' is to expense the total amount.]*.

**afskrive**
amortise
For afskrive anvendes 'amortise' om immaterielle aktiver. På amerikansk engelsk skrives 'amortize'
*[For 'afskrive' 'amortise' is applied in the case of intangible assets. In American English, the spelling is 'amortize'].*
eller *[or]*
depreciate
For afskrive anvendes 'depreciate' om materielle aktiver *[For 'afskrive' 'depreciate' is applied in the case of tangible assets]*.
Betydning 2 *[definition 2]*

At afskrive vil sige at foretage en systematisk fordeling af afskrivningsgrundlaget over et aktivs brugstid
*[To 'afskrive' is to allocate the basis of amortisation/depreciation over the useful live of an asset].*
Kollokationer *[collocations]* …………………….

Example 2 also illustrates how information fields warn the user of special American spelling and the use of two different equivalents in English (see below).

To adhere strictly to the delimitation of the subject field, we only lemmatised varieties interesting for accounting purposes of polysemous, non-LSP (ordinary) Danish terms with their English equivalents, as seen in Example 3:

**Example 3.** Equivalents distinguished by meaning discrimination

**føre**
carry
Føre oversættes i betydning (2) register; (3) include; (4) trade (in) og (5) conduct
*['Føre' translates in meaning (2) into register, (3) include; (4) trade (in) and (5) conduct].*

**Synonym**
book
Betydning *[definition]*
At føre er (1) i regnskabsmæssig betydning at indregne en post eller et beløb i regnskabet, (2) at registrere i en registrant, (3) at medtage i en fortegnelse, (4) at handle med noget, eller (5) at udføre en handling
*['At føre' is (1) for accounting purposes to recognise an item or an amount in the financial statements, (2) to register in a registry, (3) to include in a list, (4) to trade in something, or (5) to carry out an act].*
Kollokationer *[collocations]* ……………

## 5.3 Special information fields for factual and linguistic data

Example 3 shows another example of the dictionary's use of special information fields (synonyms).

The types of lexicographic data included depend on the dictionary's defined user needs. Each type of data must support or facilitate a particular function relating to language or knowledge acquisition (Simonsen 2001: 265–267). The data types are interdependent, but separately composed and arranged under each search term to enable the user to access precisely the information needed.

Through these data fields users are provided with additional, important features of the lemmata. As users of the Danish-English Accounting Dictionary are Danes, this information is provided in Danish. The data types include:

- notes to lemmata for factual information or recommended usage
- markers for English varieties: GB, US, IAS/IFRS, and DK (see above)
- comment field to the equivalent (e.g. concerning special American spelling usage or non-appropriate choice of equivalent in special contexts)
- cross-reference to the lemma in the dictionary's recommended spelling (1st Reference)
- cross-references to other lemmata (2nd Reference)
- synonyms
- antonyms
- sources (exploiting the Internet media, some are clickable providing direct access to the relevant Internet homepage).

**Example 4.** Special information fields assisting users in different use situations

**nettoomsætning**
net revenues (US)
'Net revenues' anvendes typisk af fx service-, it-, medicinal og teknologisektorerne, mens 'net sales' anvendes af handels- og fremstillingsvirksomheder.
*[The term 'net revenues' is typically applied by e.g. the service, IT, medico and technology sectors, whereas the term 'net sales' is applied by trading and manufacturing enterprises].*
eller *[or]*
net turnover (GB)
eller *[or]*
revenue (IAS/IFRS)

**Synonym**
net sales
Betydning *[definition]*
Nettoomsætningen er salgsværdien af produkter og tjenesteydelser mv., der henhører under selskabets ordinære aktiviteter, med fradrag af prisnedslag, merværdiafgift og anden skat, der er direkte forbundet med salgsbeløbet.
*[Revenue is the selling price of products and services, etc. sold in the course of the company's ordinary activities less price reductions, VAT and other taxes that are directly connected to the sales amount.]*
Kollokationer *[collocations]* ………..

The cross-references[4] guide the user from the search term to a lemma or lemmata that are relevant for comparison and additional knowledge. The user is referred e.g. to another lemma, may compare the definitions and check the collocations.

**Example 5.** Cross-references and sources giving the user additional information.

**tilknyttet virksomhed**
group enterprise
'Affiliated enterprise' er en bred betegnelse og vil fx også dække associerede virksomheder
*['Affiliated enterprise' is a broader term and will e.g. also cover associated enterprises].*

**Synonym**
group undertaking
Betydning *[definition]*
Med tilknyttet virksomhed menes en virksomhed, som i en koncern er en virksomheds dattervirksomhed, dens modervirksomhed, eller dennes dattervirksomhed. En tilknyttet virksomhed behøver ikke være konsolideret i koncernen og kan ikke være en associeret virksomhed.
*['tilknyttet virksomhed' means an enterprise which in a group is an enterprise's subsidiary, its parent or the latter's subsidiary. A 'tilknyttet virksomhed' needs not be consolidated in the group and cannot be an associated enterprise.]*
2. Henvisning *[2nd Reference]*
associeret virksomhed
koncernvirksomhed
……
Kilde *[source]*
årsregnskabsloven, bilag 1, B5

Example 5 illustrates two types of information field: (1) The cross-reference field that guides the user to the lemma '*associeret virksomhed*' (associated enterprise) mentioned in the card as well as to the lemma *'koncernvirksomhed'* (consolidated enterprise), which is a narrower concept than *'tilknyttet virksomhed'* since, contrary to the latter which is merely 'family' member of a group, the term *'koncernvirksomhed'* denotes that this enterprise is also consolidated in the group; and (2) the source field that refers the user to the source where the lemma is defined.

### 5.4 Syntagmatic information satisfying functions and user needs

A translation and production dictionary that is true to its function(s) and user needs will have to contain syntagmatic information. This type of information deals with the combination of words in collocations, phrases and sentences and provides useful and necessary information about text production and equivalence above the word (or term) level. Syntagmatic information in L1 and L2 is necessary because the lack of knowledge as to combining words in a foreign language (and also in the user's native language) within a register-specific context leads to unidiomatic texts, for instance because the collocational patterns of L1 and L2 differ. The users' needs for this type of information is reflected by the fact that the Danish-English Accounting Dictionary contains more than 15,000 collocations and phrases in its 4000 articles.

Collocations and phrases can be used to provide information about grammatical differences in L1 and L2 that are likely to occur when words are used in combination. Such differences are particularly important when they occur in a register-specific context because none of the intended user groups can be expected to know the relevant grammatical aspects in situations involving translation, production, revision or editing of accounting texts. In many cases, a Danish singular noun or noun phrase corresponds to an English plural noun or noun phrase:

**Example 6.** Phrases showing grammatical differences

**resultatføre**
…
recognise in the income statement (IAS/IFRS US)
recognise in the profit and loss account (GB)
….
• resultatføre finansielle instrumenter lineært over instrumentets løbetid
recognise financial instruments in the income statement on a straight-line basis over their lives (IAS/IFRS US)
recognise financial instruments in the profit and loss account on a straight-line basis over their lives (GB)

The Danish phrase in Example 6, preceded by a bullet, contains a noun in the singular (*løbetid*), whereas the equivalent English phrase contains a plural noun (*lives*). Furthermore, Danish uses an –s genitive (*instrumentets*) before the noun, whereas English uses a plural possessive pronoun (*their*). The space available in an Internet dictionary has also made it possible to provide two translations of the Danish phrase corresponding to the two English equivalents. This approach has been used consistently throughout the dictionary, as this provides optimal assistance to the users when translating into English or producing, revising or editing English texts.

The users need to be made aware of structural differences and options in L1 and L2, and collocations and phrases may be used to present such information explicitly and in the relevant context. It is important that the user is made aware of the availability and use of different structures, as the intended user groups will be unfamiliar with the differences and options:

**Example 7.** Phrases illustrating different linguistic structures in L1 and L2

**retvisende billede**
…
fair presentation (IAS/IFRS US)
true and fair view (GB)
…
• opfylde kravet om et retvisende billede
meet the fair presentation requirement (IAS/IFRS US)
meet the requirement of a true and fair view (GB)
…
• oplysninger af betydning for det retvisende billede
information of importance to the fair presentation (IAS/IFRS US)
information of importance to the true and fair view (GB)

The first Danish phrase in Example 7 contains a noun modified by a prepositional phrase (*kravet om et retvisende billede*) and the equivalent British phrase contains a similar structure (*the requirement of a true and fair view*). In contrast, the IAS/IFRS and American phrase contains a compound noun (*the fair presentation requirement*). This article explicitly informs the user of the different structures in English, depending on whether the phrase is British, American or IAS/IFRS English, and it also shows the different structures used in British and American. The second phrase shows a grammatical difference between Danish and English in relation to a word other than the lemma or equivalent: The Danish word *oplysninger* is countable (*in casu* plural), whereas the equivalent English word (*information*) is uncountable. Experience shows that the members of the secondary user group are unaware of this difference, so the dictionary must explicitly inform the users of this difference, for instance in a phrase containing the normal occurrence of the word in a register-specific context.

Finally, collocations and phrases within a register-specific context often require unexpected translation into a foreign language. The most important problem for the intended user groups is when the lemma in the L1 phrase is not translated directly into L2, contrary to user expectations:

**Example 8.** L2 Phrase does not contain equivalent

| |
|---|
| **ombygningsudgift** |
| … |
| rebuilding cost |
| … |
| • ombygningsudgifter vedrørende lejede lokaler |
| leasehold improvement costs |

Example 8 shows that the English phrase does not contain the equivalent to the lemma, even though the lemma is part of the Danish phrase. The user cannot know that, in this particular phrase, English uses a solution that is completely different from the Danish and, therefore, this must explicitly be stated in the article.

## 6 Conclusion

The Danish-English accounting dictionary has been compiled as a combined bilingual and extra-linguistic dictionary to satisfy user needs based on a functional approach. For LSP purposes, users will look for guidance concerning correct language as well as subject-matter information. A combination of linguistic and factual background knowledge is a prerequisite to meet the lexicographic functions of a bilingual LSP dictionary, and the compilation depends on the interaction between both types of information. It is a prerequisite for any translation situation that to achieve translation quality, the translator must understand the source text. The technical contents of LSP texts – in this case accounting texts – may be challenging to the translator. Therefore, the accounting dictionary must also communicate background knowledge about the intricate rules and procedures of financial reporting in which special terminology is embedded, besides information about the linguistic rules. To this end, collocations and phrases provide important additional information for contextual solutions.

We have shown the importance of providing combined linguistic and extra-linguistic information in an LSP dictionary. Being an Internet dictionary, the

Danish-English Accounting Dictionary benefits from the obvious advantage of the open-ended electronic dictionary: space. As LSP develops rapidly – and this is particularly true of accounting, where new rules and terminology are introduced even as these lines are written – space is of decisive importance. For an LSP dictionary it is furthermore important that the lexicographers can create links to relevant Internet websites, where the user may find additional information.

The Danish-English Accounting Dictionary is unique not only because it is the first electronic Internet accounting dictionary, but also because it is the first to combine several data types relevant to the needs of the user group in function-related situations: definitions of accounting terms, grammatical information, Danish collocations and phrases with English translations, cross-references to other relevant information, etc. In this way the dictionary meets the requirements of a state-of-the-art electronic bilingual Internet LSP dictionary.

## Notes

1. For the purpose of the accounting dictionary, international English means the English applied by the International Financial Reporting Standards (IAS/IFRS).

2. In 2001, new accounting legislation was introduced in Denmark with the new Danish Financial Statements Act.

3. (DK) is the marker for an English translation of the Danish term where no original English term (in British, American or IFRS English) would be equivalent.

4. Cross-references appear as '1. Henvisning' (1st Reference) when guiding the user from a misspelt search term to the correctly spelled lemma, and as '2. Henvisning' (2nd reference) when referring the user to another, relevant lemma.

## References


Bergenholtz, Henning and Uwe Kaufmann. 1997. "Terminography and Lexicography. A Critical Survey of Dictionaries from a Single Specialised Field". *Hermes* 18–1997: 91–125.

Bergenholtz, Henning and Sandro Nielsen. 2002. "Terms in the Language of Culture-Dependent LSP Dictionaries". *Lexicographica. International Annual for Lexicography* 18/2002: 5–18.

Bergenholtz, Henning and Sven Tarp. 2003. "Two Opposing Theories: On H. E. Wiegand's Recent Discovery of Lexicographic Functions". *Hermes* 31–2003: 171–196.

Hartmann, R. R. K. and Gregory James. 2001. *Dictionary of Lexicography*. London and New York: Routledge.

Mourier, Lise. 2003. *Translating Company Financial Statements. Study material and strategies*. Copenhagen: Samfundslitteratur.

Mourier, Lise. 2004. "Communicating financial reporting across continents". *Hermes* 32–2004: 145–165.

Nielsen, Sandro. 2002a. ”Overvejelser angående det leksikografiske grundlag for en elektronisk bilingval regnskabsordbog”. *LexicoNordica* 9–2002: 173–194.

Nielsen, Sandro. 2002b. *Lexicographical Basis for an Electronic Bilingual Accounting Dictionary: Theoretical Considerations*. www.sprog.asb.dk/sn/lexicographicalbasis.htm.

Nielsen, Sandro. 2003a. “Changes in Dictionary Subject Matter”. In: *Untersuchungen zur kommerziellen Lexikographie der deutschen Gegenwartssprache I. “Duden. Das grosse Wörterbuch der deutschen Sprache in zehn Bänden”,* H. E. Wiegand (Hrsg.), 109–115. Tübingen: Max Niemeyer Verlag.

Nielsen, Sandro. 2003b. “Towards a General Theory of Bilingual Legal Lexicography”. In *LSP Translation in the New Millennium. A Cross-Baltic Symposium on Didactics and Research*, P. Kastberg (ed.), 165–188. Aarhus: Hermes Skriftserie.

Nielsen, Sandro; Lise Mourier and Henning Bergenholtz. 2003. *Den Dansk-Engelske Regnskabsordbog*: www.regnskabsordbogen.dk. Design and layout Richard Almind. Centre for Lexicography. Aarhus.

Simonsen, Henrik Køhler. 2001. *Telelex. Theortical Considerations on Corporate LSP Intranet Lexicography: Design and Development of Telelex – an Intranet-Based Lexicographic Knowledge and Communications Management System*. PhD dissertation. Aarhus School of Business, Centre for Lexicography.

Sterkenburg, Piet van (ed.). 2003. *A Practical Guide to Lexicography*. Amsterdam/Philadelphia: John Benjamins Publishing.

Part 5

# Dictionaries and the user

CHAPTER 10

# What's so good or bad about advanced EFL dictionaries?

Ilan J. Kernerman

Monolingual dictionaries for advanced learners of English as a Foreign Language form a league of their own in pedagogical lexicography and English lexicography, and are among the most highly regarded and widely researched in the dictionary world. Considering their qualities and drawbacks, this paper attempts to assess their aims and how well suited they are to achieve them and looks at the consequences for learners of English and other languages, for dictionary use and lexicography.

## 1 Introduction

The purpose of this short paper is to question the purportedly superior characteristics of monolingual Dictionaries for Advanced-level Learners of English as a foreign language, hence DALE. Are they not, eventually, a setback for ordinary dictionary users – and for alternative lexicographical approaches? I focus mainly on what is "bad" about them as a group, mentioning their "good" features mainly as a reference, without thorough analysis of these qualities, on the understanding that the professional community is overwhelmingly familiar with them.

DALE have played more than one vital role in the evolution of lexicography and dictionaries through the second half of the twentieth century. They pioneered the adoption of state-of-the-art pedagogical principles in lexicography and introduced – notably – the use of a limited or simplified defining vocabulary in the dictionary's nomenclature, of corpus lexicography in the dictionary-making process, and the like of signposts and other helpful tips in the entry macro-structure. Their advent goes hand in hand with the worldwide expansion of the English language, of western culture in general and American values in particular, of democracy and a market economy dominated by freedom of speech and mass media, high technology and borderless communication, brandnames and sales techniques. This development of English as the world language has made it possible for DALE to enjoy exceptional investment and research, which reflects on general and learners' dictionaries for other languages as well. Yet, the result is also a counter-productive cycle that feeds on itself, draws the main attention, leaving little room for other dictionary types and lexicographical approaches.

## 2 The state-of-the-art?

Let us ask, first, what is the *English* that is so sophisticatedly advocated by DALE? Until the 1990s, this was generally termed "English as a Foreign Language" (EFL), typically defined from a point of view that distinguishes the world between *Home* and *Foreign*, as pertaining to the historical traditions of the British Empire. This is supported by bodies such as the British Council or the UK-based IATEFL (International Association of Teachers of English as a Foreign Language), although the main drive behind the immense contemporary interest in English has actually derived from the US superpower on the other side of the Atlantic. There, we have "English as a Second Language" (ESL), fitting an immigrant society with many mother tongues, promoted by bodies like TESOL (Teachers of English as a Second Language). DALE have tried to meet this challenge with inclusion of American English (and, to lesser extent, other local varieties of English) or, increasingly, by producing American versions.

However, most learners of English nowadays are practically not bothered with British or American English, neither as a Foreign nor a Second language, but in an international language of communication, a global language, "English as a Lingua Franca" (ELF), which enables speakers of any language to have a common ground with each other, more in the spirit of Esperanto, the artificial language invented by Ludwik Zamenhof at the end of the nineteenth century. This is an important difference, related to the fact that there already are today more non-native speakers of English than native ones – a development that is growing steadily. These are not humble foreigners, but people around the world with their own native languages, using English as part of their rapidly changing reality, and changing the English language therewith.

A note in passing, this concept of a culturally-neutral learning of a world language has long been described by applied linguists, who clarified the distinction between an *instrumental* motivation for foreign language learning (i.e. as a tool for communication, professions, etc) and an *integrative* one (i.e. cultural assimilation, etc).

The intention to present and teach "real" English is further undermined by use of the *corpus*, introduced into lexicography in the mid-1980s and incorporated in all DALE today. The first uses of corpora began in the 1950s, along with the appearance of computers, particularly for historical dictionaries. Doubts that might have existed about the added value of corpus evidence seem to have evaporated, and corpora have since become accepted as fundamental by lexicographers. The question remains, though, what goes into the corpus? How can we define as *scientific* decisions that are, after all, with all due respect, made by this or that person – regardless of their great knowledge of the language? Moreover, with reference to ELF as opposed to EFL or ESL etc, one may wonder who really needs to learn the so-called *real* English that is spoken – or, actually, mainly written – in English-speaking countries? And, as such, is the Internet the ultimate corpus today?

In their adherence to the Direct Method, including Total Immersion in the target English language, DALE have insistently ignored that even their assumed users with a fairly high knowledge of English still rely primarily on their own mother tongues as a prime source of reference on the world. DALE are still conceived according to circumstances of the past, when English was learnt mainly for literary (or tourist) purposes. But those golden one-size-fits-all days are over, and gone are the ease and comfort for corporations offering the same product all over the world. Welcome the "personal", "tailor-made", "customized", "localized" consumer society, struggling to preserve and nourish an individual or national or cultural identity within globalization. The first steps are being taken in the complex task of compiling learner-oriented corpora, and still more is required for developing specific bilingual learners' corpora.

The publishers of DALE are the tip of the iceberg of a huge EFL industry, hitherto mainly UK-based, with substantial commercial (and political) interests worth billions of dollars. Their immediate collaborators – whether aware or not – are teachers of English, those fortunate enough to be employed by this powerful system, or those solicited by it in order to instruct their students what to buy. Since the use of DALE has come to be considered a sign of progress, teachers often want students to use them – regardless of the effects – and thus also feel self-reassured by their own high-level teaching competence. Unfortunately, teachers often know little about dictionaries, have not the means to assess them and to determine which dictionary is most suitable for each learner group, and do not train their students in dictionary use. Users often prefer simple bilingual translation dictionaries, a "primitive" shortcut to the bottom line, rather than get lost in the valuable didactic information that is offered in DALE, which usually demands going through a long and complicated how-to-use-this-dictionary introduction, makes use of confusing abbreviations and symbols, is presented within a crowded text and unfriendly page layout, and is literally not speaking their language.

A different type of complicity stems from the professional community, mainly academics, who tend to know English on a high level, who teach it at university or are teacher trainers, etc, and for whom DALE are the natural playground, who use DALE themselves and prescribe them to their students. Incidentally, this group actually consists of users – rather than learners, in the usual sense – of English.

Lexicographers and highly experienced dictionary users discuss DALE (in English, of course) in conferences and literature, provoking an enchanted cycle of internal reference between them. To test this idea more thoroughly, I counted the number of articles and reviews devoted to DALE and to other dictionaries in the *International Journal of Lexicography*. The counts (see Appendix 2b), show that out of 36 texts concerning English learners' dictionaries that appeared from 1998 to early 2004, 31 dealt with monolingual English learners' dictionaries – of which 26 focus on DALE – and only 5 on bilingual English learners' dictionaries; another 59 dealt with other English dictionaries (33 monolingual, 26 bilingual), and 62 with non-English dictionaries.

The breakdown of these figures is remarkable. Of a total 157 items, only 40% dealt with non-English dictionaries or lexicography. Nearly a quarter (23%) concerned English learners' dictionaries – of which 72% dealt with DALE and only 14% with bilingual dictionaries. Yet, out of the texts on general English dictionaries, amounting to 37% of all texts, nearly half (44%) dealt with bilingual dictionaries.

## 3 Conclusions

Summing up, the benefits of DALE are the setting of higher standards for modern lexicography and dictionary making, the application of up-to-date and innovative methods of language analysis and pragmatics, and the creation of quality contents and products. However, a chief drawback of DALE is its becoming a role model for everything else – thus causing what narrowly concerns an elite group of users to affect the dictionary-using habits and tools of the major part of the international community, and in consequence often making dictionary use too difficult and ungratifying, and thus distancing users from other dictionaries. Adding to this the DALE model tends to ignore the vital role of the native language in any other language-learning process in general, and the special need to protect mother tongues from the dangers and threats of today's globalizing English in particular. A further potential drawback of DALE is relying on unreliable corpora and not providing the English that is *really* needed or wanted by its potential users. Finally, its somewhat unappealing design and not-easy-to-use presentation may necessitate skills that many dictionary users lack.

As an afterthought, what may be considered a positive development in this spirit is that more and more professionals are becoming disenchanted with DALE and seek alternatives. As a case in point, the 11th Euralex International Congress in Lorient (July 2004) presented four papers dealing with bilingual dictionaries for low-level learners of English.

## Acknowledgement

I am grateful to Raphael Gefen and Arleta Adamska-Sałaciak for their support and useful comments.

## Appendix 1

### Dictionaries for advanced learners of English as a foreign language (DALE):

**CIDE** (1995) *Cambridge International Dictionary of English*. (ed.) Paul Procter, Cambridge: Cambridge University Press.
**COBUILD** (2001, 3/e) *Collins COBUILD English Dictionary for Advanced Learners*. (ed.) John M. Sinclair, Glasgow: HarperCollins Publishers.
**LDOCE** (1995, 3/e) *Longman Dictionary of Contemporary English*. (ed.) Della Summers, Harlow: Longman.
**MEDAL** (2002) *Macmillan English Dictionary for Advanced Learners*. (ed.) Michael Rundell, Oxford: Macmillan Publishers.
**OALDCE** (2000, 6/e) *Oxford Advanced Learner's Dictionary of Current English*. (ed.) Sally Wehemeier, Oxford: Oxford University Press.

## Appendix 2

Key:

| | |
|---|---|
| ELDs | English learners' dictionary – monolingual and bilingual |
| BELD | bilingual English learners' dictionary |
| MELD | monolingual English learners' dictionary |
| DALE | dictionaries for advanced learners of English as a Foreign Language (CIDE, COBUILD, LDOCE, MEDAL, OALD) |
| GEDs | general English dictionaries – monolingual and bilingual |
| GMED | general monolingual English dictionary |
| GBED | general bilingual English dictionary |
| NEDs | non-English dictionaries |

### (a) A rough list of Articles / Review Articles / Reviews on ELDs in the International Journal of Lexicography from 1998 to 2004 (1)

1998 11.1 – The State of Bilingual Lexicography in Japan: Learners' English-Japanese / Japanese-English Dictionaries. Keisuko Nakao. (BELD)

1998 11.4 (A.S. Hornby, 1898–1998: Commemorative Issue)
- Hornby's Bilingualized Dictionaries. Carla Marello. (BELD)
- Recent Trends in English Pedagogical Lexicography. Michael Rundell. (MELD)
- Review of Longman Dictionary of English Language and Culture, and Oxford ALDCE Encyclopedic Edition. (MED/MELD)
- Review of Longman Essential Activator. (MELD)

| | | |
|---|---|---|
| | | – Review of Longman Interactive American Dictionary. (MELD/GMED) |
| 1999 | 12.1 | – Lexical Reference Books – What Are the Issues? Rienhard Hartmann. (MELD/BELD) |
| | | – Dictionary Use in Reception. Phil Schofield. (MELD) |
| | | – Dictionary Use in Production. Michael Rundell. (MELD) |
| | | – A User's Guide to Electronic Dictionaries for Language Learners. Hilary Nesi. (MELD) |
| 1999 | 12.2 | – A Study of Dictionary Use by ESL Students in an American University. Don R. McCreary and Fredric T. Dolezal. (MELD) |
| 1999 | 12.3 | – Metaphors and Dictionaries: The Morass of Meaning, or How to Get Two Ideas for One. Geart van der Meer. (MELD) |
| | | – Nine Learner's Dictionaries. Robert Ilson. (MELD) |
| 1999 | 12.4 | – The Treatment of Affixes in the 'Big Four' EFL Dictionaries. Tvrtko Prćić. (MELD) |
| | | – Interacting with Words and Dictionaries: The Case of Jordanian EFL Learners. Turki A. Diab and Jihad M. Hamdan. (MELD/GBED) |
| 2000 | 13.1 | – Review of the BBI Dictionary of English Combinations. (MELD) |
| 2000 | 13.3 | – Review of DAI, Dizionario di Apprendimento della lingua Inglese. (BELD) |
| 2000 | 13.4 | – Review of Cowie's English Dictionaries for Foreign Laerners: A History. (MELD) |
| 2001 | 14.2 | – Review of The Perfect Learner's Dictionary (?). (MELD) |
| 2001 | 14.3 | – Evaluating Learner Dictionaries: What the Reviews Say. (MELD) |
| | | – Review of The Use and Abuse of EFL Dictionaries. (MELD) |
| 2001 | 14.4 | – OALD6 in a Linguistic and a Language Teaching Perspective. Geart van der Meer and Rosemary Sansome. (MELD) |
| 2002 | 15.3 | – American Freshmen and English Dictionaries. Don McReary. (MELD/GELD) |
| | | – Dictionary Use and Dictionary Needs of ESP Students. Mari Carmen Campoy Cubillo (MELD + various types) |
| | | – Review on Research on Dictionary Use in the Context of Foreign Language Learning. (MELD/BELD) |
| 2002 | 15.4 | – A Study of Dictionary Use by International Students at a British University. Hilary Nesi. (MELD/BED) |
| | | – Review Article on CIDE on CD. (MELD) |
| | | – Review of Pegadogical Lexicography Today. (MELD) |
| 2003 | 16.1 | – Review Article on MEDAL: a Fifth Dictionary for Learners of English. (MELD) |
| | | – Review of Oxford Collocations Dictionary for Speakers of English. (MELD) |
| 2003 | 16.3 | – Review of Better Words, Evaluating EFL Dictionaries. (MELD) |
| 2003 | 16.4 | – Pronunciation in MEDAL on CD-ROM. Włodzimierz Sobkowiak. (MELD) |
| 2004 | 17.1 | – Review of Monolingual Dictionaries for Foreign Learners of English. (MELD) |

(b) **A rough count of Articles / Review Articles / Reviews in the International Journal of Lexicography (IJL) from 1998 to 2004 (1)**

| IJL Issue | ELDs | | | | GEDs | | | NEDs |
|---|---|---|---|---|---|---|---|---|
| | *MELD* | *BELD* | ***Total*** | *DALE* | *GMED* | *GBED* | ***Total*** | |
| 11.1 | 1 | 1 | **2** | - | - | 3 | **3** | - |
| 11.2 | - | - | - | - | 2 | 1 | **3** | 3 |
| 11.3 | - | - | - | - | 2 | 1 | **3** | 1 |
| 11.4 | 3 | 1 | **4** | 3 | 1 | - | **1** | 1 |
| 12.1 | 3 (+1) | - | **4** | 4 | 2 | - | **2** | 1 |
| 12.2 | 1 | - | **1** | 1 | 1 | 2 | **3** | 4 |
| 12.3 | 2 | - | **2** | 1 | 3 | 2 | **5** | - |
| 12.4 | 2 | - | **2** | 2 | 1 | 3 | **4** | 1 |
| 13.1 | 1 | - | **1** | - | 1 | - | **1** | 3 |
| 13.2 | - | 1 | **1** | - | - | - | - | 2 |
| 13.3 | - | 1 | **1** | - | 1 | 2 | **3** | 1 |
| 13.4 | 1 | - | **1** | 1 | 3 | - | **3** | 2 |
| 14.1 | (1) | - | **1** | - | 2 | 2 | **4** | 4 |
| 14.2 | 1 | - | **1** | 1 | 2 | 1 | **3** | 3 |
| 14.3 | 2 | - | **2** | 2 | 3 | - | **3** | 2 |
| 14.4 | 1 | - | **1** | 1 | - | 2 | **2** | 5 |
| 15.1 | - | - | - | - | - | - | - | 7 |
| 15.2 | - | - | - | - | 1 | 2 | **3** | 5 |
| 15.3 | 3 | 1 | **4** | 3 | 1 | - | **1** | 3 |
| 15.4 | 3 | - | **3** | 3 | 3 | 2 | **5** | - |
| 16.1 | 2 | - | **2** | 1 | - | 2 | **3** | 3 |
| 16.2 | - | - | - | - | 1 | - | **1** | 3 |
| 16.3 | 1 | - | **1** | 1 | - | - | - | 2 |
| 16.4 | 1 | - | **1** | 1 | - | - | - | 3 |
| 17.1 | 1 | - | **1** | 1 | 3 | 1 | **4** | 3 |
| Total | 31 | 5 | **36** | 26 | 33 | 26 | **59** | 62 |

CHAPTER 11

# Glimpses into receptive dictionary use: Proficient learners' perspective

## Rationale for the research perspective

Renata Szczepaniak

The present paper takes up the topic of receptive dictionary use from the narrow perspective of highly advanced learners performing the complex comprehension task of decoding contextually modified idioms. The results of the reported quasi-experiment with elements of the questionnaire- and observation-based study reveal a favourable impact of the monolingual learner's dictionary on performance. However, despite the fact that advanced learners turned out to be expert at extracting idioms from the macro- and microstructure, they exhibited little flexibility in the treatment of the located dictionary information and had problems when integrating it with the original context. Another factor responsible for reducing the overall positive effect of the dictionary on comprehension scores were confusing definitions. General trends in dictionary-user behaviour that emerge from the present study can be further explored in a large-scale research project.

## Introduction

Despite rapid advances in research on dictionary use, the intuitively appealing image of the pedagogical dictionary as a tool for boosting comprehension of a written text has not been confirmed beyond doubt. There are nearly as many reports on the negative influence of the dictionary on reading performance (Hosenfeld 1977, Bensoussan, Sim and Weiss 1984, Neubach and Cohen 1988, Padron and Waxman 1988) as on the positive one (Summers 1988, Laufer 1993, Knight 1994, McCreary and Dolezal 1999, Tono 2001: 75–83). Even though settling the issue remains beyond any single empirical study, it is interesting to steal yet another glance at the "very private matter" of dictionary use (Nesi and Haill 2002: 277), this time from the narrow perspective of the proficient learner performing a complex reading task.

On the face of it, advanced users make poor subjects in dictionary-user research – they look up few words (Hulstijn 1993: 144, Atkins and Varantola 1998:

34, Tono 2001: 112), mainly in order to confirm their correct guesses (Hulstijn 1993: 144, Knight 1994: 295), so the opportunities to observe their consultation process are scarce. Not surprisingly, the discussion so far has often been limited to high school students or, at best, to undergraduates not majoring in an L2 (Bensoussan, Sim and Weiss 1984, Knight 1994, Laufer 1993, McCreary and Dolezal 1999, Nesi and Haill 2002, Tono 2001: 75–83, Wingate 2002); potential difficulties of postgraduates majoring in English as a foreign language have hardly been considered at all. Actually, in view of the fact that "dictionaries are used more competently by the more linguistically sophisticated users" (Béjoint 1994: 151), setting out to test proficient learners' skills looks like nit-picking. Nonetheless, one might argue that whereas reference skills and linguistic sophistication are undoubtedly concomitant, advanced students, when given a suitably challenging task, may resort to the dictionary as often as less proficient ones, and likewise experience consultation problems.

## 1 Potential consultation problems: prior evidence

What sort of consultation pitfalls are learners likely to fall into? The most frequently reported problems are:

1. choosing the wrong entry or subentry (Nesi and Haill 2002)
2. difficulties with vocabulary or with the definition structure (Neubach and Cohen 1988; Wingate 2002)
3. problems with terminology, symbols and abbreviations (Neubach and Cohen 1988)
4. misinterpretation errors
   a. the *kidrule* (substitution) strategy, i.e. taking the familiar part of the definition for the equivalent of the headword (McCreary and Dolezal 1999; Nesi and Haill 2002; Wingate 2002)
   b. the *sham use* of dictionaries (Müllich 1990), i.e. disregarding or distorting the dictionary information so as to be able to retain one's preconceived notions
5. difficulties with applying the dictionary meaning in context (Nesi and Haill 2002)

One would expect proficient English-majoring students to be equipped with adequate searching strategies, indicative of basic metalinguistic skills and the knowledge of lexicographical conventions. On the other hand, overt optimism about the retrieval of the information and its application to the comprehension task might be unfounded.

## 2 The present study

### 2.1 Aims

This study attempts to explore the effects of dictionary use on reading comprehension performance of highly advanced learners. The following issue has been addressed:

– *Does dictionary use significantly improve comprehension scores in the target situation?*
– *How can the results be explained in view of learner- and dictionary- related factors?*

Given that exploitation of the entry presupposes mastery of searching procedures, in-depth investigation of the latter has been abandoned. Instead, special emphasis has been laid upon the retrieval of the information on meaning and integrating it with the original context.

### 2.2 Subjects

The subjects were MA students of two final years at the School of English, Adam Mickiewicz University in Poznań, Poland. Group A (no dictionary condition) comprised 75 students; Group B (dictionary condition) – 68 students.

### 2.3 Design

The experiment was designed with a view to recreating the natural situation when readers try to understand underlined fragments of short texts, mark the vocabulary items they wish to look up, and access the dictionary after the completion of the task so as to verify their inferences. (In this way, the impact of inference and the dictionary information could be separated and the disruptive effect of the search on the reading process prevented.)

#### 2.3.1 *Lexical targets*

Lexical targets were creative variants of idioms, i.e. "intentional, subjectively and stylistically motivated transformations of the meaning of the idiom in its contextual use" (Veisbergs 1997: 156), e.g.

(1) A bird in the hand is worth two in the *economic* bush [insertion]

(2) A *competent minister* in the hand is worth *many generals* in the bush [substitution] (Veisbergs 1997: 158).

Creative idiom variants guarantee the adequate level of task difficulty and thus boost the likelihood of consultations. Moreover, as they are not simple applications of dictionary meaning, it is easier to detect departures from the procedure of matching the dictionary text against the context, than if canonical forms had been used. Finally, when pre-testing the knowledge of lexical items would defeat the project, one-off forms of unfamiliar or vaguely familiar idioms make relatively safe targets.

#### 2.3.2 *Materials*

The task booklet was designed around four texts, selected from the British National Corpus on the basis of the author's subjective judgement of familiarity, transparency, and degree of transformation of the targets. The texts were eventually tailored to a manageable size of one paragraph, i.e. 7–9 lines (see Appendix 1).

#### 2.3.3 *The dictionary*

The choice of one monolingual dictionary – Longman Dictionary of Contemporary English, 3rd edition (LDOCE3) – was imposed upon the subjects so as to control the contents of the entry. This particular dictionary was selected for purely practical reasons: its popularity and availability. As there is a large number of copies of this edition in the Poznań School of English library, even those students who did not use LDOCE3 on a daily basis were likely to be familiar with its layout and conventions.

#### 2.3.4 *Method*

Apart from paraphrases of contextualised expressions, definitions elicited by means of word cues were used in the final stage of the experiment as a second means of comprehension measurement, testing for the canonical meaning of idioms. Whereas the evaluation of the difference between the "dictionary" and "no dictionary" condition for the paraphrase task was feasible within Group B (before- and after consultation), the results of the post-consultation definition task had to be compared with those in the control Group A, which replicated the experiment except the opportunity to use the dictionary. Searching procedures have been considered apart from the end-product of the consultation and registered in written protocols. Familiarity was controlled explicitly through a rating scale and self-reports of look-up motivation, to be compared with the quality of the answers.

#### 2.3.5 *Procedure*

In task one Group A was to provide English or Polish paraphrases of the underlined fragments. In the next section they extracted the idioms from the fragments and rated their familiarity on a scale of 1 (very well known) to 5 (completely unfamiliar). The last part instructed the subjects to form idioms using word-cues, and to write short definitions.

After the first task (identical with section one in Group A), Group B were allowed to use the dictionary. Meanwhile they recorded their consultations in the table (cf. Appendix 2), and entered corrections to the answers from section one underneath. After sections one and two had been handed in, the students were asked to extract the idioms from the underlined fragments and to reflect upon their motivation for the use or non-use of the dictionary. In section four the subjects formed idioms using word-cues, rated idiom familiarity, and wrote definitions.

## 2.4 Scoring procedures

### 2.4.1 *Paraphrase task*

To score paraphrases of the sentences containing modified idioms and the corrections to paraphrases (in Group B), Nippold and Martin's classification system (1989: 61) for the conventional idiom explanation task was adapted (see Table 1). Two raters scored the answers independently, checking them against the established interpretation (provided beforehand by three native speakers) and the classification system. Correct responses scored 1, partly correct ones were credited with a half point, incorrect ones got a 0. Interscorer agreement was 0.93 (Group A), 0.91 (Group B), and 0.92 (Group B – corrections to paraphrases). The contentious cases were later discussed and agreed upon.

**Table 1.** Scoring system for the paraphrase task

| **Example sentence:**<br>**"Tarring the whole of the Austrian public with the Haider brush is unfair."** | |
|---|---|
| Correct (1) | The response captures the idiomatic/ambiguous meaning of the fragment, e.g., "To think of the Austrian public as having the same faults as Haider is unfair." |
| Partly correct (0.5)<br><br>Related | The response is vague or reflects only a partial understanding of the idiomatic/ambiguous meaning of the fragment, e.g., "Treating the Austrian public like Haider is unfair." |
| Incorrect (0) | |
| Literal | The response reflects the concrete meaning of the fragment, e.g., "Covering the Austrian public with tar using the brush of the Haider type is unfair." |
| Unrelated | The response has nothing to do with the accurate idiomatic/ambiguous meaning of the fragment, e.g., "Urging the Austrian public to do sth is unfair." |
| Restatement | Part of the expression or paragraph was repeated or reworded without adding any new information, e.g., "The fact that we tar the whole Austria with the Haider brush is not fair." |
| No Response | The answer space was left blank or the student answered: "I don't know". |

#### 2.4.2 *Definition task*

The standard of correctness for the definition task was the conventional (dictionary) meaning. The scoring system for definitions draws on the criteria in Table 1.

(i) Correct (score: 1): the definition captures the conventional meaning of the idiom (as explained in reference works), e.g., "be considered to have the same faults as sb else".

(ii) Partly correct (score: 0.5): related, i.e., the response is vague or reflects only a partial understanding of the conventional meaning of the idiom, e.g., "treat sb like sb else".

(iii) Incorrect (score: 0):

 (a) Unrelated: the definition has nothing to do with the accurate conventional meaning of the idiom, e.g., "urge sb to do sth".

 (b) No Response: the answer space is left blank or the student answered: "I don't know".

#### 2.4.3 *Statistical measurement*

T-tests for correlated samples were run so as to check the difference between the "no dictionary" condition (Group B before dictionary consultation; henceforth B/1) and the "dictionary" condition (Group B after consultation, hereafter B/2) for the paraphrase task. The same type of test was carried out to check the difference between the two conditions for each of the four texts.

### 2.5 Results

#### 2.5.1 *Familiarity rating*

The results of familiarity rating (see Table 2) confirmed the adequacy of the target items. All four idioms were rated as belonging to the "low familiarity" range, with Idiom 3 only approaching "moderate familiarity" in Group B. The total means were similar, with Group A's familiarity somewhat lower than that of Group B.

**Table 2.** Familiarity rating*

| Idiom | A | B |
|---|---|---|
| Idiom 1 | 4.03 | 4.29 |
| Idiom 2 | 4.48 | 4.63 |
| Idiom 3 | 4.16 | 3.44 |
| Idiom 4 | 4.48 | 4.43 |
| Overall | 4.29 | 4.20 |

* Ranges of familiarity: 1–1.66 high familiarity; 1.67–3.32 moderate familiarity; 3.33–5.0 low familiarity

### 2.5.2 *Paraphrase task*

Considering the groups were not randomised, they show a remarkable uniformity of performance in the paraphrase task, with Group A scoring 0.20 and Group B 0.18 (see Table 3). For both groups Paraphrase 1 was the easiest; the next best in quality were Paraphrases 3 and 4; Paraphrase 2 fared badly. After look-up, performance in Group B picked up by 50%. The greatest improvement could be observed in the case of Idiom 4.

**Table 3.** Mean score and standard deviation in the paraphrase task*

| | **A** | | **B/1** | | **B/2** | | **Difference (B/1-B/2)** |
|---|---|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | Mean | SD | |
| Paraphrase 1 | 0.36 | 0.44 | 0.35 | 0.44 | 0.38 | 0.43 | 0.03 |
| Paraphrase 2 | 0.05 | 0.18 | 0.03 | 0.12 | 0.04 | 0.17 | 0.01 |
| Paraphrase 3 | 0.27 | 0.38 | 0.25 | 0.33 | 0.31 | 0.33 | 0.06 |
| Paraphrase 4 | 0.13 | 0.29 | 0.10 | 0.25 | 0.34 | 0.42 | 0.24 |
| Overall | 0.20 | 0.18 | 0.18 | 0.17 | 0.27 | 0.22 | 0.09 |

* Sample size A = 75; sample size B/1(before look-up) = B/2 (after look-up) = 68

As Table 4 illustrates, the difference of scores in the "no dictionary" and "dictionary" condition in Group B before and after dictionary consultation was significant at the 0.02% level. The difference between the scores for individual paraphrases in both conditions turned out to be statistically significant in the case of Paraphrase 4 at the 0.003% level, and for Paraphrase 3 at the 4.16% level.

**Table 4.** Paired samples t-test results (B/1 vs. B/2)

| | **Df** | **t-score** | **p-level** |
|---|---|---|---|
| Paraphrase 1 | 134 | 0.81 | 42.06% |
| Paraphrase 2 | 134 | 0.71 | 47.71% |
| Paraphrase 3 | 134 | 2.06 | 4.16% |
| Paraphrase 4 | 134 | 4.36 | 0.003% |
| Overall | 134 | 3.82 | 0.02% |

### 2.5.3 *Definition*

The dictionary favourably affected the mean score in the definition task, which increased from 0.11 in Group A to 0.51 in Group B (Table 5). The most spectacular growth occurred with definitions of Idiom 1. The remaining means rose almost as remarkably. Unfortunately, it was impossible to establish whether correct explanations reflected genuine comprehension in the case of direct quotations from the dictionary.

**Table 5.** Definition task (A vs. B)

| | Group A | | Group B | | Difference |
|---|---|---|---|---|---|
| | Mean | SD | Mean | SD | |
| Idiom 1 | 0.11 | 0.27 | 0.60 | 0.40 | 0.49 |
| Idiom 2 | 0.07 | 0.18 | 0.53 | 0.43 | 0.46 |
| Idiom 3 | 0.04 | 0.18 | 0.40 | 0.39 | 0.36 |
| Idiom 4 | 0.19 | 0.39 | 0.52 | 0.49 | 0.33 |
| Overall | 0.11 | 0.12 | 0.51 | 0.24 | 0.40 |

### 2.5.4 *Idioms searched and found*

Table 6 presents the number and proportion of searches and successful searches.

**Table 6.** Idioms searched and found*

| | Idioms searched | Idioms found | Success rate |
|---|---|---|---|
| Idiom 1 | 49 (72%) | 49 (72%) | 100% |
| Idiom 2 | 63 (93%) | 61 (90%) | 97% |
| Idiom 3 | 42 (62%) | 42 (62%) | 100% |
| Idiom 4 | 57 (84%) | 53 (78%) | 93% |
| **Overall** | 211 (78%) | 205 (75%) | 97% |

* per cent relative to the number of possible searches per idiom (68) and to the total number of possible searches for all idioms (272)

In view of the fact that numerous other words were searched, a 78%-look-up confirms high relevance of idioms to the task. 97% of all initiated searches ended in finding the idiom.

As evident from Table 7, the subjective impression of familiarity influenced look-up decisions: the less familiar the idiom the more consultations it invited. Unfortunately, the growing number of consultations did not entail increasing scores.

**Table 7.** Familiarity / Look-up / Improvement

| Idiom | ↓ Familiarity | ↑ Look-up | ? B/1-B/2 difference |
|---|---|---|---|
| Idiom 3 | 3.44 | 62% | 0.06 |
| Idiom 1 | 4.29 | 72% | 0.03 |
| Idiom 4 | 4.43 | 84% | 0.24 |
| Idiom 2 | 4.63 | 93% | 0.01 |

### 2.5.5 *Reported look-up behaviour*

Table 8 shows that, in general, the learners needed to find out the meaning of unfamiliar items rather than confirm their previous knowledge (63%), which is hardly surprising in view of familiarity rating values. Almost as many reports referred to checking the meaning of familiar items as to ignoring unfamiliar expressions (15%). Finally, 6% of decisions were not to look up familiar idioms.

**Table 8.** Reported look-up behaviour*

| | Searched | | Ignored | |
|---|---|---|---|---|
| Unfamiliar | 171 | (63%) | 41 | (15%) |
| Familiar | 44 | (16%) | 16 | (6%) |
| Total | 215 | (79%) | 57 | (21%) |

* per cent relative to the number of possible answers (4 x 68 = 272)

### 2.5.6 *"Paths"*

The subjects who decided to look up the target words followed one of the "paths", which emerge as a by-product of the scoring procedure:

Improvement: incorrect / partly correct paraphrase > look-up > correct paraphrase

No improvement: a) incorrect paraphrase > look-up > incorrect paraphrase
b) partly correct paraphrase > look-up > partly correct paraphrase

Confirmation: correct paraphrase > look-up > correct paraphrase

Deterioration: a) partly correct paraphrase > look-up > incorrect paraphrase
b) correct paraphrase > look-up > partly correct / incorrect paraphrase

Table 9 presents the total number and proportion of "paths" undertaken.

**Table 9.** "Paths"

| Paths | Total number | Proportion of path type out of all | | |
|---|---|---|---|---|
| | | Successful searches (205) | Initiated searches (211) | Possible searches (272) |
| Improvement | 46 | 22% | 22% | 17% |
| No improvement | 132 | 64% | 63% | 49% |
| Confirmation | 12 | 6% | 6% | 4% |
| Deterioration | 15 | 7% | 7% | 6% |

According to the "successful searches" column, which does justice to actual dictionary use and not merely its presence, there were 22% "improvement paths", 6% "confirmation paths", 7% "deterioration paths" and the stunning proportion of 64% "no improvement paths". Closer analysis of paraphrases shows that 35% (46) of "no improvement paths" involved no change in the original answer whatsoever. The remaining 65% (86 out of 132) corresponds to incorrectly altered answers. By combining this information with the data in the "successful searches" column, one obtains a more precise picture of the effect of dictionary use on performance in the situation under discussion:

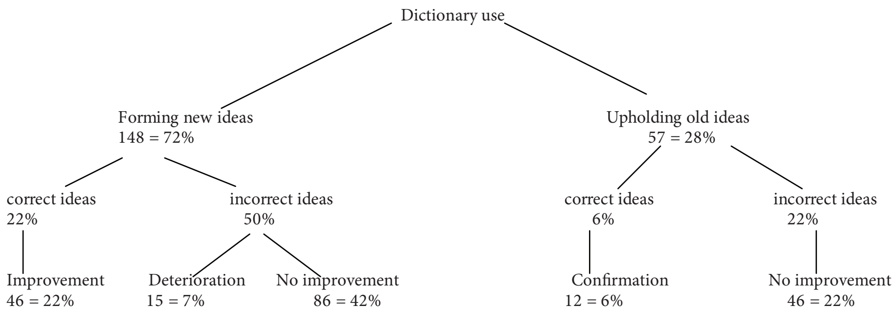


**Figure 1.** The effectiveness of dictionary use in the target situation

The dictionary helped to form new ideas in 72% of cases, and to uphold old ones in 28%. Disappointingly, dictionary use yielded negative results in 72% of cases by fostering or corroborating incorrect ideas (50% + 22%), and generated or supported positive ideas in merely 28% (22% + 6%). Out of 46 "improvement paths" only 19 involved a full-point increase, and 15 "deterioration paths" reduced the scores by one or by half a point. Consequently, the final result after consultation turned out to be as low as 0.27.

### 2.6 Discussion

Despite a statistically significant difference between the mean scores in the "dictionary" and "no dictionary" condition for the paraphrase task, the difference pertained only to two out of four targets. Although exposure to dictionary information considerably enhanced the subjects' awareness of the canonical meaning of idioms, this knowledge did not guarantee success in understanding their contextual variants. In other words, in view of high improvement rate in the definition task, the relatively poor performance in the paraphrase task cannot be unequivocally put down to the learners' ignorance of the canonical forms of idioms. Rather, the inefficient exploitation of the dictionary entry was the chief culprit, as confirmed by the "path" analysis.

What were the causes of an inefficient use of the dictionary?

First of all, the fact that 22% sustained incorrect ideas corresponds to Müllich's (1990) sham use of dictionaries. Namely, after reading the entry, users decided that the definition did not undermine their original (faulty) intuitions.

Secondly, the learners exhibited little flexibility in processing the dictionary information. They liked to incorporate fragmented definitions into paraphrases, instead of adjusting the dictionary meaning in context, e.g. *Wilson should be proud of his cooperation with Sting, Mr Simpson would charge him with as many offences as possible; Each vendor was saying sth again and again because they wanted to persuade others to accept their ideas; The USA shouldn't be blamed for crimes and faults of others*.

The difficulty in the comprehension of contextual variants is succinctly expressed in one student's words: "I'm guided by the knowledge of the idiom... but I don't know what to do with the rest". The subjects, accustomed to treating idioms as fossilised chunks, were taken aback upon encountering their innovative applications. Significantly, Idiom 4, whose form and meaning was least altered, had the highest improvement rate.

Moreover, some learners were intent on finding an equivalent at all costs, regardless of the definition or context, e.g. *a feather in one's cap = kropla w morzu potrzeb* 'a drop in the ocean of needs', *igła w stogu siana* 'a needle in a haystack'; *wierzchołek góry lodowej* 'the tip of the iceberg'; *as w rękawie* 'an ace up your sleeve'; *throw the book at = tonący brzytwy się chwyta* 'clutch at straws'; *have an axe to grind = mieć twardy orzech do zgryzienia* 'have a hard nut to crack'; *wtrącać swoje pięć groszy* 'put one's oar in'.

Finally, certain clusters of answers are indicative of what can be termed "quasi-kidrule", namely reliance on selected elements of text for interpretation, e.g. focus on the words *merely* and *one* that modified *a feather in one's cap* resulted in explanations that might be dubbed "one element out of many" (*one unimportant thing among many others; be merely an element in a larger group*, *one thing out of plenty*). Also, concentration on *increasingly crowded* yielded the interpretation "the beginning of something", reflected in *początek sukcesu* 'the beginning of success', *być zapowiedzią czegoś ważnego* 'be a prelude to sth important'. This strategy was applied to a lesser extent to the dictionary information, as evident in the example of the definition of *have an axe to grind – mieć dar przekonywania* 'to have great powers of persuasion', or in the paraphrase: *Every vendor has his/her own way of persuading others.*

Taken together, the strategies convey an impression of the subjects' adopting a superficial approach to the accessed information, be it dictionary or contextual information, and minimising the effort to integrate the two.

As far as dictionary-related factors are concerned, the unfortunate wording of two definitions became the source of considerable confusion. The definition of *have an axe to grind* ("to do or say something again and again because you want to persuade people to accept your ideas or beliefs"), by overemphasising the repetitiveness of actions, induced the belief that the idiom means 'repeat (oneself)', such an interpretation being reflected in the answers, e.g. *Every vendor has his/her own way of persuading others; Każdy sprzedawca ma tendencje do powtarzania się, gdyż zależy mu na przekonaniu do swoich pomysłów* 'Every vendor has a tendency to repeat himself/herself because they care deeply about bringing others round to their ideas'*; Każdy sprzedawca musi powtarzać w kółko to samo* 'Every vendor has to repeat things again and again', *mówić o czymś "w koło Macieju" żeby przekonać innych do swojego punktu widzenia* 'to talk about something over and over again in order to bring others round to your point of view'. Another definition – of *be/get tarred with the same brush* – does not spell out the connection one forms between the negative features of a group and an individual associated with them. The readers, left at a loss as to the reasons for blaming someone for "somebody else's faults or crimes", produced a disappointingly large number of incorrect paraphrases such as: *The US should not be blamed for the crimes and faults committed by GB as a colonial power; US shouldn't be blamed for the colonial times; US should not be blamed for crimes of colonialism.*

## 3 Conclusion

The study exposes the final processing of entry information as a weak link in the procedures directed at the retrieval of the writer-intended meaning, responsible for the reduction of the overall positive influence of consultation on reading performance. Whereas the inability to locate a multi-word item may indeed be responsible for underachievement in the case of intermediate learners, advanced users' failure is ascribable rather to insufficient attention being directed at proper understanding and application of the retrieved information. Apparently, with an increase in language proficiency, negative strategies, though largely eradicated from the macro- and microstructure search, linger on at the final stages of consultation to prevent successful exploitation of located information. As Bensoussan (1983: 345) put it, "the dictionary does not really 'give' the student the meaning of a word. The reader has to 'negotiate' that according to the context". It is those "negotiating" skills that crown the consultation act or, if they are lacking, spoil it.

Additionally, this study points to the wording of definitions as an important determinant of their readability and usability. Instances of proficient students being misled by imprecise definitions provoke questions about the suitability of the analytical definition couched in restricted vocabulary for rendering the subtleties of idiomatic meaning, and, ultimately, about the usefulness of the monolingual learner's dictionary for highly advanced readers, who normally target rare and complex vocabulary items. However, since a tiny sample of idiom entries has been considered, any conclusions regarding dictionary-related factors behind unsatisfactory comprehension performance remain tentative. A more systematic investigation, involving a larger sample of dictionary material and directed specifically at the comparison of the effectiveness of various types of idiom explanations (e.g., the analogy-based ones and definitions written in a restricted or non-restricted vocabulary) would be a valuable follow-up to the present study.

## References

*Dictionaries*

Summers, Della (ed.) 1995. *Longman Dictionary of Contemporary English.* (3rd edition.). Harlow: Longman. [LDOCE3].

*Other works*

Atkins, B. T. Sue and Krista Varantola. 1998. "Language learners using dictionaries: The final report on the EURALEX / AILA research project on dictionary use". In *Using Dictionaries: Studies of Dictionary Use by Language Learners and Translators*, B. T. S. Atkins (ed.), 21–81. Tübingen: Niemeyer.

Béjoint, Henri. 1981. "The foreign student's use of monolingual English dictionaries: A study of language needs and reference skills". *Applied Linguistics* 2: 207–222.

– 1994. *Tradition and Innovation in Modern English Dictionaries.* Oxford: Clarendon Press.

Bensoussan, Marsha. 1983. "Dictionaries and tests of EFL reading comprehension". *ELT Journal* 37: 341–345.

Bensoussan, Marsha; Donald Sim and Razelle Weiss. 1984. "The effect of dictionary usage on EFL test performance compared with student and teacher attitudes and expectations". *Reading in a Foreign Language* 2: 262–276.

Hosenfeld, Carol. 1977. "A preliminary investigation of the reading strategies of successful and nonsuccessful second language learners". *System* 5: 110–123.

Hulstijn, Jan H. 1993. "When do foreign-language readers look up the meaning of unfamiliar words? The influence of task and learner variables". *Modern Language Journal* 77: 139–147.

Knight, Susan. 1994. "Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities". *Modern Language Journal* 78: 285–299.

Laufer, Batia. 1993. "The effect of dictionary definitions and examples on the use and comprehension of new L2 words". *Cahiers de Lexicologie* 63: 131–142.

McCreary, Don R. and Fredric T. Dolezal. 1999. "A study of dictionary use by ESL students in an American university". *International Journal of Lexicography* 12: 107–145.
Müllich, Harald. 1990. *Die Definition ist Blöd! Herübersetzen mit dem Einsprachigen Wörterbuch. Das Französische und Englische Lernerwörterbuch in der Hand der Deutschen Schüler.* Tübingen: Niemeyer.
Nesi, Hilary and Richard Haill. 2002. "A study of dictionary use by international students at a British university". *International Journal of Lexicography* 15: 277–305.
Neubach, Abigail and V. Cohen. 1988. "Processing strategies and problems encountered in the use of dictionaries". *Dictionaries* 10: 1–19.
Nippold, Marilyn A. and Stephanie Martin. 1989. "Idiom interpretation in isolation versus context: A developmental study with adolescents". *Journal of Speech and Hearing Research* 32: 59–66.
Padron, Yolanda N. and Hersholt C. Waxman. 1988. "The effect of ESL students' perceptions of their cognitive strategies on reading achievement". *TESOL Quarterly* 22: 146–150.
Summers, Della. 1988. "The role of dictionaries in language learning". In *Vocabulary and Language Teaching*, R. Carter and M. McCarthy (eds.), 111–125. London: Longman.
Tono, Yukio. 2001. *Research on dictionary use in the context of foreign language learning: Focus on reading comprehension*. Tübingen: Niemeyer.
Veisbergs, Andrejs. 1997. "The contextual use of idioms, wordplay, and translation". In *Traductio. Essays on Punning and Translation*, D. Delabastita (ed.), 155–176. Manchester, UK: St. Jerome Publishing.
Wingate, Ursula. 2002. *The Effectiveness of Different Learner Dictionaries: An Investigation into the Use of Dictionaries for Reading Comprehension by Intermediate Learners of German.* Tübingen: Niemeyer.

## APPENDIX 1

Texts used in the experiment:

1) ...Sting, as a songwriter, had been signed to Virgin by the head of the publishing division, Carol Wilson. It was Wilson who had "discovered" the singer as an unknown in his native Newcastle... Under her direction, and with Virgin money, Sting had recorded a series of demos which had led to his association with the Police, a recording contract with A&M and, ultimately, the hit parade. The arrangement Sting had signed with Virgin had been for a 50–50 split, rising to 60–40 in Sting's favour after two years. It was a standard contract for an unknown songwriter, but as the Police began to enjoy enormous success, so Sting's value in the market-place rose... Sting and Carol Wilson had remained friends throughout the negotiations. In fact it was she who had advised him to get a second opinion on the matter from a lawyer... For Wilson, <u>Sting's success</u>, and the prodigious royalties it had earned for Virgin, <u>was merely one feather in an increasingly crowded cap.</u>

British National Corpus World Edition. Source description: Text id FNX; Richard Branson: the inside story. Brown, M, Headline Book Publishing plc, London (1989), 157–303. Sample containing about 42225 words from a book (domain: applied science).

2) In their dispute with Iran the British set out to close off all outlets for Iranian oil. American oil companies (fearful of the knock-on effects to themselves in the Middle East if Iranian radicals were seen to succeed) proved more enthusiastic allies than the State Department.

The British, however, remained suspicious of American motives, and resentful of their advice. Eden's private secretary, Shuckburgh, thought it "very offensive" when, following an unsatisfactory meeting on Anglo-American policy towards Iran, he was told by Acheson that the British had to learn to live in the world as it was. Yet Acheson's remarks were hardly surprising when his department was receiving so many reports of the strength of Arab feeling against the British. By 1952 he was becoming all the more determined that the United States should not be tarred with the colonial brush.

British National Corpus World Edition. Source description: Text id HY8; The special relationship. Bartlett, C J, Longman Group UK Ltd, Harlow (1992), 1–124. Sample containing about 46090 words from a book (domain: world affairs).

3) ...However, this remarkable literary work -- even given an army of fans as keen as his niece -- would not have brought in very much income, nor would the journalism, and it was to be assumed the trust provided the rest. What he was also wondering was whether Harriet had received any letters of the kind received by Tom Fearon and if so, whether she had kept them? But he had little reason yet to ask for a search warrant and Mr Simpson would go purple in the face and throw every legal book in his considerable library at him if he so much as tried. But it would be worth making contact with Miss Frances Needham-Burrell when she arrived and asking her -- when sorting out her cousin's effects -- to keep an eye open for letters with a threatening or abusive content.

British National Corpus World Edition. Source description: Text id CEB; A season for murder. Granger, Ann, Headline Book Publishing plc, London (1991), 21–134. Sample containing about 37282 words from a book (domain: imaginative).

4) Lee Rothstein, director of Southport, Connecticut-based market research firm New Science Associates Inc's advanced network computing service, takes umbrage with a lot of current thinking on open systems. He says that much of his firm's research showed, at an early stage, that open systems concepts "weren't going to cut it"; moreover, that many "were a myth". In theory, guidelines for open systems are formulated when a committee of vendors come to some sort of agreement over architectural definitions and create a standard. "However, in reality, each vendor has its own technological axe to grind and the results miss the mark", says Rothstein. The committees "never get to the leading edge because large companies usually don't understand new technology, so they drag their feet on the committees until they are able to catch up."

British National Corpus World Edition. Source description: Text id CTM; Unigram x. APT Data Services Ltd. (1993–04/1993–05). Sample containing about 8042 words from a periodical (domain: applied science).

## APPENDIX 2*

A1 No.........

Read carefully each text (they all come from different sources) and write English or Polish paraphrases of the underlined fragments in the space provided. You may disregard the formal accuracy of your paraphrases.

Text 1) ..............................................................................................................................

* The original Polish instructions have not been provided here so as to economise on space.

Text 2) ........................................................................................................................................
Text 3) ........................................................................................................................................
Text 4) ........................................................................................................................................

A2 No.......

Write the idiomatic expressions that appear in the underlined fragments. State your familiarity of the idiom before this test by circling one of the numbers 1–5: 1 – very well known, 2 – well known, 3 – vaguely familiar, 4 – rather unfamiliar, 5 – completely unfamiliar.

Text 1)
Text 2)
Text 3)
Text 4)

1. ..........................................................................................................................1 – 2 – 3 – 4 – 5
2. ..........................................................................................................................1 – 2 – 3 – 4 – 5
3. ..........................................................................................................................1 – 2 – 3 – 4 – 5
4. ..........................................................................................................................1 – 2 – 3 – 4 – 5

A3 No.......

Use the cues to build idiomatic expressions; if necessary, add articles, prepositions or other missing words. Then define the idiom (in English or in Polish). You may disregard the formal accuracy of your definitions.

1. feather, cap
idiom: ........................................................................................................................................
definition: ........................................................................................................................................
2. tar, brush
idiom: ........................................................................................................................................
definition: ........................................................................................................................................
3. book, throw
idiom: ........................................................................................................................................
definition: ........................................................................................................................................
4. axe, grind
idiom: ........................................................................................................................................
definition: ........................................................................................................................................ .

B1. No......

Read carefully each text (they all come from different sources) and circle any words/phrases that you would like to look up in the dictionary. Then write English or Polish paraphrases of the underlined fragments in the space provided. You may disregard the formal accuracy of your paraphrases.

Text
1) ........................................................................................................................................
Text
2) ........................................................................................................................................
Text
3) ........................................................................................................................................
Text
4) ........................................................................................................................................

B2 No........

You may consult the dictionary now if you wish. Register every consultation in the table: what you are looking for, under which headword, whether you have found the desired information. If you think your first paraphrase is faulty, there is space for corrections under the table. (Please do not add anything to the first part of the test.)

| Word/phrase looked up | Headword looked up | Information | | |
|---|---|---|---|---|
| | | present | partly present | absent |

Possible changes in the paraphrases:

1. ..........................................................................................................................................
2. ..........................................................................................................................................
3. ..........................................................................................................................................
4. ..........................................................................................................................................

Please circle one option that is true in your case:

1. At home I use a) the same b) a different dictionary.

Point 2 concerns people who have circled 1b.

2. Using the dictionary other than my own made arriving at the desired information

a) extremely b) rather c) a little d) not at all difficult.

B3 No.......

Fill in the table with the idioms that appear in the underlined fragments. Tick (√) the appropriate column.

Text 1)
Text 2)
Text 3)
Text 4)

| Idiom | I knew the idiom so I did not look it up. | I knew the idiom but I looked it up. | I didn't know the idiom so I looked it up. | I didn't know the idiom but I did not look it up. |
|---|---|---|---|---|
| 1 | | | | |
| 2 | | | | |
| 3 | | | | |
| 4 | | | | |

B4 No......

Use the cues to build idiomatic expressions; if necessary, add articles, prepositions, and other missing words. State your familiarity of the idioms before this test by circling one of the numbers 1–5: 1 – very well known, 2 – well known, 3 – vaguely familiar, 4 – rather unfamiliar, 5 – completely unfamiliar. Then define the idiom (in English or in Polish). You may disregard the formal accuracy of your definition.

1. feather, cap

idiom:......................................................................................................................1 – 2 – 3 – 4 – 5

definition: ......................................................................................................................................

2. tar, brush

idiom:......................................................................................................................1 – 2 – 3 – 4 – 5

definition: ......................................................................................................................................

3. book, throw

idiom:......................................................................................................................1 – 2 – 3 – 4 – 5

definition: ......................................................................................................................................

4. axe, grind

idiom:......................................................................................................................1 – 2 – 3 – 4 – 5

definition: ......................................................................................................................................

Thank you for your cooperation,
Renata Szczepaniak

CHAPTER 12

# The learner's dictionaries and grammar

## A comparison

Geart van der Meer

In this paper I have studied the way in which the five major English learner's dictionaries have managed to present and discuss the grammatical (syntactic or morphological) behaviour of words known to cause trouble to foreign learners. It turns out that learner's dictionaries in general have serious problems in incorporating grammatical information that goes beyond simple things like verb patterns and information like (un)countability. It also emerges that approaches even in one and the same dictionary can be quite diverse, in that on one occasion grammatical terminology is seemingly avoided at all costs, whereas elsewhere for related cases an apparently different editor thinks nothing of using a number of quite specific grammatical terms. There are some indications that the compilers of these dictionaries were not all of them equally conversant with formal English grammar. Likewise, there is evidence that the needs of the *foreign* learners were not always equally well recognised.

### 1 Introduction

In 1997 I published a paper on how successful two leading Dutch L1-L2 desk dictionaries were in presenting grammatical information about translation equivalents (van der Meer 1997). I did so by examining a number of Dutch words that are known to cause trouble because of the fact that their translation equivalents behave differently in English from a grammatical point of view. The word 'grammar' should here be taken in a broad sense: anything to do with syntax and/or morphology. In the present paper I shall be doing something similar by comparing the grammatical information about a number of words in what has by now become the 'big five', the five main learner's dictionaries of English (OALD6, LDOCE4, COBUILD4, CALD and MEDAL). The procedure will have to be roughly the same as in my 1997 paper, where I termed the approach 'word-based', since a dictionary does not normally enter into general grammatical discussions (like the use of the tenses) but only deals with separate words in the wordlist. The words to be dealt with have been selected partly on the basis of the 1997 paper, but also by informally browsing Michael Swan's magnificent *Practical English Usage* (Swan 1995).

## 2 The word-based treatment of grammatical problems in the five learner's dictionaries

In this section I will discuss a number of words causing grammatical problems often discussed in grammars of English. After summaries of their treatment in the dictionaries I will draw conclusions and give my own improved descriptions or at least suggestions for them. These will consist of two steps: a *semantic synopsis*, with a general semantic description, followed by the *entry proper*, with a more detailed description, examples and further information (cf. van der Meer 2002 for an example of this method of organising an entry). In section 3 I will draw some general conclusions about the feasibility of combining dictionary and grammar in the case of these and similar words.

### 2.1 The s*ome/any* problem

I will here study how the several dictionaries solved the problem of explaining the well-known distributional difference between these two words (i.e. the unstressed use).

SUMMARIES
At *any*, the most detailed information is provided by OALD6, which runs as follows (boldface emphasis by me): "*det.* used with **uncountable** or **plural nouns** in **negative sentences** and **questions**, after ***if*** or ***whether***, and after **some verbs** such as *prevent, ban, avoid*, etc. to refer to an amount or a number of sth, however large or small.... *He forbids any talking in class*". After the HELP sign it is added that in positive sentences *some* is normally used. It is a pity that the examples *prevent, ban, avoid* are not generalised to *all* words with a negative impact, including prepositions like *without*. This might have been helpful for the attentive reader. OALD6 rather uneconomically repeats more or less the same information after *any* as pronoun. MEDAL likewise refers to questions, negatives and clauses with 'if', but does not use the words '**uncountable** or **plural nouns**', saying instead it is used "for saying or asking whether there is a small amount of something or a small number of people or things", which is saying the same informally. CALD gives no useful information at all at *any*, but refers us to a note at *some*, which tells us merely that *any* is used in questions and negative sentences. LDOCE4 also restricts itself to the general remark about questions and negative sentences, but forgets about conditionals

and negative words like *prevent*. COBUILD4 finally, is characteristically wordy but manages to generalise to any negative context in addition to questions and conditional clauses.

CONCLUSION
Though OALD6 is the most detailed, one may wonder how effective it is to focus so exclusively on syntactic and semantic contexts, or the contrast with *some* [which is not discussed under *any*!], instead of the *semantic* definition of the word *any* itself, for strikingly a real direct definition is avoided. It is made to look as if the use of *any* depends on the (syntactic and semantic) context instead of the other way round: cannot we also say that *any* determines, or fits into, the context precisely because of its meaning?

PROPOSAL
The s*ome/any* problem obviously is how to describe its semantics without indulging in abstract grammatical or logical terms like 'negative polarity' and 'negative orientation' (cf. CGEL at *any*). I suggest we could describe *any* like this:

> SEMANTIC SYNOPSIS
> ***any*** (determiner, pronoun) *usu. unstressed* indicates either an indefinite **amount** or an indefinite **number** (plurals only), but these amounts or numbers (of things/persons etc.) are not necessarily asserted to exist; it is therefore a word with a non-committal or even negative feel about it; more technically, it is 'non-assertive'.
> THE ENTRY PROPER
> ***any¹*** [*often unstressed*], when used for amounts or numbers ***any¹*** is therefore used in 'nonassertive' contexts, i.e.: 1. *negative words*, either a. overt or b. implied, 2. *questions*, 3. *conditional sentences*, either with *if, in case, should* or otherwise [CGEL 10.61].
> Examples: 1a. after *overt* negative words: [speaks for itself], 1b. after implied negative words, like *fail, prevent, avoid, ban, without, hardly, rarely, seldom, hard/difficult, little/few, too.* 2. in questions [speaks for itself], 3. in conditional contexts, after *if, in case* etc. and conditional *should.*[1]

## 2.2 The indefinite article *a*

Since the entries for the indefinite article are very long and complex, and in addition differ widely from each other, I will focus on one particular problem, i.e. *a* as for instance used in *He is a teacher*, where languages like Dutch have no article at all (*Hij is leraar* 'He is teacher').

SUMMARIES (boldface emphasis mine)
OALD6 (with 10 senses) (def. 4) says "used to show that sb/sth is **a member of a group or profession**: *Their new car's a BMW. She's a Buddhist. He's a teacher. Is that a Monet (= a painting by Monet)?*" MEDAL (with 15 senses) 3 "used when you say what **class, type, or group** someone or something belongs to, or what job someone has: *Ruth's father was a lawyer. He's a liar and a cheat. Greece has been a republic since 1973…*". CALD (11 senses) "2 used to state **what type of thing or person something or someone is**: *She wants to be a doctor when she grows up. This is a very mild cheese. Experts think the painting may be a Picasso (= by Picasso).*" LDOCE4 (16 senses) 2 "used to show that you are referring to **a general type of person or thing and not a specific person or thing** *Would you like a sandwich? I want to train to be an engineer. He's a really nice man. Take a look at this. It needs a good clean…*". COBUILD4: (14 senses) 8 "You use **a** when you are saying **what someone is or what job they have**. *He was now a teacher and a respectable member of the community.* 12… in front of the names of artists to refer to one individual painting or sculpture created by them... *a Van Gogh...*".

CONCLUSION
The large number, and diversity, of the independently presented sense definitions in the five dictionaries are truly bewildering and should make us pause. It seems to me that the essential meaning of *a + NOUN* is that you refer to something of which there is at least *more than one* specimens of what the noun refers to, and that we are considering one *unidentified* member from that class. From this it would follow that *a + NOUN* may be used 1. to refer to either *one specific though unspecified and individual class member* or 2. as *an example characterising the entire class*. Concisely: *a + NOUN* establishes class membership (*to what class/group does it belong?*), and class membership at the same time *defines*, due to the definition of that class (*what is it?*). The five definitions given above either focus on the former or on the latter, but because they are all placed somewhere in a list of separate definitions the suggestion is very strong that the other senses are different. I tend to disagree. I was initially a little shocked at OALD6's juxtaposition of *He's a teacher. Is that a Monet?*, but on further consideration this can be defended: both are given as examples standing for the entire class and hence define. In *I spoke to a teacher yesterday* the stress is more on the individualising function of the article, but this is the other side of the *same* coin.

PROPOSAL
I would therefore plead for the more general part of the entry to begin with a *synoptic* definition. This is important, for no dictionary seems to realise that all its several 'senses' are special applications – or uses – of only one basic sense.

> SEMANTIC SYNOPSIS
> The indef. art. *a* (+ NOUN) means that the NOUN is one *unidentified* member out of class of similar persons/things.[2] It is therefore only used before countable nouns.
> THE ENTRY PROPER
> Here examples of the several uses can be given, possibly with glosses but without real definitions. Thus, we can avoid suggesting as e.g. LDOCE does that *a* can mean *each*, as in *once a month*. Admittedly, LDOCE and the other dictionaries do often start by saying *used to show/in…*, but a general definition is never given.

### 2.3 *the*

Once again, the entries are very complex and I will therefore focus on one particular aspect, that is the use of *the* as in *the sublime* and in *the poor*.

SUMMARIES
OALD6: "5 used with adjectives to refer to a thing or a group of people described by the adjective: *With him, you should always expect the unexpected. the unemployed / the elderly / the French*."
MEDAL: "8 used before an adjective to form a noun when you are referring to the type of thing described by the adjective: *the possible/supernatural/unthinkable etc: Are we attempting to achieve the impossible? a belief in the supernatural…* 10 used before an adjective to form a plural noun when you are referring to people of a particular type or people from a particular country: *the rich/poor/young/unemployed etc: a policy of taxing the rich to help the poor; the French/British/Japanese etc: The Japanese eat a lot of seafood*."
CALD: "used before some adjectives to turn the adjectives into nouns which refer to one particular person or thing described by the adjective: *It seems that the deceased (= this particular dead person) had no living relatives. I suppose we'll just have to wait for the inevitable (= the particular thing that is certain to happen)*. 8 used before some adjectives to turn the adjectives into nouns which refer to people or things in general that can be described by the adjective: *She lives in a special home for the elderly. The French were defeated at Waterloo in 1815*."
LDOCE4: "8 used before an adjective to make it into a plural noun when you are referring to all the people that the adjective describes. *She devoted her life to helping the poor. a school for the deaf. wars between the English and the French.* 9 used before an adjective to make it into a noun when you are referring to the particular kind of situation or thing that the adjective describes *Come on now, that's asking for the impossible. fantasy movies that make the unreal seem real*".

COBUILD4: "8 You use **the** with nationality adjectives and nouns to talk about the people who live in a country. *The Japanese, Americans, and even the French and Germans, judge economic policies by results.* 9 You use **the** with words such as 'rich', 'poor', 'old', or 'unemployed' to refer to all people of a particular type… *care for the elderly and the disabled.* 11 You use **the** in front of an adjective when you are referring to a particular thing that is described by that adjective. *He knows he's wishing for the impossible... I thought you might like to read the enclosed.*"

CONCLUSION

To me the OALD6 is at first glance the least satisfactory because of its rather unfortunate combination of two disparate things (*the unexpected* and *the elderly*), and its less than felicitous wording, calling *the unexpected*, something rather abstract, a 'thing'. The latter, however, is done by others as well, probably to keep the defining language simple, but for the rest they keep the two apart. Moreover, when OALD6 is applied literally, **the Italian* would be correct English (instead of correct *the Italians*). CALD apparently has sensed the inherent danger here and has added the word "some adjectives", which is a little vague but may at least alert us to the risks involved, though one may still frown a little at the assertion that adjectives become 'nouns' here.[3] LDOCE4 (def. 8) still allows **the Italian*, but helpfully points out that this use involves *all people*. Def. 9 uses the words "situation or thing", where the former is not quite satisfactory either: *the impossible* is neither a thing nor literally a situation. COBUILD4 9 is acceptable, but 8 and 11 would again allow **the Italian* and **the new* for *the new thing*.

PROPOSAL

Though, like *a*, the definite article is extremely complicated to describe in all its uses, it would again be wise to start with a general synopsis, for the general meaning is in fact quite simple:

> SEMANTIC SYNOPSIS
>
> ***the*** (def. art.) is used when, in whatever way, the hearer/speaker is told or simply knows who or what the following noun refers to.
>
> THE ENTRY PROPER
>
> In the entry proper a number of things need to be made clearer. For instance, we need more detailed information about the adjectives that can follow *the* without the support of a following noun. Thus, the names of nationalities need to be specified as those ending in *–(i)sh*, *-ch* or *–ese* (e.g. *the Welsh/British; the Dutch/French; the Portuguese/Chinese etc.*). Of course, syntactically these words behave as ordinary plurals.
>
> Likewise, it needs to be stated that ***the*** can be followed by adjectives without following nouns when these adjectives denote an **entire** class of people

(the 'generic' use of ***the***). Typical examples: *the old/young, the innocent, the poor, the unemployed* etc. These phrases in their entirety behave as plurals: *the poor* ***are*** *in need of help*. Since it is the adjectives themselves and **not** the article *the* that determine if a following noun may be absent, a list of the most typical adjectives will have to suffice.
Thirdly, some typical examples are required of ***the*** used similarly but then for adjectives denoting abstract qualities, as in *the unexpected, the supernatural, the inevitable, the unreal, the impossible*. These, of course, are singulars.
Finally, this section of the entry should perhaps mention some exceptional examples like *the deceased, the enclosed* (which are only singular), *the bereaved* (either singular or plural) (cf. CGEL 7.23ff.).

### 2.4 *me*

I will here discuss *me* and its uses as opposed to *I*.

SUMMARIES
OALD6 "the form of I that is used when the speaker or writer is the object of a verb or preposition: Don't hit me. Excuse me! Give it to me. Hello, it's me. You're taller than me. "Who's there?" "Only me." HELP The use of me in the last three examples is correct in modern standard English. I in these sentences would be considered much too formal for almost all contexts, especially in BrE." MEDAL "Me can be used in the following ways: as the object form of I: They invited me to lunch. In spoken English as a one-word answer, or after the verb 'to be': "Who did they blame?" "Me, of course". Jack didn't do it. It was me." CALD "me (PERSON) pronoun used, usually after a verb or preposition, to refer to the person speaking or writing: Is there one for me? She gave me some money. Could you pass me that book? It wasn't me who offered to go, it was Charlotte." LDOCE4 "[object form of 'I'] 1 used by the person speaking or writing to refer to himself or herself. Stop, you're hurting me. He bought me a drink. Give that book to me. She's two years older than me. That's me, standing on the left." COBUILD4: "... ME is used as the object of a verb or a preposition."

CONCLUSION
OALD6's phrasing is rather awkward: a speaker is not the object of a verb, to begin with, nor is it likely that all users will realise that in *It's me* the word *me* is considered an object. The editor was here apparently not shy of using grammatical terms like *object*. Of course, the examples are beyond reproach. MEDAL is in this respect better, though it fails to cover the case of *better than me*. CALD does cover this case, at least if one realises that *than* is a preposition, but misses out on one-word answers (*Me!*). It avoids the term *object*. LDOCE4 starts with "[object form of 'I'],

but after this offers no direct grammatical description and just restricts itself to giving the main contexts in example sentences. COBUILD4 obviously fails to account for cases like *It's me!* or *Me!*.

PROPOSAL
The point here, as so often, is: do we use grammatical terms or not. In this case it seems possible to avoid the word *object* altogether, though we could of course start the description in the LDOCE way: "[object form of 'I']", after which the sense definition follows:

> SEMANTIC SYNOPSIS
> [object form of 'I']; ***me*** is used when the speaker or writer refers to him/herself and is used whenever it does not precede the verb of the sentence (or when there is no verb at all).
> THE ENTRY PROPER
> Examples (those from OALD6 will do): *Don't hit me. Excuse me! Give it to me. Hello, it's me. You're taller than me. 'Who's there?" "Only me."*

In this way we can avoid grammatical terminology.

### 2.5 *she*

In connection with *she* I plan to discuss the gender problem (*he* or *she*?), and in addition give some attention to its use for inanimate objects like countries, ships, cars etc. There will be some references to *they* meaning 'he or she'.

SUMMARIES
OALD6 rather surprisingly does not mention the use of *she* for vessels etc. And we have to go to *they* to read the remark "used instead of *he* or *she* to refer to a person whose sex is not mentioned or not known: *If anyone arrives late they'll have to wait outside*". MEDAL mentions "2 MAINLY LITERARY used for referring to a nation, a ship, or a car, when it has already been mentioned or when it is obvious which one you are referring to: *There were over two thousand people aboard the Titanic when she left England. She'll do a hundred and twenty miles an hour easily.*" Other than OALD6 it also lists *s/he*: "*s/he* abbreviation *she* or *he*: used when you do not know if the subject is male or female: *A student may retake the exam if s/he does not pass first time*." At *they* it makes a remark similar to OALD6, but also helpfully adds that this is "especially when you are referring back to a word such as 'everyone', 'someone', or 'anyone' ". CALD also refers to its use ('sometimes') to refer to "especially a country, ship or vehicle, that has already been mentioned". It also lists and discusses *s/he* (used in writing) as well as *they*, the latter "used to avoid saying 'he or she' ". LDOCE4 includes "*old-fashioned* used to refer to a country, ship, or

vehicle", and also lists *s/he*. Under *they* it mentions the use to avoid saying *he* or *she*, without further specifications. COBUILD4 is the only one to mention "Some writers may use *she* to refer to a person who is not identified as either male or female. The do this because they wish to avoid using the pronoun 'he' all the time. Some people dislike this use and prefer to use 'he or she' or 'they'." It also mentions the occasional use "to refer to a country or nation" or "a car or a machine". COBUILD4 also lists *s/he*, and *they* for 'he or she', with the warning that "Some people think this use is incorrect".

CONCLUSION
OALD6 surprises because it leaves undiscussed the use for countries etc., and it is in general surprising that no dictionary gives a cross-reference to *they* as a solution to the *he/she* problem. COBUILD4 is the only one to mention the controversial use of *she* instead of the conventional *he* (meaning by implication both 'he' and 'she'). It is also helpfully adds that this use is not considered acceptable by everybody, as is also the case for *they* meaning 'he or she' when referring back to (pro) nouns that are gender-indefinite (e.g. *person*, *someone* etc.). LDOCE4, in a grammar note, also points out this latter fact.

PROPOSAL
It would be advisable to pay proper attention to the gender debate in modern society and the means to avoid hurting sensitivities here. Therefore, there will have to be some cross-references to *they, he,* and *s/he*.

> SEMANTIC SYNOPSIS
> [*pronoun, subject form*] ***she*** is used before the verb and refers to girls or women and often also to the higher female animals, esp. domesticated animals (otherwise ***it*** is used). It may also be used for countries as well as big cities (in formal language), cars, machines, ships etc. (formal, old-fashioned or as insider language). In formal writing (e.g. academic writing) ***she*** is sometimes used to mean 'he or she' to avoid using *he* or the combination *he or she*. Cf. ***they*** for a similar use. Also see ***her*** and following ***s/he***.
> THE ENTRY PROPER
> Examples: (after one or two obvious examples involving females) *There were over two thousand people aboard the Titanic when she left England. She'll do a hundred and twenty miles an hour easily* (said of a car).

### 2.6 *hardly*

I will here pay attention to the way *inversion* after *hardly* is dealt with. Because the word *inversion* is a rather technical term many people will not understand it will be interesting to see how the dictionaries have managed to avoid it.

SUMMARIES
OALD6 only manages an example, *Hardly had she spoken than she regretted it bitterly,* without direct comment, and a little unfortunately followed by *than* rather than the more regular *when*. It is only in the inset WHICH WORD that inversion is obliquely referred to by means of an example and the comment: "then the subject and verb are *turned around*" [my emphasis]. MEDAL: "In stories and in formal English, *hardly* is sometimes used at the beginning of a sentence *before an auxiliary verb* [my emphasis]: *Hardly had she begun to speak, when there was a shout from the back of the hall*." CALD only gives an example, without the characteristic *when* of the main clause: *Hardly had a moment passed before the door creaked open*. There is no comment on word order. LDOCE4, in the GRAMMAR inset (called Usage Note on the CD), only very indirectly remarks: "Do not use *hardly* at the beginning of a sentence, except in very formal writing. *I had hardly got in the house when the phone rang* is the usual way to say this. It is possible to say *Hardly had I got in the house when the phone rang* but this is very formal." The phenomenon of inversion is not explicitly mentioned. COBUILD4 illustrates this use in the sentence definition itself, with the comment in the extra column: "ADV before v.", but then the real example sentence oddly enough does not give sentence-initial *hardly*.

CONCLUSION
None of the five uses technical grammatical terms, but three of them at least manage one or more examples with inversion, plus occasionally some comment pointing out the phenomenon of inversion without using this word. LDOCE4 and CALD do not go beyond providing the example.

PROPOSAL
It would at least be wise to more explicitly draw attention to inversion, and then some mild terminology cannot be entirely avoided.

> SEMANTIC SYNOPSIS
> This will be very brief, with the sense definition.
> THE ENTRY PROPER, with at least the remark after one or two earlier examples:
> In more or less formal language *hardly* (in the combination *hardly… when/before*) may also open the sentence: e.g. in *Hardly had I got in the*

*house when the phone rang,* but then the verb immediately follows (and precedes the subject), hence not: **Hardly I had got in the house*...... [perhaps some more examples, especially with ***do/did*** replacing a main verb in addition to the auxilaries].

### 2.7 *ill*

I will here focus on *ill* in the sense 'suffering from a disease' and its syntactic behaviour in British English, i.e. that *ill* is normally not used before a noun in this sense.

SUMMARIES
OALD6 expressly says "[not usually before noun]". MEDAL manages no remark at all, and only points out that *ill* in the sense 'bad or harmful' (e.g. *ill-discipline*) comes only before a noun. CALD makes no remark at all about placement, not even in its usage note, and the same applies to LDOCE4. COBUILD4 alerts us in the extra column to the fact that *ill* usually follows the copula: 'usu v-link ADJ".

CONCLUSION
For three out of five the treatment is disappointing, and it is only the examples which have to convey the (implicit) message. Likewise, the fact that *ill* is permitted attributively in combinations like *terminally ill child* is not explicitly commented on, though in some cases the examples are provided. As far as the adjective *sick* is concerned, which is often the alternative to *ill*, none of the five makes any explicit reference to it under *ill* or remarks on its normal position (i.e. as a *premodifier*) in the sense 'suffering from a disease'.

PROPOSAL
It would be wise to keep ***ill*** and ***sick*** apart and at least point out the differences, including that between British and American English.

> SEMANTIC SYNOPSIS
> ***ill*** (*adj., mainly predicative*) unwell, suffering from a disease – not used before the noun it modifies, except in the case of ***terminally ill*** (BrE); see also at ***sick***.
> THE ENTRY PROPER
> here some examples in non-attributive positions, including one with ***terminally ill patients***, where it exceptionally occurs before the noun it modifies.

### 2.8 *sick*

I will here focus on the same problems as in the case of *ill*: how does it behave syntactically? As a student I was taught that *sick* 'physically or mentally ill' was used *before* the noun in British English and when it followed a copula it meant 'wanting to vomit', though in American English it then could also mean 'ill'. This is today still largely true.

SUMMARIES
OALD6 first gives "1 physically or mentally ill" with examples in both attributive and predicative position, though it adds "(AmE) *I can't afford to get sick (= become ill)*". Its second sense is "2 [not usually before noun] (especially BrE) feeling that you want to vomit" after which an example is given that not only contradicts the statement "not usually before noun" but is not a very apt illustration of the sense under discussion either: … *a sick feeling in your stomach*". MEDAL begins with a promising start "2 MAINLY AMERICAN if you are sick, you do not feel well" and then spoils everything by giving an example which is simply general English: "*He stayed at home caring for his sick wife*". It then does, however, provide a proper American English example and makes the right noises in the coloured inset at the end. CALD under "physically or mentally ill; not well or healthy" makes no remark about syntax or American English. In its USAGE note the difference between the two Englishes is signalled by means of examples: "*…. he was ill. (British English)*" versus "*… he was sick. (American English)*". LDOCE4 points out that *sick* is "*especially American English*", which is not true for its last two examples *a sick child. a sick animal*, and it says nothing explicitly about the differences between the two Englishes. Finally, COBUILD4 provides no information at all about this problem.

CONCLUSION
Again, the treatment of *sick* 'ill' is very disappointing. In so far as the difference between the main varieties of English is pointed out, it is done in a very vague manner and then not infrequently contradicted by some of the examples.

PROPOSAL
Though reality may be a little less clear-cut, it would be wise in *learner's* dictionaries to give clear information about the differences between ***ill*** and ***sick***.

> SEMANTIC SYNOPSIS
> ***sick*** (*adj., mainly attributive in BrE*) unwell, suffering from a disease – usually used only before the noun it modifies, though this does not apply to AmE. Hence ***sick*** in *he is sick* in BrE means 'wanting to vomit' whereas in AmE it means either 'unwell, suffering from a disease' or 'wanting to vomit'. See also at **ill**.
> THE ENTRY PROPER
> a range of examples, both British and American.

### 2.9 *number*

In languages like Dutch and German the corresponding words (Du. *aantal*, Germ. *Anzahl*) behave differently with respect to whether they should be followed by a singular or plural verb. Thus, in Dutch, the corresponding equivalent of *A number of (= some) problems have arisen* (OALD6) would usually have a singular verb, due to the fact that Dutch singular *aantal* is considered the syntactic and semantic *head* of the construction, whereas this is not the case in English.[4] I will here focus on the extent this problem is dealt with in the learner's dictionaries.

SUMMARIES
OALD6 gives examples in which the difference between two different uses of *number of* is shown, but does not really explain the difference: "*A large number of people have applied for the job. The number of homeless people has increased dramatically.*" In the first example *number* is not the head, in the second it is the head of the noun group. The HELP note "A plural verb is needed after *a / an (large, small, etc.) number of…*" is of course exactly what we need, but an added explanatory remark would have been very helpful for the foreign user. Pointing out the contrast with *the number (of...) + singular verb* might have been useful. MEDAL, under 5a, says "used for giving a general idea of how many people or things you are referring to: *A number of students (=more than a few, but not many) have part-time jobs*." There is no further comment. CALD, without pointing this out, gives an example illustrating the fact that the plural verb tends to occur *following* but not preceding: *There has been **an** increasing number **of** cases of the disease*", where *number* is clearly treated as the head of the construction. Interestingly it adds: "SLIGHTLY FORMAL ***A** large number **of** invitations has been sent*", though it fails to point out exactly *what* is formal (presumably the singular verb *has* here). LDOCE4, just like COBUILD4, does not reveal any awareness of the problem at all.

CONCLUSION
Here the dictionary makers do not seem to have realised sufficiently, or not at all, that *number* poses a problem for many *foreign* learners. OALD6's remark after the HELP note is very useful, though it fails to explain the rationale behind this, and hence gives the foreign learner no instruction how to avoid errors here.

PROPOSAL
The entry should point out that there is a difference between ***number*** meaning 'a definite figure', after which a singular verb is used, and ***number*** meaning vaguely 'some, several, etc.) in *a (large/small/significant, etc.) number of...* (before countable nouns).

SYNOPSIS
***number*** means either 'definite figure, when counting', esp. in ***the** number of...* (***the** number of people saved so far **is*** [singular!] *twelve*), or vaguely 'some, several, etc.) in ***a** (large/small/significant, etc.) number of...* (before countable nouns): ***a large number of** people **have*** [plural!] *been saved.*
THE ENTRY PROPER
with some more clear examples showing the difference.

## **2.10** *which & that*

I will here look at the *relative* pronouns and their treatment. One of the persistent topics in grammars for learners of English is the difference between *which* and *that* (I shall leave rel. *who* undiscussed here). The problem of course is how to present the difference between the two types of relative clauses: the restrictive (alternatively called identifying or defining) and the amplifying (also non-identifying or non-restrictive).

SUMMARIES (*which*)
OALD6 introduces the restrictive use as "used to be exact about the thing or things that you mean: *Houses which overlook the lake cost more*" and adds "*That* can be used instead of *which* in this meaning, but it is not used immediately after a preposition: *It was a crisis that she was totally unprepared for*." The amplifying use is defined as "used to give more information about sth: *His best movie, which won several awards, was about the life of Gandhi… That* cannot be used instead of *which* in this meaning." MEDAL defines the restrictive use first: "NOTE used for introducing information that shows what specific thing or type of thing you are talking about. It is more usual and considered better to use 'that' to introduce this type of relative clause: *I've been thinking about those questions which you asked me last week*." The amplifying use is described as "used for adding information or a comment about a particular thing or fact when it is already clear what that thing or fact is". CALD's definition of the restrictive use goes like "(USED TO REFER) *pronoun* used as the subject or object of a verb to show what thing or things you are referring to, or to add information about the thing just mentioned. It is usually used for things, not people". The amplifying use is "(ADDS INFORMATION) *pronoun* used to add extra information to a previous clause, in writing usually after a comma". LDOCE4 gives us "2 used after a noun to show what thing or things you mean *Did you see the letter which came today*?" for the restrictive use and adds a little further down "HINT sense 2 In informal and spoken English, it is more usual to use *that*". The amplifying use is defined as "used, after a comma in writing, to add more information about the thing, situation, or event you have just mentioned".

COBUILD4 rather unfortunately manages to combine both uses in one definition: "You use *which* at the beginning of a relative clause when specifying the thing that you are talking about or when giving more information about it". It is, however, the only dictionary which (under its number 4) explicitly draws attention to the fact that the antecedent can be a *sentence*. LDOCE4 did this rather implicitly by referring to this kind of antecedent as 'event'.

CONCLUSION
Under *which* all five dictionaries avoid using specific grammatical terminology, which leads to rather unfocused definitions needing the backing of examples to know what all this is about. This is a far cry from the unashamed use of terminology in e.g. the *Encarta Dictionary's* "used to introduce a *relative clause* that provides necessary information about its *antecedent*" [italics mine] (about restrictive *which*). OALD6's treatment of *which* differs from that of *that*, since in the latter case terminology is indeed used: "used as a relative pronoun to introduce a part of a sentence which refers to the person, thing or time you have been talking about:..." It continues with the remark "In spoken and informal written English *that* is nearly always left out when it is the object of the verb or is used with a preposition", where apparently also some mild terminology was allowed. MEDAL, too, is suddenly not so shy of real terminology anymore at *that*: "as a relative pronoun (forming the subject, object, or complement of a relative clause)". Further on in its description it is not afraid of using words like *CLAUSE* and *SUPERLATIVE* either. At *that*, CALD, too, allows itself to use words like *object* and *subject*. Likewise, LDOCE4 is prepared to use difficult terminology: "used after a noun as a *relative pronoun* like 'who', 'whom', or 'which' to introduce a *clause*" [italics mine]. COBUILD4 also uses 'clause' in its definition. Clearly, in all dictionaries the treatment of *that* differs widely from that of *which*, no doubt because the entries were written by different editors.

PROPOSAL
In essence, it will be possible to use a minimum of terminology as was done in the case of relative ***which***. The definitions above were a little vague and should therefore be tightened up. If the dictionary's policy allows this, more terminology can be used.

> SEMANTIC SYNOPSIS
> ***which*** (*rel pron.*) (cannot refer to persons) used in a clause [or: relative clause] saying something about a preceding noun or pronoun [or: the antecedent]. When the clause is used to identify the preceding (pro)noun [the antecedent] *that* may also be used, but *that* is not used immediately after a preposition. When the clause is used to give additional information about the (pro)noun [the antecedent], *which* is preceded by a comma in

writing and *that* cannot be used. See also ***who*** (*rel. pronoun*) and ***that*** (*rel. pronoun*).
THE ENTRY PROPER
In the entry a number of clear examples is then given, illustrating the assertions in the synopsis.

## 3 General conclusions and some remaining questions

What we have seen so far seems to warrant the conclusion that learner's dictionaries in general have serious problems in incorporating grammatical information that goes beyond simple things like verb patterns and information like (un)countability, highly useful though these things may be. Moreover, it has emerged that approaches even in one and the same dictionary can be quite diverse, in that on one occasion grammatical terminology is seemingly avoided at all costs whereas elsewhere for related cases an apparently different editor thinks nothing of using a number of quite specific grammatical terms. It is also my suspicion that the compilers of the dictionaries looked at here were not all of them equally conversant with formal English grammar. Put more cautiously, there is plenty of evidence that the needs of the *foreign* learners were not always equally well recognised. This, I admit, is difficult, since these dictionaries are used all over the world by users from greatly varying linguistic backgrounds. Yet, there might have been a greater awareness of the fact that whenever English grammar was apparently tricky to describe, the foreign learner would be even more out of his depth in such cases.

It seems to me the following issues should be addressed in future editions of English learner's dictionaries. In the first place a clear policy should be formulated as to how much specific terminology is allowed. There is evidence of a certain degree of uncertainty here and hence inconsistency. Secondly, and related to this: since it is obviously impossible for a learner's dictionary to be a complete grammar of English, policy makers should decide how much grammar should go into the compiling of this type of dictionary. And following from this, how this information – if at all – should be presented: using a minimum of terminology with a maximum of definitional clarity and clear examples, or using more terminology while still at the same time, and independently, striving for clear definitions and excellent examples. A final nagging question to be answered is whether all this would be worth the effort, seeing that all this often concerns words that (I suspect) are not often consulted by users, probably because they expect to find no clear grammatical information there, or more likely think the required information had better be looked up in a grammar book. In other words: should we prompt the users more explicitly to go to the dictionary for this kind of grammatical information or restrict ourselves and leave this kind of information to the reference grammars?

## Notes

1. Including cases of implied condition as in *Students who have* ***any*** *complaints should raise their hands* [*If students have* ***any****...*], cf. CGEL 784.

2. In this way we avoid rather irrelevant definitions like OALD6: "Used before countable or singular nouns referring to people or things that have not already been mentioned". This, in fact, is a remark about one of the possible *uses* of *a*, and is not a semantic definition.

3. Admittedly, the entire phrase behaves syntactically as a plural (*the poor* ***are****...*), but the adjective itself remains adjectival (cf. *the* ***relatively*** *poor are....*), since it can be modified by an adverb. Compare this with an example like *the reds* (= communists etc.), where *reds* has a proper plural *–s*.

4. Or, perhaps more in keeping with the truth: syntactically *number* is perhaps formally the head, but the semantics of the entire noun phrase is so strongly plural that the verb normally follows suit, except occasionally when the verb *precedes* the noun phrase, as in questions etc.

## References

*Dictionaries*:

CALD = *Cambridge Advanced Learner's Dictionary*, ed. by Patrick Gillard, Kate Woodford & Guy Jackson. Cambridge: Cambridge University Press. 2003.

COBUILD4 = *Collins Cobuild Advanced Learner's English Dictionary*. Fourth Edition, ed. by Michela Clari. Glasgow: HarperCollins. 2003.

LDOCE4 = *Longman Dictionary of Contemporary English.* Fourth Edition, ed. by Della Summers. Harlow: Longman. 2003.

MEDAL = *MacMillan English Dictionary for Advanced Learners*, ed. by Michael Rundell. Oxford: MacMillan Education Ltd. 2002.

OALD6 = *Oxford Advanced Learner's Dictionary of Current English*. Sixth Edition, ed. by Sally Wehmeier. Oxford: Oxford University Press. 2000.

*Other works*:

CGEL = Quirk, Randolph; Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1991. *A Comprehensive Grammar of the English Language* [ninth revised impression]. London and New York: Longman.

Meer, Geart van der. 1997. "Grammar, construction information, and collocations in two active bilingual dictionaries. A comparison and assessment of Wolters' Nederlands-Engels and Van Dale Handwoordenboek Nederlands-Engels". *Leuvense Bijdragen***,** vol. 86, 109–123.

Meer, Geart van der. 2002. "Dictionary entry and access. Trying to see trees and woods". *Proceedings of the tenth EURALEX international congress, Copenhagen, Denmark, August 13–17, 2002.* Anna Braasch and Claus Povlsen (eds.), 509–520.

Swan, Michael, 1995: *Practical English Usage* [second ed.]. Oxford: Oxford University Press.

CHAPTER 13

# Neuere Aspekte einer Theorie und Typologie von Wörterbuchartikeln und ihre Praxisrelevanz

Herbert Ernst Wiegand

The point of departure is the assumption that although many synonyms exist for the term *dictionary article,* neither general-language dictionaries nor technical dictionaries dealing with linguistics include it as a lemma, its definitions in technical dictionaries of lexicography are not appropriate, and in German it has no genus proximum. Following criticism of earlier definitions of *dictionary article,* a theory-based exposition of the terminological elements for a new definition are presented. In this regard the terms *accessible entry, data accessibility* and a few others are introduced. A proposal for a new definition for *dictionary article* is followed by a sample of a typology for lexicographically accessible entries as well as two typologies for dictionary articles from a related typological system. In conclusion, the practical relevance of the presented sample with regard to a theory of dictionary structures is discussed.

Es wird von der Beobachtung ausgegangen, dass es zum Terminus *Wörterbuchartikel* zwar zahlreiche Synonyme gibt, dass er aber weder in allgemeinen einsprachigen Wörterbüchern noch in linguistischen Fachwörterbüchern als Lemma angesetzt ist, dass seine Definitionen in den Wörterbüchern zur Lexikographie nicht angemessen sind und dass es im Deutschen kein genus proximum gibt. Nach einer Kritik früherer Definitionen von *Wörterbuchartikel* werden die terminologischen Elemente für eine neues Definiens theoriebezogen erarbeitet; in diesem Zusammenhang werden u. a. die Termini *akzessiver Eintrag*, *Datenakzessivität* und einige weitere eingeführt. Nachdem eine Definition für *Wörterbuchartikel* vorgeschlagen ist, wird ein Ausschnitt aus einer Typologie für lexikographische akzessive Einträge vorgestellt sowie zwei Typologien für Wörterbuchartikel aus einem zugehörigen typologischen System. Abschließend wird kurz auf die Praxisrelevanz des dargelegten kleinen Ausschnittes aus einer Theorie der Wörterbuchform eingegangen.

## 1 Vorbemerkung zum Problemhorizont

An einem Freitagabend im Januar 2003 fasste ich den Entschluss, das voraussichtlich regnerische Wochenende damit zu verbringen, für das „Wörterbuch zur Lexikographie und Wörterbuchforschung" den Synopseartikel zum Lemmazeichen *Wörterbuchartikel* zu schreiben. Am späten Abend des Sonntags hatte ich zwar keinen Wörterbuchartikel fertig, doch wusste ich definitiv immerhin Folgendes:

a. Das deutsche Kompositum *Wörterbuchartikel* ist in keinem der großen und größeren historischen und allgemeinen einsprachigen Wörterbücher des Deutschen primär gebucht.
b. In den vier derzeit existierenden Wörterbüchern zur Lexikographie und in drei überprüften Glossaren zur Lexikographie sind die Artikel zum Lemmazeichen *Wörterbuchartikel* bzw. zu seinen englischen, dänischen und spanischen Äquivalenten aus unterschiedlichen Gründen nicht ausreichend.
c. In den wichtigsten deutschen linguistischen Wörterbüchern ist das Kompositum *Wörterbuchartikel* nicht als Lemma angesetzt.
d. In der lexikographischen Werkstattsprache begegnen zahlreiche Quasisynonyme zum Kompositum *Wörterbuchartikel*, wie z.B. *Wortstelle*, *Wortartikel*, *Stichwortartikel* und *Eintrag*, wobei es dann zu den merkwürdigsten Stilblüten kommt, heißt doch beispielsweise in diesem Werkstattjargon ein Artikel zum Buchstaben *A* Wortartikel.
e. In der europäischen Metalexikographie nach 1945 wird der Wörterbuchartikel als die wichtigste lexikographische Textsorte in der Theoriebildung bis Anfang der 80er Jahre nicht wahrgenommen.
f. Ab Mitte der 80er Jahre treten die ersten Definitionen auf, die – relativ zum heutigen Wissen – allesamt mehr oder weniger große Defekte aufweisen, wie ich gleich an meiner eigenen Definition aus dem Jahre 1984 zeigen werde.
g. Es gibt keine ausgearbeitete Typologie von Wörterbuchartikeln und entsprechend keine theoriebezogenen Typbezeichnungen, sieht man von wenigen Ausnahmen, wie z.B. den Termini *Einzelartikel*, *Synopseartikel*, *Verweisartikel* und wenigen anderen ab.
h. Es fehlt ein deutsches terminologisches Hyperonym zu *Wörterbuchartikel,* das gleichzeitig hyperonym ist zu den entsprechenden Bezeichnungen im nichtlexikographischen Bereich, wie beispielsweise *Telephonbucheintrag*. Im Englischen lautet dieses Hyperonym *basic reference unit*.

Was soeben unter den Punkten a. bis h. nur behauptet wurde, ist in Wiegand (2004: 169–313) belegt und im Einzelnen diskutiert.

Im Folgenden betrachten wir, nur um das Problembewusstsein ein wenig zu wecken, meine folgende Definition aus Wiegand (1984: 566):

Ein *Wörterbuchartikel* ist ein lexikographischer Text, der als eine geordnete Menge von Textbausteinen und/oder Textelementen als lexikographischen Daten aufgefaßt werden kann, in welcher mindestens ein Lemma obligatorisch ist und für alle Textsegmente – soweit sie nicht zu den nichtkommentierenden lexikographischen Symbolen gehören – gilt, dass mit ihnen wenigstens eine Eigenschaft des Lemmazeichens beschrieben oder vorgeführt wird.

Zunächst ein kurzes Wort zu Definitionen. Definitionen sind Textausschnitte; als solche sind sie ko- und kontext- und damit theorieabhängig und nicht interessenfrei. Es gehört zu dem Minimalkonsens der Wissenschaftstheoretiker über Definitionen, dass sie beim Aufbau einer Wissenschaftssprache unverzichtbar sind. Sie dienen der Fixierung des Gebrauchs der Termini und haben damit eine gegenstandskonstitutive Funktion. Im Zuge des Ausbaus und der Weiterentwicklung einer Theorie müssen sie überprüft und gegebenenfalls an den neuen Stand angepasst werden.

Im Folgenden wird meine Definition aus Wiegand (1984: 566) überprüft, und zwar zunächst – aus Demonstrationszwecken – ohne Berücksichtigung ihres Ko- und Kontextes. Entsprechend ergibt sich folgende Argumentation: Nach dem Definiendum „ein Wörterbuchartikel" bezieht sich die Definition auf jede Art von Wörterbuchartikel.

Erster Kritikpunkt: Betrachten wir die folgenden beiden Artikel:

$wa_1$: **Sinalco**
Sinalco ist eine Limonade, die man Sinalco genannt hat, weil sie keinen Alkohol enthält: Sinalco steht für lateinisch **sin**e (»ohne«) **Alco**hol. (Lödige 2001: 280)

$wa_2$: **Rechtsnormen** setzen Recht. Sie beinhalten allgemein verbindliche Verhaltensanordnungen und zwar aufgrund der Verfassung (→ Grundgesetz) oder eines Gesetzes eines zur Rechtsetzung autorisierten Normgebers (→ Gesetzgebung). R. sind → Gesetz, → Rechtsverordnung, → Satzung. *Karin Wolski* (FsA 2002: 764)

Zur Erläuterung muss zunächst Folgendes gesagt werden: Mit *Textbausteinen* in der Definition sind nichtelementare Angaben und mit *Textelementen* elementare Angaben gemeint. In $wa_1$ und $wa_2$ treten aber gar keine Angaben auf, sondern nur Angabetexte. Entsprechend fallen die beiden Wörterbuchartikel nicht unter die Definition.

Zweiter Kritikpunkt: Weiterhin stimmt es nicht, dass in jedem Wörterbuchartikel mindestens ein Lemma obligatorisch ist. Ein Lemma ist ein Leitelementträger und gewährleistet einen direkten externen Zugriff auf die Daten des zugehörigen

Wörterbuchartikels. Neben den Leitelementträgern gibt es jedoch noch einen anderen Typ von Zugriffstextelement, nämlich den Verweisadressenträger, worauf ich noch zurückkommen werde. Wörterbuchartikel, deren Zugriffstextelement ein Verweisadressenträger ist, finden sich beispielsweise in Wörterbüchern, die eine randomisierte Makrostruktur aufweisen, wie z.B. Prosinger (1984). Einer solchen Makrostruktur ist keine äußere Zugriffsstruktur inhärent. Vielmehr ist ihr nur eine äußere mediostrukturelle Zugriffsstruktur in Form eines alphabetischen Registers zugeordnet. Entsprechend weisen die Wörterbuchartikel Zugriffstextelemente auf, auf die nur extern indirekt, nämlich mithilfe der Verweisadressen zugriffen werden kann, die im Register gefunden werden. Wörterbücher – und hierzu gehören auch die Sachgruppenwörterbücher –, die als Zugriffstextelemente Verweisadressenträger aufweisen, fallen also auch nicht unter die Definition.

Dritter Kritikpunkt: In der Definition steht, dass mit den Textbausteinen und Textelementen und damit mit den Angaben „wenigstens eine Eigenschaft des Lemmazeichens beschrieben oder vorgeführt wird".

Betrachten wir nun die Wörterbuchartikel $wa_3$ und $wa_4$:

$wa_3$: **Dämpfer,** der; *-s, - Vorrichtung zur Verminderung der Intensität der Töne, zur Erzeugung einer gedämpften Klangfarbe der Töne bestimmter Musikinstrumente*
+ umg. jmd. bekommt einen D. (*jmd. wird in seine Schranken gewiesen; jmd. erfährt eine Enttäuschung*) (HWDG: 225)

$wa_4$: **schwarz,** schwärzer, schwärzeste */Adj./* **1.** *von dunkelster Färbung und die Lichtstrahlen völlig absorbierend*, *nicht reflektierend*: s. wie Kohle; ein schwarzer Anzug; schwarze Schuhe; er hat schwarzes Haar; ein Brief mit schwarzem Rand (*Trauerbrief*); etw. s. anstreichen; in Schwarz gehen (*schwarze Kleidung, meist zum Zeichen der Trauer, tragen*); die Farbe Schwarz, ein glänzendes, mattes, tiefes Schwarz – **2.** *sehr dunkel, fast wie* 1: schwarze Augen; sie Schwarzen */oft als diskriminierend empfundene, gebrauchte Bez. für Afrikaner und Afroamerikaner/*; schwarzer Pfeffer, Tee; schwarzer Kaffee (*Kaffee ohne Milch und Sahne*); schwarze (*dunkelrote*) Kirschen; schwarze Wolken; s. wie die Nacht – **3.** umg. *dunkel von Schmutz*: er hat schwarze Hände, Fingernägel; das Hemd war am Kragen ganz s. – **4.** */nur attr./* von schwarzen (*unheilvollen*) Ahnungen geplagt werden – **5.** */adv./* etw. s. (*unter Umgehung der rechtlichen Bestimmungen*) kaufen, verkaufen (HWGD: 1030)

Der Artikel $wa_3$ ist ein rechtserweiterter Wörterbuchartikel. Im Postkommentar steht die verdichtete Phrasemangabe „jmd. bekommt einen D.". An diese Angabe sind die beiden auf sie folgenden Bedeutungsparaphrasenangaben linksadressiert. Anhand dieser beiden Angaben erfährt man aber nichts über eine Eigenschaft des Lemmazeichens. Entsprechend fällt $wa_3$ nicht unter die Definition.

In dem gekürzten Artikel $wa_4$ finden sich mehrere Glossate, z.B. „*Trauerbrief*" als Postglossat zur Kompetenzbeispielangabe „ein Brief mit schwarzem Rand". Glossate sind niemals an die Lemmazeichengestaltangabe adressiert, sondern stets artikelintern an andere Angaben. Daher gilt, dass alle Wörterbuchartikel, bei deren textueller Gestaltung von der Methode der Angabenglossierung Gebrauch gemacht wurde, nicht unter die Definition fallen.

Vierter Kritikpunkt*:* Der in der Definition verwendete Ausdruck *lexikographischer Text* ist zwar ein hyperonymer Ausdruck zu *Wörterbuchartikel,* aber kein geeignetes *genus proximum*.

Ich breche die Kritik meiner Definition hier ab. Um einen Abbruch handelt es sich deswegen, weil ich noch weitere Kritikpunkte vortragen könnte, dies aber aus Zeitgründen unterlassen muss.

Die vorstehende Kritik ist – und das hatte ich ja ausdrücklich gesagt – aus Demonstrationsgründen ohne Berücksichtigung des Kotextes der Definition erfolgt. Berücksichtigt man diesen, dann gelten die beiden ersten Kritikpunkte nicht, denn die Definition ist *expressis verbis* auf hochverdichtete und standardisierte Wörterbuchartikel in alphabetischen Printwörterbüchern bezogen.

Nach meinem heutigen Kenntnisstrand gilt, dass ich in der zitierten Definition nicht *den* Wörterbuchartikel definiert habe, sondern einen ganz spezifischen Typ von Wörterbuchartikel in Printwörterbüchern, den man wie folgt mit einer Nominalphrase charakterisieren kann: *Kondensierter und standardisierter Einzelartikel mit vollständig lemmatischer Adressierung*. Die genannte Nominalphrase gilt nicht als Terminus für einen Artikeltyp, sondern sie macht Gebrauch von folgenden Termini: *kondensierter Wörterbuchartikel, standardisierter Wörterbuchartikel, Einzelartikel* und *Artikel mit vollständig lemmatischer Adressierung.*

Jeder der vier genannten Artikeltypen hat einen typologischen Ort in einer anderen Typologie von Wörterbuchartikeln, die alle zu einem von mir erarbeiteten typologischen System von Artikeltypologien gehören. Zwei Wörterbuchartikel ($wa_5$ und $wa_6$), auf welche man die Nominalphrase *kondensierter und standardisierter Einzelartikel mit vollständig lemmatischer Adressierung* zutrifft, werden nachfolgend wiedergegeben.

$wa_5$: **beatmen**, beatmete, hat beatmet Neupräg. Med.
jmd.; jmds. Lunge b. *jmds. Lunge (mit besonderen Apparaten) durch äußere mechanische Einwirkung entweder rhythmisch zusammenpressen oder rhythmisch mit Sauerstoff aufblähen*: einen Kranken während der

Operation b.; e. Ertrunkenen, Verwundeten b.; eine Frühgeburt mit Sauerstoff b.; der Patient wurde künstlich b.; eine gut, schlecht beatmete Lunge (WDG: 442)

$wa_6$: **in|sti|tu|ti|o|na|li|sie|ren** ‹sw.V.; hat› (bildungsspr.): **a)** *in eine [gesellschaftlich anerkannte] feste Form bringen; zu einer Institution (2) machen*: die beiden Staaten wollen ihre Zusammenarbeit [noch stärker] i.; **b)** ‹i. + sich› *eine [gesellschaftlich anerkannte] feste Form annehmen; zu einer Institution (2) werden*: der Widerstand begann sich zu i.
(Duden-[4]DUW: 837)

Aus den vorgetragenen Kritikpunkten müssen folgende Lehren gezogen werden:

1. Eine Definition von *Wörterbuchartikel*, die möglichst allgemeingültig ist und als geeigneter Ausgangspunkt für ein typologisches System gelten kann, muss sich auf Beobachtungen und Untersuchungen stützen, die anhand einer erheblich größeren empirischen Basis durchgeführt wurden. Entsprechend wurden 1200 Wörterbücher untersucht, wobei darauf geachtet wurde, dass möglichst viele Wörterbuchtypen vertreten sind.
2. Es muss eine geeignete oberbegriffliche Vernetzung des Terminus *Wörterbuchartikel* zusammen mit einer handhabbaren Terminologie gefunden werden.

## 2 Auf dem Weg zu einer generellen Definition von Wörterbuchartikel

Im Folgenden stelle ich zunächst in geraffter Form die oberbegriffliche Vernetzung des Terminus *Wörterbuchartikel* dar, so dass wir auch über ein geeignetes *genus proximum* verfügen, das in einer neuen generellen Definition von *Wörterbuchartikel* Verwendung finden kann, die am Schluss dieses Abschnittes vorgestellt wird.

In der deutschen standardsprachlichen Alltagssprache verfügen wir über das Kompositum *Nachschlagewerk*. Die Menge der Bezugsgegenstände, auf die man in usuellen Texten mit diesem Kompositum Bezug nehmen kann, hat relativ viele Elemente; auf der Klassenebene sind dies z.B.: das Sprachwörterbuch, das Fachwörterbuch, das Allwörterbuch sowie alle zugehörigen Unterklassen, weiterhin gehören im nichtlexikographischen Bereich dazu: Das Faxbuch, das Telephonbuch, der Warenkatalog, die Bibliographie, der Fahrplan und zahlreiche weitere Nachschlagebücher. Nach dem alltäglichen Bewusstsein alphabetisierter Laien ist das Gemeinsame aller genannten Nachschlagewerke, dass man in ihnen etwas nachschlagen kann. Das Nachschlagen gehört zu den weit verbreiteten Kulturtechniken. Wer es kann, kann es zwar anderen zeigen, damit sie es auch können. Eine theoriebasierte allgemeingültige Erklärung des Nachschlagens gelingt allerdings keinem Laien, auch wenn er es noch so gut beherrscht. Selbst Lexikographen und Wörterbuchforschern macht eine generelle Erklärung Schwierigkeiten.

Im Folgenden gehen wir der Frage nach: Was ist allen gedruckten Nachschlagewerken gemeinsam, und zwar unabhängig davon, in welcher Sprache sie geschrieben sind und zu welchem Sachbereich sie gehören? Die Antwort lautet: Es ist die Datenakzessivität, also die Zugriffsbereitschaft der nachschlagbaren Daten; ich werde im weiteren kurz von der *Akzessivität* sprechen. Was darunter zu verstehen ist, lässt sich im Rahmen der Theorie der Ordnungsstrukturen strikt formal erklären, sei hier aber lediglich informell und auf einfache Weise, aber dennoch theoriebasiert wie folgt erläutert: Jedes Nachschlagewerk weist eine äußere Zugriffsstruktur auf. Jede äußere Zugriffsstruktur, zu welchem Typ auch immer sie gehört, ist eine lineare Ordnungsstruktur, in der alle Elemente anhand der Ordnungsgesetzlichkeiten einer anderen, nachschlagewerkexternen Ordnungsstruktur geordnet sind, wobei geringfügige nachschlagewerkinterne Modifikationen dieser Ordnungsstruktur zugelassen sind. Beispielsweise kann das lateinische Alphabet als die nachschlagewerkexterne Ordnungsstruktur um die alphabetexternen Buchstaben des Neuhochdeutschen zu einen wörterbuchspezifischen Zugriffsalphabet nachschlagewerkintern erweitert werden. Die Akzessivität der Daten für einen potentiellen Benutzer besteht nun gerade darin, dass dieser in der Lage ist, aufgrund der Kenntnis der nachschlagewerkexternen Ordnungsstruktur bzw. gegebenenfalls aufgrund der zusätzlichen Kenntnis ihrer nachschlagewerkinternen Modifikationen, auf jedes Element der äußeren Zugriffsstruktur *direkt extern zuzugreifen*, so dass er mithin jeden Leitelementträger – z.B. jedes Lemma in einem Sprachwörterbuch – gezielt suchen und gegebenenfalls finden kann. Ein Leitelementträger kann selbst (muss aber nicht) eine zugriffsbereite Dateneinheit sein, und ihm sind meistens weitere zugriffsbereite Daten zugeordnet. Letzteres ist nur dann nicht zwingend erforderlich, wenn er selbst Angabeeigenschaften aufweist. Entweder der Leitelementträger alleine – wie beispielsweise in rückläufigen Wörterbüchern – oder dieser zusammen mit denen ihm zugeordneten Daten bildet den so zu nennenden *akzessiven Eintrag*, der in jedem Nachschlagewerk – sei es nun ein lexikographisches oder ein nichtlexikographisches – in unterschiedlichen Ausprägungen auftritt. Als deutsches terminologisches Äquivalent für englisch *basic reference unit* schlage ich also den Mehrwortterminus *akzessiver Eintrag* vor, der sich in viele Sprachen leicht übersetzen lässt.

Im Folgenden erläutere ich einen kleineren Ausschnitt aus einer Typologie akzessiver Einträge in Nachschlagewerken. Dort finden wir dann als einen Untertyp des akzessiven Eintrages auch den Wörterbuchartikel.

Lexikographische und nichtlexikographische Texte vom Typ des akzessiven Eintrages sind die wichtigsten Teiltexte von Nachschlagewerken. Ein gedrucktes Werk ist nur dann ein Nachschlagewerk, wenn es *erstens* mindestens eine äußere

Zugriffsstruktur aufweist und wenn es *zweitens* eine Menge von akzessiven Einträgen hat. Es ist daher schon erstaunlich, dass es bis heute keine ausgearbeitete Typologie akzessiver Einträge gibt. Im Folgenden stelle ich – in stark geraffter Form – einen kleineren Ausschnitt aus einem von mir erarbeiteten typologischen System akzessiver Einträge vor.

Der Ausgangstyp für einen solchen Typologieausschnitt ist naturgemäß der Typ des akzessiven Eintrages. Als erstes Typologiekriterium (TK) zu seiner Unterteilung wählen wir *Klassen der Typen von Nachschlagewerken* und weisen diesem Kriterium gemäß überkommener kultureller Einteilungstraditionen eine zweifache Ausprägungsstruktur zu, so dass zwischen der Klasse der Typen lexikographischer und der Klasse der Typen nichtlexikographischer Nachschlagewerke unterschieden wird. Entsprechend ergeben sich folgende beiden Untertypen unseres Ausgangstyps:

- der Typ des lexikographischen akzessiven Eintrags und
- der Typ des nichtlexikographischen akzessiven Eintrags.

Auf den letztgenannten Typ gehe ich im Folgenden nicht ein. Für den erstgenannten Typ wird das terminologische Synonym *akzessiver Wörterbucheintrag* eingeführt. Für die Subtypenbildung zum Typ des akzessiven Wörterbucheintrages sind vor allem folgende Sachverhalte relevant: Ein Wörterbuch kann mehrere äußere Zugriffsstrukturen aufweisen, die zu verschiedenen Typen von Zugriffsstrukturen gehören. Zu berücksichtigen sind hier, neben der zentralen makrostrukturellen Zugriffsstruktur, die externen Zugriff auf die Wörterbuchartikel eröffnet und neben den äußeren Binnentextzugriffsstrukturen, die ich hier der Einfachheit halber nicht berücksichtige, vor allem die äußeren Umtextzugriffsstrukturen; diese eröffnen externen Zugriff auf die akzessiven Umtexteinträge. Entsprechend führt eine Subtypologisierung des Typs des akzessiven Wörterbucheintrages zu folgenden zwei kriterien- und stufengleichen Subtypen:

- dem Typ des Wörterbuchartikels,
- dem Typ des akzessiven Umtexteintrages.

Es ergibt sich damit folgender kleiner Ausschnitt aus einer Typologie akzessiver Einträge.

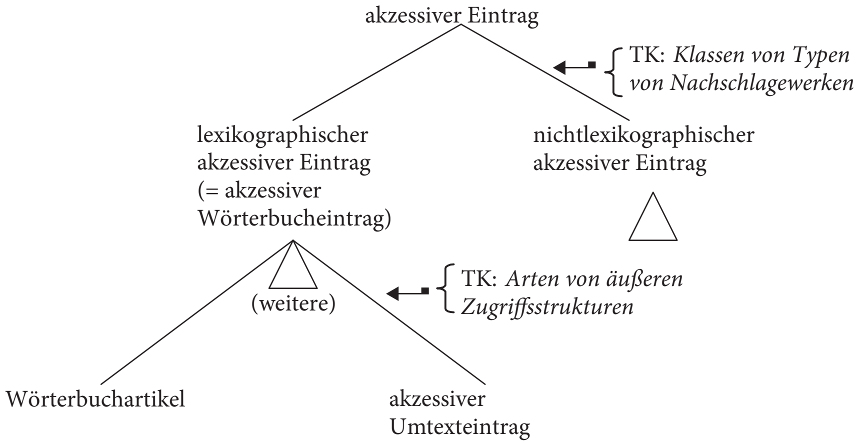


**Abb. 1.** Kommentierter Typologiegraph zum ersten Ausschnitt aus einer Typologie akzessiver Einträge; „◄━■" bedeutet soviel wie: „die Anwendung des TK führt zu der Unterteilung"; vgl. Abb. 2

Zu diesem Typologieausschnitt gehören die folgenden beiden Definitionen von *akzessiver Eintrag* und *akzessiver Wörterbucheintrag*:

Ein *akzessiver Eintrag* ist die textuelle Basiseinheit eines gedruckten Nachschlagewerkes und durch folgende drei Charakteristika eindeutig gekennzeichnet:

a. Er weist mindestens ein Zugriffstextelement auf.
b. Er ist entweder eine Konstituente des zentralen Nachschlagetextes oder eine Konstituente eines Umtextes oder Binnentextes mit äußerer Zugriffsstruktur.
c. Er enthält zugriffsbereite Daten, darunter ein Teil solcher Daten, anhand derer ein Nachschlagewerkbenutzer Informationen zu dem Nachschlagegegenstand erschließen kann, der entweder mit dem Zugriffstextelement oder aber mit dem Binnentexttitel oder dem Umtexttitel bzw. den Umtextzwischentiteln genannt wird.

Ein akzessiver Wörterbucheintrag ist die textuelle lexikographische Basiseinheit eines gedruckten Wörterbuches und durch folgende Charakteristika gekennzeichnet:

a. Er weist zumindest ein Zugriffstextelement auf.
b. Er ist entweder eine Konstituente des oder eines von mehreren Wörterverzeichnissen oder eine Konstituente eines lexikographischen Umtextes oder Binnentextes mit äußerer Zugriffsstruktur.
c. Er enthält zugriffsbereite Daten, darunter ein Teil solcher Daten, anhand derer ein Wörterbuchbenutzer lexikographische Informationen zu dem Nachschlagegegenstand erschließen kann, der entweder mit dem Zugriffstextelement oder aber

mit dem Binnentexttitel oder dem Umtexttitel bzw. den Umtextzwischentitel genannt wird.

Man könnte nun eine hinreichend generelle Definition von *Wörterbuchartikel* gleich anschließen. Dies würde aber dazu führen, dass von bestimmten Teilen der Definition nicht hinreichend deutlich ersichtlich wäre, warum sie gerade so und nicht anders formuliert wurden. Auch sei daran erinnert, dass allgemein gilt: Eine Definition des Ausgangstyps eines typologischen Systems, die am Anfang des Aufbaus eines solchen Systems steht, ist zwar notwendig. Es handelt sich bei dieser aber zunächst um eine so genannte Arbeitsdefinition, die bei der Arbeit modifiziert werden kann. Im Folgenden soll nur an einem Beispiel gezeigt werden, warum man erst bestimmte Untertypen von Wörterbuchartikeln kennen muss, wenn man die Definition des Ausgangstyps angemessen generell gestalten möchte. Zugleich soll dadurch deutlich werden, warum auch in der vorzuschlagenden Definition von *Wörterbuchartikel*, wie in den Definitionen von *akzessiver Eintrag* und *akzessiver Wörterbucheintrag*, nicht der Terminus *Leitelementträger*, sondern ebenfalls der Terminus *Zugriffstextelement* Verwendung findet, der zu den Termini *Leitelementträger* und *Verweisadressenträger* hyperonym ist.

Für die folgenden Überlegungen benötigen wir den bereits genannten lexikographietheoretischen Begriff der Datenakzessivität. Zunächst muss man zwischen interner und externer Akzessivität unterscheiden. Die *interne Akzessivität* betrifft die Zugriffsbereitschaft der Daten innerhalb von akzessiven Einträgen; auf diese gehe ich im Folgenden nicht ein. Die *externe Akzessivität* betrifft die Zugriffsbereitschaft der akzessiven Einträge als ganze Textkonstituenten. Hier muss man zunächst die *direkte externe Akzessivität* von der *indirekten externen Akzessivität* unterscheiden. Ich charakterisiere zunächst die Erstgenannte. Ein akzessiver Eintrag ist direkt extern akzessiv, wenn sein Zugriffstextelement ein Leitelementträger (z.B. ein Lemma) ist, so dass die Kenntnis eines Leitelementes als wörterbuchextern erhältliche Fundortinformation sowie die Beherrschung der allgemeinen Benutzungspraxis für Nachschlagewerke (vgl. Wiegand 1998: 370) ausreicht, um den Leitelementträger und damit den zugehörigen akzessiven Eintrag zu finden. Akzessive Einträge mit Leitelementträger heißen entsprechend (*extern*) *direkt akzessive Einträge*, wobei das Attribut *extern* weggelassen werden kann. Ein Untertyp des direkt akzessiven Eintrages ist entsprechend der *direkt akzessive Wörterbucheintrag*. Einer der Untertypen des letztgenannten Typs ist entsprechend der *direkt akzessive Wörterbuchartikel*. Ein Beispiel für diesen Artikeltyp ist $\text{wa}_6$ aus der 4. Auflage des Duden-Universalwörterbuchs. Für das Auffinden dieses Wörterbuchartikels reicht die Kenntnis des Leitelementes, also die der Buchstabenfolge |institutionalisieren|, sowie die Kenntnis der allgemeinen Wörterbuchbenutzungspraxis aus. Die direkte Akzessivität ist in diesem Fall durch nichts eingeschränkt.

Denn *erstens* ist das Duden-Universalwörterbuch in der 4. Auflage glattalphabetisch, so dass der Benutzer keine wörterbuchspezifischen Anordnungsregelungen kennen muss. Und *zweitens* gehört zum Leitelement |institutionalisieren| kein alphabetexterner Buchstabe, so dass der Benutzer keine speziellen Regelungen für das wörterbuchspezifische Zugriffsalphabet kennen muss. Es gibt – wie aus dem Gesagten folgt – mithin wörterbuchbedingte Einschränkungen der externen direkten Akzessivität, deren Berücksichtigung zu Subtypen der externen direkten Akzessivität führt, was hier nicht näher ausgeführt wird.

Im Folgenden betrachten wir die indirekte externe Akzessivität. Der Anschaulichkeit halber beginne ich mit einem – wie ich hoffe – unterhaltsamen Beispiel. Denken Sie sich bitte eine Person, sie heiße Werner, die zu den 100 schönsten Wörtern ihrer Lebensbiographie ein Wörterbuch mit nichtkondensierten Wörterbuchartikeln geschrieben hat. Die Elemente der Makrostruktur sind dabei individuell geordnet; die ersten zehn Elemente stehen in der folgenden Reihenfolge: *Trompete*, *photographieren*, *Schwester*, *Traum*, *Flucht*, *geil*, *Gott*, *Linkskurve*, *Fackel.*

Jeder der 100 Wörterbuchartikel ist genau eine Wörterbuchseite lang. Wie Sie sehen, sind die Elemente linear geordnet, nur die Ordnungsgesetzlichkeit ist niemandem bekannt. Nicht einmal Werner kann sie angeben. Die Makrostruktur in Werners Wörterbuch ist daher nichtakzessiv, d.h. ihr ist keine äußere Zugriffsstruktur inhärent. Die 100 Elemente der Makrostruktur sind damit keine Leitelementträger; entsprechend liefert z.B. das Wort *Trompete* kein Leitelement und damit keine Fundortdaten. Kurz: auf kein Element der Makrostruktur und damit auf keinen Wörterbuchartikel kann direkt extern zugegriffen werden. Dennoch weist Werners Wörterbuch eine äußere Zugriffsstruktur auf. Schon in der Überschrift zu den Benutzungshinweisen seines Wörterbuchs weist Werner, der belesene und pfiffige Softi, mit Anspielung auf Prosingers rabenstarkes Lexikon der Scene-Sprache von 1984 ausdrücklich darauf hin; denn die Überschrift lautet wie folgt:

> Benutzungshinweise zum ersten Individualwörterbuch, das den bürgerlich-lexikographischen Zwang zur alphabetischen Reihenfolge der Wörterbuchartikel lachend in den Wind schlägt (und dafür ein schönes Register am Ende hat).

Das erwähnte Register ist ein alphabetisches Zugriffsregister und hat folgende Form:

| | |
|---|---|
| Fackel | 3 |
| Flucht | 17 |
| geil | 6 |
| Gott | 44 |
| Krone | 18 |
| Linkskurve | 99 |

| | |
|---|---|
| Schwester | 40 |
| Traum | 51 |
| Trompete | 1 |

Ein Registereintrag, wie z.B. „Fackel 3" ist zugleich eine Verweisadressenangabe, und zwar handelt es sich um eine Doppeladressenangabe; mit dieser wird die numerische Hauptadresse „3" genannt, die in der Buchpaginierungsreihe ‹1, 2, 3,..., 100› als der äußeren numerischen Zugriffsstruktur liegt, sowie weiterhin die Verweisunteradresse |Fackel|. Der Artikeltitel „Fackel", mit dem die Wörterbuchseite 3 beginnt, ist daher der Verweisadressenträger für die Verweisunteradresse, die im Register genannt ist. „Fackel" auf der Wörterbuchseite 3 ist also kein Leitelementträger (damit auch kein Lemma). Denn wer nur die Buchstabenfolge |Fackel| kennt, kann weder extern direkt noch extern indirekt auf den Artikel zu *Fackel* zugreifen; vielmehr kann er nur unsystematisch im Wörterbuch herumsuchen. Er kann aber unter Zuhilfenahme des Leitelementes |Fackel| direkt auf das alphabetische Register zugreifen; nach der erfolgreichen Ausführung einer externen Zugriffshandlung, die ihn zum Registereintrag „Fackel 3" führt, kann er sodann eine Verweisbefolgungshandlung ausführen, bei der er die im Registereintrag gefundene Verweisadresse verwendet. Die Artikel in Werners Wörterbuch sind daher nur *extern indirekt akzessiv*: Um auf sie zugreifen zu können, muss der Benutzer immer über das alphabetische Register gehen.

Nach der Erläuterung des Beispiels sollte also Folgendes klar sein: Ein Wörterbuchartikel ist extern indirekt akzessiv, wenn ein Teil der Fundortdaten (oder: der Zugriffsdaten) in einem Register stehen, so dass ein Teil der für den externen Zugriff notwendigen Informationen dort erst ermittelt werden muss.

Was hier am Beispiel von Wörterbuchartikeln gezeigt wurde, gilt für alle Typen von akzessiven Einträgen. Bei den Umtexteinträgen ist die indirekte Akzessivität sogar häufiger anzutreffen als die direkte Akzessivität, und im nichtlexikographischen Bereich spielt sie ebenfalls eine bedeutende Rolle. Beispielsweise sind die meisten Warenkataloge so aufgebaut, dass die akzessiven Warenkatalogeinträge indirekt akzessiv sind. Die alphabetischen Zugriffsregister heißen in vielen Warenkatalogen sogar *Inhaltsverzeichnis.*

Weiterhin dürfte Folgendes leicht einsehbar sein: Akzessive Einträge jeder Art können sowohl extern direkt als auch extern indirekt akzessiv sein, so dass sie polyakzessiv sind, weil ihr Zugriffstextelement sowohl als Verweisadressen- als auch als Leitelementträger fungiert. Dies ist z.B. in allen initialalphabetischen Wörterbüchern der Fall, die ein Register aufweisen.

Nach diesen Überlegungen lässt sich nun der erste Ausschnitt aus einer Typologie akzessiver Einträge in Abb. 1 zu einem zweiten Ausschnitt erweitern, wie er sich in Abb. 2 findet.

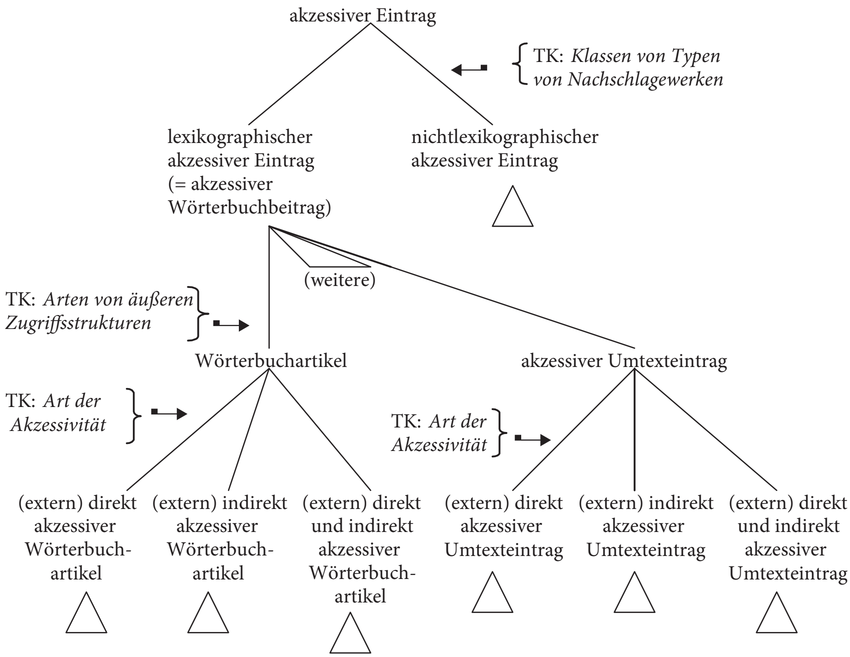


**Abb. 2.** Kommentierter Typologiegraph zum zweiten Ausschnitt aus einer Typologie akzessiver Einträge; vgl. Abb. 1

Es dürfte nun auch klar sein, warum in der nun folgenden Definition von *Wörterbuchartikel* notwendigerweise der oberbegriffliche Terminus *Zugriffstextelement* verwendet wird.

Ein *Wörterbuchartikel* ist ein lexikographischer akzessiver Eintrag (oder: ein akzessiver Wörterbucheintrag), der durch folgende drei Charakteristika eindeutig gekennzeichnet ist:

a. Er weist mindestens ein Zugriffstextelement auf.
b. Er ist eine Konstituente eines Wörterverzeichnisses.
c. Er besteht aus zugriffsbereiten lexikographischen Daten, darunter mindestens einem Teil solcher, anhand derer ein Wörterbuchbenutzer lexikographische Informationen zu dem Nachschlagegegenstand erschließen kann, der mit dem Zugriffstextelement genannt wird.

In der Formulierung des Punktes c. ist der dritte Kritikpunkt berücksichtigt.

## 3 Kurzer Einblick in ein typologisches System für Wörterbuchartikel in Printwörterbüchern

Ein typologisches System ist eine Menge von Typologien zum gleichen Ausgangstyp, wobei mindestens das jeweilige oberste Typologiekriterium (=TK), das die einzelnen Typologien des Systems unterscheidet, theoriebezogen zu wählen ist. Eine Übersicht über ein Typologiesystem, das die Theorie der Wörterbuchform als ein Modul einer Allgemeinen Theorie der Lexikographie berücksichtigt, findet sich in Abb. 3.

TYPOLOGIESYSTEM

- Artikeltypologie mit Bezug auf die innere Textkondensierung
  Oberstes TK: *Vorhandensein von standardisierten Texten*
- Artikeltypologie mit Bezug auf Makrostrukturen und Registerzugriffsstrukturen
  Oberstes TK: *Anzahl der Zugriffspfade*
- Artikeltypologie mit Bezug auf initialalphabetische Makrostrukturen
  Oberstes TK: *Zugehörigkeit zu einer funktionalen Artikelteilstrecke*
- Artikeltypologie mit Bezug auf Datendistribution
  Oberstes TK: *Bezug der artikelinternen Daten zum artikelzugehörigen Nachschlagegegenstand*
- Artikeltypologie mit Bezug auf mediostrukturelle Aspekte (ohne mediostrukturelle Orientierung)
  Oberstes TK: *Verhältnis von Wörterbuchartikeln zu mediostrukturellen Textsegmenten*
- Artikeltypologie mit Bezug auf mediostrukturelle Aspekte (mit mediostruktureller Orientierung)
  Oberstes TK: *genuiner Zweck von verweisvermittelnden Wörterbuchartikeln*
- Artikeltypologien nach Art, Anzahl und Reihenfolge der Kommentare eines Wörterbuchartikels
  Oberstes TK: *Art, Anzahl und Reihenfolge der Kommentare*
- Artikeltypologie mit Bezug auf die Datendistribution im Zweitkommentar
  Oberstes TK: *Typ der artikelinternen Datendistribution*
- Artikeltypologie mit Bezug auf die Textarchitekturen
  Oberstes TK: *Vorhandensein einer Textarchitektur*
- Artikeltypologie mit Bezug auf den Artikelumfang
  Oberstes TK: *relativer Artikelumfang*
- Artikeltypologie mit Bezug auf die Standardisierung
  Oberstes TK: *Vorhandensein von Standardisierung*

**Abb. 3.** Übersicht über das Typologiesystem für Artikeltypen (vgl. Wiegand 2004: 298)

Auch die jeweilige Ausprägungsstruktur ist nicht – wie in der Allgemeinsprache – vorgegeben, sondern wird theoriebezogen relativ zum jeweiligen Typologisierungszweck festgelegt. Neben ihren in Forschungsprozessen jeweils spezifischen Zwecken haben typologische Systeme immer eine erhebliche ordnungsstiftende Funktion und wirken terminologieprägend. Ich habe das typologische System für Wörterbuchartikel entwickelt, um die Formeigenschaften von Wörterbuchartikeln besser und in systematischer Weise übersehen zu können.

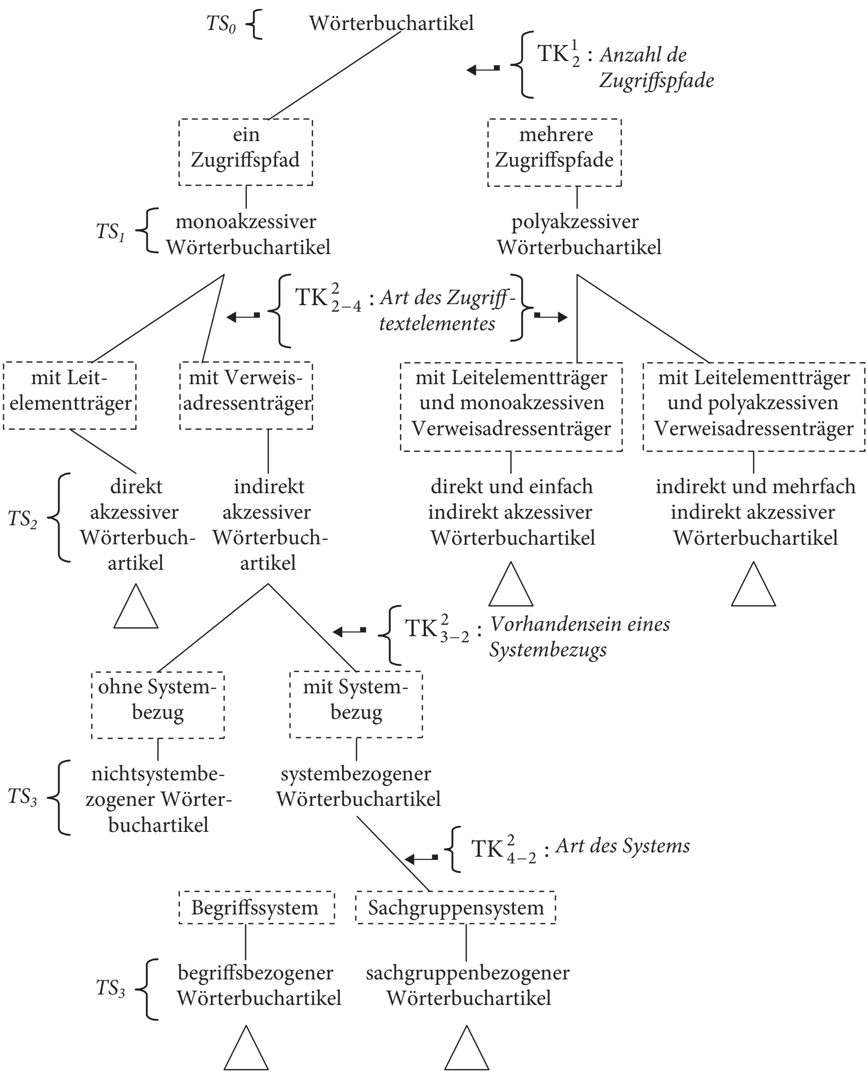


**Abb. 4.** Erweitert *kommentierter Typologiegraph* zur zweiten vierstufigen Artikeltypologie mit Bezug auf Makrostrukturen und Registerzugriffsstrukturen; TS = Typologiestufe; (nach Wiegand 2004: 221)

Im Folgenden kann nur noch ein kurzer Einblick in zwei Ausschnitte aus zwei der Typologien des Systems gegeben werden. Ich zeige zunächst einen Ausschnitt aus der Artikeltypologie mit Bezug auf Makrostrukturen und Registerzugriffsstrukturen (vgl. Abb. 3, zweite Artikeltypologie) und schließe damit partiell an meine bisherigen Ausführungen an.

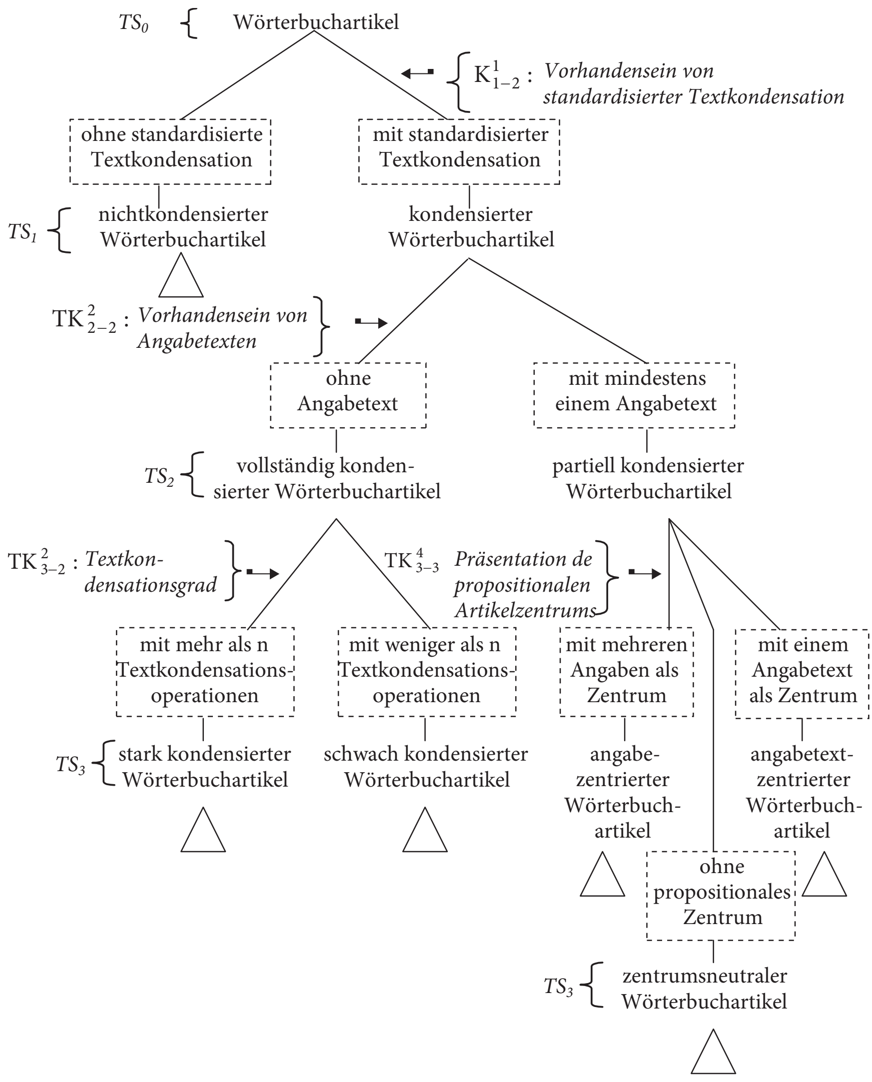


**Abb. 5.** Erweitert kommentierter Typologiegraph zur dreistufigen Artikeltypologie mit Bezug zur inneren Textkondensierung

Zu jedem Artikeltyp, der sich aufgrund der Typologiebildung in Abb. 4 ergibt, gehört eine Definition. Beispielsweise ist der Artikeltyp des direkt und mehrfach indirekt akzessiven Wörterbuchartikel nach Wiegand (2004: 218) wie folgt definiert:

> Ein direkt und mehrfach indirekt akzessiver Wörterbuchartikel ist ein Artikel, auf den sowohl direkt unter Verwendung externer Fundortdaten als auch unter verschiedenen Zugriffsaspekten mittels der Verwendung registerinterner Fundortdaten und damit extern indirekt zugegriffen werden kann.

Zwischen kondensierten und nichtkondensierten Wörterbuchartikeln bestehen hinsichtlich der Wörterbuchform bekanntlich tiefgreifende Unterschiede (vgl. Wiegand 2004: 203 ff). Ein zentraler Typologieausschnitt ist daher der in Abb. 5, der zur Artikeltypologie mit Bezug auf die innere Textkondensierung gehört (vgl. Abb. 3, erste Artikeltypologie). Zur näheren Erläuterung muss auf Wiegand (2004: 203 ff) verwiesen werden.

## 4 Schlussbemerkung

Erst wenn man anhand einer (empirisch fundierten) Typologie von Wörterbuchartikeln einen möglichst umfassenden Überblick über die realisierten Artikeltypen hat und ihre Formeigenschaften systematisch kennt, kann man bei der Planung von Wörterbüchern bestimmte Typen relativ zu Wörterbuchfunktionen systematisch auswählen und damit eine Schema-F-Bearbeitung vermeiden. Dies ist ein wichtiger Aspekt der Praxisrelevanz. Die Kenntnis der Vielfalt erlaubt die Überprüfung einer eingespielten (und damit z.T. verstaubten) Praxis und ist die Voraussetzung für neue Wege beim Schreiben von Wörterbuchartikeln. Zur Praxisrelevanz gehört auch dies: Eine systematische Kenntnis der Formeigenschaften von Wörterbuchtypen und die Verfügung über eine geeignete Terminologie ist zentral für die Lehrbarkeit des Abfassens von Wörterbuchartikeln.

## Literatur

Deutscher Verein für öffentliche und private Fürsorge. 2002. *Fachlexikon der sozialen Arbeit*. 5. Aufl. Frankfurt a.M. [FsA 2002].

Dudenredaktion. 2001. *Duden. Deutsches Universalwörterbuch*. 4., neu bearb. u. erw. Aufl. Mannheim [etc.]: Dudenverlag [Duden-⁴DUW].

Kempcke, Günther. 1984. *Handwörterbuch der deutschen Gegenwartssprache*. In zwei Bänden. Berlin: Akademie-Verlag [HWDG].

Klappenbach, Ruth und Wolfgang Steinitz. 1974. Wörterbuch der deutschen Gegenwartssprache. 1. Bd. 7. bearb. Aufl. Berlin: Akademie-Verlag [WDG].

Lödige, Hartwig. 2001. *Tesa, Tuc und Teddybär. Das große Lexikon der rätselhaften Wörter*. 2. Aufl. München: Ullstein.

Prosinger, Wolfgang. 1984. *Das rabenstarke Lexikon der Scene-Sprache*. Frankfurt a.M.: Eichborn.

Wiegand, Herbert Ernst. 1984 „Prinzipien und Methoden historischer Lexikographie". In *Sprachgeschichte. Ein Handbuch zur Geschichte der deutschen Sprache und ihrer Erforschung*, W. Besch, O. Reichmann, St. Sonderegger (Hrsg.) [Handbuch zur Sprach- und Kommunikationswissenschaft 2.1], 557–620, Berlin. New York: de Gruyter.

Wiegand, Herbert Ernst. 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilbd. Mit 159 Abbildungen im Text. Berlin. New York: de Gruyter

Wiegand, Herbert Ernst. 2004. „Überlegungen zur Typologie von Wörterbuchartikeln in Printwörterbüchern. Ein Beitrag zur Theorie der Wörterbuchform". *Lexicographica* 19: 169–313.

Part 6

# Etymology, history and culture in lexicography

CHAPTER 14

# Lexicographers as borrowers – The importance of being CAMP[2]

Arleta Adamska-Sałaciak

The paper looks at one of the lexicographic strategies for dealing with lack of interlingual equivalence, namely, sanctioning a lexical or semantic borrowing from the source into the target language. An attempt is made to isolate the factors which have to be taken into consideration when deciding whether a particular candidate for borrowing can be used as a translation equivalent in a bilingual dictionary and thereby legitimised in the target language. The discussion is illustrated with examples of recent loans from English into Polish, based on the author's experience as editor of English-Polish dictionaries.

## 1 Anisomorphism and borrowing

### 1.1 The problem

Writing bilingual dictionaries would be almost embarrassingly easy, were it not for what Zgusta (1971) calls the 'anisomorphism' of languages, a phenomenon remarked upon in the modern era at least since Locke:

> A moderate skill in different Languages, will easily satisfie one of the truth of this, it being so obvious to observe great store of Words in one Language, which have not any that answer them in another. Which plainly shews, that those of one Country, by their customs and manner of Life, have found occasion to make several complex Ideas, and give names to them, which others never collected into specifick Ideas. (Locke 1975 [1689/90] III v 8: 432f.; quoted after Law 2003: 264f.)

As a result, a large part of the bilingual lexicographer's time and effort is spent trying to circumvent the problem of lack of target language (TL) equivalents for the source language (SL) items. In addition to the usual stopgap measures (for which see Adamska-Sałaciak 2006: 120f.), there is, at least in theory, the possibility of borrowing the SL item. It is the problems involved in sanctioning such potential loans by the authors and editors of bilingual dictionaries that the present paper wishes to examine.

After looking briefly at the two general types of linguistic borrowing, illustrated with some recent loans from English into Polish, we shall concentrate on one particular case where borrowing has not (yet) taken place, and which thus typifies a situation when the options of the bilingual lexicographer – faced with the lack of a passable TL equivalent – are severely limited. The solutions to this particular problem adopted by two English-Polish dictionaries (and, for the sake of comparison, one English-French dictionary) will be critically examined, followed by some tentative recommendations.

### 1.2 Borrowing and the bilingual dictionary

Usually, when an item is borrowed, both its form and meaning are transferred from one language to the other; we call this a loan(word) or lexical borrowing. Less frequently, the SL meaning alone is borrowed and associated with an existing TL lexeme; this is called semantic borrowing.

As an example of the former, consider *cheerleaderka*, a loanword which entered Polish, together with the appearance of cheerleading teams in Poland, about a decade ago. Before the mid-1990s, an English-Polish dictionary could do no better than explain *cheerleader* by means of a definition in Polish. This is still, in fact, the solution adopted in SAPC, a bilingualised version of *Cambridge Advanced Learner's Dictionary* for Polish learners. Given that the dictionary was published as late as 2003, the avoidance of the loan testifies to the editors' preference for erring on the side of caution. As might be expected, the two major English-Polish dictionaries which came out in recent years – *PWN-Oxford Wielki Słownik Angielsko-Polski* (henceforth PWNOE) and *The New Kosciuszko Foundation Dictionary* (henceforth NKFD) – give P. *cheerleaderka* as an equivalent of E. *cheerleader*.

To take an example of a recent semantic borrowing, the Polish word *klisza* (established senses: "(photographic) film" and "(printing) plate"), is used increasingly often as an equivalent of English *cliché* (for authentic examples, see Adamska-Sałaciak 2006: 134). The fact is duly recorded in *Inny Słownik Języka Polskiego*, a recent monolingual dictionary of Polish (henceforth ISJP). The information is also present in NKFD, though not in PWNOE. The latter dictionary's resistance to this particular borrowing is understandable, given that Poles tend to perceive E. *cliché* and P. *klisza* as false friends. What is more, Polish has several words (e.g. *truizm, frazes, banał*) which can act as equivalents of E. *cliché*. Thus, strictly speaking, borrowing a new sense for P. *klisza* was not dictated by any real need – but then borrowing rarely is, especially when the donor language enjoys the kind of worldwide prestige that English does. Interestingly, *PWN-Oxford Wielki Słownik Polsko-Angielski* (henceforth PWNOP), which came out two years after its English-Polish counterpart, does contain *klisza* in the sense of *cliché*, which clearly indicates that monolingual dictionaries of Polish had been consulted more thoroughly in the making of the Polish-English side of this dictionary than during the preparation of its English-Polish side.

In general, giving one's seal of approval to a borrowing – be it lexical or semantic – which may or may not eventually be adopted by the TL speech community is not something that lexicographers undertake lightly. This is in stark contrast to the practice of translators, who employ borrowings extensively (see, e.g., Tomaszczyk 1984). There are, of course, good grounds for that. In a translated text, a borrowing is embedded in the co-text and context; if necessary, the translator may also choose to define the relevant notion somewhere in the text in order to assist the reader in the interpretation of the lexical item in question. A lexicographer deciding to include a borrowed item in a bilingual dictionary has much less room for manoeuvre. All s/he can do is specify the item's domain (by providing an appropriate subject-field label) and/or offer additional clarifying or disambiguating information in a post-gloss.

For instance, when *the Sejm* appears in an English newspaper article translated from Polish (or, for that matter, in an original English article about Poland), there is normally no danger to comprehension by the English-speaking reader, because the rest of the text will make it clear that what is being talked about is the lower house of the Polish parliament. By contrast, if we put the bare entry *Sejm – the Sejm* in a Polish-English dictionary, we risk compromising the main function of the reference work, i.e., that of explaining meaning.

Cases like P. *Sejm* are, in any event, relatively straightforward, the item in question having a very clear denotational meaning. It is incomparably more difficult to decide on a course of action when dealing with highly polysemous words or with words whose meanings are very complex or vague. It is one such problematic case of potential lexical borrowing from English into Polish that is discussed in the following.

## 2 The meaning of CAMP[2]

### 2.1 Definitions in literature

Most writers dealing with the topic stress the difficulty of defining *camp*, often quoting the character from Isherwood's (1954) novel who famously declared that *camp* is "terribly hard to define. You have to meditate on it and feel it intuitively, like Laotse's Tao... Once you've done that, you'll find yourself wanting to use the word whenever you discuss aesthetics or philosophy or almost anything" (Isherwood 1956: 106; quoted after Glenn 2000).

Despite the difficulty, Isherwood did try to shed some light on the concept, e.g., by distinguishing between high and low camp. However, the word – whose first documented occurrence, according to OED, is from 1909 – did not enter academic discourse, not to mention the general language, until the appearance of Sontag's (1964) seminal essay. Admitting the impossibility of defining *camp* precisely, Sontag opted for loose notes, which seemed best suited for capturing the essence of what she perceived as a "fugitive sensibility". Sontag identified the following characteristics of camp: its love of the unnatural and the vulgar; "the spirit of extravagance"; a taste for the androgynous; "a relish for the exaggeration of sexual characteristics and personality mannerisms"; a victory of style over content; understanding being as playing-a-role. She further noted that camp was essentially serious, even passionate, and that it was typically unintentional. Oscar Wilde was for her a transitional figure, a dandy who anticipated "the democratic *esprit* of Camp". Sontag was convinced that there was "a peculiar affinity" between camp taste and homosexual taste.

These generalisations immediately produced a strong reaction from many quarters, including such powerful cultural icons as Andy Warhol. After forty years, the essay still continues to spawn controversy, documented especially, though not exclusively, in publications produced within the gay community (see, e.g., Bergman (ed.) 1993, Meyer (ed.) 1994, Robinson 2000, Varnell 2000). For instance, while describing Sontag (1964) as brilliant, some authors nonetheless feel that much of what her "Notes..." say is no longer true, since today's camp "also has a nasty side" (Cook 1995).

The size of this paper makes it impossible to go any deeper into the complex social and historical issues involved. To round off this section, here is one contemporary characterisation of camp which I find particularly insightful:

> Combining elements of incongruity, theatricality, and exaggeration, camp is a form of humor that helps homosexuals cope with a hostile environment.Camp can refer either to a style of performance or a mode of perception. (...) Such narrative ballets as *Swan Lake* or *Giselle* have been viewed as camp although their choreographers had no such intentions. In such works, camp is in the eye of the beholder. However, other artists have intended their works to be campy – the English novelist Ronald Firbank and the American playwright Tony Kushner are two examples. But not all objects can be transformed by perception into camp, and not everyone in an audience will see camp even in the campiest performance. The camp effect requires a fit between performance and perception, between object and audience. (Bergman 1995: 130f.)

Characteristically, even the author of this seemingly confident description concedes that "[e]very student of camp must contend with its indefinability, its elusiveness, and its changeability" (Bergman 1995: 130).

### 2.2 Dictionary definitions

Given the phenomenon's inherent elusiveness, it is interesting how dictionaries of the English language propose to define it. The relevant adjective and noun (for simplicity's sake, we shall disregard the verb) are treated in three English monolinguals in the following manner:

> WNRUD:
> **camp²** *sl. n.* **1. a.** An affectation or appreciation of manners and tastes usu. considered outlandish, vulgar, or banal. **b.** Behavior displaying such affectation or appreciation. **2.** Banality, vulgarity, or artificiality when appreciated for its humor
> (the adjective *camp*, *campy* is only given here as a run-on)
>
> OERD:
> **camp²** *colloq. adj.* **1**. affected, effeminate **2.** homosexual **3.** done in an exaggerated way for effect
> *n.* a camp manner or style
>
> ENCARTA:
> **camp²** *adj.* **1**. **OVER-FEMININE** exaggeratedly or affectedly feminine, especially in a man **2. AMUSINGLY BRASH** deliberately and exaggeratedly brash or vulgar in an amusing, often self-parodying way
> *n.* **1**. **EXAGGERATED FEMININITY** exaggeratedly or affectedly feminine behaviour, especially in men **2. DELIBERATE OUTRAGEOUSNESS** deliberate outrageousness for humorous effect.

At first glance, there seems to be little to choose among these. Upon closer inspection, it turns out that the WNRUD definition does not mention effeminacy or gayness (or men, for that matter) at all. I take this as a symptom of the general reluctance on the part of the authors of WNRUD to refer to matters sexual; this is, after all, a dictionary in which the meaning "homosexual" is listed as the very last (fifth) one in the entry for *gay* (following the sense "dissolute").

While WNRUD may simply be dismissed as outdated, the same cannot be said about ENCARTA. Puritanical attitudes are also an unlikely explanation in this case, yet notice that the ENCARTA definition makes no explicit reference to homosexuality, either. Could the omission be interpreted as resulting from the editors taking a stance in the debate about the supposed link, or lack thereof, between tastes and patterns of behaviour characterised as camp and those associated with homosexual men?

OERD seems to go in the opposite direction in that it puts an equals sign (in sense 2.) between *camp* and *homosexual*. In sense 1., no attempt is made to pinpoint the – crucial, it would seem – differences between the meaning of *camp* on the one hand and that of *affected* or *effeminate* on the other. Is this oversimplification due merely to the lexicographer's helplessness in the face of a particularly difficult item or does it reflect a genuine belief in the core meaning of *camp* being reducible to the near-synonyms offered?

Tempting as it might be to pursue the question of what stands behind lexicographers' choices when they define concepts which lack an agreed-upon interpretation (and which might, in addition, be ideologically sensitive), I shall refrain from doing so. Instead, I want to move straight to the evidence of how the word *camp* is actually used in English, before proceeding to discuss its renderings in Polish and its treatment in English-Polish dictionaries.

### 2.3 Evidence from English

Here are some authentic examples, showing the kind of context in which *camp* is typically used: [1]

(1) ... a world marked by camaraderie, **camp**, humor, brotherhood and sisterhood, wit, self-help, and growing indignation at injustice.

(2) If anything, the big Citroen's styling leans more towards the quirky **camp** – there'll be no mistaking it for anything else on the road, love it or hate it.

(3) I dearly love Roger Moore, but if you talk about the best Bonds, you're talking about Sean Connery and not **camp** or sendup.

(4) ... Gunner Howerd, the man who began in **camp** concerts and became the king of **camp**.

(5) ... he emerges out of much **camp** narcissism as a figure of genuine humility.

(6) But the set he strolls through (a wedding cake floating on cloud nine), while meant to evoke the old Ziegfeld days, is just plain bizarre – too tacky to be lavish, not tacky enough to be **camp**.

(7) Something of the same quality of **camp** horror is to be found in *The Blob*, a remake of the 1958 B movie about a mysterious substance that eats people.

(8) The boy turned, in no great hurry. His name tag said LEON. He was **camp** in the strict sense that none of his gestures was useful or necessary. (Smith 2002: 367)

(9) There was something theatrical about him: almost **camp**, but distinctly heterosexual. (Fielding 2003: 231)

(1)-(7), which come from the Longman Spoken and Written English Corpus, do not exhaust the list of examples that were available.[2] I deliberately omitted all those that came close to offering a definition of *camp*, as well as some that were similar in type to the ones chosen. Example (8), which comes from a novel by Zadie Smith, is particularly telling: the author focuses on a single feature of her character's behaviour (exaggerated gesticulation), apparently confident that we will have no difficulty constructing a full picture of Leon on the basis of this one clue. The underlying assumption seems to be that the average reader is likely to have a fairly good idea of what *camp* means to start with. Example (9), from a novel by Helen Fielding, rests on a different assumption, namely that homosexuality is a necessary prerequisite of true campiness.

### 2.4 Evidence from Polish

Given the popularity of anglophone culture in Poland, one would expect to come across Polish texts discussing the phenomenon of camp and featuring the corresponding linguistic borrowing. We do, indeed, find instances of both the noun *kamp* and the adjective *kampowy*, but (so far?) almost exclusively in contexts where the so-called 'aesthetics of camp' is being talked about. Both words are also spelled (marginally less often, it would seem) with the original initial *c*.[3]

Despite the opinion-forming character of the literary magazine in which it first appeared, the Polish translation of Sontag (1964) seems to have gone largely unnoticed. Thus, *kamp* and *kampowy* hardly had a chance to assert themselves at the time. It is only in recent years that the words seem to have made a sort of comeback – a belated appearance might be a more adequate description. Even today, their domain of use remains highly restricted: if we discount explicit references to Sontag's "Notes on 'Camp'", the few occurrences of *kamp* and *kampowy* I have managed to unearth all come from papers on postmodern literature, film and poetry reviews, or announcements of art exhibitions. Here are some of those attestations:

(10) Pragniemy zaprezentować Państwu prace pierwszej polskiej artystki, która z upodobaniem uprawia świadomy **camp**. (Grzeszczyk 1997)
[We would like to present the work of the first Polish artist who delights in practising conscious camp.]

(11) Prawdziwy wysoki **camp** kryje w sobie cień powagi – nie można uznać za **campowe** czegoś, czego nie traktuje się w pewnym sensie poważnie. (Bieszk 2001)

[There is a touch of seriousness hidden in genuine high camp – you cannot regard something as campy unless in a sense you treat it seriously.]

(12) Niemniej ujawnił w swym referacie również niezwykle interesujące powinowactwo postmodernizmu i estetyki **kampu**, wskazując na ucieczkę prozy postmodernistycznej w często pusty estetyzm. (Roszczynialska 2001)
[Nevertheless, in his paper he also exposed an extremely interesting affinity between postmodernism and the aesthetics of camp, pointing to the escape of postmodernist prose into often empty aestheticism.]

(13) Stąd już krok tylko do postawy zwanej estetyką "**kampu**". (Bielenda 2002)
[From here, it is only a step to the stance called the aesthetics of "camp".]

(14) Atmosfera ta kojarzy mi się z klimatem **kampu**; może to rzeczywiście są pierwsze polskie wiersze konsekwentnie **kampowe**? (Maliszewski 2003)
[The atmosphere brings to mind the climate of camp; maybe these really are the first Polish poems that are consistently camp?]

(15) Kicz dyskredytuje i poniża, jest oskarżeniem; **kamp** uszlachetnia i wysubtelnia. (Kulpa 2003)
[Kitsch discredits and degrades, it is an accusation; camp ennobles and refines.]

Given their low frequency of occurrence and high domain-specificity, it seems that Polish *kamp* and *kampowy* are currently, at most, expert terms – unlike English *camp*, which is also a word of ordinary language.[4] Besides, a significant proportion of the Polish texts featuring *kamp* and *kampowy* (e.g., Bieszk 2001 or Kulpa 2003) are actually essays expounding on the phenomenon of camp. Accordingly, we might wonder whether *kamp* is really a fully-fledged borrowing there: it seems to be a case rather like that of *the Sejm* appearing in an English text about Polish politics (see 1.2.). Arguably, most instances of P. *kamp(owy)* thus qualify as mention of the term, not its use.

### 2.5 Treatment in English-Polish dictionaries

At the time of writing this (spring 2004), not a single monolingual dictionary of Polish contained the entries *kamp* or *kampowy*. Nonetheless, the existence of data like (10)-(15) would seem to provide sufficient grounds for including the words as equivalents of E. *camp* in an English-Polish dictionary. So far, this has not happened. While I cannot speak for the editors of PWNOE, the decisive factor for me, as editor of the English-Polish part of NKFD, was the unfamiliarity of *kamp* to educated native speakers of Polish.[5] Once the decision was taken not to record the loanword, we had to resort to the traditional solution, i.e., a combination of near-synonyms, TL definitions, and post-glosses, mirroring the various senses recognised by monolingual English dictionaries, and incorporating some information gathered from other sources. The result looks as follows:[6]

> NKFD
> **camp**[2] *a. pot.* **1**. zniewieściały; afektowany (*zw. celowo*) **2**. dotyczący subkultury gejowskiej (*np. przebierania się w kobiece stroje*) **3**. przesadnie stylowy *l.* staroświecki (*zw. w zabawny sposób*) 4. ostentacyjnie kiczowaty (*i przez to budzący zachwyt*)
> *n. pot.* **1**. zniewieściałość; afektacja; wystawianie się na pokaz; demonstrowanie swego homoseksualizmu **2**. subkultura gejowska **3**. zachwyt nad kiczem, banałem *l.* wulgarnością (*zw. dla ich walorów humorystycznych*).

This is, obviously, far from satisfactory. Let us not forget, however, that the meaning of the English lexeme is very difficult to pinpoint; to use the traditional distinction between denotation and connotation, we might say that the meaning of *camp* is largely connotational. Incidentally, this may also be responsible for the somewhat artificial multiplicity of (sub)senses into which CAMP[2] is split in the monolingual dictionaries we have looked at.

As already mentioned, PWNOE, the other recent English-Polish dictionary, does not register Polish *kamp* or *kampowy*, either. Still, its entry for CAMP[2] looks rather different from the NKFD one, apparently as a result of the lexicographer's determination to offer one-word partial equivalents rather than descriptive explanations:[7]

> PWNOE
> **camp**[2] *n* (mannered style) afektacja, teatralność, pej kabotynizm, kabotyństwo; **it's high** or **pure** ~ to czyste kabotyństwo
> *adj* pej 1 (exaggerated) *[person]* afektowany, zachowujący się nienaturalnie; *[gesture, performance]* teatralny, nienaturalny, pej kabotyński 2 (effeminate) zniewieściały; pedziowaty infml offensive 3 Theat, Cin, TV (in bad taste) kiczowaty; ~ **comedy** niewybredna burleska

The equivalents suggested by PWNOE are very imprecise: while implying (incorrectly) that E. *camp* is a strongly pejorative word, they contain no hint of the frequently deliberate character of camp behaviour or of the good-natured amusement which that behaviour is aimed at causing. Ultimately, of course, it is for the dictionary users to decide whether they prefer PWNOE's rough equivalents to the wordy explanations given in NKFD.

### 2.6 A brief look at French

For the sake of comparison, here is the relevant entry from the English-French part of *Le Robert & Collins Dictionnaire Français-Anglais, Anglais-Français* (henceforth RCD):

> RCD
> **camp**[2] * ADJ **a** *( = affected) [person, behaviour, talk]* affecté, maniéré; *( = overdramatic) [person]* cabotin; *[gestures]* théâtral; *( = affecting delight in bad taste)* qui aime le kitsch, qui fait parade de mauvais goût; *( = fashionable because of poor taste)* kitsch *inv* **b** *(= effeminate)* efféminé; *( = homosexual) [man]* qui fait pédé ** *or* tapette **; *[manners, clothes]* de pédé **, de tapette ** ▪ **to be as ~ as a row of tents** * être pédé comme un phoque ** N (also **high camp**) *(of manners)* affectation *f*; *(effeminate)* maniéres *fpl* efféminées

When considered in terms of the explanatory strategies adopted, the above seems to fall somewhere in between the corresponding entries in NKFD and PWNOE. Most importantly, just like the English-Polish entries, it contains no suggestion that there might be a loanword *camp* – or something similar – in the target language. The results of a Google search (conducted in the spring of 2003) suggest that the situation of *camp* in French may be similar to that of *kamp(owy)* in Polish, insofar as the word tends to be mentioned rather than used; cf. the following:

(16) Or, Susan Sontag définit le ***camp*** assez précisément comme un phénomène qui, bien sûr, est le plus souvent manifeste dans les milieux homosexuels urbains artistiques américains, mais qui néanmoins a surtout la caractéristique d'être inconscient et apolitique. (Dupont 1994)

(17) ... comment ne pas apprécier par exemple un écrit comme l'essai sur le ***Camp*** que Sontag a publié en 1964? (Guilbert 2002: 248)

## 3 Conclusions and recommendations

The solutions adopted by NKFD, PWNOE, and RCD, while definitely less than ideal from the point of view of lexicographic theory, seem to be justified on practical grounds. The descriptive translations and near-equivalents offered by all three dictionaries may, at least for the moment, be more user-friendly than the so-far esoteric equivalents would have been. What we have here – certainly in the English-Polish context – is a situation where the TL equivalent (which happens to be a borrowing, but could just as well be a cognate or an entirely unrelated word) is much less known in the TL culture than the SL item is in the SL culture. In fact, the equivalent's frequency is so low, and its domain of use so narrow, that it is completely unknown to the vast majority of TL speakers.

All this having been said, it seems only a matter of time before P. *kamp* and *kampowy* begin to feature in dictionaries. In fact, it is surprising that they have not yet been recorded by Polish monolinguals, given the care and attention with which terms associated with scholarly discourse are customarily treated therein. With regard to bilingual dictionaries, examination of the Polish evidence discussed in 2.4. – only a fraction of which pre-dates 2001, i.e., the time when the relevant entry in NKFD was edited – has convinced me that authors of English-Polish dictionaries should sanction the borrowing. Were the decision being made today, I would probably allow *kamp* in, while still holding on to some of the explanatory paraphrases, perhaps disguised as glosses.

To conclude, allowing a borrowing of an uncertain status into a bilingual dictionary is not an easy step to take. The candidate's presence in a monolingual dictionary of the target language should, of course, automatically decide in its favour.[8] Otherwise, the size of the evidence for the TL item must be examined (see, e.g., Čermak 2003: 20). Crucially, the evidence should involve not only the frequency of the item in TL corpora, but also TL speakers' familiarity with it. Although the two are, of course, interrelated, one can easily imagine a situation where an item which is relatively well-documented in texts of a specific type might actually be unfamiliar to most speakers of the language in question. For a few more years at least, this is likely to remain the most probable scenario for Polish *kamp* and *kampowy*, before, inevitably, the words start to spill over from academic discourse into everyday language. When this happens, English-Polish lexicographers will be able to stride through the CAMP[2] entry – only to stumble on some other case involving a possible borrowing-in-the-making and requiring careful deliberation. It is not easy to be a responsible borrower.

## Notes

1. Apart from adding bold, original formatting and punctuation has been retained in all the examples quoted in this paper.

2. Access to the Longman Spoken and Written English Corpus (of 40 million words) was granted to the team preparing LSW, an active bilingual learners' dictionary for speakers of Polish. The corpus was not consulted during the preparation of NKFD, nor indeed was it available at the time.

3. Warenstein's translation of Sontag's essay (Sontag 1979) used the *k* spelling and preserved the original capitalisation.

**4.** Exactly how familiar it is to the average (educated?) native speaker of English is a matter I do not feel qualified to comment upon (but see the remarks on example (8) in 2.3.).

**5.** A number of Adam Mickiewicz University School of English graduates confessed in informal interviews (conducted early in 2001) that they had no idea what *camp* in the relevant sense(s) meant in English, let alone being aware of a Polish equivalent.

**6.** Following the suggestion of an anonymous reviewer, rightly concerned for non-speakers of Polish, I include English translations of the entries in NKFD (immediately below) and PWNOE (following note):

> **camp**$^2$ *a. inf.* 1. effeminate; affected *(usu. deliberately)* 2. pertaining to gay subculture *(e.g. drag queens)* 3. exaggeratedly stylish or quaint *(usu. in an amusing way)* 4. ostentatiously kitschy *(and admired for it)*
> *n. inf.* 1. effeminacy; affectation; brashness; parading one's homosexuality 2. gay subculture 3. appreciation of kitsch, banality or vulgarity *(usu. for their humorous value)*.

**7.** Replacing the alleged Polish equivalents of *camp* and the Polish translations of English phrases in this entry with their closest English equivalents/translations produces the following:

> **camp**$^2$ *n* (mannered style) affectation, theatricality, pej cabotinage; **it's high** or **pure** ~ it's pure cabotinage
> *adj* pej 1 (exaggerated) *[person]* affected, behaving unnaturally; *[gesture, performance]* theatrical, unnatural, pej play-acting 2 (effeminate) effeminate; faggoty infml offensive 3 Theat, Cin, TV (in bad taste) kitschy; ~ **comedy** unrefined burlesque

**8.** Even this seemingly uncontroversial rule is not always observed in practice. As mentioned in 1.2., PWNOE does not include the semantic borrowing *klisza* as an equivalent of E. *cliché*, even though the word is recorded in the relevant sense in ISJP, a monolingual dictionary of Polish published a whole two years earlier.

## References


<u>*Dictionaries*</u>

*Encarta World English Dictionary*. 1999. London: Bloomsbury. (ENCARTA)

*Inny Słownik Języka Polskiego.* 2000. Warszawa: Wydawnictwo Naukowe PWN. (ISJP)

*Longman Słownik Współczesny Angielsko-Polski, Polsko-Angielski*. 2004. Harlow: Pearson Education Ltd. (LSW)

*The New Kosciuszko Foundation Dictionary English-Polish, Polish-English.* 2003. Kraków: Universitas. 2 vols. (NKFD)

*The Oxford English Dictionary, Third Edition: Online*. 2000-. Oxford: Oxford University Press. (OED)

*The Oxford English Reference Dictionary*. 1995. Oxford: Oxford University Press. (OERD)

*PWN-Oxford Wielki Słownik Angielsko-Polski*. 2002. Warszawa: Państwowe Wydawnictwo Naukowe. (PWNOE)

*PWN-Oxford Wielki Słownik Polsko-Angielski*. 2004. Warszawa: Państwowe Wydawnictwo Naukowe. (PWNOP)

*Le Robert & Collins Dictionnaire Français-Anglais, Anglais-Français: 21st Century Edition*. 2002. Glasgow and Paris: HarperCollins Publishers *and/et* Dictionnaires Le Robert-VUEF. (RCD)

*Słownik Angielsko-Polski Cambridge*. 2003. Warszawa: Prószynski i S-ka. (SAPC)

*Webster's II New Riverside University Dictionary*. 1984. Boston: Houghton Mifflin Company. (WNRUD)

<u>*Other works*</u>

Adamska-Sałaciak, Arleta. 2006. *Meaning and the Bilingual Dictionary: The Case of English and Polish*. Frankfurt am Main: Peter Lang.

Bergman, David. 1995. "Camp". In *The Gay and Lesbian Literary Heritage*, Claude J. Summers (ed.), 130–135. New York: Henry Holt.

Bergman, David (ed.). 1993. *Camp Grounds: Style and Homosexuality*. Amherst, MA: University of Massachusetts Press.

Bielenda, Ewa. 2002. "Proza polska końca XX wieku a postmodernizm". *Konspekt* 11. www.wsp.krakow.pl/konspekt/11/rec06.html

Bieszk, Patrycja. 2001. "Campowy bric-á-brac". *Ha!art*. www.ha.art.pl

Cook, Gareth. 1995. "Breaking camp". *Metro*, Nov. 16–22, 1995.

Čermák, František. 2003. "Source materials for dictionaries". In *A Practical Guide to Lexicography*, Piet van Sterkenburg (ed.), 18–25. Amsterdam and Philadelphia: John Benjamins.

Dupont, Eric. 1994. "Lost in space: Quelques concepts clés de la théorie queer". *Surfaces* 3.

Fielding, Helen. 2003. *Olivia Joules and the Overactive Imagination*. London: Picador.

Glenn, Joshua. 2000. "Camp: An introduction". www.hermenaut.com/a28.shtml

Guilbert, Georges-Claude. 2002. "Camille Paglia: Féministe ennemie des féministes". *Cercles* 4: 243–253. www.cercles.com

Isherwood, Christopher. 1956 [1954]. *The World in the Evening*. New York: Avon Books.

Kulpa, Robert. 2003. "Kicz – subwersja – kamp. Szkic antropologiczny". www.nic-tak-samo.uni.wroc.pl/teksty/kulpa.html

Law, Vivien. 2003. *The History of Linguistics in Europe from Plato to 1600*. Cambridge: Cambridge University Press.

Locke, John. 1975 [1689/90]. *An Essay Concerning Human Understanding*. Reprint of 4th edition. Ed. by P. H. Nidditch, Oxford: Clarendon.

Maliszewski, Karol. 2003. "Kamp i akwarium". *Tygodnik Internetowy* 107.

Meyer, Moe (ed.) 1994. *The Politics and Poetics of Camp*. London and New York: Routledge.

Robinson, M. J. 2000. "The poetics of camp in the films of Alfred Hitchcock". *Rocky Mountain E-Review of Language and Literature* 54.1.

Roszczynialska, Magdalena. 2001. "Postmodernistyczne łowy". Konspekt 7. www.wsp. krakow.pl/konspekt/konspekt7/roszczyn.html

Smith, Zadie. 2002. *The Autograph Man*. London: Hamish Hamilton.

Sontag, Susan. 1964. "Notes on "Camp"". *Evergreen Review*, Dec. 1964, New York. Reprinted in Sontag 1966. *Against Interpretation and Other Essays*. New York: Picador USA. www.pages.zoom.co.uk

Sontag, Susan. 1979. "Notatki o Kampie". *Literatura na Świecie* 9 (101): 307–323. Translated by W. Warenstein.

Tomaszczyk, Jerzy. 1984. "The culture-bound element in bilingual dictionaries". In *LEXeter '83 Proceedings*, Reinhard R. K. Hartmann (ed.), 289–297. Tübingen: Max Niemeyer.

Varnell, Paul. 2000. "Sour "Notes on Camp"". *Chicago Free Press*, May 3, 2000.
www.indegayforum.org

Zgusta, Ladislav. 1971. *Manual of Lexicography*. The Hague: Mouton.

[All Web-based sources last accessed July 27, 2007]

CHAPTER 15

# Politik und Sprachpolitik in der Lexikographie

Henning Bergenholtz und Sven Tarp

The main characteristic of lexicography and language policy is that both areas have a long history and that they have both gained increasing academic attention during the last 25 years. Furthermore, language policy has been the object of increasing public interest in the last decade or so. Within lexicography, there is no general agreement about where to draw the line between lexicography and linguistics. It depends on whether you see lexicography as a part of linguistics – a quite popular understanding, but in reality a harmful one for the development of lexicography. Likewise, there is no general agreement about where to draw the line between language decisions and language policy. The answer to this question depends on what you see as a part of politics or, more specifically, whether all language decisions are language policy decisions (cf. Bergenholtz/Gouws 2006). In this contribution we make a distinction between a general, a specific and a selecting language policy. The latter is often known as corpus planning.

Lexikographie und Sprachpolitik haben gemeinsam, dass beide Bereiche eine lange Geschichte aufweisen können, und dass sie seit etwa 25 Jahren ein zunehmendes akademisches Interesse gewonnen haben. Die Sprachpolitik hat darüber hinaus in den letzten Jahren ein großes öffentliches Interesse gefunden. In der Lexikographie herrscht keine allgemeine Übereinstimmung darüber, wo sich die Grenze zwischen Lexikographie und Linguistik befindet. Dies muss davon abhängen, ob man Lexikographie als Teilbereich der Linguistik versteht – eine Auffassung, die zwar allgemein angenommen wird, aber sehr schädlich ist für die Entwicklung der Lexikographie. Genauso ist nicht eindeutig, wo die Grenze zwischen Sprachentscheidungen und Sprachpolitik zu ziehen ist. Dies muss davon abhängen, was man der Politik zuordnet, genauer gesagt, ob alle Sprachentscheidungen letztendlich als sprachpolitisch bezeichnet werden können. Für eine solche Zuordnung wird in diesem Beitrag plädiert. Es wird unterschieden zwischen einer generellen, einer speziellen und einer selektierenden Sprachpolitik. Der letztere Bereich wird oft als Korpusplanung bezeichnet, d.h. eine phänomenselektierende bzw. normierende oder nicht-normierende Sprachpolitik, die in Zusammenarbeit mit oder von Lexikographen durchgeführt wird.

## 1 Politisch brisante und weniger brisante Wörter

Mit Wörtern kann man Politik ausüben. Mit Wörtern kann man Tatsachen so umschreiben, dass der tatsächliche Inhalt verschleiert wird. Wörter können auch so in einem bestimmten geschichtlichen Zusammenhang gebraucht, d.h. missbraucht werden, dass sie für lange Zeit oder für immer in Verruf kommen. Als Spiegelbild dazu gilt, dass die Nicht-Verwendung bestimmter Wörter, die eigentlich treffender wären, Teil einer bewussten politischen Strategie sein kann. Das ist alles bekannt. Das war, ist und bleibt ohne Zweifel auch in der Zukunft so. Einige heutige Beispiele: Wer ist ein *Terrorist* und wer ein *Widerstandskämpfer*? Ist man ein *Antisemit*, wenn man als Nicht-Jude bestimmte Entscheidungen der israelischen Regierung kritisiert? Sollte man vermeiden, das Wort *Neger* zu verwenden und stattdessen *Farbiger* sagen? Ist es im Sinne eines politisch korrekten Sprachgebrauchs, wenn man eine Frau nicht als *Idiot*, sondern als *Idiotin* beschimpft? Es geht insbesondere um Bezeichnungen für politische Einstufungen, für Rassen, für Sexus, für Sex, Ökologie und für den mentalen Zustand einer Person.

In all diesen Fällen haben wir es mit politisch brisanten Wörtern zu tun, die auch Teil einer Sprachpolitik werden können. Es geht um politisch oder gesellschaftspolitisch umstrittene Ausdrücke, die in vielen metalexikographischen Beiträgen diskutiert wurden. Dies wird typischerweise so getan, dass man eine bestimmte Anzahl von Wörterbüchern daraufhin untersucht, wie gut oder schlecht, wie politisch korrekt oder unkorrekt sie sind, oder ob sie bestimmte Wörter überhaupt in ihre Lemmaliste aufgenommen haben. So kann der metalexikographische Beitrag je nach Eigeneinstellung seine Genugtuung oder Kritik über die jeweiligen Entscheidungen in vorliegenden Wörterbüchern zum Ausdruck bringen und evt. auch einige Vorschläge für eine verbesserte Praxis machen. In der Metalexikographie, insbesondere der deutschen, kann man diese Vorgehensweise als linguistisch inspirierte Wörterbuchphilologie bezeichnen. Die relevantesten Beiträge dieser Art gehen von der Tatsache aus, dass die Verwendung bestimmter Wörter oder Ausdrücke beleidigend sein kann und fordern daher, dass die Wörterbücher auch darauf aufmerksam machen, z.B.:

> For these reasons it is imperative that dictionaries give guidance on usage and that the dictionary user is made aware of this problem. (Busse 2000: 166)

Wir stimmen dem Imperativ vorbehaltlos zu: Ein Wörterbuch sollte dem Benutzer helfende Richtlinien geben. Die Frage ist, was das heißt, bzw. wie man das am besten tut. Wie sollen Instruktionen für den Wörterbuchbenutzer formuliert werden? Es gibt hier verschiedene Strategien:

- Man kann einen bestimmten Sprachgebrauch beschreiben und hinzufügen, dass dieses Wort oder dieser Ausdruck von einigen Sprechern dieser Sprache als z.B. selten oder beleidigend aufgefasst wird. Der Lexikograph geht daraufhin davon aus, dass der Wörterbuchbenutzer nun selber entscheidet, ob er diesen Ausdruck verwenden will oder nicht.

– Man kann einen bestimmten Sprachgebrauch explizit kritisieren und den Wörterbuchbenutzer dazu auffordern, den beschriebenen Sprachgebrauch nicht zu benutzen.

Busse (2000) wählt die erste Möglichkeit. Das ist die übliche, beschreibende und zurückhaltende Praxis, die voraussetzt, dass der Wörterbuchbenutzer das Wörterbuch nicht als Orakel, sondern nur als Ausgangspunkt für eigene noch zu treffende Beschlüsse verwendet. Das klingt demokratisch und sympathisch. Wir halten diese Vorgehensweise jedoch für zu wenig explizit und in Wirklichkeit auch für pseudodemokratisch und zugleich wenig hilfreich. Die Haltung ist pseudodemokratisch, weil sie Konservierung des soeben festgestellten Sprachgebrauchs beinhaltet.[1] Ein Wörterbuch wird von den Benutzern als Orakel aufgefasst, dessen Aussagen einen eindeutigen Wahrheitswert haben und befolgt werden sollen. Das Wörterbuch hat in den Augen der Benutzer diese Orakelfunktion, und zwar zu Recht. Dies gilt insbesondere, wenn das Wörterbuch in Verbindung mit einem Textproduktionsproblem herangezogen wird. Der Benutzer sucht nicht nur einen Ratschlag, er sucht die Wahrheit.

Wir werden im Folgenden dafür argumentieren, die zweite Möglichkeit für Wörterbücher zu wählen, die nur oder auch als Textproduktionswörterbücher konzipiert werden. Allerdings sollte der Lexikograph darauf bedacht sein, dass seine Vorgehensweise als Teil einer explizit formulierten Sprachpolitik zu sehen ist. Dies gilt jedoch nicht nur für brisante Wörter, sondern für die gesamte lexikographische Arbeit.

Wir haben somit drei Gebiete: Politik, Sprachpolitik und Lexikographie. Im Normalfall sind die Abhängigkeitsbeziehungen eindeutig: Die Politik trifft sprachpolitische Entscheidungen, die Sprachpolitik richtet sich danach, die Lexikographie normalerweise auch. Entsprechend trifft man in der Sprachpolitik Entscheidungen, die die Lexikographen meist ohne Einschränkungen in ihren Wörterbüchern befolgen oder zumindest ausführlich beschreiben. Es gibt hier große Unterschiede zwischen den einzelnen Sprachen, abhängig davon, ob es eine Tradition für sprachpolitische Eingriffe gibt oder nicht. Insofern ist es nicht neu, dass man feststellen kann, dass einige lexikographische Entscheidungen sprachpolitische und manchmal auch politische Relevanz haben. Eher neu ist die Hauptthese dieses Beitrags: Sämtliche Entscheidungen der Lexikographie haben eine sprachpolitische Relevanz und daher letzten Endes auch eine politische Dimension.

Gegen diese These könnte man den Einwand einbringen, dass es bekanntlich viele und teilweise recht unterschiedliche Auffassungen davon gibt, was man unter

*Sprachpolitik* zu verstehen hat, und man daher erst diesen Terminus definieren muss, bevor die These verständlich gemacht werden kann. Das stimmt im Prinzip. Wir werden später auf die Terminologie zu sprechen kommen und eigene Definitionen und Eingrenzungen vornehmen. Wir halten dennoch an der Grundthese fest, bei der vorläufig nur *Sprachpolitik* als Terminus in jener Bedeutung verwendet wird, die dem dänischen Terminus *sprogpolitik* entspricht, bzw. der Summe der beiden deutschen Termini *Sprachpolitik* und *Sprachenpolitik* bzw. der Gemeinsamkeit der englischen Termini *language policy* und *language politics* entspricht. Diese sehr breite Auffassung von *Sprachpolitik* beinhaltet auch Gebiete wie *Sprachlenkung, Sprachplanung, Sprachkultur, Korpusplanung, Sprachausbau, Sprachmanipulation, Sprachpflege, Sprachpurismus, Sprachregelung, Sprachreinigung, Sprachnormierung*. In der deutschen Terminologie sind die ersten vier dieser zusätzlichen Termini der *Sprachpolitik*, die übrigen der *Sprachenpolitik* zuzuordnen, s. hierzu Bergenholtz (2004).

Auch mit diesem vagen und sehr allgemeinen Verständnis von Politik, Sprachpolitik und Lexikographie können nun einige Beispiele aus der Wörterbuchpraxis und zugleich des heutigen Sprachgebrauchs im Deutschen und Dänischen herangezogen werden, um die lexikographische Problematik näher zu beleuchten. In dem neuesten dänischen Wörterbuch, DEN DANSKE ORDBOG, hat der Wörterbucheintrag (1) einige Journalisten zu Zeitungsbeiträgen angeregt. Sie haben damit Probleme gehabt. Es geht um die so genannten Pendelwörter, d.h. Wörter, die zwei entgegengesetzte Bedeutungen haben. Das ist eine linguistisch interessante Beobachtung, lexikographisch ist diese Entwicklung zu berücksichtigen. Eine politische und vor allem sprachpolitische Angelegenheit ist der Wörterbucheintrag dadurch geworden, dass nicht nur Journalisten, sondern auch Politiker und viele Sprachinteressierte sich öffentlich dazu geäußert haben.

**(1) bjørnetjeneste**

1 handling som er velment, men som gør mere skade end gavn

2 handling som er til stor hjælp for nogen; stor tjeneste (*kendt fra 1992, denne brug regnes af mange for ukorrekt*)

(dt.: **Bärendienst**

1 gut gemeinte Handlung, die mehr schadet als hilft

2 Handlung, die eine große Hilfe ist für jemanden (bekannt seit 1992, diesen Gebrauch betrachten viele als nicht korrekt))

Wo einige sich bei dem Wörterbucheintrag (1) vielleicht nur leicht wundern, dass der Eintrag sprachpolitisch brisant werden kann, könnte es für viele unverständlich erscheinen, dass auch grammatische Angaben einen sprachpolitischen

Hintergrund haben können, wie es in den Einträgen des offiziellen dänischen Rechtschreibewörterbuches, Retskrivningsordbogen, der Fall ist:

(2) **hooligan** Singular:-en, Pluralis: -er
(dt.: *Hooligan* + Angabe des Morphems für Singular und Plural)

In diesem Wörterbucheintrag wird die „englische“ Pluralflexion auf *-s* nicht angegeben, obwohl sie immer und das “dänische” Flexionsmorphem so gut wie nie verwendet wird. Es ist eine politische Entscheidung, fremde, insbesondere englische Flexionsmorpheme zu verbieten. Die politische Entscheidung ist in Dänemark nicht folgenlos, da sich jeder Mitarbeiter im öffentlichen Dienst und auch jeder Schüler und jeder Student nach diesen offiziellen Angaben richten müssen und jeder Lehrer laut Gesetz jeden Verstoß dagegen als Fehler einstufen muss. Es ist aber nicht so, dass das offizielle dänische Rechtschreibewörterbuch keine Varianten zulässt. Dies geschieht zum Beispiel – wie wir es auch in deutschen Wörterbüchern kennen – bei Genusangaben. Der folgende Wörterbucheintrag ist typisch für die Feststellung, dass in vielen Variationsfällen nur die eine Variante und die andere im Sprachgebrauch nur selten verwendet wird. Der Lehrer darf die völlig unübliche Variante mit Gemeinschaftsgenus nicht als Fehler korrigieren:

(3) **cirkus** en/et
(dt.: *Zirkus* + grammatische Angabe: Gemeinschaftsgenus oder Neutrum)

Schließlich soll noch ein semantischer Streitfall herangezogen werden. Die Bedeutung für dän. *racist* (dt. *Rassist*) hat in Dänemark Anlass zu einem Rechtsstreit gegeben, zu einem politisch motivierten Rechtsstreit. In dem Rechtsstreit ging es darum, ob Wörterbucheintrag (4) oder (5) im Rahmen des dänischen Rechts in dem normalen dänischen Sprachgebrauch anzunehmen ist:

(4) **racist** person, der går ud fra, at der er kvalitative forskelle mellem racer
(dt.: **Rassist** jemand, der prinzipielle Wertunterschiede zwischen Rassen annimmt)

(5) **racist** person, som udtrykker sig kritisk om en bestemt gruppe mennesker
(dt.: **Rassist** jemand, der eine kritische Stellungnahme zu einer bestimmten Gruppe von Menschen vornimmt)

Es ist zwar eine Frage der Definition von Sprachpolitik, ob hier eine sprachpolitische Frage vorliegt, aber eine politische ist es schon. Die dänische rechtspopulistische Politikerin Pia Kjærsgaard wurde von einer politischen Gegnerin als *Rassistin* (dän. *racist*) bezeichnet. Die Politikerin hatte darauf hin einen Gerichtsprozess wegen Verunglimpfung angestrengt. Die Angeklagte wurde freigesprochen, weil das Gericht nach Anfrage beim offiziellen dänischen Sprachrat die Antwort erhielt, dass es heute öfters vorkomme, dass *racist* nicht nur in der Bedeutung 'Rassenhasser' (Wörterbucheintrag 4), sondern als 'Kritiker von Ausländern und bestimmten dänischen Gruppen' (Wörterbucheintrag 5) verwendet wird. Da dem so ist, so das Gericht, könne man mit der Bezeichnung *racist* niemanden verunglimpfen.

Selbstverständlich ist so etwas eine hochpolitische Angelegenheit, für viele politisch brisante Problemstellungen gilt dasselbe Phänomen: Sprache und Macht sind miteinander eng verbunden. Ist jemand, der etwas in Israel kritisiert, notwendigerweise ein Antisemit? Nach der üblichen Bedeutungsangabe in Wörterbüchern wohl nicht, in der politischen Diskussion, insbesondere in offiziellen israelischen Reaktionen manchmal schon. Wie soll nun ein Wörterbucheintrag *Antisemit* formuliert werden? Diese Problematik ist u.E. so explosiv, dass ihre Diskussion von der grundlegenden lexikographischen Fragestellung ablenkt, dasselbe gilt für brisante Wörter wie *Demokratie, Tortur* und *Terrorist*. Wir möchten daher auf zwei andere, zwar brisante, aber weniger explosive Beispiele eingehen. Der dänische Ausdruck

**(6)** fuck dig, din luder
(dt.: fuck you, Du Nutte)

ist ähnlich wie *Rassist* Gegenstand einer erregten Debatte gewesen. Hier ging es allerdings nicht um einen Gerichtsprozess, sondern um die Bewertung einer Doktorarbeit. Eine Doktorandin beschrieb in einem Interview im Fernsehen die Sprache der Jugendlichen, die aus Immigrantenfamilien stammen (Oktober 2003). Sie sprach von einem besonderen Ethnolekt, den sie Araberslang nannte. Der genannte Ausdruck "fuck dig, din luder" bedeutet ganz einfach "Geh!". Nicht mehr, d.h. nicht besonders grob, derb oder herabsetzend, so sagt man halt. Die Doktorandin meinte, dass wir hier einen Sprachaufruhr erleben würden, wo die Jugend aus Immigrantenkreisen ihren Selbstwert behaupten wolle, sie wollten zeigen, dass sie etwas Besonderes seien. Außerdem, fügte sie hinzu, unsere Sprache verändere sich doch sowieso die ganze Zeit.

In einer Zeitung gab es als Reaktion auf diesen Fernsehbeitrag einen Beitrag mit der Überschrift "Sprich ordentlich", hier wurde die "naive Begeisterung über latrinäre Wörter der so genannt zweisprachlichen Jugendlichen" kritisiert. In einem anderen Zeitungsbeitrag wurde die Sprache der Jugendlichen als Rattensprache, Kloaken- und Gewaltsprache bezeichnet. Statt dessen wurde die normale dänische Sprache durch die Bezeichnung Lerchengesang als positives Gegenbild hervorgehoben.

Diese Kritik wurde in anderen Zeitungsbeiträgen zurückgewiesen. Sowohl die Doktorandin als auch mehrere Linguisten meldeten sich anschließend zu Wort und hatten als Grundargument, dass der beschriebene Sprachgebrauch der Jugendlichen grammatisch einwandfrei sei, und dass die Aufgabe der Linguistik darin bestehe, zu beschreiben und nicht zu bewerten. Die Linguistik solle keine Regeln diktieren, sondern beschreiben und erklären.

Als Lexikograph möchten wir hier nicht diese Diskussion um die Aufgaben der Linguistik kommentieren, aber deutlich sagen, dass es die Aufgabe der Lexikographie ist, je nach genuinem Zweck eines monolingualen Wörterbuches, 1. Instruktionen für die Textproduktion zu geben oder 2. Beschreibungen vorliegender Texte als Hilfe bei der Textrezeption anzubieten, oder 3. Beschreibungen vorliegender Texte zu bringen, um Wissen über konkrete sprachliche Phänomene zu vermitteln. Man kann dabei diskutieren, inwiefern Beschreibungen ohne Bewertungen durch den Lexikographen machbar seien. Es steht aber außer Frage, dass Instruktionen notwendigerweise mit Bewertungen verbunden sein müssen.

Ein deutsches Beispiel kann das Schimpfwort *Drecksau* sein. Henn/Memmesheimer (2004) beschreibt, dass sowohl Kinder wie Jugendliche *Drecksau* verwenden, wo es nicht unbedingt negativ gemeint ist. Einige Male sogar eher positiv, indem man es von jemandem sagt, den man um Kleidung, Urlaub, Kleider oder ähnliches beneidet, z.B. in einem Lehrerzimmer, wo drei Lehrer Zeugnisse schreiben. Der eine schlägt seine Mappe zu, der andere sagt:

(7) Guck mal, die Drecksau ist schon fertig.

Beate Henn beschreibt anschließend sehr instruktiv die Bedingungen, die gegeben sind, wenn *Drecksau* so verwendet wird. Unsere Frage dazu aus der Sicht der Lexikographie lautet: Sollen wir so bald wie möglich auch die Wörterbücher daraufhin ergänzen? Und wenn ja, wie? Neutral beschreibend oder auch bewertend? Es hängt von verschiedenen Faktoren ab, u.a. von sprachpolitischen Entscheidungen, die für das betreffende Wörterbuch in der betreffenden Kultur gelten oder evt. speziell getroffen wurden. Selbstverständlich sind auffällige oder aufgefallene Schlagwörter, wie sie in Spezialwörterbüchern festgehalten werden, z.B. LEXIKON DER UNWÖRTER oder ZEITGESCHICHTLICHES WÖRTERBUCH, in besonderem Maße Gegenstand der öffentlichen Diskussion und somit auch auffällige Beispiele für sprachpolitische Überlegungen. Hierzu gehören auch Movierungen, wo man auch in benachbarten und sprachtypologisch sehr nahestehenden Sprachen wie Deutsch und Dänisch innerhalb von weniger als 30 Jahren ganz unterschiedliche Wege eingeschlagen hat. Vor 30 Jahren war die Verwendung von Movierungen etwa gleich in beiden Sprachen. Heute gilt es in Deutschland als politisch unkorrekt, das movierte Substantiv zu unterlassen, wenn beide Geschlechter angesprochen werden. In Dänemark gilt es in weiten Bereichen als unpassend, die movierte Form überhaupt zu verwenden. Das Grundargument ist jedoch, dass nicht nur diese auffallenden Beispiele, sondern im Prinzip jede lexikographische Entscheidung ein Teil der Sprachpolitik ist.

## 2 Politik findet nicht nur im Parlament oder in Stadträten statt

Allein die Tatsache, dass Firmen, Institutionen und Organisationen interne Sprachpolitiken festlegen, zeigt, dass sowohl politische als auch sprachpolitische Beschlüsse nicht nur in Parlamenten oder in Stadträten vorgenommen werden können. Dies ist eine generelle Entwicklung, die sich auch bei der Bildung von Komposita widerspiegelt. Im Laufe der zweiten Hälfte des 20. Jahrhunderts wurden viele neue Komposita mit dem Stamm *-politik* gebildet, z.B. *Abrüstungspolitik, Abschiebungspolitik, Agrarpolitik, Aidspolitik, Algerienpolitik*. Diese Wortbildungen beziehen sich auf bestimmte Gebiete im Rahmen von internationalen, nationalen, regionalen oder kommunalen politischen Fragestellungen. Seit den achtziger Jahren des letzten Jahrhunderts hat sich dieser Wortbildungstyp in zunehmendem Maße auf Planungen und Entscheidungen in Firmen, Vereinen, Schulen und andere nicht-politische Gruppierungen durchgesetzt, z.B. *Alkoholpolitik, Ausschüttungspolitik, Behindertenpolitik, Heiratspolitik, Informationspolitik, Kommunikationspolitik, Niedrigpreispolitik, Öffnungspolitik, Personalpolitik, Preispolitik, Sexualpolitik, Spielplanpolitik, Vereinspolitik.*

Eines der Wörter, die einen erweiterten Gebrauchsrahmen erhalten haben, ist *Sprachpolitik*. Es bezieht sich sowohl auf die traditionell politischen, als auch auf nicht-politische Gruppierungen wie Firmen und Konzerne. Es ist ein Modewort geworden, insbesondere in den letzten 5–10 Jahren. Dabei ist festzustellen, dass die verschiedenen Diskussionen, Vorschläge und Beschlüsse in vielen Fällen dadurch gekennzeichnet sind, dass keine genaue Präzisierung des Terms vorgenommen wird. Dies wäre notwendig gewesen, da weder die theoretischen Beiträge, noch vorliegende Wörterbücher nur annähernd ähnliche Definitionen oder Erklärungen bringen.

Wir werden im Folgenden insbesondere die Bedeutungsangaben in Wörterbüchern heranziehen, die Definitionsversuche in den theoretischen sprachpolitischen Beiträgen nur in der eigenen Definition berücksichtigen, vgl. dazu Bergenholtz (2003) und (2004). Es gibt nur vier gemeinsprachliche deutsche Wörterbücher, die das Lemma *Sprachpolitik* anführen. Oder eigentlich sind es nur zwei, da es sich um die drei Ausgaben von Duden-GWB und um Brockhaus-Wahrig handelt. Im Grunde kann man sich darüber wundern, dass man in Großwörterbüchern wie z.B. Wahrig, Duden-Universalwörterbuch oder Bünting/Karatas dieses Wort nicht als Lemma findet. Man kann daraus schließen, dass *Sprachpolitik* nicht zu den 100.000 wichtigsten deutschen Wörtern gehört, wenn man die gute und genaue

Lemmaselektion dieser Wörterbücher voraussetzt. In DUDEN-GWB[1] und fast identisch in BROCKHAUS-WAHRIG findet sich folgende Angabe:

(8) **Sprachpolitik** staatliche Maßnahmen im Hinblick auf eine Sprache

Es handelt sich demnach erstens um staatliche Maßnahmen, d.h. wohl auch um nationale und daher nicht um internationale und auch nicht um nicht-staatliche Sprachregelungen. Es geht zweitens um genau eine und nicht um mehrere Sprachen, sonst hätten die sonst genau vorgehenden Lexikographen wohl "im Hinblick auf eine oder mehrere Sprachen" geschrieben. Ob damit etwa *Sprachlenkung*, *Sprachnormung*, *Sprachpurismus* oder eines der vielen anderen sprachpolitisch relevanten Termini gemeint ist, die man auch in DUDEN-GWB[1] findet, wird nicht ersichtlich. In den folgenden beiden Ausgaben erhält man eine vollkommen andere Auskunft. Wenn eine neue Auflage eines Wörterbuches vorbereitet wird, führen Änderungen oft, aber nicht immer zu Verbesserungen. Im Falle von DUDEN-GWB[2] und DUDEN-GWB[3], die völlig identische Angaben machen, handelt es sich im besten Falle um eine inhaltliche Änderung verbunden mit einer sprachlichen Verschlimmbesserung. Syntaktisch gesehen ist die Erklärung so konstruiert, dass man den Inhalt nur mit sehr gutem Willen erahnen kann:

(9) **Sprachpolitik** die in einem Land gesprochene[n]Sprache[n], die in einem Land sich stellende Sprachenfrage o. ä. betreffende Politik

Wenn man voraussetzt, dass die beiden Teile der Bedeutungserklärung sich ergänzen sollen – wie auch anderswo in diesen Wörterbüchern – erfährt man durch "die in einem Land gesprochene[n] Sprache[n] betreffende Politik", dass es sich um eine nationale politische Angelegenheit handelt, die entweder eine oder mehrere Sprachen umfasst. Dies entspricht der Angabe in DUDEN-GWB[1] mit der Änderung, dass es sich nun auch um mehrere Sprachen und nicht nur um eine Sprache handeln kann. Es geht aber wahrscheinlich immer noch nur um eine nationale Angelegenheit, nicht um internationale und nicht um private Beschlüsse, sonst hätte man wohl "in einem Land, in einer internationalen oder nationalen Organisation sowie in einer privaten oder öffentlichen Firma oder Institution" geschrieben. Was genau nun die Sprachpolitik in Bezug auf diese eine Sprache oder auf mehrere Sprachen beeinflussen will ist nicht klar. Es kann sein, dass es sich um Sprachpurismus oder um Sprachnormung handelt; es kann auch sein, dass es sich um das Verbot anderer Sprachen oder um Vorschriften für die offizielle Verwendung oder Förderung einer oder mehrerer Sprachen geht. Um den zweiten Teil der Bedeutungserklärung "die in einem Land sich stellende Sprachenfrage o.ä. betreffende Politik" verstehen zu können, muss man wissen, was *Sprachenfrage* ist. Dieses Wort wird in DUDEN-GWB[2] und DUDEN-GWB[3] wie folgt erklärt:

(10) **Sprachenfrage** aus dem Zusammenleben mehrerer ethnischer Gruppen mit verschiedenen Sprachen innerhalb eines Staates herrührende Problematik

Der nicht so fleißige Benutzer, der nicht unter *Sprachenfrage* nachschlägt, wird nicht erraten können, dass Sprachpolitik sich insbesondere mit der Beziehung zwischen den Sprachen zu tun hat, die verschiedene ethnische Gruppen in einem Land sprechen. Wir haben es jedenfalls nicht gewusst, wissen aber jetzt, dass die Erklärung in Duden-GWB[1] auch daraufhin ergänzt wurde, dass es sich nicht nur um mehrere Sprachen, sondern um mehrere Sprachen handelt von Angehörigen verschiedener ethnischer Gruppen, wenn man von Sprachpolitik redet. Alles in allem sind wir zwar etwas, aber nicht sehr viel schlauer geworden. Man darf vermuten, dass Sprachpolitik kein so zentrales Thema in Deutschland ist, sondern dass die unklare Erklärung das recht geringe öffentliche Interesse und eine dementsprechend noch unklare Debatte widerspiegelt.

In der Tat ist es so, dass von Frankreich abgesehen in größeren europäischen Ländern wie Deutschland, England, Spanien und Italien keine allgemeinen und breit geführten sprachpolitischen Debatten geführt werden[2]. Dies ist in den kleineren Ländern anders, so in der Schweiz, in Österreich und insbesondere in den skandinavischen Ländern. Man kann nur Vermutungen darüber anstellen, warum das so ist. Wir meinen, dass es dafür mindestens zwei Gründe gibt: Der eine Grund ist, dass die großen Länder, und das heißt zugleich die großen Sprachen, in der EU zunehmend größeren Einfluss gewinnen – auf Kosten der kleineren Länder und somit auch kleineren Sprachen. Diese Länder sehen – auch aus diesem Grund – die Notwendigkeit sprachpolitischer Initiativen, um den vorliegenden Status der sprachlichen Vielfalt, wie es die EU festgelegt hat (Artikel 22 der Charta der Grundrechte der Europäischen Union), beizubehalten bzw. nicht weiter aushöhlen zu lassen. Deutschland unterstützt auf der einen Seite die kleineren und mittelgroßen Länder, möchte aber gleichzeitig für Deutsch als größte Sprache in der EU auch den Status erreichen, den Englisch und Französisch in der Praxis haben. Beide Richtungen gleichzeitig einzuschlagen, wird kaum gelingen können. Der andere Grund hängt mit der kulturellen Tradition in Bezug auf offizielle Sprachregelungen zusammen. Einige Länder, so z.B. Spanien und Frankreich, haben eine lange Tradition für eine prinzipiell positive Haltung zu präskriptiven Maßnahmen, in der Praxis interessiert man sich jedoch mehr für allgemeine Regelungen und nicht so sehr für Einzelheiten und für die Umsetzung von Sanktionen bei Nichteinhalten der Vorschriften. Die Entwicklung in Frankreich geht seit 1990 außerdem in eine mehr deskriptive Richtung mit der Zulassung von neuen orthographischen und flexionsmorphologischen Varianten. Die nordischen Länder, insbesondere Island, die Färöer und Norwegen haben eine vergleichbare Tradition.

Anders als in Südeuropa werden staatliche Spracheingriffe zur Änderung des vorliegenden Sprachgebrauchs hier auch durch viele offizielle Beschlüsse in die Tat umgesetzt und durch öffentliche Kontrollen unterstützt. Außerdem verfolgt die Öffentlichkeit in diesen Ländern aufmerksam die Einhaltung der offiziellen Beschlüsse. Dänemark befindet sich hier in einer Sonderrolle, da es auf der einen Seite eine lange Tradition für eine dänische laissez-faire-Haltung gibt, indem Wissenschaftler und Pädagogen einen klaren deskriptiven Standpunkt eingenommen haben. Gleichzeitig gibt es aber paradoxerweise eine Tradition für offizielle Sprachregelungen, im Prinzip nur für Orthographie und Interpunktion, aber zunehmend auch für Flexion. Kurz gesagt heißt das, dass man zwar die sprachliche Freiheitsfahne sehr hoch hängt, dass aber keiner (von einigen wenigen Journalisten und Lexikographen abgesehen) sich gegen große orthographische, flexionsmorphologischische und interpunktive Änderungen auflehnen – anders als in Deutschland nach der letzten Rechtschreibereform. Dabei gibt es solche sehr weitgehende Änderungen in Dänemark etwa alle fünf Jahre.

Es ist daher nicht überraschend, dass *Sprachpolitik* als Lemma in den beiden einzigen dänischen Großwörterbüchern zu finden ist. In dem neuesten (2003-2005) sechsbändigen Den Danske Ordbog findet man folgende Bedeutungserklärung (unsere Übersetzung, HB/ST)

(11) **Sprachpolitik** Maßnahmen einer Nation, eines Fachgebietes, einer Institution oder einer Firma, die darauf zielen, den Gebrauchsumfang einer oder mehrerer Sprachen im Verhältnis zu anderen Sprachen zu regeln, die Sprache zu entwickeln oder einen korrekten oder empfehlenswerten Gebrauch dieser Sprache zu fördern

Diese etwas umfangreiche Bedeutungsangabe besteht aus zwei Teilen: Im ersten Teil geht es um die Wahl einer oder mehrerer Gebrauchssprachen für ein Land, ein Fachgebiet, eine Institution oder eine Firma. Es ist dabei nicht ganz klar, ob es sich auch um international tätige Institutionen und Firmen handeln kann, wie z.B. die EU, NATO oder McDonald's. Sowohl national wie international hat ein kleines Land wie Dänemark einen wesentlichen Bedarf an Regelungen, da man in vielen Zusammenhängen nicht darauf bestehen kann, nur Dänisch verwenden zu wollen. Der zweite Teil der Bedeutungserklärung "die Sprache zu entwickeln oder einen korrekten oder empfehlenswerten Gebrauch dieser Sprache zu fördern" zielt eindeutig auf einzelsprachliche präskriptive oder proskriptive (empfehlende) Maßnahmen. Dies ganz im Rahmen der präskriptiven Tradition in Skandinavien, auch der dänischen mit einer deskriptiven Grundeinstellung und einer gleichzeitigen präskriptiven Praxis.

Eine inhaltlich ähnliche lexikographische Definition findet sich in Den Danske Netordbog, der sich vor allem von den bisher zitieren Wörterbüchern dadurch unterscheidet, dass hier in drei getrennte Bedeutungen aufgeteilt wird (unsere Übersetzung, HB/ST):

**(12) Sprachpolitik**

**1.** bewusste und festgelegte Wahl, Empfehlung oder Förderung einer oder mehrerer Sprachen
= generelle Sprachpolitik, Sprachwahl, Sprachförderung, Sprachstatusplanung, Statusplanung, Sprachenpolitik
ANM Eine solche Wahl liegt z.B. vor, wenn eine internationale Organisation, ein Land, eine Institution oder eine Firma den Gebrauch einer oder mehreren Sprachen ganz allgemein oder nur für bestimmte Zusammenhänge vorschreibt. In solchen Fällen liegt auch eine Abwahl anderer Sprachen vor.

**2.** bewusste und festgelegte Wahl oder Empfehlung von Kommunikationsstrategien und Stilmitteln bei der Verwendung einer gegebenen Sprache
= spezifische Sprachpolitik, Kommunikationsoptimierung, zielgerichtete Kommunikationsstrategie
ANM Man könnte sich vorstellen, dass eine Sprachenkommission, eine Firma oder eine Institution genau dieselben Kommunikationsstrategien usw. für mehrere Sprachen vorschreiben oder empfehlen würde. In der Praxis lässt sich das nur teilweise durchführen, da man eine Reihe von sprachspezifischen Beschlüssen für jede einzelne Sprache treffen muss.

**3.** bewusste und festgelegte Wahl oder Empfehlung von bestimmten grammatischen Konstruktionen, Wörtern oder Wortformen einer gegebenen Sprache
= phänomenselektierende Sprachpolitik, Sprachkorpusplanung, Korpusplanung, selektierende Sprachpolitik
ANM In der Linguistik und auch in der Lexikographie wird normalerweise zwischen einer deskriptiven (beschreibenden) und präskriptiven (vorschreibenden) Vorgehensweise unterschieden. In der Praxis wird oft eine proskriptive (empfehlende) Vorgehensweise verwendet. Das ist z.B. das Prinzip in diesem Wörterbuch.

Dieser Wörterbucheintrag baut auf Bergenholtz (2003, 2004) sowie Bergenholtz et al. (2003). Hier findet sich ebenfalls eine kritische Würdigung vorliegender Definitionen des Terminus *Sprachpolitik*. Die englische Trennung zwischen Theorie und Praxis durch die Termini *language policy* und *language politics* wird nicht übernommen. Die deutsche Trennung zwischen *Sprachenpolitik* und *Sprachpolitik* wird teilweise berücksichtigt, indem die erste Bedeutung – mit dem Synonym *generelle Sprachpolitik*

– in etwa *Sprachenpolitik* entspricht, und die dritte Bedeutung – mit dem Synonym *phänomenselektierende Sprachpolitik* – in etwa *Sprachpolitik* entspricht.

## 3 Lexikographische Implikationen sprachpolitischer Entscheidungen

Die soeben definierten drei Hauptgebiete der Sprachpolitik haben auf unterschiedliche Weise Relevanz für die theoretische und die praktische Lexikographie.

In dem ersten Gebiet geht es im Prinzip immer um die Wahl zwischen Sprachen, darunter auch um die Unterstützung einer Sprache oder das Verbot oder die Zurückdrängung einer Sprache. In internationalen Organisationen wie UN, NATO, EU und dem Roten Kreuz ist es notwendig, eine oder mehrere Arbeitssprachen, Verhandlungssprachen, Dokumentationssprachen zu wählen. Weiterhin kann auf nationaler oder regionaler Ebene darüber entschieden werden, welche Sprache(n) in offiziellen Texten zugelassen sind, in welcher oder welchen Sprachen der Unterricht auf verschiedener Ebene stattfinden darf oder muss. Entsprechend wird entschieden, welche Sprachen als Schulsprachen angeboten werden sollen. Solche Entscheidungen sind insbesondere in Ländern mit mehreren offiziellen Sprachen politisch wichtig und sind manchmal umstritten. Aber im Prinzip finden sich solche Verordnungen oder Bestimmungen in Gesetzen in jedem Land. Im Prinzip sind dieselben Entscheidungen zu treffen, wenn eine Sprache mehrere Varianten hat. Die Förderung oder das Verbot einer oder mehreren Sprachen findet als Verteidigung einer oder mehrerer Nationalsprachen statt. Hiermit ist die Wahl einer konservierenden oder puristischen Sprachpolitik gemeint, die als Ziel hat, die Sprache gegen Einfluss von anderen Sprachen zu schützen. Es kann auf entsprechende Gesetze in z.B. Polen und Frankreich verwiesen werden. Auch in Skandinavien gibt es solche Gesetze, am deutlichsten auf Island und auf den Färöern, aber auch in einigem Ausmaß in Norwegen. Damit verbunden ist oft die Unterstützung für die Erlernung und den Gebrauch einer Sprache im Ausland. Dies gilt insbesondere für die großen Länder, vgl. z.B. Alliance Française, das Goethe Institut oder the English Language Council. Dazu gehört auch die Entsendung von Lehrkräften für Schulen und Universitäten. Auch die Subvention von Wörterbüchern mit der oder den bevorzugten Sprachen kann hierzu gehören.

Insgesamt gilt, dass generelle sprachpolitische Entscheidungen meist in Form von zu befolgenden Vorschriften beschlossen worden sind. Lexikographisch gesehen haben die verschiedenen Beschlüsse wesentliche Folgen für die Lexikographie. Wir haben schon die Subvention von Wörterbüchern erwähnt. Auf nationaler Ebene gilt, dass die Wahl von Sprache(n), in der oder denen unterrichtet wird bzw. die als Unterrichtssprachen angeboten werden, den Bedarf an monolingualen und bilingualen Wörterbüchern in hohem Maße steuert. Für organisationsund konzernspezifische Wörterbücher gilt entsprechend, dass ihre Sprachenwahl entscheidet, für welche Sprache(n) die Organisation, ein Fachgebiet oder der Konzern Wörterbücher ausarbeiten lässt. Von landespolitischen Entscheidungen kann auch abhängen, welche Sprache oder Sprachen in den Umtexten und als Erklärungssprache verwendet wird. In den meisten Fällen wird dies jedoch eine Folge der Wörterbuchkonzeption sein; man kann es entsprechend der breiten Verwendung von Komposita mit *-politik* auch als Sprachpolitik im Sinne einer generellen Sprachpolitik für das Wörterbuch bezeichnen.

In dem zweiten Gebiet sind generelle Vorschläge gemeint, die darauf zielen, dass bessere Texte in einer gegebenen Sprache entstehen sollen, insbesondere geschriebene Texte. Man kann auch von Kommunikationsstrategien sprechen. Vorschläge hierfür können ausgearbeitet werden von Organisationen, Konzernen, Firmen, aber auch internationalen Organisationen wie der EU. Es geht um empfängerrelatierte Ratschläge, die für den Textaufbau, die Wahl und Abwahl der Stilmittel, Satzlänge, Passivsätze, Fremdwörter, Komposita, Anreden u.v.m. gelten. Eng verbunden mit Vorgaben für den Textaufbau sind Vorschläge zum Layout bzw. zum Webdesign, wir können hier von visuellen Kommunikationsstrategien sprechen. Im Prinzip sind alle Vorschläge zu Kommunikatonssstrategien vergleichbar mit denen, die man in Stillehren oder Briefstellern findet. Der Unterschied besteht darin, dass es hier um Vorschläge von der Leitung einer Organisation geht. Es sind zwar nur Vorschläge, weil man hier keine ganz eindeutigen Vorschriften machen kann. Die Leitung erwartet jedoch, dass die Mitarbeiter sich an die Vorschriften halten, so dass die Vorschläge eine vorschreibende Funktion bekommen.

Im Prinzip können Wörterbücher auch eine solche Sprachpolitik festlegen, insbesondere, wenn das Wörterbuch einen entsprechenden Umtext mit Ratschlägen für das Schreiben von Briefen usw. enthält. Dies findet man insbesondere in Wörterbüchern zur Geschäftssprache, aber auch in Reisewörterbüchern. Hier liegt ein wichtiges Gebiet zur notwendigen Verbesserung von Wörterbüchern vor, einige Fortschritte sind bei der Einarbeitung von Routineformeln in bilingualen Wörterbüchern festzustellen.

In dem dritten Gebiet geht es um die Wahl von konkreten sprachlichen Phänomenen, man kann auch von Kommunikationsoptimierung sprechen. Es geht hier um Empfehlungen oder aber normierende Beschlüsse auf nationaler oder regionaler Ebene oder in einer nationalen oder internationalen Organisation oder Firma für eine gegebene Sprache. Dies kann z.B. stattfinden durch Beschlüsse eines national ernannten Sprachrates oder durch ein Gremium zusammengesetzt von Repräsentanten mehrerer Länder, wie es im Falle der deutschen Rechtschreibereform geschehen ist. Entsprechend haben einzelne Firmen oder Organisationen konkrete Beschlüsse zu Einzelproblemen gefasst. Dies geschieht z.B. häufig in

Zeitungsredaktionen. In diesem Zusammenhang ist auch die internationale, nationale und firmeninterne terminologische Normierung zu sehen.

Diese Art von Sprachpolitik ist die, die am eindeutigsten für die Lexikographie wichtig ist. Auch Firmen, Zeitungsredaktionen usw. fertigen oft ihr eigenes Wörterbuch für solche Entscheidungen aus. Bekannter und allgemein wichtiger für die Lexikographie sind jedoch die offiziellen oder halboffiziellen Beschlüsse, die für eine ganze Sprachgemeinschaft gelten sollen. So in Dänemark für die Orthographie, wofür es ein eigenes Sprachgesetz gibt. Der Sprachrat hat dieses Gesetz auf die Flexionsmorphologie ausgedehnt. In anderen Ländern gibt es entsprechende Beschlüsse für die Bedeutung (wie wir es für die fachsprachliche Normung kennen). So hat man in Norwegen den Ausdruck *neger* (dt. *Neger*) offiziell abgelehnt, man soll *ein Farbiger* sagen. Entsprechend werden die Angaben in den offiziellen Wörterbüchern gemacht, in Dänemark richten sich die meisten kommerziellen Wörterbücher danach.

Man lässt in vielen Fällen mehrere Varianten zu, zwischen denen die Benutzer dann selber wählen müssen. Wir haben bereits folgendes Beispiel aus dem offiziellen Wörterbuch angesprochen:

(3) **cirkus** en oder et
(dt.: *Zirkus* + grammatische Angabe: Gemeinschaftsgenus oder Neutrum)

Der Wörterbucheintrag (3) zeigt nur die Variantenmöglichkeit für die Genusverwendung. Es werden weitere, insgesamt drei Paare mit Flexionsvarianten angeführt:

(13) **cirkus** en/et; Singular bestimmt: -sen oder -set, Pluralis unbestimmt: – oder -ser, Plural bestimt -sene oder -serne

Nicht erwähnt, und das heißt auch nicht erlaubt, sind drei zusätzliche Varianten ohne Konsonantenverdoppelung. Das sind jedoch die Varianten, die in konkreten Texten am häufigsten verwendet werden. Es wird somit nicht erwähnt, dass es diese Varianten gibt, und dass sie verboten sind. In anderen Fällen lässt man, wie anfangs dargestellt, nur eine Variante zu (in dem Fall mit der "dänischen" Flexion); die einzig verwendete Pluralvariante für dän. *hooligan* (dt. *Hooligan*) wird nicht erwähnt, d.h. sie ist verboten. Erlaubt wird:

(2) **hooligan** Singular: -en, Pluralis: -er
(dt.: *Hooligan* + Angabe des Morphems für Singular und Plural)

In allen Texten wird aber eine andere Pluralform verwendet: *hooligans*. Da diese Variante gar nicht erwähnt wird, wird der Benutzer, der die Flexion erfahren will, gar nicht auf den Sprachüblichkeitswiderspruch hingewiesen. Die sprachpolitische Entscheidung für "dänische" Flexionen ist problematisch, da niemand anscheinend nachschlägt. Im Internet gibt es mehr als 1200 Belege mit *hooligans* nur 4 mit *hooliganer* (alle vom Dänischen Sprachrat). Problematischer ist die Verwirrangabe, dass der Benutzer problemlos zwischen mehreren Varianten wählen kann, und dass beide gleich gut sind (vgl. dt. *Jahr*, wo alle Wörterbücher freie Wahl zwischen *Jahrs* und *Jahres* geben). Es ist eine problematische sprachpolitische Entscheidung, mehrere Varianten zur freien Auswahl anzubieten, ohne dem Benutzer einen Ratschlag zur Wahl oder zur Vermeidung einer bestimmten Variante zu geben. Dies gilt in gleicher Weise für grammatische und für semantische Varianten, insbesondere, wenn eine der Varianten sehr selten ist, so die eine Bedeutung von *bjørnetjeneste* (dt. *Bärendienst)* oder *racist* (dt.: *Rassist)* oder die Flexion von *cirkus* (dt.: *Zirkus),* oder wenn die vorgeschriebene Variante so gut wie nie verwendet wird (*hooligan*, dt. *Hooligan*). Natürlich ist die beschriebene lexikographische Praxis in dem dänischen offiziellen Rechtschreibewörterbuch äußerst ungenau, es liegt keine Beschreibung der Sprache vor. Es wird durch die Nicht-Erwähnung von Varianten nur implizit verboten, es werden keine klaren Ratschläge gegeben, wenn mehrere Varianten zugelassen werden. Es geht u.E. nicht darum, ob man nur beschreiben oder nur verbieten oder nur Ratschläge erteilen soll. Man soll beschreiben, evt. verbieten, und in jedem Fall auch Ratschläge explizit erteilen. Dies ist zu tun in Abhängkeit von dem vorgesehenen genuinen Zweck des Wörterbuches.

## 4 Wörterbuchfunktionen[3]

Wenn Lexikographen ein neues Wörterbuch planen, sollten sie immer die vorgesehene **Benutzergruppe** festlegen. Eine solche Gruppe kann homogen oder heterogen sein. In allen Fällen wird man eine Reihe von charakteristischen Benutzervoraussetzungen berücksichtigen müssen. Dabei werden nicht alle im Folgenden erwähnten Merkmale für jedes Wörterbuch eine Rolle spielen. Zunächst gibt es die Sprachkompetenz des Benutzers (diese Kurzform im Folgenden für 'vorgesehenen Benutzertyp'):
- Welche Muttersprache hat der Benutzer?
- Welche Kompetenzen hat der Benutzer in seiner Muttersprache, mündlich und schriftlich?

Das sind Grundfragen für jedes Wörterbuch. Für viele monolinguale und alle bi- und polylingualen Wörterbücher gilt außerdem, dass die Kompetenz in der oder den Fremdsprache(n) zu berücksichtigen ist. Es ist in vielen Fällen auch relevant, das generelle kulturelle und enzyklopädische Wissen des Benutzers zu kennen. Dies gilt nicht nur, aber insbesondere für Fachwörterbücher, wo man zwischen lexikographischen Nachschlagewerken für Laien, Semiexperten und Experten trennen kann. Es geht somit neben den beiden genannten Fragen um folgende **Merkmale des Benutzers**:

- In welchem Ausmaß beherrscht der Benutzer die jeweilige Fremdsprache?
- Über welche generellen enzyklopädischen und kulturellen Kenntnisse verfügt der Benutzer?
- In welchem Ausmaß hat der Benutzer Kenntnisse des spezifischen Fachgebiets?
- In welchem Ausmaß beherrscht der muttersprachliche Benutzer die spezifische Fachsprache?
- In welchem Ausmaß beherrscht der fremdsprachliche Benutzer die spezifische Fachsprache?

Der konkrete Benutzer hat in konkreten Situationen konkrete Bedürfnisse. Für ein neues Wörterbuch gilt es, solche Situationen zu analysieren und in Typen von Situationen zu ordnen. Solche Benutzersituationen sind nicht mit Typen von Wörterbuchbenutzungssituationen gleichzusetzen, da nicht nur ein Wörterbuch als Hilfsmittel dienen kann, wenn eine Person Probleme mit Texten oder Wissenslücken feststellt. Der Benutzer kann auch andere Bücher konsultieren, im Internet suchen, einen Freund fragen usw. Wir beschäftigen uns hier allerdings nur mit dem Fall, wo der Benutzer seine Hilfe in einem Wörterbuch sucht.

Die moderne lexikographische Funktionslehre trennt zwischen zwei Grundtypen von Benutzersituationen. Zum einen geht es um Situationen, in denen der Benutzer (aus welchen Gründen auch immer) es wünscht, Informationen über etwas zu erhalten. Es kann sich um generelle kulturelle oder enzyklopädische Zusammenhänge, um spezifische Informationen zu einer wissenschaftlichen Disziplin (Biologie, Jura usw.) oder um Informationen über eine Sprache handeln, um diese Sprache besser zu lernen. Es ist die Aufgabe des Lexikographen, diese Bedürfnisse zu analysieren und zu entscheiden, welche dieser Bedürfnisse von einem Wörterbuch befriedigt und wie die dafür nötigen Daten angegeben werden können. Es geht bei diesem ersten Grundtyp um wissensbezogene Situationen. In solchen Situationen – unter der Voraussetzung, dass ein Wörterbuch als Hilfe herangezogen wird – findet eine Kommunikation zwischen dem Lexikographen und dem Wörterbuchbenutzer statt. Der Benutzer möchte Informationen haben, der Lexikograph hat Angaben im Wörterbuch angebracht, um diese Hilfe zu ermöglichen.

Bei dem anderen Grundtyp von Benutzersituationen geht es um eine schriftliche oder mündliche Kommunikation zwischen zwei oder mehreren Personen, bei der der Lexikograph nur indirekt teilnimmt, indem sein Wörterbuch konsultiert wird, um ein entstandenes Kommunikationsproblem zu lösen. Dies nennen wir kommunikationsbezogene Benutzersituationen. In einem einfachen Kommunikationsmodell können wir zwischen muttersprachlichem und fremdsprachlichem

Sprachbenutzer ($L_1$ und $L_2$), Textproduktion, Textrezeption und Übersetzung trennen. Das Wörterbuch kann immer dann, wenn in der Textproduktion, Textrezeption oder Übersetzung ein Problem entsteht, herangezogen werden. Insgesamt ergeben sich sechs grundlegende **kommunikationsbezogene Benutzersituationen**, in denen ein Wörterbuch herangezogen werden kann:

- Textproduktion in der Muttersprache
- Textrezeption in der Muttersprache
- Textproduktion in einer Fremdsprache
- Textrezeption in einer Fremdsprache
- Übersetzung eines Textes aus der Muttersprache in eine Fremdsprache
- Übersetzung eines Textes aus einer Fremdsprache in die Muttersprache

Wenn der Lexikograph die vorgesehene Benutzergruppe und die Benutzersituationen festgelegt hat, gilt es schließlich, die Merkmale der **Benutzerbedürfnisse** einzugrenzen. Diese Festlegung sollte man jedoch nicht – wie es manche Forscher und auch Verlage tun – durch die Verteilung von Fragebögen mit Fragen nach bisherigen Wörterbuchbenutzungserfahrungen einzugrenzen versuchen. Dafür gibt es viele Gründe. U.a. kann man keineswegs sicher sein, dass die gegebenen Antworten sich mit dem tatsächlichen Gebrauch von Wörterbüchern decken. Weit wichtiger ist es jedoch, dass der Benutzer nur solche Informationen in den Wörterbüchern sucht, von denen er aus Erfahrung weiß, dass das Wörterbuch eine Lösung bietet. Selbst wenn wir bereits perfekte Wörterbücher hätten, ergeben solche Befragungen bloß ein Bild der guten und schlechten Benutzerangewohnheiten, nicht ein Bild der entstandenen wissensbezogenen oder kommunikationsbezogenen Probleme.

Nachdem diese Festlegungen erfolgt sind, kann der Lexikograph daran gehen, die so genannten Wörterbuchfunktionen festzulegen. Die lexikographische Funktion eines gegebenen Wörterbuches ist es, Hilfe zu leisten für eine spezifische Benutzergruppe mit spezifischen Merkmalen für eine Menge von Fragebedürfnissen, die in spezifischen Problemsituationen entstehen können. Ein konkretes Wörterbuch kann eine oder mehr als eine Funktion haben, d.h. mono- oder polyfunktional sein. Wie alle anderen Werkzeuge haben auch Wörterbücher einen genuinen Zweck. Dieser genuine Zweck besteht in der Summe der vorgesehenen Funktionen eines gegebenen Wörterbuches und in dem Gegenstandsbereich, den das Wörterbuch decken wird (z.B. Gentechnologie, moderne finnische Allgemeinsprache, deutsch-finnische Idiome usw.).

Die vorgesehene Wörterbuchfunktion bzw. vorgesehenen Wörterbuchfunktionen sollten Ausgangspunkt für alle anderen lexikographischen Entscheidungen sein. Sie können in kommunikationsbezogenen und wissensbezogenen Funktionen aufgeteilt werden. Die **wichtigsten kommunikationsbezogenen Funktionen** sind die folgenden:

Das Wörterbuch soll Hilfe leisten, wenn

- der Benutzer ein Problem bei der Textrezeption in der Muttersprache lösen will
- der Benutzer ein Problem bei der Textproduktion in der Muttersprache lösen will
- der Benutzer ein Problem bei der Textrezeption in einer Fremdsprache lösen will
- der Benutzer ein Problem bei der Textproduktion in einer Fremdsprache lösen will
- der Benutzer ein Problem bei der Übersetzung von der Muttersprache in eine Fremdsprache lösen will
- der Benutzer ein Problem bei der Übersetzung von einer Fremdsprache in die Muttersprache lösen will

Die **wichtigsten wissensbezogenen Funktionen** sind:
Das Wörterbuch soll Hilfe leisten, wenn

- der Benutzer allgemeine kulturelle oder enzyklopädische Informationen haben möchte
- der Benutzer spezielle Informationen zu einem bestimmten Sachgebiet haben möchte
- der Benutzer Informationen über eine Sprache oder mehrere Sprachen haben möchte

Die wissensbezogenen Funktionen können weit detaillierter aufgeteilt werden, siehe hierzu Beispiele für Sprachwissenschaft und Lexikographie (Bergenholtz 1997), für das Lernen der Muttersprache und einer Fremdsprache (Tarp 2004) und prinzipielle Überlegungen zu mehr als 200 Funktionen (Tarp 1992). Das Grundverständnis der Wörterbuchfunktionen ist somit ziemlich klar definiert, Wörterbuchfunktionen sollten wie gesagt Ausgangspunkt bilden für sämtliche andere lexikographische Entscheidungen. Es ist daher bedauerlich, dass der deutsche Theoretiker H.E. Wiegand viele Jahre nach der ersten Beschreibung dieser Zusammenhänge (Tarp 1992) eine konkurrierende Theorie vorgestellt hat (Wiegand 2001), in der er dieselben orthographischen Ausdrücke verwendet, jedoch mit ganz anderem Inhalt, teilweise mit wenig sinnvollen Unterscheidungen. U.a. verwendet er **genuiner Zweck** genau umgekehrt im Vergleich zur vorher vorliegenden Theorie, indem dieser Terminus bei Wiegand dem Terminus **Funktion** untergeordnet ist.

## 5 Vorschläge für die lexikographische Praxis

Die folgenden Einträge, die dem dänischen Internetwörterbuch entnommen sind, haben dänische Muttersprachler als Benutzergruppe und Textproduktion als Benutzersituation, auch in der besonderen Situation der Korrektur (im Verlag, bei einer Zeitung oder durch einen Lehrer) sind die Einträge so gestaltet, dass sie die Bedürfnisse der Benutzer decken können. Die Wörterbucheinträge können im Prinzip auch für dieselbe Benutzergruppe und mit Textrezeption als Benutzersituation verwendet werden. Für einen Benutzer, der im Wörterbuch Wissen über spezielle oder allgemeine kulturelle, enzyklopädische oder sprachliche Phänomene oder Elemente haben möchte, sind die Einträge nicht optimal, da sie viel ausführlicher hätten sein sollen, z.B. der Eintrag *luder* (dt. *Hure*) so umfangreich wie der fünf Seiten lange Eintrag in dem Zeitgeschichtlichen Wörterbuch, aber auch für eine solche Wörterbuchfunktion wären evt. sprachpolitische Entscheidungen oder Gesichtspunkte zu erwähnen.

Es ist in dem Wörterbuch, aus dem die Einträge entnommen sind, die Entscheidung getroffen, sich nicht nach den offiziellen Beschlüssen zu richten, dies sowieso nicht bei semantischen Auskünften, wo der Sprachrat nur Ratschläge erteilen kann (alle Einträge nur in deutscher Übersetzung):

(14) **Bärendienst**
1 gut gemeinte Handlung, die mehr schadet als hilft
2 Handlung, die eine große Hilfe ist für jemanden
ANM Diese Bedeutung, die seit 1992 belegt ist, kann zu Missverständnissen führen. Dieses Wörterbuch rät daher von diesem Gebrauch ab.

Wenn es eine offizielle Regelung gibt, wird über die Entscheidung des Sprachrats informiert, wenn dieser mehr als die empfohlene Variante zulässt oder wenn das Wörterbuch eine andere Entscheidung trifft. Wichtig für Benutzerbedürfnisse für Hilfe bei der Textproduktion ist es, dass immer klare Ratschläge erteilt werden:

(15) **cirkus** substantiv <et; -set, -ser, -serne>
ANM Der dänische Sprachrat erlaubt auch Plural unbestimmt *cirkus* und Plural bestimmt *cirkussene* und auch Gemeinschaftsgenus: *en cirkus*, aber nicht die allgemein verwendeten Formen mit nur einem -s: *cirkuset, cirkuser, cirkuserne* und *cirkusen*.

Im dem Fall mit der Verwendung von *luder* (dt. *Hure*) hat der Sprachrat keine Entscheidung getroffen. Man kann vermuten, dass der Rat analog zu dem Beispiel *Rassist* die Verwendung als ganz üblich einstufen wird. Das Wörterbuch rät von dieser Verwendung ab. Ob man dies als politische, sprachpolitische oder lexikographische Entscheidung einstuft, hängt davon ab, in welchem Kontext diese Entscheidung zur Sprache kommt.

(16) **luder (dt. Hure)**
= Prostituierte
ANM *Luder* wird nicht nur zur Bezeichnung von Prostituierten, sondern auch allgemein als abschätzige Bezeichnung von Frauen, insbesondere jungen Mädchen verwendet. Es wird dringend davon abgeraten, *luder* als Schimpfwort oder als Bezeichnung von Mädchen oder Frauen zu verwenden. Auch dieser Sprachgebrauch beinhaltet die Bedeutung 'Prostituierte' und kann daher stark beleidigend wirken.

Im folgenden Fall hatte sich der Sprachrat dafür entschieden, dass das Wort *racist* (dt. *Rassist*) nicht nur jemand ist, der Wertunterschiede zwischen Rassen und die eigene Rasse als die qualitativ bessere annimmt. Das Wort kann und wird, so der Sprachrat, auch in der Bedeutung 'jemand, der sich kritisch zu einer bestimmmten Gruppe, auch derselben Rasse oder Nationalität' verwendet. Wir halten diese Entscheidung für politisch, sprachpolitisch und auch lexikographisch falsch, da das Wörterbuch als Werkzeug für dänische Muttersprachler bei Textproduktionsproblemen konzipiert ist und die gewählte Bedeutungsangabe sowohl der historischen und sprachlichen Tradition als auch der großen Mehrheit der heutigen Verwendungen entspricht:

(17) **racist** (dt.: Rassist) jemand, der prinzipielle Wertunterschiede zwischen Rassen annimmt
ANM In einem Gerichtsprozess hat das Höchste Gericht festgestellt, dass es in Dänemark nicht strafbar ist, jemanden als Rassisten oder rassistisch zu bezeichnen oder den Betreffenden dafür zu beschuldigen. In der Urteilsbegründung wird auf eine Auskunft vom Dänischen Sprachrat hingewiesen. Der Sprachrat stellt fest, dass *racist* heute nicht nur zur Bezeichnung von Personen verwendet wird, die die Überlegenheit einer bestimmten Rasse annimmt, sondern auch etwas ungenauer von Personen gesagt wird, die eine Geringschätzung einer anderen Gruppe von Menschen zum Ausdruck bringen. Dieses Wörterbuch kann bestätigen, dass es einen solchen Sprachgebrauch gibt. Er wird jedoch nicht als eigene Bedeutung angeführt. Es ist bekannt, dass sich jede Sprache wandelt. Es ist auch nötig, dass ein Wörterbuch diese Wandlungen erfassen muss. Wir meinen jedoch, dass die große Mehrzahl der Dänen immer noch die Wörter *racist*, *racistisk*, *racisme* mit dem Inhalt 'Hass auf andere Rassen' verbindet. Es wird von dieser großen Mehrzahl von dänischen Sprechern als

stark beleidigend aufgefasst, wenn jemand als Rassist bezeichnet wird. Wir raten daher stark von diesem Sprachgebrauch ab.

## Noten

1 Man spricht auch von der präskriptiven Kraft deskriptiver Aussagen. Wenn viele Lexikographen diese Tatsache übersehen, ist dies dadurch zu erklären, dass sie in Wirklichkeit als Linguisten und nicht als Lexikographen argumentieren. Die große Mehrheit von Linguisten sieht es als ihre ausschließliche Aufgabe, den Sprachgebrauch zu beschreiben. Ein Lexikograph muss auch beurteilen, um dem fragenden Benutzer gute und klare Ratschläge geben zu können.

2 Dies heißt keineswegs, dass gar keine Debatte in Deutschland stattfindet. Die Gründung eines deutschen Sprachrates, in dem u.a. das Institut für Deutsche Sprache in Mannheim, die Gesellschaft für deutsche Sprache, das Goethe-Institut und Inter Nationes vertreten sind, ist ein Zeichen dafür. Über weitere Initiativen informiert IdS-Mannheim (2003). Der entscheidende Unterschied besteht in der Intensität und der Breite der öffentlichen Diskussion z.B. in Dänemark, Island und Norwegen im Vergleich zu der in Deutschland.

3 Dieses Kapitel geht in hohem Maße von Bergenholtz/Tarp (2005) aus.

## Literatur

Bergenholtz, Henning. 1997. „Linguistische und lexikographische Fachwörterbücher in Skandinavien“. *Lexicographica 13*, 3–16.

Bergenholtz, Henning. 2003. “Bryder Dansk Sprognævn den danske sproglov? Sprogpolitik i teori og praksis”. In *Från Närpesdialekt till EU-svenska. Festskrift till Kristina Nikula*, udg. af Harry Lönnroth. Tammerfors: Tampere University Press, 17–31.

Bergenholtz, Henning. 2004. „Sprachpolitik: Also: Deutsch sprechen Deutsche in deutschen Landen. Aber warum schreibt man manchmal Deutsch und manchmal deutsch?“ In *EliS_e, Jahrgang 4, Heft 1, Juli 2004 – gewidmet K.-D. Bünting zum 65sten Geburtstag.* Essen 2004, 43–54. www.elise.uni-essen.de/elise/elise_04_01/el_04_01_bergenholtz.pdf

Bergenholtz, Henning; Jonna Bisgaard, Majken Brunsborg und Kamilla Kvist Kwichmann. 2003. “Sprogpolitik: So ein Ding müssen wir auch haben”. In *Hermes 31,* 135–165.

Bergenholtz, Henning und Rufus Gouws. 2006. How to do policy with dictionaries. *Lexikos 16,* 13–45.

Bergenholtz, Henning und Sven Tarp. 2005. “Wörterbuchfunktionen”. In *Schreiben, Verstehen, Übersetzen und Lernen: Zu ein- und zweisprachigen Wörterbüchern mit Deutsch,* hrsg. von Irmhild Barz/Henning Bergenholtz/Jarmo Korhonen. Frankfurt a.M./Bern/New York/ Paris: Peter Lang, 11–25.

Bergenholtz, Henning und Uwe Kaufmann. 1996. “Enzyklopädische Informationen in Wörterbüchern”. In *Semantik, Lexikographie und Computeranwendungen.* Hrsg. von Nico Weber. Tübingen: Niemeyer, 168–182.

Brockhaus-Wahrig 1980–1984 = *Brockhaus-Wahrig. Deutsches Wörterbuch in sechs Bänden.* Hrsg. von Gerhard Wahrig, Hildegard Krämer, Harald Zimmermann. Wiesbaden: Brockhaus/Stuttgart: Deutsche Verlagsanstalt.

Busse, Ulrich. 2000. "Recent English Learners' Dictionaries and their Treatment of Political Correctness". In *Symposium on Lexicography IX. Proceedings of the Ninth International Symposium on Lexicography April 23–25, 1998 at the University of Copenhagen,* ed. by Jens Erik Mogensen, Viggo Hjørnager Pedersen and Arne Zettersten. Tübingen: Niemeyer, 165–201.

Bünting/Karatas 1996 = Karl-Dieter Bünting und Ramona Karatas: *Deutsches Wörterbuch. Mit der neuen Rechtschreibung.* Chur: Isis.

Den Danske Netordbog 2004 = Henning Bergenholtz og Vibeke Vrang under medvirken af Lena Lund, Helle Grønborg, Maria Bruun Jensen, Signe Rixen Larsen, Rikke Refslund og Jette Pedersen: *Den Danske Netordbog.* Database og layout: Richard Almind. http://netordbog.asb.dk.

Den Danske Ordbog 2003-2005 = *Den Danske Ordbog.* København: Det Danske Sprog- og Litteraturselskab/Gyldendal.

Duden-GWD[1] 1976–1981 = *Duden. Das große Wörterbuch der deutschen Sprache in sechs Bänden.* Hrsg. u. bearb. vom Wissenschaftlichen Rat und den Mitarbeitern der Dudenredaktion unter Leitung von Günther Drosdowski. Mannheim/Wien/Zürich: Bibliographisches Institut.

Duden-GWD[2] 1993–1995 = *Duden. Das große Wörterbuch der deutschen Sprache in acht Bänden.* 2., völlig neu bearb. und stark erw. Aufl. Hrsg. u. bearb. vom Wissenschaftlichen Rat und den Mitarbeitern der Dudenredaktion unter Leitung von Günther Drosdowski. Mannheim/Wien/Zürich: Bibliographisches Institut.

Duden-GWD[3] 1999 = *Duden. Das große Wörterbuch der deutschen Sprache in zehn Bänden.* 3., völlig neu bearb. und erw. Aufl. Hrsg. vom Wissenschaftlichen Rat der Dudenredaktion. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.

Duden-Universalwörterbuch = *Duden. Deutsches Universalwörterbuch.* Hrsg. von der Dudenredaktion. 5. überarb. Aufl. Mannheim/Leipzig/Wien/Zürich: Dudenverlag 2003.

Henn-Memmesheimer, Beate. 2004. "Handlungsspielräume im sprachlichen Variationsfeld". In *Der Deutschunterricht* 1/2004, 26–40.

Hjorth, Ebba. 2003. "Yes". *Mål og Mæle 4,* 5–9.

IdS-Mannheim 2003: *Sprachpolitik*. www.ids-mannheim.de/aktuell/sprachpolitik.

Lexikon der Unwörter 2000 = Horst Dieter Schlosser: Lexikon der Unwörter. o.O.: Bertelsmann Lexikon Verlag.

Retskrivningsordbogen 2001 = Dansk Sprognævn: *Retskrivningsordbogen*. 3. udgave. København: Aschehoug.

Tarp, Sven. 1993. *Prolegomena til teknisk ordbog*. Ph.D. dissertation. Aarhus: Spanish Institute. Aarhus School of Business.

Tarp, Sven. 2004. "Reflections on dictionaries designed to assist the users with text production in a foreign language". *Lexikos 14,* 222–252.

Wahrig 2000 = Gerhard Wahrig: *Deutsches Wörterbuch*. Neu hrsg. von Dr. Renate Wahrig-Burfeind. Mit einem "Lexikon der deutschen Sprachlehre". Gütersloh/München: Bertelsmann.

Wiegand, Herbert Ernst. 2001. “Was eigentlich sind Wörterbuchfunktionen? Kritische Anmerkungen zur neueren und neuesten Wörterbuchforschung”. In *Lexicographica 20*, 217–248.

Zeitgeschichtliches Wörterbuch 2003 = Stötzel, Georg und Thorsten Eitz unter Mitarbeit von Astrid Jährling-Marienfeld, Lea Plate u.a. *Zeitgeschichtliches Wörterbuch der deutschen Gegenwartssprache.* 2. erweiterte und aktualisierte Auflage. Hildesheim/Zürich/New York.

CHAPTER 16

# Revising German Etymologies in the *Oxford English Dictionary*

Andreas Gröger

The present paper shows in which respects the etymologies of English loanwords from German (including the many English words formed after German models) have been improved in the revised third edition of the *Oxford English Dictionary* through a critical overview of relevant etymological resources, and provides a comparison of selected revised etymologies with their precedents in the second edition.

## 1 Introduction

In the course of the revision of the *Oxford English Dictionary*, all etymologies are being given a thorough overhaul, among them those of German loanwords and English words formed on German models. The number of German loanwords in English might at first seem paltry compared to the huge contribution of French and Latin, yet it is by no means negligible.

The published range of the third edition of the *Oxford English Dictionary* (hereafter referred to as *OED3*) extended from M to OLLYCROCK in March 2004. It contains ca. 450 loanwords from early modern and modern German and ca. 530 words formed on German models. By way of comparison, the same range contains over 2000 French loanwords from all periods of French, plus at least 600 words formed on modern French alone, or 40 loanwords from Dutch.

After a survey of the lexicographical resources on which *OED3*'s treatment of German loanwords draws, the present paper will compare selected entries from the published range of *OED3* with their predecessors in *OED2* in order to show in what respects their etymologies have changed.

It is not possible within the scope of this paper to give background information about the general principles and methods of the revision of *OED* as a whole; for these see the *Preface to the Third Edition* (Simpson 2000) and Gilliver (2005).[1] For an overview of the principles and methods of etymological revision in *OED3*, profusely illustrated with draft entries, see Durkin (1999). Finally, a recent short description of *OED3* in 24 key points (in German) is found in Durkin/Thier (2003).

## 2 Resources available for work on German etymologies at *OED3*

The most important scholarly resources in our work on German etymologies are, of course, the historical and etymological dictionaries of German. A synopsis of these is given in Table 1.

**Table 1.** Resources for work on German etymologies[2]

| Dictionary | Language stage |
|---|---|
| **Historical dictionaries** | |
| *DWB* | (in general) early modern German to 19th cent.[3] |
| *DWB* revised (A-F) [4] | as in *DWB*, but extended to 20th cent. |
| *DGWB* | modern German (19th/20th cent.) |
| **Period dictionaries** | |
| Paul | chiefly 18th cent. to present, especially strong on Goethe |
| *Ahd. Wb.* | Old High German |
| Schützeichel | Old High German, excluding glosses |
| Starck/Wells | Old High German, glosses only |
| Schade | OHG, MHG, Old Saxon |
| Lexer | Middle High German |
| *BMZ* | Middle High German |
| *FWB* | early modern (High) German |
| Lasch/Borchling | Middle Low German |
| Schiller/Lübben | Middle Low German |
| **Etymological dictionaries** | |
| Kluge/Seebold | N/A |
| Pfeifer | N/A |

It is a principle of our etymological work that all foreign-language forms quoted in *OED3* are not just gleaned from etymological dictionaries covering several languages, but always verified in dictionaries of the respective language which meet scholarly standards.

If necessary, forms are further checked against scholarly editions and text corpora, which of course may also serve as a source of foreign-language information on their own. As yet there seem to be no large-scale on-line historical corpora of

German, in sharp contrast to the excellent ones for Romance languages, all of which are extensively used by *OED3*.[5] The situation with regard to modern German is much better. At the scholar's disposal are, for example, the rich collection of corpora of the *Institut für Deutsche Sprache* in Mannheim, the huge corpus of the University of Leipzig project *Deutscher Wortschatz*, and the emerging *Digitales Wörterbuch der deutschen Gegenwartssprache*. For literary texts the German version of *Project Gutenberg* provides a large but philologically unreliable corpus (which thus is of only limited use to *OED*).

Of course, *OED3*'s etymology editors also make use of many *OED* in-house resources, especially *OED* dictionary slips, e.g. quotation slips or slips containing fruits of the Scholarly Reading Programme or correspondence from external consultants. These slips are passed on to the Etymology Group by the Revision Group editors, who complete work on the semantic side of an entry before the etymologists start working on it. In this way, the etymology editors can take into account any changes made to the entry at a previous stage of revision, e.g. antedatings, which may invalidate a whole etymology.

The German dictionaries have some drawbacks and limitations which the OED3 etymologists need to be aware of. Many of the above dictionaries do not give first dates of attestation for words, still less for senses. The available Old and Middle High German dictionaries do not date their quotations at all, and Lasch/Borchling and *DWB* also rarely supply dates (which in the case of the latter must be worked out using the *Quellenverzeichnis*; Hass-Zumkehr 2001: 130, 138).[6] Even Pfeifer (1993), which endeavours to give dates of first attestation as far as possible, often cannot do so in the case of Middle High German words, but this is hardly surprising in view of the parlous state of Middle High German lexicography (Pfeifer 1993: vii). In fact, the relative paucity of data in all older Germanic languages, the preferred style of scholarly dictionaries for each of these languages, and the fact that Middle English has so much fewer Germanic loans than Romance ones, speak very much in favour of *OED3*'s policy, viz., that older Germanic lexical items are not given a date at all, unlike Romance words.

In contrast, we try to give dates for early modern German and modern German words whenever possible, though this will often take the format (e.g.) "1850 or earlier", using the earliest hit of a corpus or library catalogue search as *terminus ante quem*.[7] Table 2 lists some library catalogues which are especially useful for this purpose, first and foremost – in a German context – the *Karlsruher Virtueller Katalog* (*KVK*) which allows the user to search all union catalogues in the German-speaking countries.[8]

**Table 2.** Some major library catalogues providing useful (ante)datings

| **Catalogue** | **(library and) country** |
|---|---|
| *Karlsruher Virtueller Katalog* (*KVK*) | union catalogue of scholarly libraries in German-speaking countries and some other countries |
| *COPAC* | union catalogue of UK scholarly libraries |
| *EUREKA* | union catalogue of US scholarly libraries[9] |
| *MELVYL* | union catalogue of University of California libraries, US |
| British Library Public Catalogue (*BLPC*) | British Library, UK |
| *BNF OPALE-PLUS* | Bibliothéque Nationale, France |
| Library of Congress Catalogue | Library of Congress, US |
| *HOLLIS* (Harvard On-Line Library System) | Harvard University, US |
| *NYPL-OPAC* | New York Public Library, US |

Among the German etymological dictionaries, Pfeifer's dates are especially trustworthy, so much so that the 23rd edition of Kluge/Seebold often preferred them to the dates found by Professor Seebold's editorial team (Kluge/Seebold 1995: vii). In addition, Nes (1999: 12–13) points out that Pfeifer's Old High German forms are the most reliable found in any general etymological dictionary.

Although *DWB* contains quotations from more than 25,000 sources (Kirkness et al. 1991b: vii), the great majority of these quotations come from a limited range of ca. 2,000 sources in 4,000 volumes which had been read systematically (Schlaefer 1996; Hass-Zumkehr 2001: 138). In the same vein, the 5th to 8th editions of Paul's *Deutsches Wörterbuch*, which were edited by Werner Betz, contain so many quotations from Goethe´s works (including *hapax legomena*) that this dictionary has been called a "kleines Goethe-Wörterbuch" (Hass-Zumkehr 2001: 195).

Definitions were not a major concern of Jacob and Wilhelm Grimm either, who often deemed it sufficient to provide one or more Latin glosses of a German lemma; only from the 1930s did *DWB* treat semantics in detail (Hass-Zumkehr 2001: 124, 137). Finally, much of the scholarship in the first edition of *DWB* is elderly. This also applies to the available Middle High German dictionaries and to Schade's *Altdeutsches Wörterbuch* (Schade 1872–1882). The latter was already available to the editors of *OED1*, and is even now sometimes used to verify Old High German and Old Saxon forms which were given in *OED1* but cannot be found in any other dictionary. Special care must be taken when falling back on it, however, since it is highly out of date and contains forms from three quite distinct language stages: Old High German, Old Saxon and Middle High German.

It is well-known that for puristic reasons hardly any *Fremdwörter* were included in the first edition of *DWB*, and Jacob Grimm's original policy was followed by other German dictionaries of the 19th and early 20th centuries. Its revised edition is thankfully much more open to *Fremdwörter*, as Strauss (1991: 652–654) demonstrates by an impressive selection of them taken from entries in A and D, but will only become of use to *OED3* when the latter reaches the letter A.[10] Neither do specialized technical terms of the natural sciences, for example, receive much attention in German historical and etymological dictionaries[11] and specialist historical dictionaries or databases for this field are woefully lacking. To my knowledge, a work similar to Cottez' (1985) extremely useful *Dictionnaire des structures du vocabulaire savant* does not exist in German. The *Deutsches Fremdwörterbuch* does not necessarily close such gaps with regard to *OED3* as it does not cover highly specialist vocabulary either. As a result, it is very often necessary to send out time-consuming library requests to trace a foreign-language coinage, and dates must once again be given approximately or even cannot be obtained at all. *DGWB* provides good coverage of the modern language, including non-standard registers and a good many technical terms, but is a synchronic dictionary.

On the plus side, *DWB* is invaluable especially for the early modern German period because only a small part of *FWB* (1986-) has been published to date; it must also routinely be consulted for Germanic cognates and, last but not least, provides many German regional forms, though it was never intended to cover German dialects exhaustively (see Niebaum 1991). That said, *OED3* does not normally cite German regional forms unless they are of special significance for the etymology.

## 3 Comparison between OED2 and OED3: kinds of improvement

For the purposes of this paper, four categories of *OED3* entries on German loanwords can be distinguished, depending on the amount and nature of changes in their etymologies from the second to the third edition.

**3.1.** The first group comprises entries which already were main entries in *OED2* and whose etymology did not change significantly. Yet even here the quality of etymological treatment has improved: the style of presentation has become more consistent, and *OED3* often gives much more detailed information than *OED2*.

As the etymology of MANDELSTEIN *n*. demonstrates, coinage details for scientific terms are given in *OED3* whenever possible, and the reader can now follow up the ulterior etymology of German *Mandel* and *Stein* through added cross-references in the form of hypertext links:

*OED3*: MANDELSTEIN *n.* [etymology; first attested in English in 1799]

> **mandelstein,** *n.*
> [< German *Mandelstein* (A. G. Werner *Kurze Klassifikation u. Beschreibung der Verschiedenen Gebirgsarten* (1787) 13) < *Mandel* almond (see MANDELIC *a.*) + *Stein* STONE *n.* Cf. AMYGDALOID *n.*
> A form *mandelstone* is also recorded, with substitution of STONE *n.* for the second element:
> 1890 *Cent. Dict.*, Mandelstone.]

Another example of the wealth of added detail is the *OED3* etymology of NOODLE *n.*²:

*OED3*: NOODLE *n.*² [etymology; first attested in English in 1779]

> **noodle,** *n.²*
> [< German *Nudel* (16th cent.), further etymology uncertain: prob. a variant of *Knödel* dumpling (14th cent. in late Middle High German in sense 'small knot'; a diminutive formation (cf. -EL¹) < a variant of *Knoten* KNOT *n.*¹), although an etymology has also been suggested < Rhaeto-Romance (Grödnerisch) *menùdli* small dough piece in soup, prob. ult. < classical Latin *minūtulus* (see MINUTULOUS *a.*). Cf. (< German) French *noudle, nudeln* (1765), *nouilles* (1767), Swedish *nudel* (1686 in a compound).]

However, it needs to be emphasized that *OED3* usually treats remoter etymologies in less detail (understandably, since it is an English rather than a comparative etymological dictionary), and that the information given in *OED3* for remoter etymologies is to a large extent derivative.

**3.2.** The second group comprises entries which were subordinate (nested) entries in *OED2* but have now been upgraded to main entries. Such entries did not require a full etymology in *OED2*, where indeed many apparently self-explanatory derivatives were not given a full etymology, merely "from prec[eding]" or similar. In *OED3* they are presented with an explicit, full etymology, like any main entry, and this is especially welcome in the case of formations after German models, since their origin can now be made much more explicit. Cf. e.g. the etymology of NATIONAL SOCIALIST *a.* and *n.*, which includes useful encyclopedic information and complements the etymology of NAZI *n.* and *a.*; one of its chief sources is Schmitz-Berning (1998), a comprehensive historical-etymological dictionary of Nazist ideological terms:

*OED3*: NATIONAL SOCIALIST *a.* and *n.* [etymology; first attested in English in 1916 as adjective with reference to the National Social Union, in 1923 as adjective with reference to the Nazi party, and in 1931 as noun]

> **National Socialist,** *a. and n.*
> [< NATIONAL *a.* + SOCIALIST *a.* (cf. SOCIALIST *n.* 2), after German *nationalsozial* (1881 in E. Dühring designating racist nationalist ideology, 1896 in *der Nationalsoziale Verein* the National Social Union: cf. sense 1) and *nationalsozialistisch* (1896 as *national-sozialistisch* designating the National Social Union (cf. sense 1); 1920 in *die Nationalsozialistische Deutsche Arbeiterpartei* the National Socialist German Workers' Party (cf. sense 2)). With use as noun, cf. German *Nationalsozialist* (1887 in general sense 'popular national leader'; 1930 or earlier denoting a member of the National Socialist German Workers' Party). Cf. French *national-socialiste* (1923 as noun; 1928 as adjective), Italian *nazionalsocialista*, adjective (1932).
> The National Social Union referred to at sense 1 was unrelated to the later National Socialist German Workers' Party.]

> *OED3*: NAZI *n.* and *a.* [etymology; first attested in English in 1930 as noun and adjective]

> **Nazi,** *n. and a.*
> [< German *Nazi* (*c*1920), shortened < *Nationalsozialist* or *Nationalsozialistisch* (see NATIONAL SOCIALIST *a.* and *n.*). Cf. French *Nazi* (1930).
> The spelling with *z* prob. arose by analogy with *Sozi* (shortened < *Sozialist* SOCIALIST *n.*).
> The term was originally used by opponents of the National Socialist German Workers' Party and may have been influenced by Bavarian *Nazi*, a familiar form of the proper name *Ignatius* and used to refer to or characterize an awkward or clumsy person. The German form *Inter-Nazi* (shortened < *Internationale* INTERNATIONALE *n.*) which is attested much earlier may also have contributed to the adoption of the term *Nazi*.]

**3.3.** A third, not very numerous, group consists of entries entirely new to OED3. Many of them are unassimilated rare 19th- or 20th-century German loanwords denoting German Realien from various fields, e.g. politics (MITBESTIMMUNG *n.*), food and culinary culture (EISBOCK *n.*, MAHLZEIT *n.*, NUSSTORTE *n.*), or medieval German culture (MEISTERSÄNGER *n.*, MEISTERGESANG *n.*; these complement MEISTERSINGER *n.* and MASTERSINGER *n.*, which were already in *OED2*, and are perhaps best known in English through Richard Wagner´s opera). However, their etymologies are usually transparent and again draw heavily on the standard German etymological dictionaries.[12]

**3.4.** The most interesting and last group contains entries which are main entries both in OED2 and OED3 and whose etymologies have significantly changed. Table 3 gives a "telegraphese" summary of the etymological formulae in OED for each specimen entry in this group, before and after revision.

**Table 3.** Significant changes in the revised etymology of *OED* entries

| ***OED3* entry** | ***OED2* etymology type** | ***OED3* etymology type** |
|---|---|---|
| MAJORAT *n.* | < French < Latin | < German or its French etymon, after Spanish |
| MANGLE *n.*[3] | < Dutch, ult. < Greek | Partly < Dutch, partly < its MHG etymon, this ult. < Latin (where further etymology is given) |
| MARZIPAN *n.* | "from various Continental sources", further etymology unknown | **(entry upgraded from MARZIPAN, MARCHPANE)** < German or its Italian etymon, this perh. < Persian or Arabic |
| MARCHPANE *n.* and *a.* | --- | **(new entry: split from MARZIPAN, MARCHPANE)** < Italian or German, after PAIN *n.*[2] |
| MASTERSINGER *n.* | "anglicized form of MEISTERSINGER" | compound, after German |
| MAW SEED *n.* | "half-translated ad. Ger. dial." | < German (obs.) + SEED *n.*, prob. after reflex of a MHG word; regional German cf´s given |
| MEDIATIZATION *n.* | suffix formation | detailed etymology given, senses 1 and 2 split |
| MEDIATIZE *v.* | < French | either < French, or < German < French verb + German verb suffix *-iren* (now *-ieren*) |
| MELANGE *n.* (sense 3) | no etymology given for sense 3 | in sense 3 < German (chiefly Austrian); this sense app. not in French |
| MESITENE *n.* | suffix formation | suffix formation after German |
| METALANGUAGE n. | "see META-" | compound, after German (Tarski), itself after Polish |

| ***OED3* entry** | ***OED2* etymology type** | ***OED3* etymology type** |
|---|---|---|
| MOGUL *n.*[2] | perh. < German regional (southern) | < German regional (Austria), app. remodelled after MOGUL *n.*[1] |
| MONOSACCHARIDE *n.* | compound | compound, after German |
| MOSSE *n.* | < German *Maß* n. | < early modern German *Maße* f.; cf. German regional (Bavaria, Austria) *Maß* f. |
| MOTORIC *a.* | suffix formation | suffix formation, in sense 2 after German |
| MUCKER *n.*[2] | prob. < German | perh. < German |
| MULCH *a.* | internal Middle English formation, prob. related to MELCH; cf. German regional | either reflex of unattested Old English adjective (cognate with early modern German, German regional), or < early modern German |
| MUMCHANCE *n.* and *a.* | < Middle Low German | < Middle Dutch or Middle Low German or early modern German |
| MURKY *n.* | "see quots." | < German, further etymology unknown |
| NAZI *n.* and *a.* | "representing the pronunciation of (German) *Nati-*" | < German, shortened; perh. influenced by German regional (Bavarian) *Nazi*, pet-form of male forename *Ignaz* |
| NIX *n.*[1] and *int.* | in sense 1 < colloquial Dutch, German | < German colloquial shortened form; cf. Dutch |
| NOCKERL *n.* | < German regional (Austrian) | < German, orig. German regional (southern, Austrian, Swiss) diminutive formation; further etymology uncertain |
| NOODLE *v.*[5] | < NOODLE *n.*[3] | perh. < German regional (Leipzig), or perh. the same word as NOODLE *v.*[2] |
| NOSTALGIA *n.* | < mod. Latin < Greek | < post-classical Latin (Hofer 1688) < ancient Greek, after German; cf. French, Italian |
| NOUMENON *n.* | < Greek; introduced by Kant | < German (Kant 1783) < ancient Greek (Plato) |
| NOVELETTE *n.* | either suffix formation, or < Italian | suffix formation, after Italian; in sense 2 after German (Schumann 1838) |

As is seen in Table 3, some of the etymologies have been improved out of all recognition. Two of the three etymologies discussed in detail below come from the semantic field of music.

The obsolete English noun *minikin* is first attested in the 16th century. According to *OED2*'s entry MINIKIN *n.* and *a.* this word was of Dutch origin and meant both 'sweetheart, darling' and 'a kind of lute-string'. The link between these senses is far from clear.

*OED2*: MINIKIN *n.* and *a.* [etymology and selecxted quotations]

> **minikin**, *n. and a.*
> Forms: 6 **mynykyn**, **menekyn**, **-in**, **menyking**, 6-8 **miniken**, **minnekin**, 7 **mynnikin**, 7-8 **miniking**, 6-9 **minnikin**, **minikin**. Also 7 **minckins** (*sing.*).
> [ad. early mod.Du. *minneken* (MDu. *minnekijn*), f. *minne* love + *-kijn* -KIN.]
> **A.** *n.*
> **1.** A playful or endearing term for a female. *Obs.* exc. *dial.* (see quot. 1879).
> **a1550** (...)
> † **2.** A thin string of gut used for the treble string of the lute or viol. Also *attrib.*, as ***minikin string***.
> **1541** (...)
> **b.** esp. in the phrase *† **to tickle (the) minikin***, to play the lute or fiddle. (Frequently used by early 17th c. dramatists, often with allusive suggestion of sense 1.) **1601** (...)
> **B.** *adj.*
> **1.** Dainty, elegant, sprightly. Now contemptuously: Affected, mincing.
> **a.** of a person; formerly of a girl or woman, but now applied to a person of either sex. Also in jingling combinations, ***minikin-finikin***, ***-finical***. **1573** (...)
> **b.** of a person's actions, attributes, etc. **1545** (...)

Several years ago a scholarly correspondent supplied a quotation from the travel journals of Sir Thomas Hoby, an Elizabethan diplomat and author, which provided the key to the solution of the *minikin* puzzle: "Miniken..this towne standethe uppon a river that goethe into the Danubius, called Yser...It is a famous towne for the best lutestrings in all Germanie." As the river names already suggest (Danube and Isar), the town in question is Munich. Hoby's form *Miniken* is a 16th-century English variant of the name of the Bavarian capital, deriving from the contemporary German form *Münichen* and showing the unrounding of [Y] to [I] typical of the Bavarian regional dialect.[13] Two articles in (non-linguistic) scholarly journals independently confirmed the etymology, stating that minikins were manufactured in Munich and were an expensive, imported item in 16th-century England. Such a drastic change of the etymology of one sense meant that the entry must be split. Consequently, *OED3* features two homonymic entries – MINIKIN *n.*[1] and *a.*[1] ‘sweetheart’ and MINIKIN *n.*[2] and *a.*[2] ‘lute-string’.[14]

*OED3*: MINIKIN *n.*[1] and *a.*[1] [etymology and selected quotations]

> **minikin**, *n.*[1] and *a.*[1]
> Now *rare*.
> Forms: 15 **mynikin**, **mynnykyn**, **myntkin**, **mynykyn**, 15–17 **miniken**, **minnekin**, 15- **minikin**, **minnikin**, 16 **minckins**, **miniking**, **mynnikin**, 18 **minnikon**, 19- **minniken**.
> [< Dutch *minnekijn*, *minneken* sweetheart, beloved, darling, friend (now rare; Middle Dutch *minnekijn*, *minnekin*, *minneken*) < *minne* love (see MEAN *v.*[1]) + *-kijn* -KIN. Cf. MINIKIN *n.*[2] and *a.*[2]
> In form *minckins* prob. showing –S[2]; cf. -KINS.
> Sense A. 3 is perh. influenced by MANIKIN *n.* (cf. quot. 1761).]
> **A.** *n.*
> **1.** A young girl or woman (usually as a term of endearment). Also: a minion, a favourite (*obs.*). Now *rare*.
> Sometimes with allusion to MINIKIN *n.*[2] 1, as in quot. 1608. ***c*1540** (...)
> **B.** *adj.*
> **1. a.** Originally, of a girl or woman: dainty, elegant, sprightly. In later use, applied to both sexes: affected, mincing, or (of a man) effeminate. Also in reduplicated compounds, as ***minikin-finical***, ***-finikin*** adjs. *Obs.*
> In quot. 1545: delicate, having a woman’s daintiness. **1545** (...)

*OED3*: MINIKIN *n.*[2] and *a.*[2] [etymology and selected quotations]

> **minikin**, *n.*[2] and *a.*[2]
> Forms: 15 **menekin**, **menekyn**, **menykinge**, **mynyken**, 15–16 **miniken**, **minkin**, 15- **minikin**, 16–17 **miniking**, 17 **minnekin**, 19- **minnikin**, (in sense 2) **manikin**.
> [< *Miniken* Munich (16th cent. in this form in English contexts) < German *Münichen*, obsolete form of *München* (see MUNICH *n.*), with *-i-* in the first syllable prob. reflecting unrounded pronunciation of *ü* in the Bavarian regional dialect. Cf. -KIN and MINIKIN *n.*[1] and *a.*[1] (cf. sense 1b).
> In the 16th cent. Munich was renowned for the production of high-quality treble strings for the lute. See quots. 1676 and 1976 at sense A. 1a, and cf.: ***a*1566** T. HOBY *Trav.* (1902) 112 Miniken..this towne standethe uppon a river that goethe into the Danubius, called Yser... It is a famous towne for the best lutestrings in all Germanie.]

**A.** *n.* **I.** Simple uses.
**1. a.** A thin strand of catgut used for the treble strings of a lute or viol. More fully ***minikin string***. Now *hist.*
**1541** (...) **1676** T. MACE *Musick's Monument* 65 Be carefull to get Good Strings, which would be of three sorts, viz. Minikins, Venice-Catlins, and Lyons. (...) **1976** *Early Music* 4 433/1 More uniform treble strings were available from Munich from Capirola's time... They were called 'minikins' in England and were very expensive.
† **b.** ***to tickle the minikin***: to play the lute or viol. *Obs.*
Freq. used humorously by early 17th cent. dramatists with allusive suggestion of MINIKIN *n.*[1] 1.
In quot. **1601** apparently used of a fiddle.
**1601** J. MARSTON et al. *Jacke Drums Entertainm*. I. 14 When I was a yong man and could tickle the Minikin,..I had the best stroke, the sweetest touch, but now..I am falne from the Fidle, and betooke me to thee [sc. the Pipe]. (...)

Another "musical" entry turned out to be a loanword from German rather than from ancient Greek. The earliest unambiguously English quotation for MELISMA *n*. dates from 1880 (it has not been antedated in the course of revision). It comes from *Grove's Dictionary of Music*, which notes: "Mendelssohn employs the term in order to distinguish the Mediation and Ending of a Gregorian Tone from the Dominant, or Reciting Note." The ultimate source language of *melisma* certainly is ancient Greek, but note the following footnote in Grove: "See his [*scil.* Mendelssohn-Bartholdy's] letter to Zelter, dated Rome, June 16, 1831". This letter was duly tracked down; in it, the composer writes "... machen sie ein kurzes Melisma".[15]

*OED2*: MELISMA *n*. [etymology and selected quotations]

**melisma**. *Mus.* Pl. **melismata**, **melismas**.
[Gr. μέλισμα song, air, melody.]
(See quot. 1880.) Also, in singing, the prolongation of one syllable over a number of notes.
[**1611** T. Ravenscroft (*title*) Melismata; Musical Phansies fitting the Court, Citie, and Countrey Humours.] **1880** W. S. Rockstro in Grove *Dict. Mus.* II. 248/2 *Melisma*. Any kind of Air, or Melody, as opposed to Recitative, or other music of a purely declamatory character. Thus, Mendelssohn employs the term in order to distinguish the Mediation and Ending of a Gregorian Tone from the Dominant, or Reciting Note. Other

writers sometimes use it (less correctly) in the sense of *Fioritura*, or even *Cadenza*. (...)

*OED3*: MELISMA *n.* [etymology and selected quotations]

**melisma**, *n. Music.*
Plural **melismata**, **melismas**.
[< German *Melisma* (in a letter of 1831 by Felix Mendelssohn) < Hellenistic Greek μέλισμα song, air, melody < ancient Greek μέλος song, melody (see MELOS *n.*) + -ισμα (see -ISM). Cf. French *mélisme* (1837), and earlier MELISMATIC *a.*
Cf. earlier occurrence of the plural of the Greek word in an English context:
**1611** T. RAVENSCROFT (*title*) Melismata. Musical phansies. Fitting the court, citie, and countrey humours.]
A melody or melodic sequence of notes. Usually *spec.* (in singing and vocal composition): the prolongation of one syllable over a number of notes; an instance of this.
**1880** W. S. ROCKSTRO in G. Grove *Dict. Music* II. 248/2 (...)

Finally, OKOUME *n.*, denoting the wood of a West African tree, was presented as simply a "native name" in *OED2*, like so many non-European loans in English. From the earliest quotations in *OED2* and *OED3* it is clear that the word cannot have come into English directly from an African language: the first quotation occurs in an English translation of a book by Albert Schweitzer, *Zwischen Wasser und Urwald* (1921). Although this is written in German, the problem of the immediate donor language must remain open, as Schweitzer – a native of Kaysersberg in Alsace – was bilingual in German and French and is likely to have first encountered *okoumé* in French sources, Gabon being a French colony at the time. The earliest French date for *okoumé* is given by Robert (1986) as 1914, in an article in the journal *L' année scientifique et industrielle*; this was confirmed by library research.[16] Chronologically both German and French are possible etyma, as both are attested earlier.

*OED2*: OKOUME *n.* [etymology and selected quotations]

**okoume**. Also **okoumé**.
[Native name.]
= GABOON, GABOON.
**1922** C. T. CAMPION tr. *Schweitzer's On Edge of Primeval Forest* vi. 108 The chief sorts [of wood] dealt in are mahogany,..and okoume (Aucoumea klaineana), the so-called false mahogany. (...)

*OED3*: OKOUME *n.* [etymology and selected quotations]

> **okoume**, *n.*
> Forms: 19- **okoume**, **okoumé**, **oukoume**.
> [< German *Okoume* (1921 in the source translated in quot. 1922) or its etymon French *Okoumé* (1914; also in form *oukoumé* (1956)), either < Myene (a language of the part of Gabon from which the wood is exported) *okume*, or < Tsogo (a language of the interior of Gabon, where the trees grow) *okume*.]
> The wood of the West African tree *Aucoumea klaineana* (family Burseraceae); the tree itself. Also called *Gaboon mahogany*.
> [**1920** A. L. HOWARD *Man. Timbers of World* 140 Gaboon. *Boswellia klaineana*... This extremely useful wood deserves a name of its own.., though it should not be called mahogany. This fact is recognised on the Continent, where it is known by the name of Okumé.] **1922** C. T. CAMPION tr. A. Schweitzer *On Edge of Primeval Forest* vi. 108 (...)

## 4 Conclusion

To sum up, although what *OED3* can achieve is always limited by the quantity and quality of the available research tools (not to mention the strict deadlines set for editorial work), I hope to have illustrated what can now be done in the field of German loanwords and more generally German influence on the English language, and to have made it clear that, in this respect just as well as in many others, the Third Edition of the *Oxford English Dictionary* is greatly improved from the Second Edition.

## Notes

1. This originated as one of the plenary lectures at the *9th International Symposium on Lexicography* in 1998.

2. See References for full titles. Many scholarly dictionaries recently published or in the process of revision are presented in the articles in Städtler (ed.) (2003), a good introduction.

3. Jacob Grimm declared in 1838 that his planned dictionary should cover the vocabulary of literary German from Luther to Goethe (Gottzmann 1991: 170), but *DWB* never consistently kept to these limits; it does indeed contain a good deal of Old and Middle High German words. However, these are usually given in a modern German guise (ibid.: 177), and in any case *DWB* cannot be used as a substitute for the specialist Old and Middle High German (still less Middle Low German) dictionaries.

**4.** Cf. Schlaefer (1996), which also presents a concise overview of the history of *DWB*, and – for a more detailed analysis of *DWB*'s revision principles – Strauss (1991).

**5.** One need only mention *Frantext* or *Gallica* for French, the *Corpus Diacrónico del Español* (*CORDE*) for Spanish or the *Opera del Vocabolario Italiano* (*OVI*) for Italian.

**6.** The revised edition of *DWB* always dates quotations but does not at present cover any part of *OED3*'s working range.

**7.** Although the Romance dictionaries are more concerned about dating, one still has to resort to such approximate dates very often.

**8.** See also the catalogue links page of the *Bayerische Staatsbibliothek* (www-opac.bib-bvb.de/subbvb/ordbvb/info/links1.htm).

**9.** *EUREKA* is an excellent source for identifying the source of English translations and for locating early non-English publications and microform publications.

**10.** The most recent (9th and 10th) editions of Paul's *Deutsches Wörterbuch*, published in 1992 and 2002, are also refreshingly atypical of German historical lexicography in that they include a large amount of *Fremdwörter*.

**11.** On the coverage of specialist and scientific terms in *DWB* see Schiewe (1991).

**12.** On the semantic fields of German loanwords in English, Cannon/Pfeffer (1994) is invaluable, even though its datings largely derive from *OED1* and *OEDS* and are thus partly no longer correct. Lehnert (1988) classifies the German loanwords in *OEDS*. Stubbs (1998) also used *OED2* – more exactly the *OED* on CD-ROM – as a database, but much of this short article is taken up by a case study of collocations of English *angst* ( < German *Angst*), so its title is somewhat misleading.

**13.** A modern form of the name widespread in spoken usage among the locals is *Minka*.

**14.** The expression *to tickle the minikin* (s.v. MINIKIN *n.*[2] A 1 b) still provides a bridge between the two homonyms: in *OED3*'s words (definition note) it is "freq. used humorously by early 17th cent. dramatists with allusive suggestion of MINIKIN *n.*[1] 1" (= 'sweetheart').

**15.** *OED3* also habitually compares French or other Romance cognates, but usually only if these are attested earlier than the English word and of some help for its etymology. French *mélisme* is first attested in 1837 (see *DDL*), but is unlikely to be the etymon of MELISMA *n.*, as the earliest English quotations do not show any French connection.

**16.** The datings of Robert (1986) are often inaccurate, unlike those of *Französisches Etymologisches Wörterbuch* or *Trésor de la langue française*; therefore first dates given only in Robert, but no other historical or etymological dictionary of French, are not accepted by *OED3* unless checked. If Robert is the only source of a date for a French word and it is impossible to verify Robert's dating, it is given in *OED3* in the style "dated [year] in Robert *Dict. Alphabétique et Analogique* (1986)".

## References

*Editions of the Oxford English Dictionary*

[*OED1*] = Murray, James A. H. et al. (eds.). 1884–1928. *A New English Dictionary on Historical Principles*. Oxford: Oxford University Press. Reissued in 12 volumes with a Supplement, 1933.

[*OEDS*] = Burchfield, Robert W. (ed.). 1972–1986. *A Supplement to the Oxford English Dictionary.* Oxford: Oxford University Press.

[*OED2*] = Weiner, Edmund S. C. and John A. Simpson (eds.). 1989. *The Oxford English Dictionary. Second Edition*. Oxford: Oxford University Press.

[*OED3*] = Simpson, John A. and Weiner, Edmund S. C. (eds.). 2000-. *The Oxford English Dictionary. Third Edition.* Oxford: Oxford University Press. On-line at http://dictionary.oed.com.

*German Dictionaries*

[*Ahd. Wb.*] = Karg-Gasterstädt, Elisabeth, Frings, Theodor et al. (eds.). 1968-. *Althochdeutsches Wörterbuch*. Berlin: Akademie-Verlag.

[*BMZ*] = Benecke, Georg F.; Wilhelm K. H. Müller and Friedrich Zarncke. 1854–1866. *Mittelhochdeutsches Wörterbuch*. Leipzig: Hirzel. On-line at http://gaer27.uni-trier.de/MWV-online/MWV-online.html.

Cannon, Garland and Pfeffer, Alan J. 1994. *German Loanwords in English: An Historical Dictionary*. Cambridge: Cambridge University Press.

[*DGWB*] = Scholze-Stubenrecht, Werner (ed.). 1999. *Duden: Das große Wörterbuch der deutschen Sprache in 10 Bänden*. 3rd ed. Mannheim, Leipzig, Wien, Zürich: Dudenverlag. [1st ed. 1976–1981]

[*Duden-Etym.*] = Drosdowski, Günther (ed.). 1989. *Duden: Das Herkunftswörterbuch. Etymologie der deutschen Sprache*. 2nd ed. Mannheim, Leipzig, Wien, Zürich: Dudenverlag. [1st ed. 1963]

[*DWB*] = Grimm, Jacob, Grimm, Wilhelm et al. (eds.). 1854–1961. *Deutsches Wörterbuch*. Leipzig: Hirzel. On-line at http://germazope.uni-trier.de/Projects/DWB.

[*DWB* revised ed.] = Berlin-Brandenburgische Akademie der Wissenschaften, Akademie der Wissenschaften zu Göttingen (eds.). 1965-. *Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm: Neubearbeitung*. Leipzig, Stuttgart: Hirzel.

[*FWB*] = Reichmann, Oskar et al. (eds.). 1986-. *Frühneuhochdeutsches Wörterbuch*. Berlin: de Gruyter.

[*Kluge/Seebold*] = Seebold, Elmar (ed.). 1995. *Kluge Etymologisches Wörterbuch der deutschen Sprache*. 23rd ed. Berlin: de Gruyter. [1st ed. 1883]

Lasch, Agathe; Carl Borchling and Gerhard Cordes (eds.). 1928-. *Mittelniederdeutsches Handwörterbuch*. Neumünster: Wachholtz.

Lexer, Matthias. 1872–1878. *Mittelhochdeutsches Handwörterbuch* (including *Nachträge*). Leipzig: Hirzel. On-line at http://gaer27.uni-trier.de/MWV-online/MWV-online.html.

[*Paul*] = Henne, Helmut et al. (eds.). 2002. *Hermann Paul: Deutsches Wörterbuch. Bedeutungsgeschichte und Aufbau unseres Wortschatzes*. 10th ed. Tübingen: Niemeyer. [1st ed. 1897]

Pfeifer, Wolfgang (ed.). 1993. *Etymologisches Wörterbuch des Deutschen*. 2nd ed. Berlin: Akademie-Verlag. [1st ed. 1989]

Schade, Oskar. 1872–1882. *Altdeutsches Wörterbuch*. 2nd ed. Halle: Verlag der Buchhandlung des Waisenhauses. [1st ed. 1866]

Schiller, Karl and Lübben, August. 1875–1881. *Mittelniederdeutsches Wörterbuch*. Bremen: Kühtmann.

Schützeichel, Rudolf. 1995. *Althochdeutsches Wörterbuch*. 5th ed. Tübingen: Niemeyer. [1st ed. 1969]

Starck, Taylor and John C. Wells (eds.). 1971–1990. *Althochdeutsches Glossenwörterbuch*. Heidelberg: Winter.

*Other Works*

Cottez, Henri (ed.). 1985. *Dictionnaire des structures du vocabulaire savant*. 4th ed. Paris: Larousse. [1st ed. 1978]

[*DDL*] = Quémada, Bernard. 1972–1998. *Documents et datations lexicographiques*: *matériaux pour l'histoire du vocabulaire français, publié par le Centre National de la Recherche Scientifique, Institut National de la Langue Française*. Paris: Klincksieck.

Durkin, Philip N. R. 1999. "Root and Branch: Revising the Etymological Component of the *Oxford English Dictionary*". *Transactions of the Philological Society* 97, 1–49.

Durkin, Philip N. R. and Thier, Katrin. 2003. "*Oxford English Dictionary* (*OED*)". In Städtler (ed.) (2003), 529–540.

Gilliver, Peter. 2005. "Materials and Methodologies for the New *Oxford English Dictionary*". In Gottlieb et al. (eds.), 241–248.

Gottlieb, Henrik; Jens Erik Mogensen and Arne Zettersten (eds.). 2005. *Symposium on Lexicography XI: Proceedings of the Eleventh International Symposium on Lexicography. May 2–4, 2002 at the University of Copenhagen* [Lexicographica. Series Maior 115]. Tübingen: Niemeyer.

Gottzmann, Carola. 1991. "Das Alt- und Mittelhochdeutsche im *Deutschen Wörterbuch*: Zum Stellenwert der älteren Sprachstufen". In Kirkness et al. (eds.) (1991a), 169–223.

Hass-Zumkehr, Ulrike. 2001. *Deutsche Wörterbücher – Brennpunkt von Sprach- und Kulturgeschichte*. Berlin: de Gruyter.

Kirkness, Alan; Peter Kühn and Herbert E. Wiegand (eds.). 1991a. *Studien zum Deutschen Wörterbuch von Jacob Grimm und Wilhelm Grimm*. 2 vols. Tübingen: Niemeyer.

Kirkness, Alan; Peter Kühn and Herbert E. Wiegand 1991b. "Zur Einführung: Von der philologischen zur metalexikographischen Beschreibung und Beurteilung *des Deutschen Wörterbuches*". In Kirkness et al. (eds.) (1991a), vii-lxi.

Lehnert, Martin. 1988. "The German Element in the Four Supplementary Volumes to the *Oxford English Dictionary* (*OED*) edited by R.W. Burchfield". *Zeitschrift für Anglistik und Amerikanistik* 36, 5–25.

Nes, Oddvar. 1999. "Etymologiske ordbøker over germanske språk". *Maal og Minne* 91, 19–56.

Niebaum, Hermann. 1991. "Zur Rolle der Mundarten im *Deutschen Wörterbuch*". In Kirkness et al. (eds.) (1991a), 435–475.

Robert, Paul (ed.). 1986. *Dictionnaire alphabétique et analogique de la langue française*. Paris: Larousse.

Schiewe, Jürgen. 1991. "Fach- und Wissenschaftssprachen im *Deutschen Wörterbuch*". In Kirkness et al. (eds.) (1991a), 225–263.

Schlaefer, Michael. 1996. "*Deutsches Wörterbuch* Göttingen: Ein Leitfaden für den Benutzer". On-line at www.grimm.adw-goettingen.gdwg.de.

Schmitz-Berning, Cornelia. 1998. *Vokabular des Nationalsozialismus*. Berlin: de Gruyter.

Simpson, John A. 2000. "Preface to the Third Edition of the *Oxford English Dictionary*". On-line at http://dictionary.oed.com/about/oed3-preface/.

Städtler, Thomas (ed.). 2003. *Wissenschaftliche Lexikographie im deutschsprachigen Raum*. Heidelberg: Winter.

Strauss, Günter. 1991. "Die Bände I und VI der Neubearbeitung des *Deutschen Wörterbuches*: Unterschiede in der lexikographischen Bearbeitung". In Kirkness et al. (eds.) (1991a), 627–702.

Stubbs, Michael. 1998. "German Loanwords and Cultural Stereotypes". *English Today* 53: 14: 1, 19–26.

CHAPTER 17

# Extracting usage information from 18th century English-Danish and Danish-English dictionaries

Marcin Overgaard Ptaszyński

The paper focuses on difficulties with identification and extraction of usage information in 18th century dictionaries. Four dictionaries between English and Danish are analysed, in order to determine (1) what kind of data can be regarded as usage information, and (2) how it can be classified. Due to problems stemming both from the nature of the dictionaries (no explications of labels used, inconsistent labelling, inconvenient typography, misprints and mistranslations) and from the absence of relevant 18th century language corpora, extracting usage information is difficult – as is classifying it, since modern typologies seem unsatisfactory for the purpose. The paper offers possible solutions to the problems, though some turn out to be of limited applicability – for two reasons: (1) it is impossible to accurately assess how well the dictionaries reflect past linguistic realities, and (2) modern lexicography is still unable to avoid all ambiguities in usage information, notably subject field labels.

## 1 Introduction

This paper focuses on several problems which researchers may encounter when studying usage information in dictionaries. My main interest lies in bilingual dictionaries from earlier centuries, and my research has concentrated on English-Danish and Danish-English dictionaries. Of the multitude of dictionaries that involve English and Danish,[1] for the purpose of this paper I have chosen only four works: Berthelson (1754) (English-Danish), Wolff (1779) (Danish-English), and Bay (1796 and 1798) (English-Danish and Danish-English, respectively.) The first two of these are the first dictionaries between English and Danish to contain any usage information, and as such they constitute a natural starting point for a diachronic investigation of the phenomenon. Bay's works were the only other English-Danish and Danish-English dictionaries that appeared in the 18th century. Thus, on the one hand the analysis involves all the works of the period. On the other hand, it reflects part of a broader study, aimed at description and classification of usage information provided in dictionaries between English and Danish, from the 18th century up to the present.

The tradition of supplying information on usage in dictionaries between Danish and English is nearly as old as the history of these dictionaries. Wolff's dictionary was the first Danish-English one ever published, and Berthelson's work had only one predecessor. As the overall structure and content of dictionaries for this language pair have been subject to changes over the past 250 years, so has the usage information provided in them. Therefore, any attempt at applying a modern typology to such information given in the dictionaries investigated here is bound to be problematic. Rigid systems of labelling usage according to various categories, e.g. subject fields, style, (lack of) politeness, or regional usage, are known in today's dictionaries but did not exist in the 18th century. Moreover, with regard to typography, usage labels were often identical with other "additional" information provided within entries. This in turn poses considerable difficulties not only in classifying, but also in extracting usage information from the entries. Below I discuss both problems, as well as possible solutions to them.

To help solve problems with classifying usage information I have adopted Landau's typology as a starting point. It employs nine categories: currency, frequency, regional usage, technical or specialized terminology (field labels), restricted and taboo usage, insult, slang, style or functional variety, and status or cultural level (1984: 175). Since nearly all usage information present in the dictionaries investigated here seems to fall under the category of field labels, this paper will only briefly touch upon the other categories.

## 2 Berthelson's English-Danish dictionary (1754)

Berthelson attempts to provide a fairly uniform system of labelling words that belong to specialized language. In all, Berthelson marks 433 lexemes with various subject field labels. Since the dictionary aimed primarily at merchants and sailors (Berthelson 1754: address to the reader), the two most frequent labels are *Lov-Term* ('legal term') and *Skibs-Term* ('nautical term'). A sample containing several examples is presented in Table 1.

**Table 1.** Berthelson 1754: subject field labels

| Lexeme | Information | Category |
|---|---|---|
| abaft | Skibs-Term ('nautical term') | subject field |
| abalienate, to | Lov-Term ('legal term') | subject field |
| adjutant-general | Mil.-Term ('military term') | subject field |
| alay | Jagt-Term ('term in hunting') | subject field |
| arm a shot, to | Artil. Term ('term in artillery') | subject field |
| beat, to | Jagt-Term ('term in hunting') | subject field |
| dearly beloved | Eccl. Term ('ecclesiastical term') | subject field |
| camisade | Krigs-Term ('term in war') | subject field |
| capot | Term i Piquet-Spil ('term in the game of piquet') | subject field |
| level-coil | Spil-Term ('term in games') | subject field |

Apart from the above examples, the dictionary contains 36 other lexemes which are supplied with semantic information that potentially reveals facts about their usage. A few cases are presented in Table 2.

In the dictionary, the information of immediate concern to us (found in the "information" column in the table above) is, with few exceptions, provided according to the same typographical convention as the "proper" usage information included in Table 1, that is, within brackets. This, together with the fact that no explanation as to the labelling of usage appears in the front matter of the dictionary (or anywhere else), justifies considering the above examples as potential usage information. The problem lies, however, in finding a method that would allow us to extract from them only usage information, thereby distinguishing it from any other semantic data, notably (1) semantic specification which serves merely to discriminate between various senses of the same lexeme,[2] and (2) quasi-definitional information.

We can begin solving this problem by asking the following question: how is it possible to tell "legitimate domains" of field labels from any other semantic information? Although a satisfactory answer will not eliminate all difficulties connected with extraction of usage information,[3] it can help us delimit our analysis to the area of subject field labels and discard other additional semantic information provided in the entries.

**Table 2.** Berthelson 1754: potential usage labels

| Lexeme | Information | Equivalent | Category |
|---|---|---|---|
| demise | bruges allene om Kongelige Personer ('is used solely with reference to the royal family') | Bortlejelse, Bortfor-pagtning | ? |
| juke, to | siges om en Falk ('is said about a falcon') | at sætte sig til at sove | ? |
| kinswoman | siges om en Qvinde ('is said about a woman') | Slægting; Söskendbarn | ? |
| knuckle down to your taw | er en Talemaade som smaae Drenger bruger i Marmor Spil ('is a phrase which small boys use in the game of marbles') | læg Haanden paa samme sted, hvor din Marmor ligger | ? |
| leap, to | siges om en Hengst ('is said about a stallion') | at bedække | ? |
| man the ladder | Commando-ord til Skibs ('an order at the ship') | folk paa Fald-reebet | ? |
| racy | siges om Viin ('is said about wine') | gammel og listig | ? |
| set a paper, to | i Trykkerie ('in printing') | at sætte en Side | ? |
| spade is trump | i Kort Spil ('in card games') | Spader er Trumph | ? |
| stale, to | om Hester ('about horses') | at pisse | ? |
| to our trusty and well-beloved | Cancellie-stil ('Civil Service style') | til vor troe og elskelige | ? |
| uncloyed | om en Canon ('about a cannon') | ufornagled, uforstopped | ? |

First, we could postulate that the domains of field labels are defined in abstract rather than concrete terms. This postulation would eliminate such cases (excluded from Table 2.) as:

**Table 3.** Berthelson 1754: "genus proximum" information

| Lexeme | Information | Category |
|---|---|---|
| mullet | fisk ('fish') | ? |
| ox-tongue | urt ('herb') | ? |

Here, the additional semantic information seems to be clearly the specification of the *genus proximum* of the concepts denoted by the lexemes. In a monolingual dictionary it would most probably be incorporated into the respective definition rather than appear a usage label.

We might also assume that domains of subject field labels should be perceptible as domains, e.g. branches of knowledge, activities, etc. However circular this argument might sound, it is apparently rather difficult to reject in reality. This is reflected in the fact that probably no two dictionaries apply an identical set of subject field labels. Judging the semantic information provided in Table 2 on the basis on my own intuition,[4] I can distinguish between two basic kinds: one that possibly could be interpreted as a domain / subject field, and all other information. The first kind is (more or less) clearly represented by only one case, namely that of *i Trykkerie* ('in printing'), found within the entry of 'set a paper, to'. This leaves us, however, with the remaining 35 cases in the "residue".

In the light of the above, we could be tempted to say that these cases are not domains and, by extension, not usage information. However, we would quickly run into a problem here, since in the case of some examples from Table 2 a potential domain indeed seems to be more easily perceptible than in others. Let us look closer at such examples as *siges om en Falk* ('is said about a falcon') and *siges om en Qvinde* ('is said about a woman'), attached to the lexemes 'to juke', and 'kinswoman', respectively. The form of the provided semantic information, especially the use of the verb *siges* ('is said'), might imply that it pertains to the *use* of these lexemes, rather than to their denotation. Here, however, we seem to be running into a dead end, as the two examples are clearly different from each other. Even if such domain as *falconry* could be legitimately postulated now for a dictionary from 250 years ago, and even if we could assess whether the lexeme it is attached to was part of specialized language at the time, there is no domain we could establish for the other piece of semantic information. And apparently there should not be any, because this data does not seem to contain any usage information at all: indeed, it would be hard to imagine it as information specifying that the lexeme 'kinswoman' or its Danish counterparts: *slægting* or *söskendbarn* belong to any specialized language. Still, the forms of the provided information are identical in both cases, which makes it difficult for the researcher to distinguish such quasi-definitional information from clear subject field labels.

One might simply claim here that usage information in Berthelson's dictionary was presented inconsistently, and try to check whether any of the "problematic" pieces of information listed in Table 2 might be reasonably incorporated into any domain already present in the dictionary. A closer look reveals that some lexemes could be marked with field labels otherwise used by Berthelson: 'spade is trump' could, for example, fall under the heading of *Spil-Term* ('term in games') and 'man the ladder' under *Skibs-Term* ('nautical term').

On the other hand, one might say that maybe Berthelson was, after all, right when he refused to attach field labels to the above-mentioned terms. Perhaps they were not part of specialized language, but of common vocabulary. This is, however, rather hard to assess. First, contemporary corpora of neither English nor Danish existed when Berthelson was writing his dictionary, so he was not able to obtain such information with any degree of reliability. Second, even if they existed now, data obtained from them might not be of much use for investigating subject field labelling in bilingual dictionaries. Bo Svensén (1993: 183) claims that subject field labels are used to indicate both that a given word belongs to specialized language and that a concept denoted by a particular word belongs to a given domain. In practice, the dictionary user is unable to distinguish between the two kinds of data, since the labels are identical in both cases, though only the former contain usage information, the latter serving as sense discriminators or quasi-definitional information.[5] Such a state of affairs is obviously problematic for the researcher, who is compelled to treat both kinds of subject field labels as one. In line with Svensén's argument, however, Berthelson could have easily attached the relevant field labels to 'spade is trump' or 'man the ladder', be they technical words, or merely words denoting concepts within particular domains.

Another problematic case in Berthelson is that of the note *Cancellie-stil* ('Civil Service style'), probably a style label, attached to the phrase 'to our trusty and well beloved'. As we have seen in Table 1, a similar phrase, 'dearly beloved' is labelled *Eccl. Term* ('ecclesiastical term'). – definitely a subject field label. A question worth asking is why the two similar phrases are supplied with different kinds of usage information. The answer is probably simple: inconsistency.

The above attempts at extracting and classifying usage information provided in Berthelson's work show that it is no easy task. Let us now compare its difficulty with a similar analysis of the first Danish-English dictionary.

## 3 Wolff's Danish-English dictionary (1779)

Published in 1779, this dictionary was a pioneering work. According to its author, it was meant to be a counterpart to Berthelson's dictionary (Wolff 1779: preface), and even a quick look at both works reveals striking similarities in their macro- and microstructure. There is however, one major exception: any discernible system of field labels, such as that found in the English-Danish dictionary, is absent from its counterpart.

In Wolff's dictionary, all semantic data concerning the lemmata are provided within one and the same convention: they invariably appear right after the headword and the grammatical category marker (if present), and always in brackets. Kinds of data range from synonyms to short definitions and from sense discriminators to usage information. The difficulty in distinguishing the last of these from all the others is thus apparent. Table 4 contains some of the 86 pieces of semantic data that could potentially be interpreted as usage information.

The examples of semantic information included in the "information" column of Table 4 appear to be similar to those in the corresponding column in Table 2, referring to Berthelson's dictionary. None of them is a clear field label, and some of them might not contain any usage information at all. Those which do contain it must therefore be extracted and classified.

A possible step towards solving the first task could be to check whether any of the data from Wolff's dictionary might be incorporated into the subject field domains present in Berthelson's work. Such a procedure appears to be legitimate insofar as we bear in mind the fact that the two dictionaries are meant to be counterparts. A thorough look at Table 4 shows that this solution might indeed work for some of the labels. Consequently, 'bratspil' could, for example, be labelled as *Skibs-Term*, and 'en dobbelt örn' as *Herald-Term*.

Such a solution to the problem, however, is only partial, even within the area of subject field labels. First, if we agree to place 'en dobbelt örn' in the domain of *Herald-Term*, should 'skiold' be granted the same status? After all, the only difference in the semantic data attached to these two lexemes is the absence of the preposition *i* ('in') the latter case. Should this automatically mean that *vaaben* is not a usage label? Second, in the case of 'manillie', the semantic information is: *i l'ombre spil* ('in the game of omber'). Would it be legitimate then to establish a subject field label *Term i l'Ombre-Spil*, that could be treated on a par with Berthelson's *Term i Piquet-Spil*, attached to only one lexeme ('capot', see Table 1) in his dictionary? If so, then we would find ourselves adding to the original set, like in the case of the analysis of Berthelson's dictionary, where we postulated an additional domain of *printing*.

**Table 4.** Wolff 1779: (potential) usage information

| Lexeme | Information | Equivalent | Category |
|---|---|---|---|
| bratspil | til söes ('at sea') | a windlass, a draw-beam, the capstan of a ship | ? |
| bryst | fejl ('mistake') | *only cross-reference to 'bröst'* | ? |
| frakiendelse | i ret ('in law') | condemnation, a finding guilty | ? |
| haler (v.) | blandt söefolk ('among seamen') | to hale or hail | ? |
| jan | i spil ('in games') | the having no trick at cards | ? |
| indrykker (v.) | i en bog ('in a book') | to insert or interline | ? |
| löber i rid (v.) | siges om hunde ('said about dogs') | to desire the male | ? |
| madsfos | skiældsord ('term of abuse') | a sluggard | insult? |
| manillie | i l'ombre spil ('in the game of omber') | Manill | ? |
| nordbagge | en liden nordske hest ('a small Norwegian horse') | a little strong Norway-horse, a Norway-nag | regional? |
| nævner | i regnekonst ('in arithmetic') | a denomination | ? |
| omsats-stem-pel | hos bogbinderne ('at [the] bookbinder's' OR: 'among bookbinders') | a bookbinder's roll or tool | ? |
| qvint | i piket spil ('in the game of piquet') | a quint | ? |
| rie | norsk ('Norwegian') | an access, a fit | regional? |
| roe | næpe paa norsk ('*næpe* in Norwegian') | a turnip | regional? |
| skattere | hos skildrere ('at [the] painter's' OR: 'among painters') | to shadow a picture | ? |
| skiold | vaaben ('coat of arms') | an escutcheon or scutcheon, a coat of arms | ? |
| skrupel | i vægt ('in weight') | a scruple, a third part of a sham | ? |
| svær | paa slesk ('in Schleswig Danish') | the sword of bacon | ? |
| en dobbelt örn | i vaaben ('in a coat of arms') | an eagle with two heads | ? |

A slightly more complex problem with extracting usage information from Wolff's dictionary is the case of all semantic data consisting of phrases that begin with the Danish word *hos* (see e.g. the data for 'omsats-stempel' in Table 4). As seen in the table, such data can be understood in two different ways, e.g. *hos bogbindere* can be interpreted as 'at [the] bookbinder's' or 'among bookbinders'. To find out whether such a label signals that we deal with a concept occurring in a given domain, or a technical word associated with this domain, one would have to consult both contemporary corpora and contemporary users of Wolff's dictionary. As in the absence of both it is impossible to do so, I have decided to conduct a relevant query among native speakers of Present-Day Danish. These tend to perceive the label as a marker of specialized vocabulary, i.e. usage information "proper".

In Wolff we see only few items of usage information which are other than subject field labels. There are, for example, regional labels, mostly attached to Norwegian words. Two of them are particularly interesting as regards their form. The first one is that attached to 'nordbagge': *en liden nordski hest* ('a small Norwegian horse'). It is not clear whether this is only quasi-definitional information, or at the same time also a regional label. Here it may be worth mentioning that such a problem is not confined to Wolff's dictionary, but appears often with regard to modern dictionaries, as well.[6] The other case is that of *næpe paa norsk*, the information attached to 'roe' ('turnip'). Given that (1) the word *næpe* is not lemmatized in the dictionary, in which (2) misprints abound, one feels tempted to mistake this word for *næppe* ('hardly'), and consequently understand the whole label as 'hardly in Norwegian'. Should this be correct, then providing such information in Wolff's dictionary would seem methodologically questionable for two reasons. First, the dictionary is not supposed to involve Norwegian as the source language. Secondly, even if we consider Norwegian to be a regional variety of Danish,[7] such data could imply that we deal with a standard form that is supplied with a (reverse?) regional label. That the above interpretation of *næpe paa norsk* is flawed is obvious only to those who know both Norwegian and Danish, and are aware of the fact that *næpe* is the Norwegian word for the Danish *roe* (turnip), and the purpose of the semantic data provided for the latter in the dictionary is to acquaint the user with the regional variant of the headword. The question that remains open is of how much use such information might be for a user translating from Danish into English.

Other examples of usage information provided in Wolff's work are two easily identifiable cases of insult terms. Another one is the data attached to the lexeme 'bryst', specifying that the form of the headword is incorrect.[8] This marker of usage information would be classified as a status label (probably: *nonstandard*) according to Landau's typology.

Compared with the task of identifying and classifying usage information provided in Berthelson's dictionary, a similar analysis with regard to Wolff's work is much more difficult. Indeed, it becomes a little easier if we, to some extent legitimately, apply the labels used by Berthelson, but it still leaves us with more than a few cases that cannot readily be either unambiguously identified or placed within any modern typology.

## 4 Bay's English-Danish dictionary (1796)

This dictionary was to a significant extent not an original work, but a translation of Theodor Arnold's English-German dictionary (1757).[9] It consists of three parts: the dictionary proper and two lists of (mainly nautical) terms. In the first part, 59 lexemes are marked with more or less easily identifiable usage information, mostly subject field labels.

The subject field labelling here is in some ways similar to what we have seen in the two earlier dictionaries. On the one hand, it resembles the system found in Berthelson's work, as some of Bay's labels include the word *term* or *ord* ('word'), thus (ideally) referring to the lexeme rather than to the concept it denotes. On the other hand, Bay is as inconsistent as Wolff, in that several different labels are used to refer to the same domain, which is especially the case with words related to navigation. In the case of Wolff's dictionary we have solved this problem by trying to fit some of the apparently irregular labels into Berthelson's system. There is, however, no reasonable argument for doing the same with regard to Bay's labels, since the dictionary is in no way overtly related to Berthelson's work. The same problem pertains to field labels in the form of *i* + [domain], such as *i Spil*. We cannot be sure whether they carry usage information, or serve as sense discriminators or quasi-definitional information, or perhaps perform both functions.

Apart from such cases, Bay contains other problematic pieces of semantic data. Among these, there are 11 cases of what we have seen in Berthelson's dictionary, namely a piece of information in the form of *(siges) om...* ('(is said) about...'). An analysis of such data, however, will reveal that for 9 out of the 11 cases it would be pointless to establish any subject field domains with a view to identifying the data as usage information: they are most probably quasi-definitional information. Of the remaining two cases the more challenging one is that of the lexeme 'irrepleviable'. It is a derivative of the verb 'replevy', which is labelled in the dictionary as *Lov-Term* ('legal term'). Why is then 'irrepleviable' given a different label, namely: *siges om Gods belagt med arrest* ('is said about forfeited goods')? Whether the word was part of specialized language or not is difficult to assess, but a clear subject field label would be definitely of more use than the one provided by Bay.

Other problematic examples of semantic data included within the entries in the dictionary include three cases where reliable identification of the information is nearly impossible.

Judging by the content of the provided information, the first case seems to be a usage note. However, both the data and the lexeme it is attached to appear a bit unusual for the lexicographic practice: hardly any general bilingual dictionary could be expected to include words that are part of someone's (in this case, the Crown Prince of England's) idiolect. Perhaps a more credible suggestion would be that 'Ich-Dien' belongs to formal language of English Crown Princes in general, and by extension the provided information is a style label, but then the presence of this lexeme in a dictionary like Bay's would still be out of the ordinary.

**Table 5.** Bay 1796 (English-Danish): problematic pieces of semantic data (i)

| Lexeme | Information | Attached to | Equivalent | Cat. |
|---|---|---|---|---|
| Ich-Dien | Kron-Prindsen af Engellands Bal-Sprog ('feast language of the English Crown Prince') | equivalent | jeg tiener, eller see! Der er Manden, paa Bal, lisk | ? |
| swab, to | skibskost ('swabber') | headword | [en Svabbert] | ? |
| thane | et Angelsaksisk Ord ('an Anglo-Saxon word') | headword | [herre] | ? |

The second example, that of *skibskost* ('ship-mop', i.e. 'swabber') attached to 'to swab' is also puzzling. It appears to be, if anything, a synonym of the Danish equivalent, which in turn is not an adequate translation of the headword, as each of these belongs to a different grammatical category.[10] For some reason, however, the data seems to be attached not to the equivalent but to the lemma. Perhaps a plausible explanation as to why the label appears in the dictionary at all is that Bay might have wanted, by using the word *skib* ('ship'), to point to the fact that the English lexeme belongs to the domain of navigation. Whether it is part of specialized language, and the label is, in consequence, a piece of usage information proper is another matter.

The last example is even more misleading. Does 'Anglo-Saxon word' serve here as (the dictionary's only example of) etymological data, or is it to be interpreted as a piece of usage information, namely a currency label, specifying that the word is *so* old that it apparently is obsolete? Using both common sense and linguistic evidence from Present-Day English, it should, however, be concluded that the former of the above interpretations is improbable, whereas the latter is plainly incorrect: it is the concept of "thane", not the word that is obsolete. In this case, the provided label is, at best, unnecessary; at worst, it is misleading.

The last set of misleading pieces of semantic data found in Bay's dictionary which I will discuss here comprises 8 examples of what at first sight appears to be usage information in the form of status labels, informing the dictionary user that the forms of the lexemes they are attached to are substandard/nonstandard.

One must examine the whole dictionary and reach the last headword of 'won't', as well as consult a contemporary or historical dictionary to find out that what Bay possibly meant here by *i stedet for* ('instead of') were variant forms of the headwords. It is not improbable, though, that he considered at least some of these forms as substandard at the same time. The substandard status of 'prithee' and 'pox' is confirmed in Dr Johnson's dictionary (1785), one of the sources for Bay's work. As regards the other examples, here the *OED* comes in helpful: with the exception of the case of 'won't', nearly all quotations provided for these lexemes come from long before Bay's dictionary was published. We could therefore possibly assume that these words were archaic or obsolete in Bay's time. This indeed makes the label *i stedet for* appear to be multifunctional: as a marker of variant spelling, as a status label, and as a currency label. The last two are certainly usage information.

**Table 6.** Bay 1796 (English-Danish): problematic pieces of semantic data (ii)

| Lexeme | Information |
|---|---|
| fixt | i stedet for Fixed ('instead of fixed') |
| mo | i stæden for More ('instead of more') |
| pox | isteden for Pocks ('instead of pocks') |
| prithee | isteden for pray thee ('instead of pray thee') |
| tost | istedenfor tossed ('instead of tossed') |
| tother | istedenfor the other ('instead of the other') |
| vext | i stedet for vexed ('instead of vexed') |
| won't | istedetfor will not ('instead of will not') |

As regards the appendixes to Bay's dictionary, i.e. the lists of (mainly) nautical terms, identifying and classifying usage information provided there should not appear to be difficult. After all, these lists should by definition include words supplied with a subject field label if included in any regular dictionary. However, a closer look at them reveals that the task is far from simple.

**Table 7.** Bay 1796 (English-Danish): potential usage information in the list of nautical and mercantile terms

| Word/Phrase | Information |
|---|---|
| helm a lee | commando ('command') |
| nip the cable | [unmarked, but the Danish equivalent appears in the imperative] |
| holla! | Svar, at man har hørt Commandoen ('a reply confirming that one has heard the order') |
| roomer | saa kalder Søefolket et stort Skib ('so is called a big ship by seamen') |
| port the helm | [unmarked, but listed together with 'helm a lee'] |

The first list includes 1412 of "some of the most necessary terms of trade and navigation". Which of these expressions were part of specialized vocabulary of 18th century English is hard to assess, but at least a partial analysis is possible here. Those expressions which lend themselves to an analysis are presented in Table 7. Some of them bear the label *commando* ('command'), which clearly refers to the term, not the concept, and thus might imply that it was part of seamen's jargon. Some other phrases are provided with other information that refers to the given expression and not the concept it denotes. Still other ones are verb phrases that appear in the imperative. They are not marked with any labels, but their form implies that they are commands, and thus logically, the subject field label 'nautical term' that is implicitly attached to the whole list, should be interpreted here as a piece of usage information. Yet other phrases are listed together with some of the expressions marked with *commando*, and their Danish translations have indeed the form of commands, so they also seem to belong to the nautical jargon. All these can be thus considered usage information proper. However, in the case of a vast majority (1381) of the terms in the list, we cannot be sure of whether they are part of the nautical jargon or just denote concepts associated with the navy.

The second list comprises 320 "words of command and sea-terms by various occurrences". The list is not Bay's own, but was in its totality copied from Jacobus Serenius' English-Swedish dictionary (1757). Here, claiming that they are marked for usage rather than only for belonging to a particular domain should be entirely unproblematic. However, after a closer look we see that this is not so. First of all, some of the "commands" are non-lexicalized verb phrases which clearly belong to common vocabulary, like 'to work and command a man of war in an engagement'. Others are whole sentences, which are hardly part of specialized language or even lexicographic units, e.g. 'up youths in the tops, and mind your great chests and your small arms'. Finally, some expressions, like 'yea, yea', are hardly perceivable as any commands or sea-terms.[11]

Extracting and classifying usage information provided in Bay's English-Danish dictionary seems to share some of the difficulties we have encountered in the case of Berthelson's and Wolff's works, notably the inconsistency in labelling. Additionally, some of Bay's data are hardly decipherable. Let us now conduct a similar analysis with regard to the Danish-English dictionary by the same author.

## 5 Bay's Danish-English dictionary (1798)

This work appears to include a typological sum total of the difficulties encountered in the investigation of the other three dictionaries.

Like in the dictionaries discussed before, also here the editor did not seem to be consistent in labelling. Again, the inconsistency is most conspicuous with regard to subject field labels relating to navigation.

An extract from other problematic semantic data is presented in Table 8.

As in his English-Danish dictionary, also here Bay provides a number of labels in the form of *i* + [domain], e.g. *i Spil* ('in games'). The problem with identifying such information has already been touched upon. In the case of this dictionary, the dilemma of whether to treat such labels as usage information or other semantic data is, however, even more clearly visible, since some of the lexemes marked with such field labels, notably 'construction' and 'componist' ('composer'), definitely do not belong to specialized language.

Even a cursory glance at Table 8 reveals yet another similarity to Wolff's system of providing (potential) usage information, and the problems stemming from its nature, namely the presence of labels including the word *hos*. Here, however, an assumption that all such labels contain usage information would be absurd, since such data as *hos hedningene* ('among the pagans') definitely do not refer to the use of the lexemes they are attached to, i.e. 'krigens Gud' ('Mars, the God of war') but to the concepts these words denote.

Another example of a problematic, or rather misleading, field label included in Table 8 is that of *bibelsk* ('biblical'), attached to the headword 'concordance'. In modern dictionaries this label is normally placed next to words and expressions that either denote a concept associated with the Bible, or occur there in a unique sense. Here, however, neither seems to be the case. And even if the only concordances found in 18th century Denmark were those of the Bible, then the label is to be understood as quasi-definitional information rather than a usage note.

**Table 8.** Bay 1798 (Danish-English): potential usage information

| Lexeme | Information | Attached to | Equivalent | Category |
|---|---|---|---|---|
| baghaand | i Spil ('in games') | headword | the younger hand | ? |
| befrielseseed | i Kirkesager ('in ecclesiastical matters') | headword | a canonical purgation | ? |
| blinde | hos Krigsfolk ('among the military') | headword | a blind to open trenches | ? |
| componist | i Musik ('in music') | headword | a composer or setter to music | ? |
| concordants | bibelsk ('biblical') | headword | a concordance of the Bible, an alphabetical index of all Biblewords and phrases and it's places | ? |
| christne | geistligt Ord ('spiritual word') | headword | to christen, baptize a child | ? |
| flonel | fiint engelskt ('fine English') | equivalent? | swanskin | ? |
| indslag | hos Skræderne ('at [the] tailor's' OR: 'among tailors') | headword | the turning down | ? |
| krigens gud | hos hedningene ('among the pagans') | headword | Mars, the god of war | ? |
| messing | gammelt ('old') | headword | shruff | ? |
| poet | gammel engelsk ('old English') | equivalent | a bard | ? |
| sirup | i daglig Tale kaldes Sirup ('in everyday speech is called sirup') | headword | treacle | ? |

In the case of a few pieces of semantic data provided in the dictionary it is difficult to determine what they actually refer to. Two such pieces, namely *gammel engelsk* ('old English') and *gammelt* ('old') resemble modern currency labels. Only a closer examination reveals that neither of them can be interpreted as *archaic* or *obsolete*. The former refers to the concept of the equivalent, while the latter specifies the headword. Indeed, 'bard' was not gone from the English language in the 18th century, and 'shruff' could be rendered in Danish as 'gammelt messing'.

Like in his English-Danish dictionary, also here Bay includes a few cases of "cryptic" semantic data. One example is that of the label *geistligt Ord* ('spiritual word') attached to the lexeme 'christne' ('to christen, baptize a child'). No doubt, its form suggests it is a usage note. But what Bay actually meant by 'spiritual word' we can only guess at. In another case, the word 'sirup' ('treacle') is provided with a label *i daglig tale kaldes sirup* ('in everyday speech is called *sirup*'). It appears to be a style label, but its illogical content needs no further comment.

## 6 Conclusion

Extracting and classifying usage information provided in the four dictionaries investigated here is far from easy. Some of the problems the researcher is likely to face are rather general and stem from the structure of the dictionaries. First of all, none of the four works has a rigid system of marking usage information. Due to identical typographical features of all semantic information provided there, it is thus often a hard task to distinguish usage information from other semantic data. Secondly, in all these dictionaries misprints and mistranslations (all of which have been retained in the presentation of the data in this paper) are not uncommon, which sometimes makes it difficult to identify the provided information correctly.

Other problems are more dictionary-specific, like "cryptic" information and obscure reference of certain labels. Lastly, the controversies over the nature of the subject field label and the regional label have not been solved until today. In the case of the former, Kalverkämper postulates two alternative sets of labels to mark the fact that a given word either denotes a concept associated with a certain domain or that it is part of the specialized vocabulary associated with this domain (1989: 685). Such a solution is, however, seldom applied, mostly for reasons of the space it would take up in the dictionary. A similar problem concerning the regional label was touched upon, among others, by Norri in his article (1996) on regional labelling in some British and American dictionaries. A method, or rather labelling practice, that would enable the dictionary user to distinguish between regional words and regional concepts is still to be developed.

Some of the questions that may be worth asking in the light of the analysis I have undertaken are: is "borrowing" of a system of labelling from one dictionary for the investigation of another one legitimate, even if the two works are meant to be counterparts? Is it justifiable to postulate new domains of subject fields? Last but not least, is Svensén's policy of labelling subject fields fully reliable for the purpose of our investigation?

Any answer to the above questions would be much more credible if we had corpora of 18th century Danish and English, and even more so if the authors of the investigated dictionaries had had access to them. However, given their absence, the researcher must rely to a considerable extent on his/her own intuition and common sense. Therefore, the answer is probably: yes.

## Notes

**1.** For a complete bibliography of such dictionaries up to the year 1980 see Haugen (1984).

**2.** Svensén (1993: 181, 183) argues that subject field labels can also perform this function. Still, both kinds of information are clearly not identical in their purposes, especially if a given lexeme has only one sense listed in the dictionary.

**3.** See Svensén's approach to usage labels outlined below.

**4.** Reliable support for such a postulate would definitely require a comprehensive psycholinguistic analysis, which is far beyond the scope of this paper.

**5.** For a similar observation of the problem and comments on the lexicographical practice connected with it, see Kalverkämper (1989: 685).

**6.** See, for example, Norri (1996: 17–20).

**7.** The treatment of Norwegian words as dialectal in this Danish dictionary is fully legitimate for several reasons. First of all, at that time Denmark and Norway were politically and to a large extent linguistically uniform, Danish being the official written language. Secondly, Wolff's (and Berthelson's, too) mother tongue was the Danish spoken and written in Norway, as both authors were Norwegians (Kabell and Lauridsen 1994: 283).

**8.** Incidentally, this form is the only correct one in modern Danish.

**9.** See Ferrall and Repp (1845: ix-xv) for (exaggerated) criticism of all Bay's dictionaries on account of their being mere acts of piracy, as well as Kabell and Lauridsen (1994: 308–310) for a response to this criticism.

**10.** Both Danish words are nouns, whereas the English headword is a verb.

**11.** Cf., though, the PDE 'aye, aye (, Sir)!', clearly marked with usage labels in various dictionaries.

## References


Arnold, Theodor. 1757. *A complet vocabulary English and German oder vollständig kleines Wörterbuch englisch u. deutsch.* Leipzig: n.p.

Arnold, Theodor and Christian Friderich Bay. 1796. *A compleat vocabulary, English and Danish: Fuldstændigt engelsk og dansk haand-lexicon, hvorudi indeholdes alle i det engelske sprog brugelige ord med deres saavel korte som lange accent, samt tilføyede udtale; Udgivet paa engelsk og tydsk af Theodor Arnold; nu paa dansk oversat, forbedret og forøget med en samling af handels- og søe-termini og talemaader af Christian Friderich Bay*. [*Complete English and Danish manual-lexicon in which are contained all such words as are used in the English language, with their short as well as long emphasis, to which is added the pronunciation; edited in English and German by Theodor Arnold; and now translated into Danish, improved and* *enlarged with a collection of mercantile and naval terms and phrases by Christian Friderich Bay*.] (2nd impr. ed.). Copenhagen: Gyldendal.

Bay, Christian Friderich. 1798. *Fuldstændigt dansk og engelsk haand-lexicon, udarbejdet efter de nyeste og bedste ordbøger af Christian Friderich Bay* [*Complete Danish and English manual-lexicon, compiled from the newest and best dictionaries by Christian Friderich Bay*.] Copenhagen: Gyldendal.

Berthelson, Andreas. 1754. *An English and Danish dictionary: containing the genuine words of both languages with their proper and figurative meanings, interspersed with a large variety of phrases, idioms, terms of arts and proverbial sayings, collected from the most approved writers*. London: Printed for the author by John Haberkorn.

Ferrall, James Stephen and Thorleifr Gudmundson Repp. 1845. *Dansk-engelsk ordbog. A Danish-English dictionary*. Copenhagen: Gyldendal.

Haugen, Eva L. 1984. *A bibliography of Scandinavian dictionaries. (With an introduction by Einar Haugen)*. New York: Kraus International Publications.

Johnson, Samuel. 1785. *A dictionary of the English language in which the words are deduced from their originals and illustrated in their different significations by examples from the best writers to which are prefixed a history of the language and an English grammar: Vol. 1–2.* (6th ed.). London: Printed for J. F. and C. Rivington [etc.].

Kabell, Inge and Hanne Lauridsen 1994. "English-Danish/Danish-English dictionaries in the second half of the 18th century. A discussion of the authors, their lives and their production". In *Symposium on Lexicography VI. Proceedings of the Sixth International Symposium on Lexicography May 7–9, 1992 at the University of Copenhagen*, Hyldgaard-Jensen, Karl and Viggo Hjørnager Pedersen (eds), 283–314. Tübingen: Max Niemeyer Verlag.

Kalverkämper, Hartwig. 1989. „Diatechnische Markierungen in allgemeinem einsprachigen Wörterbuch". In: *Wörtherbücher. Ein internationales Handbuch zur Lexikographie.* Vol 1., Hausmann, Franz Josef (ed.), 680–688. Berlin: de Gruyter.

Landau, Sidney I. 1984. *Dictionaries. The art and craft of lexicography*. Cambridge: Cambridge University Press. [A second edition was published in 2001.]

Norri, Juhani. 1996. "Regional labels in some British and American dictionaries". *International Journal of Lexicography* 9, 1: 1–29.

*OED* = *Oxford English Dictionary*. On-line edition. Accessed at: http://dictionary.oed.com

Serenius, Jacobus. 1757. *An English and Swedish dictionary: wherein the generality of words and various significations are rendered into Swedish and Latin*. (2nd ed.). Harg: Stenbro.

Svensén, Bo. 1993. *Practical lexicography. Principles and methods of dictionary-making.* (Translated from Swedish by John Sykes and Kerstin Schofield). Oxford: Oxford University Press. [A thoroughly revised Swedish edition appeared in 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. 2nd ed. Stockholm: Norstedts Akademiska Förlag.]

Wolff, Ernst. 1779. *En dansk og engelsk ordbog: sammenskrevet af Ernst Wolff, medlem af det höylovlige lands-huusholdnings sælskab udi Kiöbenhavn, og kiöbmand i London*. [*A Danish and English dictionary: compiled by Ernst Wolff, member of the most illustrious Agricultural Society of Copenhagen, and merchant in London*.] London: Frys, Couchman, and Collier.

CHAPTER 18

# Dictionaries and culture

Roda P. Roberts

Dictionaries have as their primary purpose to provide information on a language or languages. As such, they are first and foremost linguistic tools. But, as scholars such as Dubois & Dubois have clearly indicated (1971: 8), dictionaries are also "cultural objects, integrated as such into a culture: they bear witness to a civilization." In other words, dictionaries present not only language, but also culture. Language represents culture because words refer to a culture. Therefore, dictionaries, which constitute an archive of the words of a language, present, *de facto*, the culture underlying the language.

In this paper, I will first consider different perspectives on and aspects of culture. In particular, I will focus on the distinction between national culture, transnational culture (or "superculture") and subculture. I will then examine how these different aspects are reflected in dictionaries.

## 1 Introduction

Dictionaries, which are organized collections of lexical items accompanied by pertinent information (meaning, pronunciation, examples of use, translation equivalents if the dictionary is bilingual, etc.), have as their primary purpose to provide information on a language or languages (Rialland-Addach 1995: 91). As such, they are first and foremost linguistic tools. But, as scholars such as Dubois & Dubois have indicated (1971: 8), dictionaries are also "cultural objects, integrated as such into a culture: they bear witness to a civilization."[1] In other words, dictionaries present not only language, but also culture.

If the relationship between dictionaries and language is obvious, the link between dictionaries and culture is much less so. This link is indirect in that it is established through language. It is based on two postulates: (a) that language and culture are intimately related (Rialland-Addach 1995: 91) and (b) that culture is mediated by language (Nida 2001: 25). The first postulate is based on the fact that language is a part of culture (a cultural institution) and shares several similarities with it. Some of the similarities are the following: both language and culture are acquired very early in life and in the informal settings of home and schoolyard; no one person ever completely controls a language or a culture, which are both collective enterprises; both language and culture vary from person to person and from group to group; both these entities change over time, although often slowly and sometimes superficially. (Nida 2001: 14–23) The second postulate, that culture is mediated by language and that language is therefore essential for the functioning and perpetuation of culture is based on the fact that information about the processes and values of a culture can only be provided through the use of language. Indeed, Nida (2001: 25) outlines eight different uses of language by culture, including the one just mentioned.

Language represents the culture because words refer to the culture. Therefore, dictionaries, which constitute an archive of the words of a language, present the culture underlying the language. In this paper, I will first consider different perspectives on and aspects of culture and then examine how these different aspects are reflected in dictionaries.

## 2 Perspectives on and aspects of culture

### 2.1 Definition of culture

What exactly constitutes culture? The complexity of this concept has already been revealed by the statements above that language is a part of culture but that language represents culture, which would not survive without it. First, we have to distinguish between the concept of culture as it is often used in everyday language, and the concept as it is used by sociologists. In the former case, the term "culture" normally refers, in a relatively narrow way, to activities relating to music, art, literature, cinema and the like, and more specifically to the cultural products created in these areas (e.g. Shakespeare's plays, which have come to symbolize English cultural development).

Sociologists, on the other hand, have adopted a much wider definition of the concept of culture (although activities such as music, art and so forth are included within this definition), mainly because they want to use it to say something about the overall nature of a society. They use the term to refer to every aspect of a society that involves the production and consumption of ideas about the nature of society, the individual, their relationship and so forth. In basic terms, therefore, a wider interpretation of the concept involves the idea that a "culture" is a "way of life", transmitted from one generation to the next. But although sociologists generally agree about the need for a wider definition of culture than that used in everyday social interaction, this is not to say that there are not differences of interpretation involved in relation to the most significant features of the concept. In other words, there is little consensus on a definition of culture, even among specialists.

### 2.2 Types of culture

However, a clearcut definition of culture is not essential for our purpose. More relevant for our needs is to distinguish between different types of culture.

The first distinction to be considered is one that most sociologists make between *material culture* and *non-material culture*. The former covers the material things a society creates, the physical objects (cars, telephones, computers, etc.) that a society produces to reflect its interests and preoccupations. The latter covers the non-material things a society creates, the knowledge, habits and beliefs that influence people's behaviour and which they learn, share and transmit from generation to generation. Among the non-material or intangible elements of culture lies language, which has been characterized as "the most distinctive feature of a culture" (Nida 2001: 13).

The close ties between language and culture presented above has led scholars to refer to a "language-culture" (Nida 2001: 15). This raises the obvious question: is there one single culture underlying every language? Just as there is often more than one nation with a given language, so too there is often more than one culture per language, and in other cases, a culture may be unrelated to a specific language. A distinction therefore needs to be made between *national culture*, *superculture* (or *transnational culture*), and *subculture*.

The notion of national culture has long been a dominant concept in the discussion of cultural products. It is based on the ideology of nationalism, according to which human societies should be organized in politically sovereign states populated by a single nation. The nation is in turn defined as a spontaneous association of humans bound together by shared language, culture, ethnicity and beliefs. Nationalist theory stipulates a normative congruence of political institutions, economic activity, and cultural identity and experience. Nationalists tend to believe that without such congruence communities are unstable and their members do not feel "at home." It is a belief which, as Gellner states, strives "to make culture and polity congruent, to endow a culture with its own political roof and not more than one roof at that" (Gellner 1983: 43). Examples of national culture would be Japanese culture or Danish culture or American culture. However, today fewer and fewer nation states are as homogeneous ethnically, socially and linguistically as Japan. Even the United States, with its melting-pot approach to immigrants, is culturally quite diverse. And American culture has also extended beyond its political borders and influenced other national cultures. Thus, although the concept of national culture still exists, it must be redefined to some extent as the culture of a group sharing a geographical location, a political identity, and one or more "official" languages, but not necessarily ethnicity or native language. The concept of national culture also needs to be supplemented by those of superculture (or transnational culture) and subculture.

A superculture (or transnational culture) is, quite simply, a culture that transcends national boundaries. Thus, Hispanic culture is the culture of Spanish-speaking peoples wherever they may live. This superculture is based primarily on linguistic unity. However, the fact that we can speak about a Western European culture despite the fact that many different languages are spoken in Western Europe points to the existence of a superculture that transcends language per se. Moreover, the concept of superculture does not have to be linked with either language or contiguous geographical area. Kenneth Boulding, an American economist, postulated a modern superculture based upon technology and technological innovations (1968: 170). And, in his paper entitled "Global Culture, Local Cultures, and the Internet: *The Thai Example*" (1998), Soraj Hongladarom addresses the all-important question of whether the Internet represents an all-unifying force, turning all cultures within its domain into one giant superculture where everything becomes the same.

At the other end of the spectrum from supercultures are subcultures, i.e. cultures that are more limited in scope than national cultures. The sociological term "subculture" has become a widely used concept in recent years; yet, with all of this usage, it has received very little critical attention. It has often been construed as synonymous with deviants and their activities (e.g. punk culture), or the "sub" has somehow been interpreted to mean inferior. There is no denying that the subculture concept is applicable to the study of deviance. However, here "subculture" refers to the culture of any group of people within a national culture or even a superculture, whose behaviors and beliefs include some of the dominant features of this culture or superculture, and also include certain features not found elsewhere in the given national culture or superculture. As Thomas Lasswell (1965: 211) has pointed out:

> (...) every group that is at all functional must have a culture of its own that is somewhat similar to the cultures of other groups with whom it interacts. Such a group culture is not partial or miniature, it is a complete, full-blown set of beliefs, knowledges, and ways for adjustment to the physical and social environment.. The culture itself is not smaller than the great culture (...) the group which enacts it is smaller than the great society.

Subcultures are usually characterized by the fact that individuals can function within more than one subculture at one time, with varying degrees of commitment, and sometimes pass through several stages as they grow older or their attitudes change. Subcultures can be identified by age (e.g. youth subculture), ethnicity (Black subculture), religion (Baptist subculture), language variety (Canadian French subculture), and geographical area within a given country (e.g. the Canadian North), among other criteria.

## 3 Objectives and methodology of study

In the following sections, I will first examine what dictionaries say regarding their coverage of culture and then what they actually do in terms of treatment of culture. More particularly, I will analyze what types of culture (material and non-material, superculture, national culture and subculture) are presented in dictionaries and how they are presented in dictionaries. I will limit the analysis of superculture in dictionaries to that related to language or contiguous geographical area and the examination of subcultures to those related to geographical areas within a given country.

My analysis will be based on the examination of two English dictionaries: Random House Webster's Unabridged (RHWUD, 1999) and Oxford Canadian (OXCAN = Barber (ed.) 2001) and two bilingual English-French Dictionaries: Robert-Collins Senior (RCS = Atkins et al. 1996) and Oxford Hachette (OXHA = Corréard & Grundy (eds.) 2001).[2]

## 4 The Dictionary's intentions regarding culture

The four dictionaries examined all make some reference to coverage of culture-related words in their front matter (preface, introduction, guide to using the dictionary, etc.), although not all to the same extent. The unilingual dictionaries are more explicit on this question than the bilingual dictionaries, but even the former vary in their emphasis on culture.

Of the two unilingual dictionaries, the OXCAN provides the most information on its approach to culture – national culture, superculture and subculture. As far as national culture is concerned, the OXCAN, a Canadian dictionary, focuses primarily on Canadian culture, although its list of geographical labels indicate that other national cultures such as that of the US and Britain as well as Australia, New Zealand and others are not totally ignored. With regard to Canadian culture, the editors indicate that special attention has been paid to the vocabulary of economic activities (e.g. logging and mining) and sports and pastimes (e.g. hockey and hunting) of particular interest to Canadians, although other distinctively Canadian vocabulary such as obscure or obsolete English words used to describe political realities (e.g. *reeve, riding*) are by no means ignored. The distinction between the national culture and a superculture blurs to some extent, however, since "the broader base of Canadian ethnicity demands a broadening of vocabulary" (OXCAN 2001: x): thus, the OXCAN presents words borrowed from other cultures (e.g. those denoting foods such as *bruschetta*), due to Canada's highly multicultural society. One specific superculture presented is that of North America – this is indicated by the fact that the geographical label *N Amer* is used to identify the vocabulary used chiefly in Canada and the US. Finally, supercultures that go beyond geographical areas, such as that of technology and more especially the Internet, are represented by the presence of new words or the addition of new meanings, as exemplified by *Web site, bookmark, browser,* and *firewall*. In addition to presenting Canada's national culture, the OXCAN also aims to present the subculture of certain groups within Canada. It claims, for instance, to have thoroughly researched the culture of Canada's aboriginal peoples. Moreover, its detailed list of geographical labels, which includes a large number of regions of Canada (e.g. Cape Breton, BC, North), indicates that it is interested in the language of groups residing in different areas of the country.

The American dictionary RHWUD provides far fewer details on its intentions regarding culture. As far as national (US) culture is concerned, the only (and very limited) insight is found in the presentation of geographical labels, where it is indicated that entries not labelled are considered to be in general use throughout the US. More specifics are provided for supercultures, primarily in the context of new words in the dictionary. Thus, we are told that neologisms such as *intranet* and *V-chip* have largely arisen as a result of scientific and technological advances. Specific reference is made to cultural movements and their influence on language:

> During the past two decades significant and influential social and cultural movements have taken place, reverberating through our language not only as new words and meanings but in our attitudes toward language and its use. Thus recent influences that have expanded and reshaped our vocabulary and its use come not only from such obvious fields as science and technology (...), but also from history itself and from such social and cultural movements as concern with the environment, the women's movement, and a new awareness of and respect for ethnic diversity.

As far as subcultures are concerned, the only mention is of regional subcultures through reference to specific regional labels as well as the label Dial., which is used to mark a few terms, especially those with a rural flavor like *agin*, which are too widespread to warrant a specific regional label, but not in general use throughout the US. More significant, however, is the reference to Regional Variation Notes included at suitable points in the dictionary, which are described as brief dialect studies discussing different regional terms for the same item or different regional forms of the same word.

As indicated above, the front matter of bilingual dictionaries is much more reticent on the topic of culture than the unilinguals. The more vocal of the two examined for this paper, the OXHA, manifests briefly, although clearly, its interest in culture, through the following statement in the revised 2001 edition:

> Thousands of new words, phrases and meanings have been added, reflecting the rapid changes in language and culture of the last few years. New developments in every area of contemporary life require new means of expression. (OXHA 2001: viii)

However, while some mention is made of the superculture of the Internet, little interest is shown in subcultures. The primary focus seems to be on the national cultures of Great Britain, the US and France: this is indicated by mention of the language varieties covered and of specific Cultural Notes devoted to aspects of French, English and American society, institutions and culture.

If the front matter of the OXHA speaks only briefly of its interest in culture, that of the RCS is virtually silent on the matter. It contains only a passing reference to maps, added to the new version of the dictionary, that concentrate largely on the French-speaking world in the French-English volume and on the English-speaking world in the English-French volume, thus indicating some minimal interest in the anglophone and francophone supercultures. The only other allusion to cultural issues is found in the list of abbreviations, which includes abbreviations for countries (tied to national cultures) such as Australia, Canada, Belgium and the US, and for more limited regions (linked to subcultures) such as Scotland and the North of England.

To sum up, dictionary front matter is not very revealing when it comes to the dictionary's focus on culture. The link between culture and language is generally made in the discussion of new words added to the dictionary. While supercultures, national cultures and subcultures are alluded to, the emphasis is on national culture(s), with the only subcultures of interest being those related to geographical regions of the dictionary's country of origin. And no clearcut distinction is made between cultural realities per se, i.e. the institutions, habits or practices, concepts, and attitudes of a cultural group, as well as the persons who belong to that culture, and linguistic particularities (especially words and senses) that distinguish the group but which do not designate cultural realities specific to the group.

## 5 Cultural words in the dictionary

A study of dictionary entries for the letter A in the four dictionaries mentioned above is a little more revealing about each dictionary's approach to culture. I looked for entries revealing the culture of the dictionary's country of origin, for those dealing with geographical supercultures and geographical subcultures. I first examined the nomenclature to establish the number and type of cultural words included.

## 5.1 RHWUD

To begin with, I examined the CD-ROM version of the RHWUD, searching for all entries under A that contained "US" or "America" (national culture), "North America" (geographical superculture), and "US" with a qualifier such as "Southern" (geographical subculture).[3]

### 5.1.1 *National culture*

There are 224 entries devoted to some aspect of the US as a whole, although the total number of references is greater since some of these entries contain more than one subdivision referring to the US. Among these, those dealing with plants, animals and physical locations were eliminated, since these are not normally considered cultural realities. The same applies to famous Americans or people involved in US activities. This left a total of 106 entries, which cover the following main areas:

- US armed forces: divisions, ranks, weapons/equipment
  e.g. *air division* – US air force unit of command
  *adjutant general* – US army rank
  *Apache* – US army helicopter
- US history: events, political parties, members of groups
  e.g. *abolition* – the legal prohibition and ending of slavery, esp. of slavery of blacks in the US
  *Antifederal party* – party that first opposed adoption of the US constitution and later favored its strict construction
  *Antifederalist* – members or supporters of the Antifederal party
- US government: divisions, members, procedures
  e.g. *administration* – the executive branch of the US government as headed by the President and in power during his or her term of office
  *Attorney General* – head of the US Department of Justice
  *apportionment* – determination of the number of members of the House of Representatives
- US space industry: spacecraft, rockets
  e.g. *Apollo* – one of a series of US spacecraft
  *Aerobee* – US rocket
- Artistic/literary movements originating in the US
  e.g. *abstract expressionism* – movement in painting originating in the US

- US organizations
  e.g. *AAA* – American Automobile Association
- US attitudes and ideals
  e.g. *American Dream* – 1. the ideals of freedom, equality, and opportunity traditionally held to be available to every American. 2. a life of personal happiness and material comfort as traditionally sought by individuals in the US

Relatively few of the cultural words refer, surprisingly, to material culture. Those few are primarily related to rockets and military awards. The majority of the cultural words deal with non-material culture, related to the major categories indicated above and especially to the US armed forces and government.

While we have eliminated biographical entries as not being cultural in themselves, one has to take them into account if one wants to obtain a better idea of American life and society, since the activities of these Americans reveal, albeit indirectly, the vigour of American literature, music, painting, and even government. Thus, the very number of famous American writers (13 among the A's) shows the vibrancy of the literary scene.

#### 5.1.2 *Superculture*

If adequate attention is certainly paid to the national culture of the US, the same does not seem true of the superculture of North America. In fact, of the 26 entries containing the word "North America", only two (those for *America* and *American*) can be considered cultural entries per se, although there are two biographical entries that also mention the vaster region. Another dozen or so entries that we have identified as cultural mention the US along with Canada or Mexico; thus, in the entry for *ABC Powers*, we are told that this item refers to "Argentina, Brazil, and Chile, usually with reference to their mediation between the US and Mexico in 1914". The limited number of entries related to the superculture of North America reveals the lack of importance that this superculture has to the editors of the RH-WUD and possibly to Americans as a whole.

#### 5.1.3 *Subcultures*

However, they seem to have a much greater interest in their regional subcultures, as denoted by the 30 plus entries containing the word "US" along with a modifier such as "Northern", which we have identified as culturally pertinent. What is particular about these entries is that they all focus on linguistic particularities of the region, i.e. they present the way things are said in the region, although the concepts designated are not necessarily – or even usually – regional. Thus, the word *asequia* is marked as being used in Southwestern US to denote an irrigation ditch;

although irrigation ditches are found throughout the US (and indeed, throughout the world), the use of this word of Spanish origin in Southwestern US reveals the Chicano influence in that region.

To sum up, then, the RHWUD seems interested in everything American, both national and regional, but much less so in the wider North American context.

## 5.2 OXCAN

As we have seen above, the Canadian dictionary OXCAN speaks the most in its front matter about culture and claims to pay particular attention to the culture of Canada. I examined its entries in A using the keywords "Canada" for national culture, "North America" for superculture, and "Canada" with a modifier such as "Western" or the names of provinces and territories for subcultures.

### 5.2.1 *National culture*

As far as entries directly related to national culture, I identified a total of 52, after excluding those designating places, plants, animals and people. These cover the following main categories:

- Canadian politics: voting, officials
  e.g. *ADM* – Assistant Deputy Minister
  *advance poll* – an early poll for voters who expect to be absent from their riding on election day
- Canadian armed forces: primarily ranks
  e.g. *able seaman* – a non-commissioned officer of the second-lowest rank in the Canadian Navy
  *air command* – the official name for the Canadian air force
- Canadian history: events, political parties, members of groups
  e.g. *air chief marshal* – the highest rank in what used to be the Royal Canadian Air Force
- Canadian organizations
  e.g. *ACCC* – Association of Canadian Community Colleges
  *Academy of Canadian Cinema and Television*

As in the case of the RHWUD, the majority of the cultural words deal with non-material culture, related to the major categories indicated above. In the OXCAN, at most two cultural words can be considered as referring to material culture:

> *Arborite* (Cdn proprietary) a plastic laminate used in countertops, tables, etc.
> *advanced green* (Cdn) a flashing green traffic light in advance of the steady green light, indicating that oncoming traffic is halted

Finally, the OXCAN, like the RHWUD, contains some biographical entries which, although they are not included in this analysis, can serve to give a better idea of Canadian life and society.

#### 5.2.2 *Superculture*

At first sight, it seems as if the OXCAN gives better coverage to North American culture than Canadian culture. Indeed, my search for entries containing the indication "North American" revealed 82 such entries. And various combinations of "Canadian", "British" and "Australian" turned up another 6 entries. 47 of them only provided a North American linguistic equivalent for realities not unique to North America: for example, the word *apartment* for the British *flat*. The rest covered cultural realities as diverse as *Abolition* (the abolition of slavery in the British Empire or the US), *Abenaki* (a member of an Algonquian-speaking Aboriginal people, as well as the language and culture of the Abenaki), and *American Sign Language*.

#### 5.2.3 *Subcultures*

28 entries were labelled with the name of one or more Canadian provinces or regions and were thus considered entries representative of subcultures. Again, a large number of them (11) provided linguistic equivalents used in a given province or region to designate a reality that is not, however, typical of that area; for example, the word 'angle' is used in Newfoundland to designate a curved inlet in a lake or pond. The rest of the entries covered cultural realities ranging from ethnic groups to ranks in provincial police forces to specific regional practices, as the following examples show:

- *Abegweit* – a member of an Algonquian band living on Prince Edward Island
- *Assistant Director General* – an officer of the Sûreté du Québec ranking above chief inspector and below director general
- *annual allowable cut* – the volume of wood that may be cut each year in a specified area (in British Columbia)

To sum up, the OXCAN is interested in all types of cultures: national culture, superculture and subcultures. Although the relative number of entries seems to suggest that more attention has been paid to the North American superculture than the national Canadian culture, one should not jump to the conclusion that this dictionary is less interested in Canadian culture. The more limited number of Canadian entries can be explained by the fact that Canadian culture outside of the French-speaking province of Quebec is to a large extent synonymous with North American culture, with only a few elements particular to Canada. My brief analysis of cultural entries beginning in A merely confirms this reality.

## 5.3 RCS

Since the RCS is published in Great Britain and France, I presumed that the national cultures given priority would be related to those countries. As I was using the English-French section of the dictionary I searched for entries designating the national culture using "British" and "Great Britain" as the key words. For the superculture, "British/Great Britain" and "US" were sought in combination to see if there was any presentation of a Western anglophone superculture, while for the subcultures "British/Great Britain" and a narrower regional focus (e.g. Wales) were looked for.

### 5.3.1 *National Culture*

A total of 84 entries contained the indication "British/Great Britain", although the total number of cultural references is greater since sometimes there is more than one reference per entry. At least 44 of these entries present British linguistic peculiarities (e.g. *advert*, British abbreviation of "advertisement"; *aerial*, for telecommunications "antenna"). Therefore, only 40 entries present other facets of British culture. The main areas covered are the following:

- British education system: exams/diplomas, institutions
  e.g. *A levels* – qualification to enter university
  *academy of music* – music school
- British government: divisions, royalty-related matters
  e.g. *Ministry of Agriculture* – a government department
  *oath of allegiance* – oath to the Queen
- British organizations
  e.g. *AA* – British Automobile Association
  *Advertising Standards Authority* – organization controlling advertising standards

It should be noted that none of the cultural words refer to material culture. All deal with non-material culture. Also, in this bilingual dictionary as in almost all such dictionaries, there are no biographical entries to give a better idea of British life and society.

### 5.3.2 *Superculture*

As far as superculture was concerned, I wanted to see if a British dictionary covered what was common to the Anglophones of the Western hemisphere. So I searched for entries containing mention of both Great Britain and the US. There were 19 such entries. However, none of them presented what was common to both cultures. Some of them showed usage of a word that was different in Britain and in the US (e.g. *apartment* – refers to a room in British usage and to a flat in US usage). Others related to a similar reality referred to by different words (e.g. *Ministry of Agriculture* – British institution, vs. *Department of Agriculture* – US institution). Overall in this dictionary there seems to be more of a desire to show what is different between Britain and the US than what is the same. Thus, the idea of a geographical superculture does not seem to exist here.

#### 5.3.3 *Subcultures*

Although I searched hard for coverage of geographical subcultures, by scanning for "British/Great Britain" and a modifier such as "Southern", I found no such entry at all.

In conclusion, one can state that the RCS places more focus on language than on other cultural realities. Overall it does not seem to be interested in a geographical superculture and ignores regional subcultures. Thus, from a cultural standpoint, this dictionary is disappointing.

### 5.4 OXHA

Since, like the RCS, the OXHA is published in Great Britain (and also in France), the search for entries designating the national culture used "Great Britain" as the key word. For the geographical superculture, "Great Britain" and "US" were sought in combination, while for the subcultures "Great Britain" and a narrower regional focus (e.g. Wales) were looked for.

#### 5.4.1 *National culture*

A record 183 entries containing the label "GB" are found under the letter A in the OXHA. 119 of this total presented British linguistic equivalents (e.g. *advert* for American *ad*). The rest of the entries covered the following main areas:

- British associations and organizations
  e.g. *ABTA* – Association of British Travel Agents
- education
  e.g. *access course* – a course that allows candidates who do not have the required prerequisites to enter university
- social security
  e.g. *attendance allowance* – allowance paid by the State to someone looking after a severely handicapped relative

The OXHA, like the RCS, contains no biographical entries that could help the dictionary user form a clearer idea of British culture.

#### 5.4.2 *Superculture*

Although there were a certain number of entries containing both the GB and US labels, none of them presented any aspect of a geographical superculture. In all cases, the entry contained both British and American senses or usages.

#### 5.4.3 *Subcultures*

No entries containing GB plus an indication of a region (e.g. Midlands, South, North) were found. Hence it is safe to assume that subcultures are not important for the OXHA.

The OXHA makes a real effort to present national culture, but like most bilingual dictionaries ignores geographical subcultures and also neglects the geographical superculture.

### 5.5 Conclusions on dictionary nomenclature

On the basis of the study of the cultural nomenclature beginning with A in four dictionaries, the following tentative conclusions can be drawn.

In all dictionaries, the cultural nomenclature consists primarily, if not uniquely, of words pertaining to non-material culture. Few words, if any, designate material culture at the national, transnational or subculture level. The only reason that I can think of for this state of affairs is that objects of material culture are no longer limited today to a given region or people but are found globally. Given that I have not attempted such a global analysis, material culture seems to have little or no importance in this study.

In all four dictionaries, a fairly large proportion of cultural words show linguistic particularities rather than designate other facets of culture. These linguistic particularities include the following:

- word (e.g. *antenatal* GB = *prenatal* US)
- spelling (e.g. *aluminium* GB, as opposed to *aluminum* US)
- pronunciation
- sense (e.g. *agency* – esp. NAmer: a government office providing a specific service)
- usage (e.g. *all* – GB *what with the heat and all*)

Linguistic particularities dominate in the presentation of elements of regional subcultures, perhaps because language varies more from one region to another than other aspects of culture. However, linguistic particularities do not seem to be an important part of geographical supercultures on the basis of three of the four dictionaries studied, the only exception being the OXCAN.

## 6 Treatment of culture in the dictionary entry

Now that the types of cultural words included in the dictionary have been identified, I will examine how the entry proper deals with cultural words specifically and with culture more generally. The treatment of cultural words will be studied through geographical labels, definitions, encyclopedic notes, linguistic notes, equivalents, and examples. The use of examples to propagate cultural information unrelated to a given cultural word will also be analyzed.

### 6.1 Cultural words and geographical labels

For dictionary users who are seeking to identify cultural words in a dictionary, the most useful guide is a geographical label that accompanies a given word and categorizes it as belonging to a given national culture, superculture or subculture. However, if I had relied only on this method of identification, I would have missed a number of cultural words, because the use of geographical labels varies considerably from one dictionary to another and even within a given dictionary.

This variation in the use of geographical labels is due, first and foremost, to a difference in opinion on what they stand for. In principle, geographical labels in general dictionaries serve to mark certain linguistic elements and usages that are particular to a given region, i.e. what are commonly called regionalisms in traditional French lexicography. More specifically, they provide metalinguistic information that indicates the geographical extension of a given word or sense or usage in relation to the totality of linguistic space described in the dictionary (Office de la langue française 1998: 115).

But a word that constitutes a regionalism does not necessarily present a sense or reality that is specific to that region, and a sense or reality that is typical of a given region is not necessarily presented by a word that is specific to that region. Thus, for example, the word *ambulance technician* is a word typically used in the Canadian province of Quebec to designate a paramedic: however, while the word itself may be a regionalism, the concept designated is by no means particular to Quebec or even Canadian culture.

On the other hand, *acclamation* is not a word that is uniquely Canadian, but it has a particular sense in Canadian politics – that of "act or (...) instance of election by virtue of being the sole candidate" – as opposed to the political sense it has in the US of vote by voice or applause rather than by formal ballot. Then, there are words like *allophone*, which are both Canadian in usage (as a noun referring to a person) and in sense ("an immigrant whose first language is neither French nor English"). Finally, there are words such as *admiral*, which designates a rank in the navy, but a rank that does not correspond exactly in the naval forces of every country: in

Canada, it is the highest rank in the Maritime Command, while in the US, it is the second highest rank in the navy.

Some lexicographers would mark all four words using the geographical label *Cdn*. whereas others would make at least a twofold distinction between Canadianisms (linguistic elements particular to Canada) and Canadian realities.

The OXCAN tries to make such a distinction by using the geographical label *Cdn* to mark *ambulance technician*, *acclamation* and *allophone* as *Cdn* on the one hand, and by adding the geographical note *(in Canada)* preceding the definition of *admiral*. However, another Canadian dictionary, the *Gage* (1997), labels *acclamation Cdn*, precedes the definition of *allophone* by the words "especially in Quebec", and uses a combination of a geographical and a field label, *Canadian Forces*, to identify *admiral*. (This dictionary does not include an entry for *ambulance technician*.) The lack of uniformity between these two dictionaries seems to be due in part to different methods of marking cultural words. The most logical way to do so might be to use the geographical label for words used only in a given region and for general words with a regional sense which does not designate a cultural reality (i.e. for Canadianisms of all kinds), and to use "in NAME OF REGION" preceding a definition (or even within a definition) for words used elsewhere but designating a reality particular to a given culture. However, there are definitely words such as *Cegep* (a post-secondary educational institution in Quebec that provides both pre-university and vocational training) which are both Canadian in form and also designate a Canadian reality and therefore could be identified using either or both of the proposed methods. Moreover, bilingual dictionaries, which do not normally provide definitions, would find it more difficult to implement any such system. This probably explains why only geographical labels are used in such dictionaries for all types of cultural words.

However variable the geographical identification might be, it is still better than little labelling. The latter trend, which has long been visible in American dictionaries, is clearly seen in the RHWUD. In this dictionary, for example, only the American sense of *acclamation* is provided and it is not labelled *US*; nor is its definition of *admiral* as "a naval officer of a high rank" preceded or followed by the geographical note (in the US). However, the phrase *advice and consent*, which is defined as "a phrase in the Constitution (Article II, Section 2) allowing the Senate to restrain presidential powers of appointment and treaty-making" is marked by the combination label *US Govt*. The tendency for American dictionaries is to assume that dictionary users understand that what is being presented is American English and by extension American culture. This overall attitude explains why geographical marking is given lesser importance in these dictionaries.

### 6.2 Cultural words and definitions

Cultural words which are merely linguistic and do not designate a cultural reality per se are defined like any other word. Presented below are a few examples:

| | |
|---|---|
| *angry 3.* | *Chiefly New Eng. and Midland US* inflamed, as a sore; exhibiting inflammation. (RHWUD) |
| *any more 2.* | *N Amer. informal* nowadays; at the present time (OXCAN) |

In cases where such words are simply spelling variants of words used elsewhere, the definition is generally replaced by a cross-reference:

| | |
|---|---|
| *aeradio* | *Cdn var. of* AIR RADIO (OXCAN) |

However, the definitions of cultural words which designate a specific cultural reality have to make the latter clear. This is done in the following examples by a definition that clearly indicates the country where it applies:

| | |
|---|---|
| *administration 6.* | (*often cap.*) the executive branch of the US government as headed by the President and in power during his or her term of office (RHWUD) |
| *apportionment 2.* | the determination of the number of members of the US House of Representatives according to the proportion of the population of each state to the total population of the US (RHWUD) |
| *armed forces 1* | *(Armed Forces)* the official name of the united military services in Canada (OXCAN) |
| *astronaut 2* | *Cdn (BC)* a (usu. Asian) immigrant to Canada who commutes back to Hongkong, Taiwan, etc. frequently to work, while leaving dependants in Canada (OXCAN) |

In other cases, however, such as the following, the definition does not include the cultural reference, but relies on a label to make that reference clear:

| | |
|---|---|
| *apprentice 4.* | *US Navy.* an enlisted person receiving specialized training. (RHWUD) |

In the case of definitions, as in that of geographical marking, use of a consistent method would best serve users.

### 6.3 Cultural words and encyclopedic notes

Although the unilingual dictionaries I have chosen to examine are not as encyclopedic as others, they do include some encyclopedic elements. Encyclopedic notes are found more particularly in the RHWUD. The entry for *admiral* illustrates how they are used (the underlined part is what I consider an encyclopedic note):

> *admiral 3.* a naval officer of a high rank: <u>the grades in the US Navy are fleet admiral, admiral, vice-admiral, and rear admiral</u>.

However, in the entries under A, such notes were not very frequent. They could be usefully added to words designating cultural realities, but should be made to stand out typographically from the definition *per se*.

### 6.4 Cultural words and linguistic usage notes

More frequently found, especially in the OXCAN, are linguistic usage notes such as the following (underlined part):

> *army fatigues* *esp. N Amer.* loose-fitting clothing, usu. khaki, olive drab, or camouflaged, of a sort worn by soldiers on field duty or when engaged in manual labour, e.g. ditch-digging, sandbagging, etc. <u>In the Canadian Forces, the word combat dress is used</u>.

Linguistic usage notes often help to differentiate usage in one national culture from that in another. Regional variation notes, found at the end of certain entries in the print dictionary RHUD, serve to present different ways of expression in different regional subcultures. More such notes would impress upon users the major role that linguistic elements play in culture.

### 6.5 Cultural words and equivalents

Bilingual dictionaries have the difficult task of rendering the cultural words of one language into another language in which there may be no exact equivalent. Equivalents in bilingual dictionaries fall into one of two categories: translational equivalents and explanatory equivalents.

> A translational equivalent is a lexical unit which can be immediately inserted into a sentence in the target language; e.g. in an English-French dictionary, boy = garçon. An explanatory or descriptive equivalent is one which cannot be always inserted into a sentence in the target language; e.g. boyhood: état de garçon. (Al-Kasimi 1977: 60).

Given that national cultures and subcultures differ to some extent at least, one would expect that cultural words designating specific cultural realities would be rendered more by explanatory equivalents than by translational equivalents, although those cultural words that are purely linguistic could be rendered by the latter. A study of equivalents for certain cultural words in both the RCS and the OXHA will help to determine if this is indeed the case.

(1) *A levels* ≈ *baccalauréat* m (RCS)
**A level** ≈ baccalauréat m (dans une matière) (OXHA )

Both dictionaries have provided a translational equivalent, with a typographic indication that the equivalent is approximate. However, the translational equivalent for the singular *A level* cannot be *baccalauréat*, because the latter diploma involves a series of subjects.

(2) *angels-on-horseback* GB brochette d'huîtres à l'anglaise (OXHA)

This equivalent for a typically English dish is translational.

(3) *antimarketeer* GB adversaire mf du Marché commun (OXHA)

This equivalent is halfway between translational and explanatory, because, while it can readily be used in a sentence, it explains the concept of "market"

(4) *austerity* modifier GB [furniture, clothing] *fabriqué par l'État pendant la Deuxième Guerre mondiale*. (OXHA)

This is clearly an explanatory equivalent, as the typography confirms.

(5) *assurance* (Brit: insurance) assurance *f* (RCS)
*assurance* GB insurance assurance f (OXHA)

A translational equivalent is provided in both dictionaries.

(6) *autumn* automne *m* (RCS)
*autumn* surtout GB automne m (OXHA )

Again, a translational equivalent is found in both dictionaries.

What the above examples reveal is that almost all the equivalents provided, including those for words designating specific cultural realities, are translational. In fact, only the equivalent for *austerity* (as a modifier) is clearly explanatory. The trend towards translational equivalents even for words denoting realities not found in the target culture can no doubt be attributed to the desire of bilingual dictionary editors to provide equivalents that can be directly used in target language sentences. However, the dangers of such equivalents can be seen from the example of *A level*. It might be better to provide a good understanding of the cultural reality rather than a poor translation of it.

### 6.6 Culturally-based examples for non-cultural words

So far I have considered only cultural words and their entries. However, cultural allusions can be brought in through examples for non-cultural words. This has been done in all the dictionaries we examined.

Two of the A entries in the RHWUD fall into this category. The non-cultural word *air freight* is illustrated by the example "*to airfreight French bread to the US*", which seems to indicate a fondness for French bread in the country.

At least eight OXCAN entries bring Canada in through examples provided for words that do not represent Canadian culture in themselves. E.g. in the entry for *answer: "Quebec's answer to the Champs-Élysées"*

The bilingual dictionaries are no exception in this regard, for they add the local colour of Great Britain through examples provided for words that do not themselves represent British culture. For example, in the RCS the entry for *area* contains the example *"the London area"*.

Thus, culture can be readily integrated into a dictionary by means other than entries for specifically cultural words.

## 7 Conclusion

The number of cultural words in relation to the total number of words in a dictionary is not necessarily very high (less than 500 for more than 3000 words beginning with A in the OXCAN, for example). However, this is understandable, first because many words are common to all cultures using a given language, and second because dictionary editors often choose the nomenclature that is most widely used.

What is more important than the number of cultural words presented is their treatment. And, as the analysis above reveals, this is not always as systematic or complete as it could be. What needs to be further explored is better methods of communicating cultural content, not in the form of separate cultural notes (which I have not considered here), but using the various lexical information categories found in the entry.

## Notes

1. This is my translation of the following citation: "Les dictionnaires (...) sont (...) des objets culturels, intégrés en tant que tels à [une] culture: ils témoignent d'une civilisation".

2. I have generally (but not exclusively) used the CD-ROM version of these dictionaries for entry analysis because of the rapidity of consultation, but the printed version for the front matter, which is normally not found in the CD-ROM format. The two versions are not necessarily identical, nor have they been published the same year. In the references section below, I have given the reference for the printed version used.

3. In addition to these keywords, as well as those used for searches in the other dictionaries, various abbreviations of them were also employed.

## References

Al-Kasimi, Ali. 1977. *Linguistics and Bilingual Dictionaries.* Leiden: E.J. Brill.

Atkins, Beryl T. et al. 1996. *Le Robert & Collins Dictionnaire français-anglais, anglais-français senior*. New York: HarperCollins Publishers; Paris: Dictionnaires Le Robert.

Barber, Katherine (ed.). 2001. *The Canadian Oxford Dictionary*. Toronto: Oxford University Press. [A second edition was published in 2004.]

Boulding, Kenneth. 1968. *Beyond Economics: Essays on Society, Religion and Ethics.* Ann Arbor: University of Michigan Press.

Corréard, Marie-Hélène and Valerie Grundy (eds.). 2001. *The Oxford-Hachette French Dictionary*. Oxford: Oxford University Press.

Dubois, Jean and Claude Dubois. 1971. *Introduction à la lexicographie: le dictionnaires*. Paris: Larousse.

*Gage Canadian Dictionary.* 1997. Toronto: Gage Educational Publishing Company.

Gellner, Ernest. 1983. *Nations and Nationalism*. Ithaca: Cornell University Press.

Hongladarom, Soraj. 1998. "Global Culture, Local Cultures, and the Internet: *The Thai Example*", pioneer.netserv.chula.ac.th/~hsoraj/web/Global_Culture.pdf [last accessed July 27, 2007]

Lasswell, Thomas. 1965. *Class And Stratum: An Introduction To Concepts And Research*. Boston: Houghton Mifflin.

Nida, Eugene A. 2001. *Contexts in Translating*. Amsterdam and Philadelphia: John Benjamins.

Office de la langue française. 1998. "Marques lexicographiques". In *Les Marques lexicographiques en contexte québécois*, Louis Mercier and Claude Verreault (eds.), 111–145. Quebec: Office de la langue française.

*Random House Webster's Unabridged Dictionary*. 1999. New York: Random House. [A second edition was published 2001.]

Rialland-Addach, Valérie. 1995. "Culture partagée et dictionnaire monolingue de français langue étrangère". *Études de linguistique appliquée* 97: 91–103.

CHAPTER 19

# Glimpses of the future of English-based lexicography

Arne Zettersten

The present paper, originating in the Otto Jespersen Memorial Lecture of 2004, discusses future aspects of English-based bilingual lexicography. The writer tries to provide answers to questions like the following: Will electronic dictionaries take total command over traditionally printed dictionaries? What kind of electronic techniques will dominate in the future? Will completely automatic dictionaries exist? How important will corpus linguistics be? What type of electronic assistance will be available for translators and interpreters? What type of English can we expect to be dominant in the future? In addition to answers to such questions, some practical advice on how to deal with intricate future dictionary problems will be provided. Some of these views are based on the author's experience as chief editor of the Politiken English-Danish Dictionary (1999).

## 1 Lexicography, English and Otto Jespersen

Otto Jespersen was the first *professor ordinarius* of English language and literature at the University of Copenhagen and, as we all know, the author of the majestic 7-volume Modern English Grammar – in addition to 822 other publications dealing with practically all central aspects of the English language, like grammar, phonetics, language acquisition, history of language, classification of language, etc. He even wrote about the comma in English. He was a pioneer in fields such as child language, the language of men and women, the philosophy of language, even the language of Europe, and, not surprisingly, world language.

In 1993, the English Department held a symposium to celebrate the 100th anniversary of Otto Jespersen's appointment as professor, and simultaneously the 50th anniversary of his death.[1]

The first Otto Jespersen Memorial Lecture was given on the same occasion by a fellow grammarian, Professor Randolph Quirk, now Lord Quirk, London, the second in 1994 by Professor John Sinclair, Birmingham, the creator of the Cobuild

Corpus of the English Language, and the third Memorial Lecture by Professor David Crystal, Bangor, in 1995.

Later on, we enjoyed memorial lectures by Professor Ladislav Zgusta, Cleveland, Ohio, Professor Herbert Wiegand, Heidelberg, Professor Gabriele Stein, also from Heidelberg, Professor Norman Blake, Sheffield, and in 2002 by Professor Jean Aitchison, Oxford. My plan is to edit these Otto Jespersen memorial lectures in a separate volume within the near future.

Some previous Otto Jespersen memorial lecturers started their talks by explaining that Otto Jespersen was a grammarian, not a lexicographer, and although he had written on practically any conceivable aspect of the English language, he had not produced much connected with lexicography.

It is a pleasure for me to emphasize that Otto Jespersen was actually a very diligent writer for Salmonsen's encyclopedia (*Konversationslexikon*), responsible for entries such as *Accent, Affix, Alphabet, Americanisms, Assimilation, Bilabial,* etc. We can observe items here which often correspond to headwords found in large dictionaries, although they were often carried out in greater detail using more space than in traditional monolingual dictionaries. Otto Jespersen's interest in lexicography manifested itself also in his Preface to Pitman's *International English Dictionary*, in his reviews of, for example, Schmidt's *Shakespeare Lexicon*, and his various glossaries of English texts for school purposes.

I mentioned the fact that Jespersen took an interest in world language, and that particular concept gives us a clue as to what vocabulary we should aim at when producing English dictionaries in the future. World English is the type of 'lingua franca' we need to look for in order to find out what category of dictionaries we should focus on, whether we are lexicographers or general users of English monolingual or bilingual dictionaries.

## 2 English in interaction

The new millennium has turned out to be much more eventful and sensational and even more frightening than anyone could have expected. The world has been struck by international disasters, by much-debated warfare, by deeds of terror (cf. Aitchison 2005: 1–15). Intensified media coverage of all disasters has made the global village look much smaller, the globalization has increased at an already amazing speed, and the place where the CNN cameras move in is where the future comes first.

It is under such circumstances that we now often find newly created vocabulary for our new dictionaries. This is the ever-changeable background or scenario of modern linguistic change. This is where we should find our evidence for predicting the future of the English language, for understanding if and how English dictionaries could have a future in printed form.

A statement by the Danish humourist Storm Petersen regarding the future has been frequently quoted. He is supposed to have said: "It is hard to make predictions, especially about the future." This witty and ironic remark has not dissuaded scholars from guessing wildly about the future of English.

In the past two decades, we have observed how various types of regional standard or standardizing Englishes have emerged, for example, Caribbean Standard English and East Asian Standardizing English. They are likely to continue to keep their specific regional identity and develop their specific vocabulary, idioms and pronunciation (cf. the discussion of the 'Circle of World English' in McArthur (1998: 97)).

Some scholars, notably David Crystal in his *English as a Global Language* (1997), have put forward the idea that we may foresee the emergence of a 'World Standard Spoken English' (WSSE), which will draw on new vocabulary, new idioms, and new speech habits of regions where English is a second language or foreign language. The emergence of such a WSSE is certainly debatable. We shall probably live with both regional standards and WSSE when representatives of many nations meet at conferences, etc. On a more popular level, the regional characteristics will be more marked, and the fragmentation of English into more mutually unintelligible variants, predicted by several scholars, may need many more specialized dictionaries than today.

Since the late 1990s, a good number of scholars have published quite scary and provocative articles concerning the continuous and possible threat of English in the future. Terms and concepts like language death, linguistic complacency, linguistic imperialism, killer language, cannibal language, hegemonic aggressor, aggressive expansion, language suicide and many other frightening phrases have perhaps enriched our vocabulary a little – but are they justified or relevant?

One aspect of the debate concerns the relation of the spread of English to loss of other languages in the world. While some scholars have perceived in this form a cultural imperialism or hegemony, others have questioned the cause-and-effect relationship between the two phenomena, arguing that the spread of English may even help to strengthen local linguistic identities. This aspect of the debate could be termed 'the impact of English upon other languages' (see, for instance, Phillipson 2003). Another aspect of the debate could be termed 'the impact of the spread of English upon the English language itself'. Will the increasing development of New Englishes in the world serve to fragment the language into mutually unintelligible dialects? Already the number of native speakers of English is outnumbered by those who use English as a second-language or foreign language. In this regard, David Crystal and other scholars have raised the issue of 'international standards for spoken English', the ISSE issue.

A third aspect of the debate concerns the impact of demographic factors and technological factors upon the future status of English. Within fifty years, the number of speakers of Mandarin Chinese will increase sufficiently to reinforce the dominance of that language in the world. The number of speakers of Hindi-Urdu will also increase considerably, as will the number of speakers of Spanish and Arabic. With the advent of a much increased use of the internet by non-English speakers and the development of viable translation tools, the use of English as the internet language will not be so dominant proportionally as before.

While countries such as France, Russia, Portugal, and Italy have introduced or enacted certain legislation designed to counter the infiltration of English words into their national languages, and while other countries have evinced some concern about this same issue, I maintain – in agreement with a number of linguists – that the anxieties are much exaggerated, out of proportion to the "danger". In fact, historically there have been larger mutual borrowings, and at present the average number of borrowings from English by other European languages usually amounts to no more than 1500 to 2000 Anglicisms, sometimes up to 4000.[2]

A different matter is that many individuals of the younger generation in many countries overload their conversations with English business, computer or media jargon, with citations of terminology or phrases which are not yet established Anglicisms or loanwords.

It is notable that scholars have studied thoroughly the impact of Anglicisms on many European languages in late years. A few years ago, a major comparative project, focusing on the influence of English on sixteen other European languages, yielded a impressive series of books (Görlach 2001 (ed.), 2002a, 2002b (ed.), and 2003). More recently, these works were accompanied by others, involving additional aspects of the linguistic anglification taking place in Europe: Aijmer & Melchers (eds.) 2004, discussing the English influence on the Nordic languages, and Anderman & Rogers (eds.) 2005, including papers on more pan-European aspects of the impact of English.

It is quite natural that computer terminology, business English and the English-dominated mass media, including the ever-increasing variation of English pop lyrics, are the primary sources of new loanwords in many languages. It is no wonder that a large percentage of Anglicisms in most languages may nowadays add obvious prestige to the speech-habits of young people all over the world. Are the increased number of English words and phrases, on the other hand, really detrimental to these languages as many people often claim? Certainly not in all cases where there is a loanword. An indigenous language accepting many borrowings often becomes functionally enriched, and innumerable, new – often sophisticated – nuances will be added to that language. Sometimes the user of a great number of Anglicisms will therefore be provided with a flavour of internationalism or globalism.

Admittedly, there are cases where we may talk about actual hybrid languages for which there are slightly jocular or derogatory terms like 'Japlish' or 'Janglish' (for Japanese and English), 'Franglais' (for Francais and Anglais) and many others.

Prestige is, however, not the only cause of the rapid impact of English on some languages. On a more pragmatic basis, hybrid languages are sometimes created to make communication easier and faster in communities or areas where two languages compete and two major layers of the population speak two different languages. The obvious example is 'Spanglish', the hybrid language spoken in some parts of the US, where the Spanish-speaking population is particularly large, for example in New Mexico, Florida, Southern California, and also New York. 'Spanglish' or varieties of 'Spanglish' may certainly be regarded as a kind of 'lingua franca' or 'common language'.

## 3 English as a Lingua Franca

Based on the situation during the Roman Empire, when Latin was the 'lingua franca', the term has been used in modern times for many major languages, for example, Spanish, Arabic, Hindi, French, in addition to English. In Africa Swahili is an excellent example of a 'lingua franca', as are some pidgins and creoles. In Europe the so-called 'Euro-English' could be an example of a 'lingua franca'. It is obvious that 'English as a lingua franca' (= ELF) is now "developing into a vibrant research area in applied linguistics" (see Jenkins 2004: 63).

With all the conflicting views on how and why and where to promote the expansion of English in the world, I regard the notion of 'lingua franca' as the starting point for a discussion of why the active promotion of English is justified. It is easy to predict that the English language will stay on as the only workable 'lingua franca' in the world, even if we consider a long-term perspective of about fifty years from now on. According to the prediction of the British Council (See Graddol 1997: 49), the leading five languages in 2050 will be Chinese, Hindi-Urdu, English, Spanish and Arabic. The latest report by the United Nations shows that India will have reached a population of about 1.600 million and China about 1.400 million by 2050.

At the same time, it is getting more and more obvious that larger numbers of people will – in the future – become bilingual, and multilingual speakers (in whatever format) will form an ever increasing market in the world, whether we think of book-form dictionaries or electronic dictionaries. On the background of the prospective large migration of people in the world tomorrow, we can also expect a great variability of accents or shades of pronunciations within the emerging WSSE (World Standard Spoken English). We may assume with conviction that English will consist of a large family of varieties, forming WSSE.

Due to the extended mobility and integration of people all over the world in the future, I do not hesitate to suggest that we could look upon the future WSSE in a new and modified way. We may find variants of 'Multicultural English' or 'Integration English' influencing WSSE. In this future 'Multicultural English', we may summon all kinds of varieties of English such as:

- German English, Danish English etc. due to varying mother tongues.
- English learners' language due to the use of a mixed vocabulary.
- Administrative English like Euro-English.
- Pidgins and creoles.
- Tourist English, like "European Tourist Pidgin".

A question of importance here is of course whether there will be a need for only one major WSSE dictionary or for numerous distinctive variant WSSE dictionaries. My view is that we shall need special dictionaries for all such varieties of English, and that new technology will allow us to publish them.

Again, we are back to the main question: "What kind of English or Englishes do we expect in tomorrow's world?" We can, for example, expect a large variety of language corpora like the various sub-corpora from the Linguistic Data Consortium (LDC) at the University of Pennsylvania, Philadelphia, which can be accessed on the Web and which give a clear indication of how language corpora may develop during the next few years (see Appendix 1). Some further sources that can be accessed (according to Biber et al., 1998: 286) via the World Wide Web are, for example, the following: ALEX: A Catalogue of Electronic Texts on the Internet, the BNC (British National Corpus), Cobuild Direct World Wide Web Service, and International Computer Archive of English (ICAME).

## 4 English and the Internet

Within the field of dictionaries and encyclopaedias the Internet offers innumerable possibilities. For example, *The Literary Encyclopedia*, released in 2001, is a free Internet reference work which provides biographical profiles for English-language literary writers, critical profiles of all major texts, articles on important theories, in fact a portal for all valuable Internet resources in the humanities (see Clark 2000).

In the early 1990s, we knew about videodisks, viewdata and satellite communication, but we were not yet familiar with the CD-ROM technique or the Internet.

Synthetic speech was known for computer-assisted pronouncing dictionaries, but artificial intelligence was just as enigmatic as it is today. Videoconferencing and digital television were known but took a long time to develop. Multimedia did not mean the same as today.

Since the Internet has developed at such extraordinary speed, it is most likely that multimedia techniques and online services will change so much over the next ten years that the technical presentations of dictionaries will look astoundingly different on computers, television screens or mobile telephones. Encarta World English Dictionary is an example of how an Internet concept (by Microsoft) is presented in a traditional book format (Encarta 1999). On the Internet one can now find a number of Encarta activities, besides Microsoft Encarta Encyclopedia, for example, Encarta World English Dictionary Online, Encarta Thesaurus and Encarta Reference Library. There is also the CD-ROM version of Encarta World English Dictionary. The Encarta Reference Suite is an example of multimedia information.

It has been clear at many PC expos over the last few years that PCs and handheld computers are getting smaller and smaller. Computers and human beings are getting closer and closer and the most futuristic view is that computer-like implants in the human brain may be a reality within the next 50 years. When will artificial intelligence have a break-through, when will automatic translation or interpreting be completed, when will multimedia techniques make all necessary lexical linguistic and cultural information integrated in our intelligent and productive work-stations? When will we get automatic lexicography? May we eventually produce written texts just by thinking in the environment of our work-station? With what dictionaries will we be assisted? Will the techniques of OED Online, the *Longman Web Dictionary* or Microsoft's Encarta appear outdated already by 2010? How will the papers at the International Symposium on Lexicography in Copenhagen be presented in 2014?

If we take a look at the words relevant for dictionary use, we will certainly find complications. For example, popular words come and go and are sometimes linked only to one specific year or period. One example is 'The Millennium Bug'. This word was supposed to be relevant at 00.00 hours from the 1st of January, 2000. But it actually started to be used in 1997, and a craze for the word continued for a few years and then died out.

The ten most popular words towards the end of the Millennium were voted on in England. Words like 'hope', 'love', 'peace', were among the 'top 10', but the winner was actually 'serendipity'. If you 'google' for serendipity, you get no less than 732.000 hits.

Another list is rather significant. Every year, the 'American Dialect Society' votes on Words of the Year, meaning those words that reflect the concerns and preoccupations of the past year. Participants in this voting are usually American dialect or language experts and do not represent a cross-section of people in general, but the experts' views are topical. This is one of the lists, covering "the word of the year" from 1990 to 2003:

- 1990 Bushlips
- 1991 Mother of all
- 1992 Not!
- 1993 Information superhighway
- 1994 Cyber, morph
- 1995 Web, Newt
- 1996 Mom
- 1997 Millennium bug
- 1998 E-
- 1999 Y2K
- 2000 Chad [re the elections in Florida ]
- 2001 9–11
- 2002 Weapons of mass destruction, WMD
- 2003 Metrosexual

One may observe the tendency of favourite words becoming rather political. There are other votes taken, too, like Most Useful: Google (web); Most Creative: Iraqnophobia; Most Euphemistic: regime change (forced change in leadership), etc.

I should also like to mention a different source of information on topical words that come and go, namely the Top Ten Word List according to the website 'yourDictionary.com'.

The Top 10 words of 2003 were the following: 1. Embedded; 2. Blog; 3. SARS; 4. Spam; 5. Taikonaut; 6. Bushism; 7. Allision; 8. Recall; 9. Middangeard; 10. Celibacy.

The corresponding words for 2004 were: 1. inCivility; 2. Red States/Blue States; 3. Blogsphere; 4. Flip-flopping; 5. Esrever; 6. Fahrenheit; 7. iPod; 8. IM; 9. Liberal; 10. Eurosceptic.

The top 10 from the Cambridge Dictionary Online gave a different result, of course, because people wish to look up words they are interested in. These were the top 10 words of 2004 in the Cambridge Dictionary, with the results of 2003 in brackets: 1. Advice (1); 2. Liaise (2); 3. Effect (3); 4. Regard (5); 5. Comply (6); 6. Appreciate (9); 7. Commit (12); 8. Assess (8); 9. Endeavour (16); 10. Acquire (13).

What I have tried to indicate by these examples here in particular is that topical words change according to the criteria for the questionnaires.

In this electronic age – where we are so used to online dictionaries, DVD dictionaries, CD-ROM dictionaries, hand-held dictionaries of all kinds – linguists tend to ask themselves if the speed of internet activity is detrimental to language expansion and to personal creativity in language or not. Physicist Robert Logan, at the University of Toronto, claims that the Internet is the sixth link in an evolutionary chain of languages, including speech, writing, mathematics, science, and computing. These six languages all have developed their own distinct semantics and syntax. Mathematics and writing came into existence about 3,100 BC, and science about 1,000 years later. Towards the end of the second Millennium, two more languages appeared, computing and the Internet (Logan 2000). The new language on the web has been termed "Netspeak" in David Crystal's book *Language and the Internet* (2001).

The newest figures indicate that as many as 1.9 billion people, one third of the world's population, know some kind of English. Half of Europe's business deals are conducted in English, three quarters of the world's mail is written in English. Due to the fact that so many new speakers of English seem to exist, there is naturally a strong prospective demand of monolingual as well as bilingual dictionaries.

At the same time, some of the minority languages threatened by extinction have become strengthened by the help of the Internet. Some of these languages have constructed large-scale dictionaries and language-learning projects online. The newest invention is the new way of handling writing systems and translations of complicated characters.

Not long ago, computers did only handle writing systems that could be translated in ASCII (American Standard Code for Information Interchange). First we got Extended ASCII, which could recognize accents and various symbols used in European languages. Now – suddenly – we got "Unicode", which can translate any character into 16 bytes. The 8-byte Extended ASCII could handle 256 characters before, but now 'Unicode' can manage more than 65,000 unique characters. Maybe this covers all the writing systems in the world, making multi-dictionary online systems a reality for the future.

Having reached the last two languages out of six, according to Robert Logan's view, we are ready to look at the type of dictionaries at our disposal and what kind of lexicographical processes we can identify.

## 5 Future English Lexicography

Tom McArthur – in his discussion of 'Tomorrow's dictionaries' in *Living Words* (1999) – presents a series of eight processes in dictionary making, all terms ending in –ization:

1. Globalization
2. Localization
3. Bilingualization
4. Semibiligualization
5. Nationalization

6. Regionalization
7. Thematization
8. Electronicization

All these are relevant for the future of English lexicography.

1. The first, *globalization*, refers to dictionaries showing English as an international language for all people and places. McArthur rightly maintains that the emergence of the learners' dictionaries in the 1930s was a turning-point in the development of English lexicography. Particularly the *Advanced Learner's Dictionary of Current English* in all its editions has had a vast impact on perceptions of English as an international language, according to McArthur.
2. The term *localization* means that the international dictionaries may be customised for one country or one group of countries. For example, the *Times-Chambers Essential Dictionary*, published by Federal Publications, Singapore (1997), contains the 'core English' standard (standard British and American), English words specific to Singapore and Malaysia (called 'SME') and regional words adopted into SME.
3. *Bilingualization* indicates that a major dictionary is translated into another language and made available to users in that language. McArthur explains how this has been a success with Chinese learners in the case of the *Longman Dictionary of Contemporary English* (1976 onward) and the *Longman Lexicon of Contemporary English* (1981).
4. *Semibilingualization* is a process whereby the main principle is that all translation equivalents are inserted for all dictionary headwords. Examples can be provided by the pioneer in this field, the Kernerman Semi-Bilingual Dictionaries, for example, *Passport: English Hebrew Learner's Dictionary* (1996) based on *Passport English Learner's Dictionary*. A step forward is the English-Danish dictionary published by Politiken, Copenhagen (Zettersten & Lauridsen (eds.) 1999). It is based on the entries, definitions and authentic examples of the Collins Cobuild English Dictionary. This type of English-based bilingual dictionary is the closest one can get to a combination of a monolingual English dictionary and a bilingual one. This means that the dictionary may be used not only for translating purposes (i.e. from English into Danish) but also as a production dictionary, for producing English text.
5. The fifth process is called *nationalization,* indicating that major English-speaking countries like Australia or Canada have produced indigenous Australian or Canadian dictionaries, such as the *Macquarie Dictionaries* and the *Canadian Oxford Dictionary*, respectively. There are now also national style guides, such as the *Cambridge Australian English Style Guide* (1995), and the *Guide to Canadian English Usage* (1997).

6. The next process is *regionalization*, which refers to regions larger than a single state, such as, for example, South and East Asia. McArthur suggests that a native-speaker-style dictionary for this area may be foreseen, as well as one fitting Australian English more firmly into the Asian context. Whether this means that English will be characterized as an Asian language is too premature to verify.
7. The next process, *thematization*, indicates formats like, for example, *Roget's Thesaurus*, in which words and meanings are presented thematically, not alphabetically. Further examples are the *Longman Essential Activator: Put Your Ideas into Words* (1997) and the *Oxford Learner's Wordfinder Dictionary* (1997).
8. The last process could be called *electronicization*, including all varieties of electronic tools for dictionary making and electronic media for presenting dictionary items and complete dictionaries.

In all these eight categories, one would hope that dictionary producers always bear in mind the wishes and attitudes of the users (cf. Stein 1984: 6 and Hartmann 1987).

It is under this category 'electronicization' that we should place one of the keys to the prospective expansion of dictionary-making. The combination of electronic competence and advanced corpus linguistics is such a key to the future expansion that I want to predict. I am a total optimist in my prediction that, as new variants of English emerge, there will be a new need for further dictionaries on the basis of the picture of varieties you saw.

But what about the formats? Will there be any printed hard-cover dictionaries left for us to caress in our future bookcases? Will we keep only virtual bookcases and libraries in the future? Is there any substance in anything we have heard lately of the death of languages, death of the dictionary, etc? In the case of languages, this is all true. It has been predicted that about half of the present 6,000 languages will die out. One estimate even indicates that one language will die every 2 weeks. About the printed dictionary, I am optimistic, because the techniques of printing and reproducing dictionaries will probably develop in new and spectacular ways. The new types of electronic dictionaries serve as eminent complements to the printed dictionaries and add to my optimism.

## 6 Changing times: The old and new OED

The new formats of the old venerated *Oxford English Dictionary* (OED) is one particular reason for my optimism. The notion of 'death' and even 'murder' connected with the OED is, on the other hand, what has recently put special emphasis on some aspects of the the history of the OED and certainly given a boost to the great recent interest in all the versions of the OED.

The following story is the fascinating background to the renewed interest in the dictionary and its famous chief editor, James Murray:

"American murderer helped write Oxford Dictionary" was the staggering headline of an article published in the *Sunday Star* in Washington DC in July, 1915, by an American journalist called Hayden Church. The story about Dr. William Chester Minor, who constantly helped the editors of the *Oxford English Dictionary* for about 20 years with nearly ten thousand definitions during the last few decades of the nineteenth century and the beginning of the twentieth century, also appeared in the *Strand* magazine in September, 1915.

The close cooperation between the Chief Editor, Dr. James Murray, and Dr. Minor is well documented through Dr. Murray's own letters. The first meeting between 'the Professor and the Madman' is however known through the fictionalised account in Hayden Church's story which runs as follows, starting here at the point when Dr. Murray first arrived at Broadmoor, Crowthorne in Berkshire, and introduced himself to a man behind a director's desk:

> "I, Sir, am Dr. James Murray of the London Philological Society", he said in his finely modulated Scottish voice, "and editor of the *Oxford English Dictionary*."
> "And you sir, must be Dr. William Minor. At long last. I am most deeply honoured to meet you".
> There was a pause. Then the other man replied: "I regret not, sir. I cannot lay claim to that distinction. I am the Superintendent of the Broadmoor Criminal Lunatic Asylum. Dr. Minor is an American, and he is one of our longest-staying inmates. He committed a murder. He is quite insane." (Winchester 1998: 169–71)

Dr. Murray asked to be taken to Doctor Minor, and this lead to a long friendship between the two men, and to further cooperation. When Minor was an old man, Murray succeeded in obtaining a pardon for Dr. Minor who could return to the US as a free man in 1910. The minister for the Interior, who signed this decision, happened to be Winston Churchill.

This remarkable story is based on a combination of facts and fiction, and it has created new interest in the fascinating background of the world's probably greatest lexicographical achievement so far, the *Oxford English Dictionary*, published between 1884 and 1928 in originally ten volumes. The renewed interest in the OED these days is due partly to the recent publication of Winchester's book, cited above, partly to the newest formats of the OED, recently introduced by Oxford University Press.

The history of the OED and its dedicated editors has been well documented, particularly by publications such as KM. Elisabeth Murray's *Caught in the Web of*

*Words* (1977), Donna Lee Berg's *A Guide to the Oxford English Dictionary* (1993), and Simon Winchester's *The Meaning of Everything* (2003).

However, the book by Simon Winchester emphasizes the extraordinary circumstances under which the collection of the vast material was carried out. It also shows us in which ways the Chief Editor, Dr. James Murray, had to improvise and to make unprecedented decisions. Even today, with the use of so much sophisticated computer hardware and software, do we admit that some early dictionary projects are quite impressive. By comparison, it is quite difficult to imagine how a small team of OED scholars could mastermind so much with so little technology at the time of the first ten volumes.

This brief introduction to the history of the OED is meant to put the latest formats of this dictionary in perspective, namely the CD-ROM version of 1999 and the online edition of 2000, being examples of how databases of words, expressions and texts can be utilised with new technology whether it is for monolingual or bilingual purposes. The upgraded CD-ROM version came in 1999 with search techniques like the following:

1. Search for headwords or compounds at any time through the "Find word" function.
2. Search for words or phrases across the entire text or narrow your parameters (e.g. search only quotation text, definition text or etymologies).
3. Search by date, title of work or author quotation.
4. A powerful new proximity search allows you to find words which occur near, before, or after one another in an entry.
5. Use wildcards to search for any number of unknown characters and to search for any individual character you are unsure of.

The OED Online Edition (http://dictionary.oed.com) was made available in 2000 by annual subscription. In addition to the Second Edition of the OED (1989) and three Addition volumes, there will be quarterly releases and new revised material up to maybe 2018. The publishers have also advertised the following benefits of the OED Online:

1. Select how entries are displayed by turning pronunciations, etymologies, variant spellings, and quotations on and off.
2. Everything from simple word look-ups to proximity searching, using any of the fields in the dictionary, can be done with speed and ease.
3. Find a term when you know the meaning but have forgotten the word.
4. Use wildcards if you are unsure of spelling, or if you want to search for words with common characteristics.
5. Search for quotations from a specified year, or from a particular language.
6. Search for words which have come into English via a particular language.

7. Compare revised entries with entries from the Second Edition to see how language has changed and how new scholarship has increased understanding of our linguistic and cultural heritage.

The example of the development from manpower to machine power is only one of many similar developments in the history of lexicography. The reason for using the OED for this exemplification is of course the exceptional bulk of both the original project and the latest venture to double the size of the dictionary by 2010. This plan to bring the OED completely up to date will probably make the revision programme one of the world's largest humanities research projects.

This Third Edition of OED will thus develop continuously through the quarterly updates. The new revised Online Edition will be presented with the older entries included in a pop-up window. The present view of the editors is that this revision process is unlikely to finish before 2018.

## 7 Technical advances in lexicography

As the storing capacity of computers developed during the 1980s and 90s, it became more meaningful to utilise new scanning techniques and the new storing capacity of computers in order to create large databases of language text (linguistic corpora) for new types of dictionaries.

The real pioneer project in this field was the Bank of English, composed at the University of Birmingham, under the directorship of Professor John Sinclair. The first corpus, containing 20 million words, formed the basis for the first edition of the *Collins COBUILD English Language Dictionary* of 1987. The Bank of English continued to grow in the 1990s, and the corpus used for the second edition in 1995 comprised ca. 200 million words, divided into several subcorpora. Now the Bank of English contains well over 500 million words.

These days there are many innovations in corpus linguistics. The software has become more sophisticated, the searches are quicker and the corpora much more diversified, as online resources continue to grow. One promising thing for the future is the way that some dictionaries now allow access to corpus citations directly from a specific dictionary sense. It was suggested already in 1994 by Krista Varantola and introduced by Collins Cobuild on CD-ROM with a searchable Word Bank. Recently, access to speech corpora as an option of the dictionary look-up procedure has been suggested (Sobkowiak 2002).

For Internet dictionaries and for CD-ROM versions there will be a variety of retrieval programs or tools in order to make the search for linguistic data automatic. John Sinclair described such a tool in 1994 in his Otto Jespersen memorial lecture called "Prospects for Automatic Lexicography" (Sinclair 1996), delivered at the International Symposium on Lexicography in Copenhagen. Nothing is, however, fully automatic yet in this age of multimedia, CD-ROMs and virtual reality.

In a festschrift to Martin Gellerstam, a new article was published by John Sinclair, in which he introduced the term *Floating Dictionary*. Sinclair says:

> A few years ago I felt that the time was ripe to plan a new type of dictionary, one that would never exist on paper, but would be automatic or almost automatic in its self-updating. It would, so to speak, float on top of a corpus, rather like a jelly-fish, its tendrils constantly sensing the state of the language […]. This may sound an extremely optimistic venture, stretching far into the future and unlikely ever to be successfully implemented. However there is a floating dictionary published, which has been on sale since 1994, and is now in its third edition. The title is *Collins Cobuild CD-ROM* (2003), and the dictionary floats on top of a corpus of five million words in such a way that the user can call up examples from the corpus to check the information in the dictionary entry. (Sinclair 2001: 407).

I am not going to open up a full discussion of the possible linguistic situation as far ahead as a hundred years from now on, or of the scientific or medical developments that we may postualate for the distant future. Therefore I will not speculate in new gene technology here, because a full understanding of how genes give rise to brain structure and how our brain structure gives rise to language lies, according to experts, 50 to 100 years ahead. Likewise, any hopes or fears of an implant of a computer chip in the brain for simultaneous translation and usage of a data bank can be given up for the moment. Also, we should give up any hopes or fears this century that human beings may ever be able to understand the thoughts or ideas of people of any nationality or language and thus bypass the intervening spoken language.

Before I get even more tempted to lead you any further into the future non-reality, I should like – finally – to mention a new technology for the future which will revolutionalize book printing and dictionary production, and explain why I may be able to predict the slow death of the paper dictionary and the survival of the electronic dictionaries: The answer is *electronic paper*.

Electronic reusable paper employs a technology invented at the Xerox Palo Alto Research Canter (PARC), called 'Gyricon'. It is electronically writable and erasable and can store images. One Gyricon sheet is a thin layer of transparent plastic in which millions of small beads are randomly dispensed. Voltage is applied to the surface to create images as text and pictures, including dictionary entries. With a built-in input scanner, a kind of wand becomes a hand-operated multi-function device – a printer, copier, fax, scanner etc.

A high-resolution monochrome electronic paper was invented in 1998 by a Massachusetts-based company called E-Ink in collaboration with Philips in

Holland. Some nano-technologists now claim that there will be newspapers that update themselves constantly and can be carried in your pocket.

E-Ink has also reported a big step forward – an ultra-slim computer display that bends and rolls up into a narrow tube, just about an inch in diameter. The present prototype of the screen is a little larger than a business card and has a resolution of 96 dots per inch (Chen et al. 2003).

It is not only newspapers that will be constantly updated. Your own dictionary for special purposes will be able to store new words and expressions automatically, not by printing a new edition once in a while, but just by sending new information continually to your electronic paper dictionary.

The papers in this magic-sounding dictionary can be updated in several ways. One way is to use a computer or a pager. Another method developed by Xerox is a 'magic wand'. The wand is a kind of wireless transmitter, moving from one page of the prospective dictionary to another.

## 8 Dictionaries forever

Consequently, my optimism is justified. There will exist an immense number of dictionaries in the future, not necessarily in the conventional paper formats, but in electronic versions of many kinds. They will be more tied to fluctuations of interest and to political changes in the world. It will be justified for the future synthetic voices of tomorrow's dictionaries to say – using the famous quotation from Mark Twain after he learnt that his obituary had been published prematurely in the New York Journal: "The reports of my death are greatly exaggerated".

I should like – finally – to quote a totally different language, suitable for the climate of our Nordic symposium and of Otto Jespersen. You recall that Otto Jespersen invented hos own language called Novial. So did my older colleague in Oxford, whom I cooperated with regarding editions of medieval texts, Professor J. R. R. Tolkien. He invented 13 languages or more, one was called Arctic, a little similar to the elvish tongue Quenya, with some inspiration from Finnish, a truly Arctic language. The quotation is:

> Mára mesta an nu vela tye en to, ya rato nea.

It means the following:

> Good bye till I see you next, and I hope it will be very soon!

## Notes

**1.** The lectures delivered at that Otto Jespersen Symposium were published in Nielsen & Zettersten (eds.) 1994.

**2.** On the difficulty of assessing the number of English borrowings in a language, see Gottlieb 2002.

## References

*Dictionaries of English*

Collins COBUILD English Dictionary. 2nd ed. 1995. London: HarperCollins Publishers.

Collins COBUILD Advanced Learner's Dictionary. 4th ed. 2003. London: HarperCollins Publishers.

Collins Cobuild CD-ROM Resource Pack. 3rd ed. 2003. London: HarperCollins Publishers.

Encarta World English Dictionary. 1999. New York: St. Martin's Press. *Corpus Linguistics*. Cambridge: Cambridge University Press.

The Longman Dictionary of Contemporary English. 3rd ed. 1995. Harlow: Longman.

The Oxford Advanced Learner's Dictionary of Current English. 5th ed. 1995. Oxford: Oxford University Press.

The Oxford Advanced Learner's Dictionary of Current English. 6th ed. 2000. Oxford: Oxford University Press.

The Oxford English Dictionary. 2nd ed. 1989. Oxford: Oxford University Press.

*Other works*

Aijmer, Karin & Gunnel Melchers (eds.). 2004. *The influence of English on the languages in the Nordic countries*. Special issue of *Nordic Journal of English Studies* 3 (2).

Aitchison, Jean. 2005. "Absolute Disasters: The Problems of Layering", In *Symposium on Lexicography XI* [Lexicographica. Series Maior 115]. Gottlieb, Henrik; Jens Erik Mogensen & Arne Zettersten (eds.), 1–15. Tübingen: Max Niemeyer Verlag.

Anderman, Gunilla & Margaret Rogers (eds.). 2005. *In and Out of English: For Better, For Worse?* Clevedon: Multilingual Matters.

Berg, Donna Lee. 1993. *A Guide to the Oxford English Dictionary*. Oxford: Oxford University Press.

Biber, Douglas; Susan Conrad & Randi Reppen. 1998. *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Chen, Y.; J. Au, P. Kazlas, A. Ritenour, H. Gates & M. McCreary. 2003. "Electronic paper: Flexible active-matrix electronic ink display". *Nature* 423: 136.

Clark, Robert. 2000. "Literacy, the Book and the Internet: Retrospects and Prospects". *The European English Messenger* 9 (2), 57–61.

Crystal, David. 1997. *English as a Global Language*. Cambridge: Cambridge University Press.

Crystal, David. 2001. *Language and the Internet*. Cambridge: Cambridge University Press.

Görlach, Manfred (ed.). 2001. *A Dictionary of European Anglicisms*. Oxford & New York: Oxford University Press.

Görlach, Manfred. 2002a. *An Annotated Bibliography of European Anglicisms.* Oxford & New York: Oxford University Press.

Görlach, Manfred (ed.). 2002b. *English in Europe.* Oxford & New York: Oxford University Press.

Görlach, Manfred. 2003. *English Words Abroad.* [Terminology and Lexicography Research and Practice 7]. Amsterdam / Philadelphia: John Benjamins Publishing

Gottlieb, Henrik. 2002. "Four Germanic dictionaries of Anglicisms: When definitions speak louder than words". In *Symposium on Lexicography X.* [Lexicographica. Series Maior 109]. Gottlieb, Henrik; Jens Erik Mogensen & Arne Zettersten (eds.), 125–43. Tübingen: Max Niemeyer Verlag.

Graddol, David. 1997. *The Future of English*. London: The British Council. Full text available at: www.britishcouncil.org/learning-elt-future.pdf.

Hartmann, R.R.K. 1987. "Dictionaries of English: The User's Perspective". In *Dictionaries of English*, Richard W. Bailey (ed.), 121–135. Cambridge: Cambridge University Press.

Jenkins, Jennifer. 2004. "EFL at the Gate. The Position of English as a Lingua Franca", *The European English Messenger* 13 (2), 63–69.

Jespersen, Otto. 1909–49. *Modern English Grammar on Historical Principles* (7 vols.)**.** London: Allen & Unwin. Copenhagen: Einar Munksgaard.

Logan, Robert. 2000. *The Sixth Language: Learning or Living in the Internet Age.* Toronto: Stoddart Publishing.

McArthur, Tom. 1998. *The English Languages*. Cambridge: Cambridge University Press.

McArthur, Tom. 1999. *Living Words. Language, Lexicography and the Knowledge Revolution.* Exeter: Exeter University Press.

Mogensen, Jens Erik; Viggo Hjørnager Pedersen & Arne Zettersten, (eds.) 2000. *Symposium on Lexicography IX* [Lexicographica. Series Maior 103]. Tübingen: Max Niemeyer Verlag.

Murray, K. M. Elisabeth. 2001. *Caught in the Web of Words. James Murray and the Oxford English Dictionary.* New Haven: Yale University Press.

Nielsen, Jørgen Erik & Arne Zettersten (eds.). 1994. *A Linguistic Miscellany, P.O.E.T.*, Vol. 20. Department of English, University of Copenhagen.

Phillipson, Robert. 2003. *English-Only Europe? Challenging Language Policy*. London: Routledge.

Sinclair, John. 1996. "Prospects for Automatic Lexicography", In *Symposium on Lexicography VII.* [Lexicographica. Series Maior 76]. Zettersten, Arne & Viggo Hjørnager Pedersen (eds.), 1–10. Tübingen: Max Niemeyer Verlag.

Sinclair, John. 2001. "The Floating Dictionary", In *Gäller stam, suffix och ord: Festskrift till Martin Gellerstam den 15. oktober 2001.* Sture Allén et al. (eds.), 393–419. University of Göteborg: Meijerbergs institut för svensk etymologisk forskning.

Sobkowiak, Włodzimierz. 2002. "On the phonetics of *Trans-* in EFL dictionaries", In *Symposium on Lexicography X,* [Lexicographica. Series Maior 109]. Gottlieb, Henrik; Jens Erik Mogensen & Arne Zettersten, (eds.), 293–301. Tübingen: Max Niemeyer Verlag.

Stein, Gabriele.1984. *The English Dictionary: Past, Present and Future.* Exeter: Exeter University Press.

Varantola, Krista. 1994. "The dictionary user as a decision-maker". In Martin, Willy et al. (eds.), *Euralex 1994 proceedings. Papers submitted to the 6th Euralex International Congress on Lexicography in Amsterdam, The Netherlands,* 606–611. Amsterdam: Vrije Universiteit.

Winchester, Simon. 1998. *The Professor and the Madman. A Tale of Murder, Insanity, and the Making of the Oxford English Dictionary*. New York: HarperCollins Publishers.
Winchester, Simon. 2003. *The Meaning of Everything. The Story of the Oxford English Dictionary.* Oxford: Oxford University Press.
Zettersten, Arne. 2000. "Aspects on the Future of Lexicography within Interdisciplinary Research", In *Artefakte/Artefiktionen. Transformationsprozesse zeitgenössischer Literaturen, Medien, Künste, Architekturen. Festschrift für Christian W. Thomsen zum 60. Geburtstag.* Angela Krewani (ed.). Heidelberg: Universitetsverlag Winter.
Zettersten, Arne & Viggo Hjørnager Pedersen (eds.) 1994. *Symposium on Lexicography, VII,* [Lexicographica. Series Maior 76]. Tübingen: Max Niemeyer Verlag.
Zettersten, Arne & Hanne Lauridsen (eds.). 1999. *Politikens Engelsk-Dansk med betydningsforklaringer*. Copenhagen: Politikens Forlag.

## Appendix 1

Examples of corpora offered by the Linguistic Data Consortium (LDC):

*Lexicon corpora*
Microphone lexicon:
LDC99L23 American English Spoken Lexicon

Pronunciation lexicon:
LDC97L20 CALLHOME American English Lexicon (PRONLEX)
LDC97L19 CALLHOME Egyptian Arabic Lexicon
LDC97L18 CALLHOME German Lexicon
LDC96L17 CALLHOME Japanese Lexicon
LDC96 L15 CALLHOME Mandarin Chinese Lexicon
LDC96L16 CALLHOME Spanish Lexicon
LDC99L22 Egyptian Colloquial Arabic Lexicon

*Speech corpora*
Broadcast speech:
LDC97S66 1996 English Broadcast News Dev & Eval (Hub-4)
LDC97S44 1996 English Broadcast News Speech (Hub-4)
LDC98S71 1997 English Broadcast News Speech (Hub-4)
Mobile-radio speech:
LDC94S14B Air Traffic Control BOS
LDC94S14A Air Traffic Control Complete
LDC94S14C Air Traffic Control DCA

*Text corpora*
Broadcast text:

LDC98T31 1996 CSR Hub-4 Language Model
LDC97T22 1996 English Broadcast News Transcripts (Hub-4)
LDC98T28 1997 English Broadcast News Transcripts (Hub-4)
LDC98T24 1997 Mandarin Broadcast News Transcripts (Hub-4NE)
LDC98T29 1997 Spanish Broadcast News Transcripts (Hub-4NE)
LDC99T36 USC Marketplace Broadcast news Transcripts
Conversation text:
LDC97T14 CALLHOME American English Transcripts
LDC97T19 CALLHOME Egyptian Arabic Transcripts
LDC97T15 CALLHOME German Transcripts
LDC96T18 CALLHOME Japanese Transcripts
LDC96T16 CALLHOME Mandarin Chinese Transcripts
LDC96T17 CALLHOME Spanish Transcripts
LDC98T26 Hub-5 Mandarin Transcripts
LDC98T27Hub-5 Spanish Transcripts
LDC99T33 SUSAS Transcripts
LDC93S7-T SWITCHBOARD-1 Transcripts
Newswire text:
LDC2000T43 BLLIP 1987–89 WSJ Corpus Release 1
LDC95T6 CSR-III Text
LDC2000T48 Chinese Treebank (preliminary release)
LDC95T11 European Language Newspaper Text
LDC99T34 Japanese Business news Text Supplement
LDC95T8 Japanese Business News Text
LDC2000T45 Korean Newswire
LDC96T10 MUC-VI Text Collection
LDC95T13 Mandarin Chinese News Text
LDC95T21 North American News Text Corpus
LDC98T30 North American news Text Supplement
LDC99T40 Portuguese Newswire text
LDC95T9 Spanish News Text
LDC99T41 Spanish Newswire Text, Volume 2

# Index

In the series *Terminology and Lexicography Research and Practice* the following titles have been published thus far or are scheduled for publication:

10 **GOTTLIEB, Henrik and Jens Erik MOGENSEN (eds.):** Dictionary Visions, Research and Practice. Selected papers from the 12th International Symposium on Lexicography, Copenhagen 2004. 2007. xi, 321 pp.

9 **YONG, Heming and Jing PENG:** Bilingual Lexicography from a Communicative Perspective. 2007. x, 229 pp.

8 **ANTIA, Bassey Edem (ed.):** Indeterminacy in Terminology and LSP. Studies in honour of Heribert Picht. 2007. xxii, 236 pp.

7 **GÖRLACH, Manfred:** English Words Abroad. 2003. xii, 189 pp.

6 **STERKENBURG, Piet van (ed.):** A Practical Guide to Lexicography. 2003. xii, 460 pp.

5 **KAGEURA, Kyo:** The Dynamics of Terminology. A descriptive theory of term formation and terminological growth. 2002. viii, 322 pp.

4 **SAGER, Juan C.:** Essays on Definition. With an introduction by Alain Rey. 2000. viii, 257 pp.

3 **TEMMERMAN, Rita:** Towards New Ways of Terminology Description. The sociocognitive approach. 2000. xvi, 258 pp.

2 **ANTIA, Bassey Edem:** Terminology and Language Planning. An alternative framework of practice and discourse. 2000. xxiv, 265 pp.

1 **CABRÉ CASTELLVÍ, M. Teresa:** Terminology. Theory, methods and applications. Edited by Juan C. Sager. Translated by Janet Ann DeCesaris. 1999. xii, 248 pp.