

# Principles and Practice of South African Lexicography

woord<sup>6</sup> woorde<sup>6a</sup>

:word<sup>6</sup> words<sup>6a</sup> lentšu<sup>18</sup> izwi<sup>19</sup> igama<sup>20</sup>

*Thembeka kan enige woord spel want sy's (=sy is) baie sli  
word because she's (=she is) very clever.*

woordeboek<sup>6</sup> woordeboeke<sup>6a</sup>

:dictionary<sup>6</sup> dictionaries<sup>6a</sup> pukuntšu<sup>18</sup> isichamazwi<sup>19</sup> isich

# Principles and Practice of South African Lexicography

R.H. Gouws and D.J. Prinsloo



*Principles and Practice of South African Lexicography*

Published by SUN PReSS, a division of AFRICAN SUN MeDIA, Stellenbosch 7600  
www.africansunmedia.co.za  
www.sun-e-shop.co.za

All rights reserved. Copyright © 2005 R.H. Gouws & D.J. Prinsloo

No part of this book may be reproduced or transmitted in any form or by any electronic, photographic or mechanical means, including photocopying and recording on record, tape or laser disk, on microfilm, via the Internet, by e-mail, or by any other information storage and retrieval system, without prior written permission by the publisher.

First edition 2005  
ISBN: 978-1-919980-90-4  
e-ISBN: 978-1-919980-91-1  
DOI: 10.18820/9781919980911

Cover design by Dewald van Zyl  
Image used in cover design was sourced from: *Juta's Dictionary Afrikaans/English – English/Afrikaans* (1988) by Peter Grobbelaar

Typesetting by SUN MeDIA Stellenbosch  
Set in 10/12 Palatino Linotype

SUN PReSS is a division of AFRICAN SUN MeDIA, Stellenbosch University's publishing division. SUN PReSS publishes academic, professional and reference works in print and electronic format. This publication may be ordered directly from [www.sun-e-shop.co.za](http://www.sun-e-shop.co.za)

*Aan Ilse Gouws en Michele Prinsloo:  
woordeboekweduwees sonder weerga.*

# Table of contents

---

Preface .....	i
CHAPTER 1: The development of lexicography .....	1
1.1 Introduction .....	1
1.2 Within a linguistic fold .....	3
1.3 The Wiegand era .....	4
1.4 Specialised lexicography .....	6
1.5 Lexicographic functions .....	7
1.6 Electronic dictionaries .....	8
CHAPTER 2: The lexicographic process .....	9
2.1 Introductory remarks .....	9
2.2 A primary comprehensive lexicographic process .....	10
2.3 A secondary comprehensive lexicographic process .....	11
2.4 Dictionary specific lexicographic processes .....	12
2.5 The dictionary plan .....	13
2.6 The use of the dictionary .....	19
CHAPTER 3: Material collection and corpus building .....	21
3.1 Building text corpora .....	21
3.2 Querying text corpora .....	25
3.3 The use of corpora on macrostructural level .....	30
3.4 The use of corpora on microstructural level .....	33
CHAPTER 4: The user-perspective .....	39
Chapter 5: Dictionary typology .....	45
5.1 Making realistic choices .....	45
5.2 Different types of dictionaries .....	46
CHAPTER 6: The frame structure of dictionaries .....	57
6.1 Dictionaries as carriers of text types .....	57
6.2 Data distribution .....	58
6.3 Integrated and non-integrated outer texts .....	58
6.4 The planning of outer texts for South African dictionaries .....	61
6.5 The central list .....	62
6.6 Macrostructural issues regarding the central list .....	64

CHAPTER 7: Macrostructural aspects .....	67
7.1 Lemmatisation strategies .....	67
7.2 Different types of lemmata and articles .....	85
7.3 The ordering of lemmata .....	96
7.4 Balancing article stretches .....	111
 CHAPTER 8: Microstructural aspects .....	 115
8.1 Introduction .....	115
8.2 Categories of entries .....	115
8.3 The article structure and the data categories in general monolingual and bilingual dictionaries .....	118
8.4 The addressing structure .....	134
8.5 Types of microstructures .....	138
 CHAPTER 9: Definitions .....	 143
9.1 Introduction .....	143
9.2 Basic types of definitions .....	145
9.3 Enhancing the quality of definitions .....	147
9.4 Electronic corpora as the key to writing better definitions .....	148
 CHAPTER 10: Equivalent relations in bilingual dictionaries .....	 151
10.1 Introduction .....	151
10.2 Translation equivalent versus paraphrase of meaning .....	152
10.3 Relations between lemma and translation equivalents .....	153
10.4 Different types of equivalent relations .....	154
10.5 Communicative equivalence .....	161
 CHAPTER 11: The access and search area structure .....	 165
11.1 General remarks .....	165
11.2 The access structure .....	165
11.3 The search area structure .....	172

CHAPTER 12: Cross-referencing as a Lexicographic Device .....	177
12.1 Introduction .....	177
12.2 Some basic terms relating to a theory of mediostructures .....	177
12.3 The internal cross-reference address .....	179
12.4 The external cross-reference address .....	179
12.5 The dictionary external cross-reference address .....	180
12.6 Combatting decontextualisation .....	181
12.7 Cross-references aimed at additional information retrieval .....	184
12.8 Dead cross-references .....	185
12.9 Failure to utilise cross-referencing where needed .....	186
12.10 Cross-references to the wrong cross-reference address .....	187
12.11 Cross-references that misguide the user in respect of information retrieval .....	188
12.12 Using cross-referencing to avoid a full treatment of the lemma .....	189
12.13 Unidirectional versus bi-directional cross-referencing .....	189
12.14 Implicit versus explicit cross-referencing .....	191
12.15 In conclusion .....	192
 Bibliography .....	 193
 Index .....	 205

# Preface

---

The last decade has witnessed the emergence of a new wave of professional lexicographers in South Africa. The establishment of the Pan South African Language Board's National Lexicography Units, one for each of the eleven official languages, and the growing interest in lexicographic principles and practice among academics, freelance lexicographers and publishers, created the need for a theoretically-based book directed at the needs of experts, both in the lexicographic practice and in the field of theoretical lexicography. *Principles and Practice of South African Lexicography* is a response to this need. The book endeavours to cover significant phases and features of a lexicographic process, with a focus on the historical orientation of theoretical lexicography, the preparation phases in the lexicographic process, including material collection and corpus building, and various aspects of the data distribution programme, e.g. different lexicographic structures. Different facets from the general theory of lexicography are applied to the lexicography of the South African languages and from the perspective of lexicographic challenges confronting the South African languages suggestions are made to enhance the general theory of lexicography. This book wishes to promote the desperately needed interaction between theory and practice in the field of lexicography.

The authors wish to thank their respective Departments and Universities, i.e. the Department of Afrikaans and Dutch at the University of Stellenbosch and the Department of African Languages at the University of Pretoria, for research opportunities and assistance without which they would not have been able to write this book. They are also indebted to many colleagues, both locally and abroad, and students who have contributed in various ways and over many years to the ideas put forth in this book.

*Rufus Gouws and Danie Prinsloo*

# The development of lexicography: A brief historical perspective

---

## 1.1 Introduction

In this modern age, characterised by a knowledge explosion and a sophisticated information highway, dictionaries are still used as utility tools and their users rely on them as authoritative containers of knowledge. Although dictionaries are not an invention of the twenty first century, the twenty first century still sees them as household products, fulfilling their role as practical instruments as they have done for many centuries, cf. McArthur (1986) and Al-Kasimi (1977) for a discussion of some aspects of the history of dictionaries.<sup>1</sup> Today the field of lexicography is seen as having a twofold nature, i.e. a theoretical component and a practical component. The theoretical component focuses on research regarding e.g. the form, contents and functions of dictionaries whereas the practical component leads to the compilation of dictionaries. Lexicography has not always had this twofold character and an overview of the development of lexicography will give ample evidence that the theoretical component can be regarded as a relative late-comer because lexicography has originally only been associated with the practice of dictionary-making. Although it is today widely accepted that any good dictionary needs a sound theoretical basis, this has not always been the case. Dictionaries are much older than lexicographic theory and for many years dictionaries developed in a pretheoretical era, cf. Gouws (1989).

One of the salient features of dictionaries throughout many centuries is their function to assist users with real problems. This tradition of practical assistance had already been introduced in the early dictionaries, e.g. those compiled on clay tablets by the Assyrians to assist children in understanding Sumerian writings, the early Egyptian dictionaries written on papyrus leaves and the Arabic dictionaries giving their users access to the holy scriptures of Islam. As practical instruments these dictionaries were sources of knowledge, directed at the specific needs of specific user groups. Likewise the *glossae collectae*, compiled in Christian monasteries during the medieval era, were attempts to assist the students, qualifying themselves for a clerical career, in their confrontation with the biblical and classical languages. The compilation of these reference works was motivated on practical grounds and no theoretical framework or model had an influence on either their content or their structure.

The practical component of lexicography developed well into the second half of the twentieth century before it was complemented by a theoretical component. The advent of theoretical lexicography led to a number of early publications but by far the most important work and the first major publication to establish theoretical

---

<sup>1</sup> This chapter is a revised, shortened and adapted version of Gouws (2004b).

lexicography as a research field was the monumental *Manual of Lexicography*, authored by Ladislav Zgusta and published in 1971.

Soon after the publication of the *Manual of Lexicography* the influence of Zgusta's ideas was already noticeable, resulting in the rapid growth of theoretical lexicography but also in an improvement in the quality of new dictionaries.

The *Manual of Lexicography* clearly linked lexicography with linguistics, e.g. when Zgusta (1971:15) states it categorically that "lexicography is a very difficult sphere of linguistic activity." Zgusta also argues that a lexicographer needs to be familiar with linguistics in a much broader sense and has to take into consideration not only the whole structure of the language in question but also the culture of the respective linguistic community. By referring to the culture Zgusta makes way for an approach which compels lexicographers to contextualise the language in terms of the more general world of the relevant speech community. The advent of theoretical lexicography had been positioned within the broader linguistic framework. This would demand that lexicographers take cognisance of developments in linguistic theory and that the data presented in a dictionary should result from a sound linguistic analysis.

The publication of Zgusta's book heralded a new approach towards lexicography. The first four chapters of his book are not primarily concerned with lexicography but rather with linguistics, focusing on topics like lexical meaning, formal variation of words, combinations of words and variation in language. By including chapters on the formal variation of words and variation in language Zgusta gave a clear signal that linguistic influence does not only, or even primarily, run along the lines of formal grammar but the dictionary needs to reflect the real language usage and not only the language of the ideal speaker-hearer. In this regard lexicography constituted a form of opposition to the ideas of the Transformational Generative Grammar and could rather be seen as a forerunner of some of the ideas of sociolinguistics. Zgusta (1989) yet again focuses on the role of dictionaries in the development of the standard and in reflecting linguistic change. This emphasises a descriptive approach in lexicography in stead of the prescriptive approach that played such a dominating role in dictionaries.

Given the then scepticism prevalent among some linguists regarding the position of lexicography as a subdomain of linguistics this effort by Zgusta gave a clear signal that a sound lexicographic theory utilises sound linguistic principles. Zgusta's book, however, went further than this. Already in the introductory chapter Zgusta indicates the bivalent approach needed by a lexicographer when he says that the lexicographer is doing scientific work but publishes it for users whose pursuits are always more practical (Zgusta 1971:16). Once again one has to negotiate the very real distinction between the theoretical lexicographer and theoretical lexicography on the one hand and the practical lexicographer and the lexicographic practice on the other hand. Important in this statement by Zgusta is the fact that lexicography may not be regarded as a theory merely for the sake of theory. The broad and inclusive domain of lexicography has to be regarded as directed primarily at the process of dictionary compilation. Theoretical lexicographers devise theories aimed at enhancing the efforts of the practical lexicographer in his/her process of dictionary compilation. A dictionary can, among other things, rightfully be regarded as the display-window

of linguistics. The people looking at these display-windows are not trained linguists but rather the average members of the relevant speech community.

By referring to the fact that a dictionary is prepared for users whose pursuits may be more practical Zgusta introduced a point of view which would become a driving force in the lexicographic research of the nineties, i.e. the user-perspective. Lexicographic theory would contribute to models which would allow practical lexicographers to compile dictionaries aimed at a well-identified target user group, taking cognisance of their specific needs and reference skills. Yet again, lexicography is not working on an abstract level but provides in the real needs of real users. In a much later publication Zgusta (1988:vi) says that lexicography is one of the few areas in which linguistic activity has an immediate impact on many people. Lexicographic theory has to negotiate this issue in a very real way.

## **1.2 Within a linguistic fold**

Zgusta's book heralded a period which saw lexicography moving into a linguistic fold. Unfortunately some lexicographers did little to ensure an optimal utilisation of this situation and rather tried to maintain a theory-free practice. Even the title of a popular text book on lexicography, i.e. Landau (1989), does not help the cause of lexicography very much but plays into the hands of sceptics by referring to the "art and craft of lexicography". One of the reasons why the lexicographic practice still partially eschewed a stronger theoretical linguistic influence could be found in the commercial success of dictionaries. As long as it sells there is no need to change it. That this approach diminishes the linguistic authority of dictionaries did not convince enough publishers to give serious attention to a stronger linguistic and theoretical basis for their dictionaries.

Since 1971 varying degrees of proximity has prevailed between lexicography and linguistics with different theories and schools of thought in linguistics having a lesser or a bigger influence on both lexicographic theory and the lexicographic practice. Different linguistic theories, cf. Geeraerts (1984; 1986) and Gouws (1989), had an influence on e.g. the explanation of meaning in monolingual dictionaries. Different approaches to lexical semantics, the way in which the influence of structural linguistics led to a distinction between semantic and encyclopedic data and the much more lenient approach following from cognitive linguistics had an impact on the contents of the lexicographic definition.

The varying influence of linguistics on lexicography as seen in general dictionaries has primarily been noticeable in the nature and extent of the presentation of semantic data. General language dictionaries, both monolingual and bilingual ones, have displayed a strong semantic bias, cf. Gouws (1996), and this has been to the detriment of other data categories. Burkhanov (1998:136) states a widely accepted belief that lexicographic practice belongs to the domain of applied linguistics whereas metalexicography forms part of theoretical linguistics. He indicates that this belief has led to the assumption that linguistic semantics should provide the theory for lexicography. This would virtually equal linguistic semantics and metalexicography.

Variation was also noticeable in the nature, extent and treatment of other data types in general dictionaries. Data types like pronunciation, morphology, etymology and

even syntax have been presented and treated in a fairly consistent way. Although the outer texts of a dictionary displaying a frame structure, cf. Kammerer & Wiegand (1998), Gouws (2001a; 2002; 2004), often include a mini-grammar or even a brief explanation of some of the productive rules of word-formation or syntactic constructions, the articles in the central list still display an insufficient account of syntactic and morphological data. However, in line with the lexicographer's assignment to record the real language and not to set its style, cf. Sledd & Ebbitt (1962:92), modern-day lexicography has emphasised the importance of examples as an integral part of the treatment of a given word. Fox (1987:137) says that the use of examples forms an integral part of the learning of a word. These examples have to come from a corpus and have to represent real language (Fox 1987:138). The use of examples plays an important role in dictionaries and the influence from sociolinguistics has made a definite impact on the way in which lexicography deals with this type of entry.

During the seventies and eighties theoretical lexicography was performed and studied largely within a linguistic context. Many publications in the field of metalexicography focused on linguistic aspects of dictionaries resulting from a situation where many researchers working in the field of metalexicography were linguists by training and attached to university departments of linguistics or languages.

### 1.3 The Wiegand era

Metalexicography in the eighties and nineties was dominated by the work of the German scholar Herbert Ernst Wiegand. In his early work he already signalled the importance of the formulation of a general theory of lexicography, cf. Wiegand (1983; 1983a; 1984). Wiegand (1984:13-15) argues that lexicography is neither a branch of applied linguistics nor a branch of lexicology and it is by no means theoretically determined by lexicology alone. He regards linguistic lexicography as a scientific practice aimed at the production of reference works on language, whereas the field of metalexicography is constituted by the four components, i.e. the history of lexicography, a general theory of lexicography, research on dictionary use and the criticism of dictionaries. This approach of Wiegand to the lexicographic practice is ascertained and confirmed in Wiegand (1998:62, 254). Wiegand (1989:251) maintains that lexicography is a practice, aimed at the production of dictionaries in order to initiate another practice, i.e. the cultural practice of dictionary use. Wiegand (1998:256) also confirms the status of theoretical lexicography, he uses the term *dictionary research*, as a scientific research area and a discipline with a clearly identifiable academic existence, and maintains that dictionary research can be divided into four research areas, i.e. research on dictionary use, critical, historical and systematic dictionary research.

Although he recognises the importance of linguistics for lexicography Wiegand has initiated an approach that lexicography has to be regarded as a discipline which is influenced among others by linguistics but not to such a degree that it should still be regarded as a subdiscipline of linguistics. Linguistics has language as its study object. Practical lexicography is aimed at the process of dictionary-making whereas theoretical lexicography deals with dictionary research, cf. Hartmann & James (1998), Wiegand (1984; 1998) and Hausmann & Wiegand (1989). Although

linguistics is an important influence in lexicography, the object of lexicography is not language but dictionaries. Consequently lexicography cannot be regarded as a branch of linguistics, although it does overlap with various subdisciplines from the field of linguistics.

In his prolific portfolio of publications Wiegand has focused dictionary research not only on the contents of dictionaries and dictionary articles but also on the structure of dictionaries. Since Wiegand (1983b) numerous of his publications have dealt with wide-ranging issues regarding the structure of dictionaries. This, yet again, ascertained his approach that metalexicography is no branch of linguistics. By analysing and discussing the structure of dictionaries Wiegand has added a component to his successful attempts of formulating a general theory of lexicography that emphasises the formal features of dictionaries. The Wiegand era has been characterised by the identification of the different components of dictionary articles and by a meticulous description of their specific structure and function. In this process Wiegand has also made numerous suggestions in order to improve the quality of the lexicographic practice and to ensure that the genuine purpose of a specific dictionary can be achieved. The description of the structure of dictionaries has not been done in such a way that a theoretical model is formulated and then imposed on the lexicographic practice. Wiegand rather took a critical look at existing dictionaries to identify and describe their structural features. He has moved from the practice to the theory so that the theory could be applied to enhance the practice.

Although lexicography is not regarded as a subdiscipline of linguistics the strong link between linguistics and lexicography, both theoretical lexicography and the practice of the compilation of especially language dictionaries, may never be ignored. Consequently, Wiegand had been instrumental in seeing to it that lexicography also featured as a topic in the authoritative HSK series (*Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science*) resulting in the impressive state-of-the-art three volume *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography* (Hausmann et al 1989-1991). This work focuses on a number of relevant topics in lexicography, e.g. dictionaries and their public, dictionaries and their users, the history and theory of lexicography, components and structures of dictionaries, problems of description in the general monolingual dictionary, dictionary types, dictionaries dealing with language varieties, procedures in lexicographical work, lexicography of individual languages and the theory of bilingual and multilingual lexicography. A fourth volume of this book is currently being compiled and focuses on recent developments in lexicography, with special reference to computational lexicography, cf. Gouws et al (in preparation).

One of the noticeable features of developments in theoretical lexicography during the Wiegand era has been the strong bias towards the needs and the reference skills of the target users of dictionaries. In this regard the attention given by Wiegand to the structure of dictionaries played no mean role. His research regarding the access structure of a dictionary, i.e. the search route a user follows to reach the desired data, and the detailed discussion of aspects like the data distribution structure, the micro-architecture and the different search fields in dictionary articles, cf. Bergenholtz, Tarp & Wiegand (1999), as well as the use of integrated and non-integrated outer

texts, helping to constitute the frame structure of a dictionary, cf. Kammerer & Wiegand (1998); Gouws (2002), places the focus yet again on the user-perspective, so prevalent in modern-day metalexigraphy. Wiegand's contribution to the development of a general theory of lexicography has culminated in Wiegand (1998), the first volume of a comprehensive account of some of the most salient aspects of his research in the field of lexicography. The biggest part of this first volume focuses on research in dictionary use, including an exposition of the methodology of usage research and different types of dictionary consultation situations.

The focus on the structure of dictionaries during the Wiegand era emphasised the fact that as containers of knowledge, cf. McArthur (1986), both the contents and the form must be regarded as extremely important. From a purely linguistic perspective little interest exists in the structure of dictionary articles or the use of front and back matter texts and inserted inner texts. Neither does the layout of a dictionary fill mainstream linguists with excitement. The metalexigrapher, however, may not ignore the significance of these formal properties of a dictionary. Too often in the past dictionaries have not achieved an optimal transfer of data due to an insufficient presentation and a less than satisfactory form to accommodate the good contents. During the last decade the emphasis on both the contents and form of dictionaries has found another realisation, as pursued in publications like Bergenholtz (1995) and Almind & Bergenholtz (2002), i.e. a focus on problems relating to dictionary layout.

Almind & Bergenholtz (2002:261) indicate that layout is not a cosmetic issue but rather contributes to the access structure of a dictionary. A good layout enhances both the outer and the inner access structure. The choice of various typographical and non-typographical structural indicators is employed to ensure an optimal retrieval of information. Yet again the theory of lexicography goes further than purely linguistic procedures.

## 1.4 Specialised lexicography

Wiegand's arguments, cf. Wiegand (1984; 1989; 1998), that linguistics is only one of many disciplines influencing lexicography is illustrated by the attention in theoretical lexicography to the development of special-field lexicography, cf. Bergenholtz & Tarp (1995). Bergenholtz & Tarp make a distinction between language for general purposes (LGP) and language for special purposes (LSP). General dictionaries primarily deal with LGP, although some LSP items will also be included and treated in these dictionaries. Specialised dictionaries treat the various special fields of the lexicon. The compilation of LSP dictionaries presupposes collaboration between the lexicographer and the subject expert. The influence from the relevant subject field will determine the nature of the specific dictionary.

In the development of dictionaries for special purposes, theoretical lexicographers have been involved in devising models for a number of special field dictionaries. In this regard the work done by Henning Bergenholtz and Sven Tarp in the Centre for Lexicography at the Aarhus School of Business in Denmark needs to be mentioned. Their theoretical publications, cf. among others Bergenholtz & Tarp (1995), Tarp (2000), have played an important role in making practical lexicographers aware of salient theoretical issues. Their theoretical work has been complemented by a

variety of high quality and exemplary LSP dictionaries – both in printed format and on the internet. These publications display a sound theoretical base and reiterates the importance of a continued interactive relation between theoretical lexicography and the lexicographic practice.

The theoretical models devised by metalexicographers have been applied successfully in the lexicographic practice. Lexicography should also benefit from this expertise. At present a project is underway to compile a special-purpose dictionary with lexicography as its object: the *Wörterbuch zur Lexikographie und Wörterbuchforschung / Dictionary of Lexicography and Dictionary Research*, cf. Beißwenger et al (in preparation). Theory is once more put to practice. An important contribution of this dictionary will not only be the explanation and standardisation of more than four thousand lexicographic terms but one of the front matter texts of this dictionary will be a systematic introduction to the field of lexicography and dictionary research, cf. Wiegand (2003). This will enable the target user to ascertain a brief overview of the subject field of this dictionary.

## 1.5 Lexicographic functions

The development of lexicography shows some interesting themes appearing, disappearing and, sometimes, reappearing. One such a theme which is particularly relevant in modern-day lexicographic thought is that of lexicographic functions. The notion of functions is nothing new – it had already been introduced in 1940 by the Russian linguist Scerba. Following the suggestions made in Scerba (1940) some theoreticians in the field of dictionary research hold the view that for any given language pair at least four and perhaps even eight bilingual dictionaries have to be compiled to meet the diverse needs of the users coming from both language groups. According to them provision has to be made for separate dictionaries aimed at the active and passive use by source and target language users respectively, cf. Kromann et al. (1984, 1984a). The active/passive principle focuses on the function of dictionaries in text production and text reception respectively, cf. Hausmann (1977; 1986). Certain applications of the active/passive principle imply that four different functions can be identified for each member of a language pair and that each one of these functions should be dealt with in a separate dictionary. Even the most sophisticated system cannot work with four to eight dictionaries per language pair. From a user-perspective it also is an unattainable objective.

Lexicographic theory may not be formulated at the cost of a successful lexicographic practice. Therefore Wiegand (1996a:XV) emphasises the fact that the formulation of a theory for bilingual dictionaries, and it also applies to monolingual dictionaries, may never be isolated from the lexicographic practice. Consequently Wiegand (1996:2) pleads for the accommodation of different functions within one dictionary and even one dictionary article. He argues convincingly in favour of the compilation of only one polyfunctional bilingual dictionary for any given language pair.

During the last few years lexicographic functions once again came to the fore in lexicographic research, cf. Tarp (1994; 2000; 2002; 2002a), Bergenholtz & Tarp (2002), Wiegand (2001) and Tarp & Gouws (2004). In their respective publications on lexicographic functions Bergenholtz and Tarp have a different approach compared to that of Wiegand.

Modern-day lexicographic theory has an underlying assumption that dictionaries are utility products. Consequently Tarp (2002:67) argues that the methodology for planning a dictionary should make a typology of potential users, user situations and problems that might arise for each type of user in each type of user situation. The profile of the users must be determined, and eventually the relation between the needs of each type of user in each type of user situation and the data included in a dictionary to satisfy these needs constitute the basis for the theory of lexicographic functions. According to Tarp (2002:70) a lexicographic function represents the assistance that a dictionary provides to a particular type of user to cover the needs of that user in a specific user situation. Bergenholtz & Tarp (2002) distinguish between knowledge- and communication-orientated functions. These functions are discussed in par. 2.5.2.

The function of a dictionary is not only determined by the users but also by the usage situation and this situation in which a dictionary is used should have a definite influence on the data distribution programme and on the function of that dictionary. The focus on lexicographic functions emphasises the user-directed approach in modern lexicographic theory ever so strongly.

## **1.6 Electronic dictionaries**

The last decade has witnessed tremendous developments in the field of electronic dictionaries. The electronic medium has become increasingly important for the transfer of knowledge and lexicography had to respond to this. Developments in the field of lexicography saw numerous dictionaries being produced on CD ROM and on the internet. One of the major problems in the production of electronic dictionaries is an insufficient utilisation of the possibilities offered by this medium. Too often electronic dictionaries are little more than electronic variants of printed dictionaries. The electronic medium allows lexicographers a wholly new approach to dictionaries without space restrictions and the limitations which macrostructural ordering and the access structure put on printed dictionaries. A challenging endeavour for metalexigraphers is the development of a tailor-made model for the compilation of electronic dictionaries. A comprehensive research project focusing on such a model is currently in progress at STIAS, the Stellenbosch University Institute for Advanced Study.

# The lexicographic process

---

## 2.1 Introductory remarks

The publication of any dictionary should not only be the result of the preceding compilation activities but it has to be regarded as the culmination of a much more comprehensive set of activities, the so-called lexicographic process. The compilation and eventual publication of any dictionary form part of at least one lexicographic process. One of the major reasons why a dictionary often fails to be the kind of linguistic and communication instrument it is supposed to be, is the arbitrary way in which it has been compiled, due to a lack of planning prior to commencing with the compilation. This failure can be regarded as a direct result of an ill-founded lexicographic process. Lexicographers are often not aware of the fact that the work on any dictionary constitutes a lexicographic process, and that such a process compels them to adhere to certain planning and organisational criteria.

In recent lexicographic research, cf. Wiegand (1998), Gouws (1999a; 2000b, 2001a), much attention has been given to the extent and nature of a lexicographic process. A lexicographic process, cf. Wiegand (1998), is part of a comprehensive historical process which coincides with the development of a language. A lexicographic process is constituted by all the activities leading to the publication of a dictionary as a text. These activities do not only include the actual compilation but also the planning, data collection, etc.

One of the most annoying experiences in the process of dictionary consultation is to be confronted with an inconsistent presentation of data. A dictionary user wants to find the data (s)he is looking for as quickly as possible, and each consultation procedure should be done with the knowledge that the data is presented in a systematic way and that the lexicographers have compiled the dictionary according to a meticulous and consistently applied pattern. The successful application of a well-devised lexicographic process leads to dictionaries characterised by specific features, e.g. predictability, calculability, analysability and controllability, cf. Wiegand (1997). To ensure this, a lexicographic process should always contain a reflexive component, which enables the editors to reflect continuously on the finished work as well as the continuing activities. However, the most important aspect is to devise the entire lexicographic process long before the compilation of a dictionary commences and to ensure that all the activities and details relevant to the planned dictionary have been sufficiently identified and addressed in the formulation of this lexicographic process. A lexicographic process represents the blueprint for the compilation of a dictionary and although different dictionary projects often share the same goals and objectives, each dictionary project needs to be the product of its own lexicographic process, formulated uniquely for that dictionary. Such a lexicographic process has to focus on all the activities resulting in the eventual publication of that dictionary.

The establishment of a lexicographic process leads to the formulation of a dictionary plan, which forms the basis of all the decisions regarding the compilation of the

dictionary. The dictionary plan includes two main components, i.e. the *organisation plan* and the *dictionary conceptualisation plan*. The organisation plan is primarily directed at the management and logistics. The dictionary conceptualisation plan is concerned with the more direct lexicographic issues and focuses on aspects like the lexicographic functions, dictionary typology, the target user, the structure of the dictionary, the lexicographical presentation, etc.

The present situation in South Africa sees various dictionary projects undertaken by the different national lexicography units (NLUs), established by PanSALB (the Pan South African Language Board), commercial publishing houses and private lexicographers. The National Language Service of the Department of Arts and Culture (DAC) undertakes projects in the field of dictionaries compiled for languages for special purposes. The overall planning of lexicographic activities in South Africa should preferably be governed by a comprehensive lexicographic process so that the different role players do not duplicate unnecessarily and can focus their efforts on projects that are really needed by the different speech communities. This means that there has to be liaison between the different participants in this process. Such an overall lexicographic process is known as the *primary comprehensive lexicographic process*. Within each NLU a variety of projects will eventually be launched. Not only should each individual project be planned well in advance but each NLU will necessarily also have to negotiate the range of activities to be undertaken within that NLU to ensure that each project has its rightful place within the spectrum of activities. Although each dictionary project has to be the focus of a separate lexicographic process, each NLU will need a plan according to which the different projects have to be tackled. This comprehensive plan of a NLU is called the *secondary comprehensive lexicographic process*. It includes the formulation of a lexicographic process, a *specific lexicographic process*, for each separate project, cf. Gouws (1999a; 2000b).

The following hierarchy can be suggested for the different lexicographic processes in the present South African environment:

*The primary comprehensive lexicographic process*

(with PanSALB, commercial publishers and DAC as role players)

*Secondary comprehensive lexicographic process 1 ..... Secondary comprehensive lexicographic process n*

(with the NLUs, commercial publishing houses and other lexicographic institutes as role players)

*Specific lexicographic process 1 ... Specific lexicographic process n*

(for each secondary comprehensive lexicographic process).

## **2.2 A primary comprehensive lexicographic process**

PanSALB has to regard the establishment and maintenance of the NLUs as the most important part of the primary comprehensive lexicographic process in South Africa, and they have to be well aware of the vital role they play in this process. All the work which PanSALB had done in preparation of the establishment of the NLUs falls within the domain of the primary comprehensive lexicographic process. The primary comprehensive lexicographic process also makes provision for a continued

PanSALB commitment to ensure the ongoing success of the NLUs. This is a vital component of the basis of a theoretical model for the compilation of the dictionaries in the NLUs. The secondary comprehensive lexicographic processes to be hosted by each NLU are subordinate to the primary comprehensive lexicographic process. PanSALB represents all the NLUs in the primary comprehensive lexicographic process and PanSALB's position as dominant role player should imply the responsibility they have towards all the NLUs. One of the most important issues in this regard is the training of staff members. Although the staff members of each NLU need training directed at language-specific issues to be addressed in the dictionaries they plan to compile, all the NLU staff members also need some general non-language specific lexicographic training. To save time and money and to prevent the duplication of training sessions, the primary comprehensive lexicographic process should accommodate an ongoing training programme for all the NLUs. Training is not only a part of a comprehensive lexicographic process but also a part of the model for the compilation of any dictionary.

### **2.3 A secondary comprehensive lexicographic process**

The secondary comprehensive lexicographic process forms the basis for the planning and co-ordination of all the dictionary projects undertaken in a NLU or in any given publishing house. This ensures that no project is the result of an arbitrary decision. The identification of short, medium and long term dictionary projects is one of the first assignments in establishing a secondary comprehensive lexicographic process. The prioritisation of dictionary projects as short, medium or long term objectives has to reflect a user-driven approach which is sensitive to the reference needs and reference skills of the speech communities, and an identification of the real lexicographic needs of a speech community has to be seen as a prerequisite for this prioritisation activity. Working towards a model for the compilation of any dictionary implies a project management infrastructure. This is a basic component of the secondary comprehensive lexicographic process and it needs to be put into place at the earliest possible opportunity.

One of the telling differences between older dictionaries and their modern-day counterparts, compiled within the framework of a well-devised lexicographic process, is the fact that the latter dictionaries are directed at very specific target user groups, and each dictionary fits into a very specific niche within the reference collection provided for the target users. Dictionaries used to be compiled without a specific target user in mind. This may no longer prevail. Hartmann (1989) so aptly indicates that the compilation of any new dictionary has to be preceded by an in depth analysis of users' needs. The user is central to almost every aspect of lexicographic planning. User-driven lexicography has an influence on the typological nature, the structure, contents and presentation of every dictionary, i.e. on every aspect of the lexicographic process of any given dictionary and any secondary comprehensive lexicographic process designed to co-ordinate different lexicographic projects.

As a result of the prominence of the user-perspective in modern-day dictionary research a variety of methods have been devised to assist the lexicographers in identifying the target user and establishing the typological preferences, needs and reference skills of these users. Relying on the results of this research can help to acquire the necessary expertise in this component of a secondary comprehensive

lexicographic process. The following publications could be of great value in this regard: Wiegand (1998), Van der Merwe-Fouché (1999), Hartmann (1989), Otto (1989).

Adhering to the criteria of user-driven lexicography implies the need for the secondary comprehensive lexicographic process to recognise the importance of user-friendliness and a dictionary culture in the relation between the dictionary projects and their potential target users. In terms of Hausmann (1989:13) a dictionary culture implies that the members of a given speech community are familiar with different types of dictionaries and with the contents and presentation of these dictionaries. It also implies that they have reached a certain level of dictionary using skills. These skills do not come instinctively but need to be acquired. User-friendliness refers to the way in which dictionaries are directed at the dictionary culture level of the target users. The target users of a general language dictionary are not academics and students but the average members of the speech community who can be empowered by access to a dictionary. The formulation of the model of any dictionary depends on the ability of the target users to use the dictionary and the secondary comprehensive lexicographic process has to negotiate the proper relation between user-friendliness and a dictionary culture to ensure that the lexicographic products to be published will empower the intended target users.

## 2.4 Dictionary specific lexicographic processes

In modern-day metalexicography the notion of lexicographic planning is emphasised and re-emphasised. However, this is not a new concept. The publication of Samuel Johnson's famous dictionary in 1755 was preceded in 1747 by his *Plan of a dictionary of the English language*. Partly due to the fact that this dictionary was the result of a proper planning exercise, it became one of the most influential dictionaries which, for a lengthy period, played an extremely important role as reference instrument in England and later also in America.

The compilation of each individual dictionary has to be done within the framework of a lexicographic process, specifically devised for that dictionary. Where a dictionary forms part of a wider range of dictionaries, its dictionary specific lexicographic process must function subordinate to the overall secondary comprehensive lexicographic process of that body.

Each lexicographic process leads to the formulation of a dictionary plan which can be divided into an organisation plan and a dictionary conceptualisation plan. This dictionary plan plays a vital role in determining the model of the dictionary to be compiled. Just as each dictionary project needs its own lexicographic process, each dictionary needs its own model. A lexicographic process and a dictionary model both display a project-specific nature. A theoretically motivated model for the compilation of dictionaries has to make provision for different models applicable to each individual dictionary type. However, these dictionary specific models will necessarily have a lot of features in common as well as some unique features. The dictionary specific lexicographic process should emphasise both these aspects.

## 2.5 The dictionary plan

### 2.5.1 The organisation plan

The organisation plan is primarily directed at the logistics of the project and all the managerial aspects.<sup>2</sup> This planning is essential for the success of any dictionary project and the logistic and managerial infrastructure must precede any editorial work. The editor in chief, in co-operation with the Board of Directors, where applicable, will usually be responsible for the formulation and implementation of the organisation plan of a specific dictionary. The organisation plan should include a budget as well as a programme to indicate the nature and extent of the duties of each member of staff involved in the specific dictionary project.

One aspect regarding a dictionary plan which has been neglected in the planning of so many dictionaries but which forms an integral part of the organisation plan is the identification and formulation of the *genuine purpose* of the intended dictionary. This is of extreme importance to ensure a sound theoretical point of departure for the compilation process.

### 2.5.2 The genuine purpose of a dictionary and lexicographic functions

Dictionaries are practical instruments and are compiled to be used by a specific target user group and to fulfill a specific purpose and specific functions in the different situations of usage. The identification and formulation of this purpose and of the specific lexicographic functions have to precede the compilation process because the compilation process should be steered by the purpose and functions of the specific dictionary project. In metalexicographical terms this purpose of a dictionary is known as the *genuine purpose* and the functions are known as lexicographic functions. The organisation plan of every dictionary project has to include a clear and unambiguous exposition of the genuine purpose and the lexicographic functions of the dictionary to be compiled. The genuine purpose is reached when all the functions of the dictionary have been achieved successfully.

The genuine purpose of a dictionary is co-determined by, among others, its typological nature and its intended target user group. Desk or standard bilingual or monolingual dictionaries, the typological categories with a high usage frequency, belong to the broader category of linguistic dictionaries. Their genuine purpose is to transfer, by means of lexical data, information regarding the set of lexical items included as treatment units in order to ensure the linguistic empowerment of the intended target user. The genuine purpose of a dictionary implies that a dictionary is produced so that the target user who uses the dictionary in a typical usage context will have an instrument to assist him or her in achieving a successful dictionary consultation procedure by reaching the goals that motivated the search. The genuine purpose of a dictionary should therefore be to ensure successful dictionary consultation procedures. A successful dictionary consultation procedure depends on the way in which the needed linguistic information can be retrieved, and this depends largely on the way in which the dictionary is steered by its functions.

---

<sup>2</sup> The discussion in par. 2.5 is an adapted version of a section of Gouws (2001b).

Although the genuine purpose of a dictionary has major implications for the dictionary conceptualisation plan because of its direct impact on the structure and contents of the dictionary articles and the data distribution pattern, the organisation plan must make provision for a managerial infrastructure able to cope with the compilation of a dictionary that fits the requirements implied by its identified genuine purpose. The extent of the data transfer and the eventual information retrieval to be achieved by the intended user will differ from project to project, and during the planning phase when the organisation plan is established, the genuine purpose has to be formulated to form a basis for the dictionary conceptualisation plan.

The formulation of the genuine purpose of a dictionary has to be regarded as a response to the needs of the potential target users. General dictionaries, e.g. desk and standard bilingual and monolingual descriptive dictionaries, should be compiled for the average member of a speech community and not for academics. The needs of the speech communities have to be put before the more sophisticated reference needs of e.g. linguists and academics. Reference needs have to determine the genuine purpose of the dictionaries. This will necessarily have an influence on the typological choice as well as the structure, contents and presentation of the intended dictionaries.

The notion of lexicographic functions is not new in the field of dictionary research. The research done by the Russian lexicographer Scerba (1940) provided the base for this approach. Recent developments in the field of lexicographic theory, spearheaded by the Danish lexicographers Henning Bergenholtz and Sven Tarp, saw lexicographic functions taking a central position. Today the compilation of every dictionary needs to be done in accordance with one or more specific lexicographic functions, cf. Tarp (2000). Bergenholtz & Tarp (2002) distinguish between knowledge- and communication-orientated functions. Knowledge-orientated functions assist the user by providing

- general cultural and encyclopedic data
- special data about the subject field
- data about the language.

Communication-orientated functions assist the user to solve problems related to

- text production in the native language
- text production in the foreign language
- text reception in the native language
- text reception in the foreign language
- translation of texts from the foreign to the native language
- translation of texts from the native to the foreign language.

Along with the formulation of the genuine purpose of a dictionary the identification and planning of the relevant functions play a central role in the lexicographic process.

### **2.5.3 The dictionary conceptualisation plan**

In the planning of a dictionary project the dictionary conceptualisation plan, as part of the dictionary specific lexicographic process, has the most direct influence

on the compilation process. According to Wiegand (1998:151) the dictionary conceptualisation plan can be divided into five subdivisions, i.e.

- the general preparation phase
- the material acquisition phase
- the material preparation phase
- the material processing phase and
- the publishing preparation phase.

This should form an integral part of the dictionary plan in any lexicographic process. In the following paragraphs a brief indication will be given of the contents of each phase.

### 2.5.3.1 *The general preparation phase*

The general preparation phase of the dictionary conceptualisation plan lays the foundation for the structure, contents and presentation of the final product. One of the first assignments the staff members of a dictionary project will have when commencing with the general preparation phase, is the compilation of a lexicographic *instruction book*, also known as a lexicographic style guide, cf. Bergenholtz (1990). The instruction book should contain a comprehensive description of the system applied in the dictionary. This document is the most important instrument in the hand of the editorial staff members to ensure a consistent and systematic presentation and treatment in the compilation process. The lack of a well-devised instruction book is bound to lead to a chaotic dictionary, especially in the case of a project where more than one person is responsible for the lexicographic treatment. Bergenholtz (1990) gives a good idea of the typical issues to be dealt with in an instruction book. In the South African lexicographic environment the style guides used by the Bureau of the WAT and the Dictionary Unit of South African English are good examples of instruction books. Each NLU should devise instruction books for each individual dictionary project. A specific unit can devise a general instruction book to form the basis of instruction books for all their individual dictionary projects.

Issues to be dealt with in an instruction book, cf. Bergenholtz (1990), include the lemmatisation process (with reference to e.g. the influence of initial capital letters, diacritics, the order of word and stem forms), the use of typographical and non-typographical structural indicators in the articles, the marking of different senses of a lemma, the use of abbreviations in the metalanguage of the dictionary, the positioning and marking of new search fields in the article, etc.

A second issue to receive attention in the general preparation phase is the *microstructural programme* of the dictionary. From a very early stage of the lexicographic process, the staff members should be well-aware of the microstructural programme of a dictionary, i.e. the different data categories to be included in the treatment of the lemmata and the typical article slots allocated to these categories. This planning has to be done as part of the general preparation phase because it will assist the staff to determine the quantitative extent of the dictionary. A dictionary always has space limitations and an early identification of the microstructural programme is necessary to ensure a functional space budget. A further value of an early identification of the microstructural programmes is that it assists the lexicographers during the early phases of the compilation process to focus on

those data categories that will be included in the dictionary. If a lexicographer knows beforehand that no article slot will be reserved for a specific data category, the allocation of editorial tasks can be simplified because no one has to acquire the expertise to deal with the treatment of that data type.

One of the typical consequences of a dictionary planned without attention to the microstructural programme is the situation where the lexicographers decide in a haphazard way to include a certain data category in a specific article and omit it from the next. This implies that the dictionary cannot adhere to the predictability criterion and functions in an unsystematic way. The formulation of a microstructural programme is a vital component of the planning of a dictionary compiled according to the norms and criteria of a well-devised model.

A third issue to be dealt with in the general preparation phase is the planning of the *frame structure* of the dictionary. The frame structure is discussed in more detail in Chapter 6. At this stage it is sufficient to say that modern dictionaries typically display a frame structure, consisting of the central list, i.e. the alphabetical component of a general dictionary or the so-called dictionary proper, the front matter and the back matter. When planning and compiling a dictionary, special care should be given to the use of outer texts, i.e. the texts included in the front and back matter. These texts assist the user to ensure successful dictionary consultation procedures and they allow the lexicographer to include a more comprehensive variety of data in the dictionary. It is important that the frame structure should be planned well in advance so that the lexicographers know which texts and what kind of data should be included in the front and back matter texts. Because the frame structure expands the function of the dictionary as a container of knowledge the lexicographer has the opportunity to include data categories not typically accommodated in the central list of a dictionary in the outer texts. This structure gives a definite added value to a dictionary.

A fourth issue to be dealt with in the general preparation phase is the identification, establishment, nature, extent and description of a *dictionary basis* which suits the relevant dictionary project in the best possible way. A dictionary basis, cf. Wiegand (1998:139), can be described as the total of the source language material for the specific lexicographic process. This includes all the possible sources which accommodate such material, as well as informants and mother-tongue speakers of the language who can assist the editorial staff in the building up of a material collection. The dictionary basis will differ from dictionary project to dictionary project according to the typological nature of the dictionary.

The dictionary basis of a general monolingual or bilingual dictionary can be compiled from three types of sources, cf. Wiegand (1998:140). The *primary sources* of the dictionary basis will be all the written material reflecting typical communication situations. Although the primary sources will usually be texts, the dictionary basis of a dictionary compiled for a language with a strong oral tradition can also use recordings of the orature as primary sources. The *secondary sources* are all the available dictionaries in the specific language. However, in this regard lexicographers have to be very careful not to perpetuate lexicographic failures of the past in the new dictionary projects. The *tertiary sources* comprise of all other linguistic material that can be used, e.g. linguistic monographs, papers and grammars.

An early identification of the dictionary basis enables the lexicographers to apply a well-directed material collection policy which in its turn allows a more rapid macrostructural selection. This is of special importance for the compilation of dictionaries for the African languages. One of the problems when compiling a corpus for these languages is the relatively limited amount of written sources. In contrast to their limited written sources the African languages have a rich orature. In many cases a corpus compiled only from written sources will not be fully representative of the lexical stock of the language. Having determined the target user and the typological category of a dictionary to be compiled, the lexicographers can ascertain whether the written sources will be sufficient to render the dictionary basis needed for the specific project or whether other sources, e.g. mother-tongue speakers who can convey some orature, should be consulted to complement the written sources as a component of the dictionary basis.

### *2.5.3.2 The material acquisition phase*

A logical phase to follow the identification and formulation of a dictionary basis in the general preparation phase of the dictionary conceptualisation plan, is the material acquisition phase. This phase is primarily aimed at establishing the dictionary basis identified during the general preparation phase. It precedes the compilation process and focuses on the gathering of speech material from the sources earmarked for the dictionary basis. A result of the material acquisition phase is the compilation of the lexicographic corpus – the collection of items gathered from the primary, secondary and tertiary sources of the dictionary basis.

In modern-day lexicography the material acquisition phase will inevitably lead to a corpus.

No modern dictionary can be representative if it is not based on a reliable corpus. The compilation of corpora has to be regarded as a highly skilled activity and dictionary units have to make ample provision in their planning for this important aspect of their lexicographic endeavour. In the material acquisition phase the infrastructure for corpus building and therefore also the infrastructure regarding computational aspects, have to be in place. A dictionary specific lexicographic process focuses on much more than the mere compilation process.

### *2.5.3.3 The material preparation phase*

During this phase the lexicographers have to prepare the collected material for the next steps of the lexicographic process. Especially in the case of oral material, the recordings have to be transcribed and scanned into the computer for eventual inclusion in the corpus. This phase also gives the staff the opportunity to sort the material in order to omit material that cannot be used. By the end of the material preparation phase the corpus should be in good order and the lexicographers must be in a position where they can utilise the corpus to select the citations and examples to be included as verbal illustrations in the dictionary article. This is one of the beginnings of the compilation phase because some of the activities are already directed at filling specific article slots.

Once the corpus is in order the lexicographers can proceed with the macrostructural selection to present the lexical items to be included as lemmata in the dictionary. This is done in accordance with the typological criteria of the specific dictionary.

The lemmata function as the most typical treatment units of a dictionary and once these treatment units have been selected, ordered and presented as guiding elements of their respective articles, the lexicographers are in a position to apply the lexicographic treatment by activating the microstructural programme.

The proper execution of the material preparation phase establishes a good basis for the actual compilation process which forms the central activity of the next phase of the dictionary conceptualisation plan.

#### *2.5.3.4 The material processing phase*

Prior to the material processing phase the frame structure and the microstructural programme have been devised and the macrostructural selection has been completed. This implies that the lexicographers know what data to include in the dictionary, where to include each entry and at which primary treatment units the entries have to be addressed. The material processing phase comprises the application of the data distribution structure and the writing of the dictionary texts. The data distribution structure of a dictionary, cf. Bergenholtz, Tarp & Wiegand (1999), determines the specific position of each data type in the dictionary as a so-called carrier of text types. Some data will be included in the texts accommodated in the front and back matter while other data will be included in the articles, i.e. the texts constituting the central list of a dictionary.

Once the microstructural programme has been formulated and the macrostructural selection has been completed the lexicographers are in a position to pursue the construction of the dictionary articles as texts in the central list of the dictionary.

#### *2.5.3.5 The publishing preparation phase*

The final phase of the dictionary conceptualisation plan is the preparations for the publishing of the dictionary. From an editorial point of view this phase is directed at the various stages of proofreading and final adjustments to the manuscript. The dictionary plan has to make provision for a rapid and functional execution of this part of the lexicographic process. It is important that this phase of the lexicographic process, albeit the final part of the lexicographic process, should be planned during the early phases of the lexicographic process. The publishing preparation phase determines that lexicographers should know right from the start of their project for what medium of publication the dictionary has to be prepared, as a printed dictionary, an electronic dictionary or in both formats. This has definite implications for the data distribution structure and actually every aspect of the compilation process. If the choice is made for an electronic dictionary a further decision regards the specific format, e.g. as a version on CD ROM or as an internet dictionary.

The publishing preparation phase for a printed dictionary demands a lot of work throughout the different phases of the lexicographic process. Lexicographers need the assurance that the dictionary will be published. Therefore it is important that negotiations with printers and publishing houses should be done at an early stage to ensure the eventual publication of the dictionary. It is often necessary that the production of the manuscript should adhere to guidelines from the publisher and if these guidelines can be followed from the first manuscript work, a lot of time can be saved.

In the past the publishing preparation phase had too often been regarded as something outside the scope of the lexicographic process, being the responsibility and assignment of a printer or publisher. Lexicography was previously regarded as a subdiscipline of linguistics and consequently the focus in lexicographic research and discussions used to be on the linguistic contents of dictionaries. Changes and developments, cf. Chapter 1, led to a focus shift from an exclusively contents-directed approach to an inclusive approach which puts the structure, functions and usage situations of dictionaries within the scope of lexicographic research and which elevates lexicography to a discipline in its own right. This focus shift has had definitive implications for the nature and extent of the lexicographic process and, also, for the publishing preparation phase. As a carrier of text types, cf. Wiegand (1996), and a container of knowledge, cf. McArthur (1986), the success of a dictionary also depends on the ease with which it can be used and the swiftness with which the data on offer can be accessed. The presentation of the data becomes increasingly important when one evaluates a dictionary as a utility instrument.

The layout of dictionaries forms an integral part of the publishing preparation phase. This is no longer a task solely for someone in the printing division but it is a vital assignment and a part of the lexicographic process in which the lexicographers should have a deciding say. The layout includes the design of the cover, the entries on the spine, the title page, etc. and this constitutes an important part of the outer access structure of the dictionary. Therefore the dictionary conceptualisation plan should also focus on various aspects regarding the layout as part of the publishing preparation phase, cf. Almind (2005) and Almind and Bergenholtz (2000). If the layout issues are settled at an early stage of the lexicographic process a lot of time is saved in the publication preparation phase.

## **2.6 The use of the dictionary**

The scope of the lexicographic process includes every aspect of the planning, compilation, production and publication of a dictionary. Wiegand (1997) has indicated the need that any lexicographic process should contain a reflexive component that compels the lexicographer at any particular time of the process to look back at the completed work in order to ascertain whether it had been done in a satisfactory way. Although the formal task of a lexicographer can be regarded as completed once a dictionary has been published the completion phase does not give evidence of the success or failure of the specific dictionary as a utility instrument. This can only be evaluated once the dictionary has been used by its target users in typical usage situations. The reflexive component of a lexicographic process should therefore not be limited to the section of the lexicographic process representing the planning, compilation, production and publishing phases. It also needs to be applied once the completed product is put to use. As indicated in Chapter 1, Wiegand (1989:251) maintains that the production of dictionaries initiates a further practice, i.e. the cultural practice of dictionary use. This practice is the final test for the success or failure of any given dictionary and consequently the lexicographic process should include an application of the reflexive component which looks back on the results of dictionary use. This component should focus on real instances of dictionary consultation in typical situations of usage. The application of the reflexive

component should constitute a formal part of the lexicographic process and the results should be evaluated by the lexicographers in order to enhance the quality of future revisions of the dictionary or the lexicographic process of other dictionaries.

# Material collection and corpus building

---

## 3.1 Building text corpora

It has been stated in Chapter 2 that the compilation process is preceded by the collection of written and spoken material from the sources earmarked for the dictionary basis. Data is compiled and stored as a lexicographic data basis which should preferably be an *electronic corpus*. An electronic corpus can be defined in an oversimplified way as a computerised collection of texts. Such a collection of texts can, for example, consist of tape recordings of conversations and written texts which have been typed into the computer. The following paragraph is an extract from one of the texts from a Sesotho sa Leboa electronic corpus.

KAONAFALO YA DIOFISI TŠA METSE YA LEBOWA

Kgwele bjale gona o ka re e dinong, ga e sa le kgauswi le dino mo metseng ya Lebowa mabapi le kaonafalo ya diofisi tša mebušo ya ditšhaba. Metse ya Lebowa e a phadišana ka go aga diofisi; ga go na le wo o ratago go šalela morago. Rena bagatiši ba pukwana ya “Tšwelopele” re akela phadišano ya mohuta wo. Re a tseba gore moo go nago le phadišano, go na le tšwelopele, go na le kgatelopele ya mannete. Ka go aga diofisi tša Mmušo wa Setšhaba, re bona go se na le kgoši yeo e ratago ga phalwa le go šalela morago. Seo se re kgahlago gape ke gore diofisi tše kgolo di tlile go nyaka bahlankedi ba bantšhi. Moo go molaleng. Klereke e tee e ka se šome e nnoši diofising tša go feta tše lesome. Ka gona le taolo ya Mmušo wa Setšhaba e tlo phakišwa le go laolwa gabotse. Wena Kgoši o eme kae ka diofisi tše kgolo? Le ge o se o thome, tseba, lebelo ga le na le motlogapele.  
Nkadimeng (1985:5)

This paragraph in itself can be regarded as a small electronic corpus containing 182 running words or *tokens*. Some of these 182 words occur more than once in the paragraph, e.g. *le* (14), *go* (13), *ya* (10). There are thus only 87 *different* words or *types* in this little corpus. Words occurring only once in a corpus (in most cases roughly 50% of the words) are referred to as *hapax legomena* or *hapaxes* for short. *Le*, occurring the most times in the corpus has the highest *rank* and the *hapaxes* have the lowest *rank*.

Kennedy defines a corpus more formally as follows:

In the language sciences a corpus is a body of written text or transcribed speech which can serve as a basis for linguistic analysis and description.  
(Kennedy 1998: 1)

Ideally a corpus should contain large quantities of both spoken and written data. Spoken data is absolutely essential especially for those African languages which do not as yet have many written sources. The study of oral data can pinpoint words which tend to be used more frequently in oral *versus* written communication. Unfortunately most corpora around the world lack sufficient data from spoken sources. The reason for this is that there are many logistical problems and ethical

factors involved in the collection of spoken data. It is also much more expensive and time consuming to enlarge the corpus with spoken data compared to data available in electronic, printed or even handwritten format. Extending the corpus with data already in electronic format such as texts downloaded from the internet or texts already available on computer disk is relatively easy. Printed matter which is not available in electronic form can also relatively easily be computerised by means of Optical Character Recognition (OCR), commonly referred to as 'scanning'. Scanned documents often contain a variety of predictable and unpredictable mistakes or scanning errors which have to be rectified manually.

Typical scanning errors are given in Table 1; The symbol ' $\leftarrow \rightarrow$ ' indicates that the scanning error can occur in both directions, e.g. ' $c \leftarrow \rightarrow o$ ' means an 'o' is incorrectly scanned as 'c' and 'c' is incorrectly scanned as 'o'.

**Table 1: Typical scanning errors**

1 $\leftarrow \rightarrow$ l	c $\leftarrow \rightarrow$ o	n $\leftarrow \rightarrow$ ri	š $\rightarrow$ s
0 $\leftarrow \rightarrow$ O	lc $\leftarrow \rightarrow$ k	Š $\rightarrow$ S	c $\leftarrow \rightarrow$ e
m $\leftarrow \rightarrow$ in	hl $\leftarrow \rightarrow$ ki	i $\rightarrow$ l	B $\leftarrow \rightarrow$ 8

The corpus compiler will soon learn which typical scanning errors occur for a specific language and it is advisable to study the first few test pages of especially a major source to be scanned for the nature of such errors. Some typical scanning errors detected beforehand could be reduced by 'teaching' the software how to handle them if the software package has a so-called training option. A good example for scanning Sesotho sa Leboa texts with *Omnipage* is training the software beforehand on a number of pages from the source to be scanned not to scan š as s, or ê as e or ô as o. Such action will at least reduce the number of scanning errors that have to be corrected afterwards. Correcting scanning errors, however, is unavoidable – the most practical strategy in addition to proofreading the text word by word is to use the normal 'search and replace' function of a word processor either manually (correcting errors one by one) or automatically if a recurring error pattern is unique.

An additional or alternative strategy is to use available spellcheckers for English, Afrikaans and the African languages. No affordable scanning solution currently exists for handwritten material such as numerous card collections of dictionary articles, and retyping them is the best option. Retyping even typed sources is often a better option than scanning them if the quality of the material is bad, e.g. very old typed card collections of dictionary articles. Copyright issues should always be considered before embarking on the scanning of sources, especially in the case of the use or potential use for any other purpose than purely academic research, e.g. publication of commercial products. Copyright laws and prosecution acts regarding corpus material in South Africa seem to be vague or lacking but that is no reason not to follow proper procedures before scanning material or downloading material from the internet.

As for the size of the corpus, it is generally assumed that 'bigger' means 'better' although there are many counter arguments to such a viewpoint. What is true is that corpus sizes have constantly grown in the past decade. The earliest major electronic corpus, the *Brown University Standard Corpus of Present-Day American English* (also known as the *Brown Corpus*), contained roughly one million words

(Francis & Kucera, 1964). At that stage and for a few decades to follow, one million words had been regarded as a kind of norm or the 'going rate' for corpus size. Since the early nineties, however, huge corpora were built such as the *Collins Birmingham University International Language Database (COBUILD)*. Between 1991 and 1995 the *British National Corpus (BNC)*, contained about 100 million words and the *Bank of English*, more than 320 million words in 1998. Since 2000 corpus sizes continued to grow by several hundreds of millions of words. Collections such as the *Media24 Archive for Afrikaans* is estimated at 800 million words and could well be one of the biggest corpora in the world.

Apart from the issue of corpus size, corpus compilers paid particular attention to the nature, types and quantities of material collected and included in their corpora. This generally revolves around two major concepts 'balanced corpora' *versus* 'representative corpora'.

A general corpus is typically designed to be **balanced**, by containing texts from different genres ... including spoken and written, ... (Kennedy, 1998: 20)

For a corpus to be 'representative' there must be a clearly analysed and defined population to take the sample from. (Kennedy, 1998: 52)

Questions associated with 'representativeness' and 'balance' are complex and often intractable. (Kennedy, 1998: 62)

The idea of representativeness has been central to our thinking about the structure of the corpus. We believe that unless the corpus is representative, it is *ipso facto* unreliable as a means of acquiring lexical knowledge. Our answer to the question: 'Representative of what?' would be 'Representative of the standard language in a very general sense, not restricted to a regional variety [...] or a narrow range of text types' [...] What we mean by *representative* is covering what we judge to be the typical and central aspects of the language, and providing enough occurrences of words and phrases for the lexicographers [...] to believe that they have sufficient evidence from the corpus to make accurate statements about lexical behaviour. (Summers, 1993: 186, 190)

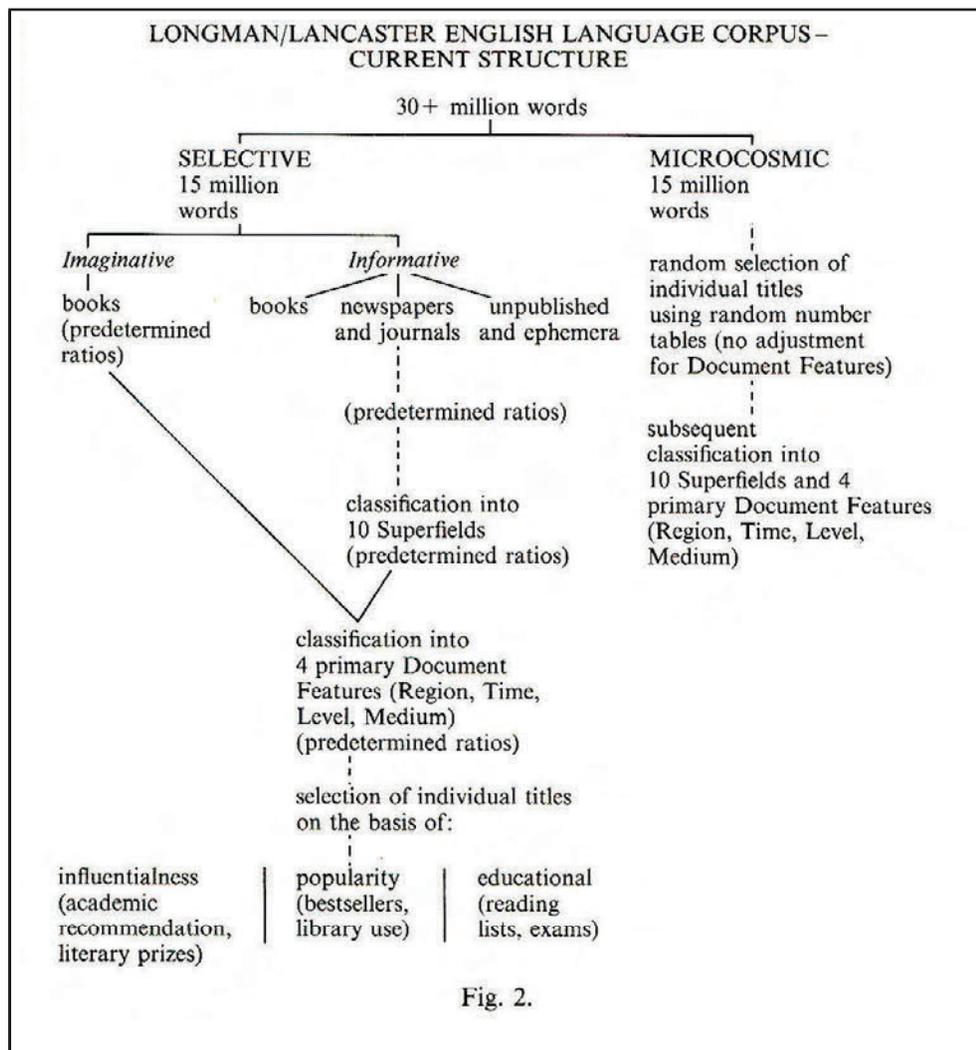
... to be representative of general language. This is a bold ambition – some say one that is impossible to fulfil. (Summers, *s.d.* [1996-1998]: 6)

Kilgarriff relates the issues 'size', 'balance' and 'representativeness' as follows:

COBUILD have always insisted that it is impossible to create a corpus that is truly representative of the language, and have focused on size of corpus rather than balance. (Kilgarriff, 1997: 150)

Corpus compilers also paid much attention to corpus planning and design - the design of the *Longman Lancaster English Language Corpus* is a good example of such efforts. Two strategies are apparent: a 'selective half', chosen through a mixture of pragmatic measures to gather a broad range of objectively defined 'document types', and a 'microcosmic half' compiled by a random selection of books.

## Design of the Longman Lancaster English Language Corpus (Summers, 1993: 201)



The reality for most of the African languages is such that a neatly designed collection strategy is not possible and the whole selection process eventually boils down to the collection of all *available* texts for the specific language. In many instances available texts have to be heavily supplemented by sub-corpora compiled from oral data collections in order to reach corpus sizes of a few million running words.

Important for lexicographic work in South Africa is that corpus compilers should be sensitive to all of these aspects. i.e. to build as far as possible, corpora that are big enough, well balanced and representative so that valid conclusions for lexicographic purposes can be drawn.

Lexicographers traditionally aim at a 'representative' or 'balanced' corpus, that is, the corpus should be appropriate as the basis for generalizations concerning the language as a whole (Kruyt & Dutilh, 1997: 230)

An interesting approach to the compilation of corpora, and one that fits the situation for the African languages like a glove, is the concept of *organic corpora*, introduced by Sue Atkins:

A corpus builder should first attempt to create a representative corpus. Then this corpus should be used and analysed and its strengths and weaknesses identified and reported. In the light of experience and feedback the corpus is enhanced by the addition or deletion of material and the circle repeated continually. This is the way to approach a balanced corpus. One should not try to make a comprehensive and watertight listing [...] rather, a corpus may be thought of as organic, and must be allowed to grow and live if it is to reflect a growing living language [...] In our ten years' experience of analysing corpus material for lexicographic purposes, we have found any corpus – however unbalanced – to be a source of information and indeed inspiration. *Knowing that your corpus is unbalanced is what counts.* (Atkins, oral communication at Salex'97, (Atkins et al 1997))

In following the organic approach to corpus building it is true that the corpus compiler runs the risk of creating a skewed corpus. It is often found that media publications such as texts from newspapers and journals mostly available in large quantities can skew a corpus. Compare in this regard MacLeod & Grishman (2000). In this presentation the authors illustrated how an increase in the Brown Corpus (which is generally regarded as well balanced) of 1,329% (thus more than thirteen times) resulted in a skewed or inadequate corpus:

... the make-up of the POS corpus, with its preponderance of newspaper text, skewed the choice of high-frequency verbs. This can be seen by comparing the frequency-ranked list from this corpus with that from Brown, a more balanced corpus. Among the top 50 verbs from our corpus, quite a few (business-related) verbs were not in the top 50 from Brown, including *sell, rise, buy, pay, and increase*. In fact, some were not even in the top 750 from Brown, such as *post, boost, invest, value, and resign*. (MacLeod & Grishman 2000: 142)

This urged Prinsloo and De Schryver (2001) to do random tests for Sesotho sa Leboa and Xitsonga in order to test corpus stability for African language corpora built according to the organic approach. They concluded that these corpora were indeed well balanced and also found that corpora of one million words or even less for the African languages are already big enough to be used for corpus based dictionaries for African languages. This issue will be discussed in more detail below.

### 3.2 Querying text corpora

Corpora in itself are of little use unless tools are available to manipulate the data in different ways. For lexicographic purposes such corpora are mainly queried to obtain/generate alphabetical word lists, frequency lists reflecting overall and/or

comparative counts or cotexts reflecting the actual use of a specific word in context. There are quite a number of software packages available to perform these tasks, like *Corpus Bench* from Denmark, *MonoConc* from the US, and *WordSmith Tools* from England. *WordSmith Tools* is widely used in South Africa and highly recommended for the day-to-day tasks of the lexicographer. Consider a number of examples reflecting such typical tasks performed by *WordSmith Tools* in Table 2.

**Table 2: Section of an alphabetical word list generated from the Sesotho sa Leboa corpus**

Word	Freq.	Word	Freq.	Word	Freq.
dijo	107	dikakanyo	14	dikantoro	2
dijong	3	dikamogakgopolo	5	dikantorong	1
dikadika	14	dikamogelo	1	dikaonafatšo	1
dikadikago	1	dikaneditšego	2	dikapolelo	1
dikagare	1	dikaneditšwe	2	dikarabo	11

**Table 3: A frequency word list reflecting overall counts for the 100 most frequently used words in Sesotho sa Leboa.**

Rank	Word	Freq.	Rank	Word	Freq.
1	a	330,123	51	moo	9,398
2	le	255,891	52	gago	9,300
3	go	248,408	53	bjale	9,230
4	ka	245,650	54	bolela	9,051
5	ba	205,161	55	tseba	8,927
6	o	169,268	56	mme	8,563
7	ke	166,454	57	dira	8,535
8	e	147,893	58	morena	8,419
9	ya	118,109	59	wena	8,374
10	re	91,129	60	monna	8,271
11	ge	85,800	61	taba	8,262
12	se	85,748	62	kua	8,256
13	wa	76,459	63	lego	8,210
14	gore	71,612	64	tšona	8,144
15	ga	66,079	65	fao	8,134
16	sa	60,845	66	mola	8,073
17	di	57,788	67	rena	7,987
18	mo	56,796	68	kgoši	7,843
19	be	55,346	69	bana	7,740
20	tša	47,707	70	mongwe	7,634
21	la	44,522	71	gwa	7,590
22	bona	34,300	72	leo	7,508
23	ye	31,809	73	bao	7,024
24	tla	27,076	74	ao	6,875
25	gona	26,618	75	bile	6,781

Rank	Word	Freq.
26	tše	25,978
27	gagwe	25,541
28	ile	24,094
29	yo	21,508
30	na	20,987
31	yena	19,339
32	yeo	17,478
33	fela	16,715
34	motho	16,353
35	gomme	16,095
36	goba	14,797
37	bjalo	14,615
38	tlo	14,187
39	batho	13,720
40	bjja	13,537
41	nna	12,179
42	yona	12,053
43	tše	11,741
44	pele	11,443
45	bego	11,280
46	wo	11,167
47	gobane	10,923
48	moka	10,910
49	bo	10,854
50	seo	10,811

Rank	Word	Freq.
76	ra	6,765
77	šetše	6,716
78	eng	6,709
79	tsebe	6,540
80	woo	6,367
81	tloga	6,236
82	kwa	6,235
83	no	6,226
84	mang	6,174
85	lena	6,154
86	gape	6,147
87	morago	6,145
88	swanetše	6,110
89	mokgwa	6,099
90	selo	6,025
91	ngwana	5,915
92	fihla	5,762
93	mosadi	5,749
94	banna	5,693
95	thoma	5,670
96	pelo	5,614
97	modimo	5,373
98	eupša	5,351
99	nako	5,252
100	tee	5,207

In Table 4 the total count of occurrences of some Setswana words in all the texts taken together are given in the second column and the spreading or distribution of each word across the different texts/sources are reflected in the next 5 columns. So, for example, *letlapa* occurred 31 times in the corpus as a whole, not in the first source but 16, 2, 7 and 6 times in the other sources respectively, adding up to the total of 31.

**Table 4: Setswana: Total counts and spreading across sources S1, S2, ... S5**

Word	Total	S1	S2	S3	S4	S5
letlapa	31		16	2	7	6
letsatsi	168	53	47	29		39
letshogo	16	3	3	2	4	4
mafura	25		15	3	3	4
mainakgopolo	10					10

Another most significant corpus query output is the generation of concordance lines, also called corpus lines, or actually, 'KWIC concordance' for 'keyword-in-context

concordance'. The occurrences of a word or phrase is extracted from the corpus and presented with cotext usually about 5-10 words preceding and following the keyword.

**Table 5: Concordance lines for isiZulu verbs occurring with the prefixal cluster wayesezo-**

Lachamusela isu likaMjike-Joe Umona usuka esweni <i>Mjike-Joe's plan hatched. Jealousy lies in the eye of the beholder</i>	wayesezofika <i>He would have arrived</i>	ekhaya Bambuyisela eGoli <i>at home but they let him go back to Johannesburg</i>
khona ePrince of Wales Training College. UJabulani <i>there at Prince of Wales Training College. Jabulani</i>	wayesezothola <i>would have received</i>	izincwadi zokufundisa ekupheleni <i>his study material at the end of</i>
Sathi sehlukana noDolly wayengitshela ukuthi <i>Just when we said goodbye to Dolly she told me that</i>	wayesezoqala <i>she now began</i>	ukumemezela ukuthi uphethwe yisisu <i>to proclaim that she was pregnant</i>
UDlaba akafundanga okutheni, wayeka phakathi <i>He did not learn much and gave up in the middle</i>	wayesezosebenza <i>He would by now have worked</i>	kwaVukusebenze. Ufike exova udaga <i>at Vukusebenze. He then started mixing mortar</i>
nje ukuthi okwakuyikhona kumphethe kabi yikuthi <i>in this manner, that which existed made him bad, it is because</i>	wayesezolahlekelwa <i>he would have lost</i>	ngabantu labo ababeza kuye <i>those people who had come to him</i>
umuntu wayephumelele yini eLuhloleni njengoba <i>someone was successful or not in the adjudication since</i>	wayesezoqala <i>he would have begun</i>	nje uNhlolanja. Ngazo lezozinsuku ng <i>in January. In those specific days</i>

By looking at these concordance lines running from top to bottom on the computer screen with the keyword in the centre, the lexicographer can make a number of valuable conclusions regarding the different senses of the words, typical examples of usage, typical collocations of the word, etc. See De Schryver and Prinsloo (2000b) for a detailed discussion.

The corpus query program also gives vital statistical information on the corpus, e.g. the size (number of tokens), types (different words), average word length, sentence length, etc.

N	1	2	3	4	5
Text File	OVERALL	ZWIRENDO.TXT	IRENDO.DOC	ZWIKOL~1.TXT	ZWIKOL~1.DOC
Bytes	6,845,956	90,255	330,752	3,487	27,648
Tokens	396,980	16,340	10,824	664	2
Types	15,679	2,564	1,951	269	2
Type/Token Ratio	3.95	15.69	18.02	40.51	100.00
Standardised Type/Token	37.74	36.37	35.45		
Ave. Word Length	4.33	4.26	4.27	4.20	1.50
Sentences	16,496	937	575	1	0
Sent. length	21.84	15.93	18.29	440.00	
sd. Sent. Length	64.62	14.33	37.99		
Paragraphs	1,128	98	0	2	0
Para. length	195.43	166.68		332.00	
sd. Para. length	781.68	312.21		465.28	
Headings	0	0	0	0	0
Heading length					
sd. Heading length					
1-letter words	39,762	1,523	986	78	1
2-letter words	67,569	2,767	1,892	101	1
3-letter words	82,542	3,709	2,426	149	0
4-letter words	48,059	2,156	1,398	84	0
5-letter words	44,737	1,901	1,227	76	0
6-letter words	35,648	1,345	917	59	0
7-letter words	27,208	811	562	43	0

Figure 1: Statistical analysis of a Tshivenda corpus in WordSmith Tools.

It is important to note that corpus query output should not be restricted to the eyes of the lexicographer, i.e. utilised only ‘behind the scenes’ – selected information should find their way into the dictionary. Frequency counts, for example, should be indicated in the dictionary itself, preferable not as raw data or actual numbers but stratified e.g. on the basis of high, medium, low, etc. Consider the strategies employed by COBUILD and LDOCE as excellent examples in this regard. In COBUILD a system of ‘frequency bands’ is used. Five filled diamonds indicate that the lemma sign occurs within the 700 most frequently used words in English, four that it occurs within the first 1,900, three that it occurs within the first 3,400, two that it occurs within the first 6,600, and one that it occurs within the first 14,700. Consider an example from each category:

COBUILD 2	
walk	◆◆◆◆◆
length	◆◆◆◆◇
meaning	◆◆◆◇◇
elephant	◆◆◆◇◇
opium	◆◆◆◇◇

In LDOCE frequencies are indicated in a similar way with additional distinction between spoken and written communication.

leg	S1W1
legal	S3W1
leisure	W3

From this example it is clear that *leg* is used with very high frequency in both written and spoken English, *legal* is very frequent in written sources but comparatively less frequent in spoken English while *leisure* only obtained a frequency rating in written sources.

In dictionary compilation the data provided by electronic corpora shown in Tables 2 to 5 assist the lexicographer in several ways on both the macrostructural and microstructural levels.

### 3.3 The use of corpora on macrostructural level

On the macrostructural level word-frequency counts is an extremely useful tool in the compilation of a lemmalist for a new dictionary. Publishers are normally very specific and prescriptive regarding the number of lemmas to be treated. If the required number of lemmas is say, 10,000, the lexicographer could for instance isolate the top 10,000 types from a frequency list, lemmatise it and supplement the lemmalist with lower frequency items from the frequency list until the desired number of lemmas has been met. A major advantage of such an approach is that on the one hand frequently used words will not accidentally be omitted and on the other hand that precious dictionary space will not be taken up by lemmas less likely to be consulted by the target user. See De Schryver and Prinsloo (2000a) for a detailed discussion.

The significance of frequency as an important criterion is sometimes debated but the following statistics for English and Sesotho sa Leboa for example underlines the significance of frequency in the selection of lemmata. The analysis of log files reflecting the actual lookups by dictionary users, (De Schryver and Joffe 2004), strongly support the assumption that frequently used words are in principle the most likely to be looked up.

If one compares the top 100 Sesotho sa Leboa searches with the ranks of the corresponding items in a frequency list derived from a 6.1-million-word Sesotho sa Leboa corpus, then one notices that 30 of the top 100 searches can also be found in the corpus top 100, while as many as 63 can be found in the corpus top 1 000. Clearly, users indeed look up the frequent words of the language...

An analogous study of the top 100 English searches reveals a similar pattern... (De Schryver and Joffe 2004:190)

**Table 6: Summary of frequency band values in Cobuild**

Number of filled diamonds	Lemmas per category	Totals	% of all written and spoken English
5	700		
4	1200		
(Total 5+4)		1900	75
3	1500		
2	3200		
1	8100		
(Total 3+2+1)		12800	20
(Total 5+4+3+2+1)		14700	95

From Table 6 it is clear that the top 1,900 lemmas represent 75% of English (tokens) and the top 14,700 an astonishing 95%.

For Sesotho sa Leboa the top 1,000 types represent 77.5% of the tokens and the top 10,000 types 91.7%:

**Table 7: Types versus tokens in Sesotho sa Leboa**

Types (Number of different words)	Total frequencies (Sum of all counts)	Tokens (Total number of words in the corpus)	% of tokens
Top 1,000	4,615,053	5,957,553	77.5
Top 10,000	5,462,500	5,957,553	91.7

For the revision of existing dictionaries, frequency lists can play a vital role in ascertaining that frequently used words were not accidentally omitted and on the other hand that dictionary space is not occupied by articles of lemmas unlikely to be looked for by the target users.

Consider Table 8 as an example of obvious gaps in the lemmalist of a dictionary compiled without a corpus. Frequently used words such as *dad*, *daily*, *dairy*, etc. are not included simply because they did not come to mind in the compilation of these dictionaries.

**Table 8: Comparison of dictionaries reflecting gaps in the macrostructure**

DS	SED	POP	KW
<b>dab</b>		<b>dab</b>	<b>dab</b>
	dabble	dabble	dabble
	dad	dad	dad
			daffodil
	daft		daft
dagga	dagga	dagga	
dagga-pipe	dagga-pipe		
	dagger	dagger	dagger
		dahlia	dahlia
	daily	daily	daily
	dainty	dainty	dainty
	dainties		
	dairy	dairy	dairy
	dais		dias
	daisy	daisy	daisy
	dale	dale	dale
	dally		
<b>dam</b>	<b>dam</b>	<b>dam</b>	<b>dam</b>

Compare also a second example in this regard for an African language. In Table 9 the same type of inconsistency namely obvious gaps in the lemmalist is reflected but this time in terms of inconsistency in respect of verbal derivational forms. Verbs given in boldface and in capital letters, in most cases occurring frequently in the corpus, were omitted from the lemmalist of a Sesotho sa Leboa dictionary.

Table 9: Obvious omissions of frequently used verbal derivations

root → ↓ derivation	bolela (5,735)	dira (5,475)	hwetša (3,371)	rata (2,786)	reka (551)	tseba (5,851)
+ applicative	bolelela (76)	direla (508)	— (0)	ratela (11)	rekela (88)	tsebela (47)
+ passive	bolelwa (408)	dirwa (636)	hwetšwa (260)	ratiwa (5), ratwa (126)	rekwa (122)	tsebja (441)
+ applicative & passive	bolelelwa (6)	direlwa (40)	— (0)	— (0)	rekelwa (19)	— (0)
+ perfectum	boletše (767)	dirile (910)	hweditše (671)	ratile (151)	rekile (90)	tsebile (234)
+ perfectum & passive	boletšwe (44)	dirilwe (137)	hweditšwe (57)	ratilwe (13)	rekilwe (17)	tsebilwe (10)
+ causative	bolediša (72)	diriša (200)	— (0)	— (0)	rekiša (223)	tsebiša (376)
+ causative & passive	boledišwa (45)	dirišwa (72)	— (0)	— (0)	rekišwa (27)	tsebišwa (63)

Frequency counts obtained from the corpus thus assist the lexicographer in solving one of the basic problems in dictionary compilation, namely what to include and what to exclude from the dictionary. Ideally the corpus lexicographer should be able to motivate the inclusion or omission of each and every lemma in the dictionary. Such motivation becomes quite relevant especially when lemmalists have to be compiled for very specific or narrowly defined target user groups when the number of lemmas are severely restricted. Say, for example, a lemmalist restricted to 3,000 lemmas has to be compiled for a dictionary for primary school children mainly to be used for reception and production purposes in respect of their prescribed text books. The lexicographer has to find a sound balance in terms of the selection of lemmata between words likely to be looked up by the target users from their prescribed work and from the general language.

A sound strategy proved to be to compile a so-called dedicated corpus for the prescribed material and then to compare frequency counts from this dedicated corpus with frequency counts from the general corpus of the language in order to select a lemmalist. De Schryver and Prinsloo (2003) in preparation of a suggested lemmalist for the compilation of *Nuwe woordeboek sonder grense* (NWSG) selected all words occurring 9 times or more in the dedicated corpus and those occurring 100 times or more in the general corpus. This means that even words with zero occurrence in the general corpus were considered for inclusion in the lemmalist on the basis of relative frequent occurrence in the dedicated corpus. This strategy has since been applied for a few other dictionary projects with similar target user groups. Once again corpus query software plays an important role. The *Keyness* function in *WordSmith Tools*, for example, is ideal for selecting so-called key words by comparing, say, a dedicated corpus with a general corpus. Compare for example, a dedicated corpus of English prescribed textbooks for junior learners against a general English corpus. All the words in Table 10 especially *learner(s)*, *assess(ment)*, and *outcomes* occur much more frequently than expected in the dedicated corpus compared to the general corpus, and will be exhaustively treated.

Table 10: Positive keys in a comparison, dedicated versus general corpus, calculated with WordSmith Tools

WORD	FREQUENCY Dedicated corpus	FREQUENCY General corpus	KEYNESS
LEARNERS	10,722	4	46,363.0
ACTIVITY	6,461	375	25,150.3
LEARNER	5,289	6	22,797.2
ASSESSMENT	2,580	30	10,841.3
ANSWERS	2,721	295	9,912.1
WRITE	3,190	1,455	8,381.4
HOW	7,123	12,403	8,230.9
GROUP	2,736	810	8,223.4
SCIENCES	2,064	147	7,883.6
QUESTIONS	2,468	1,002	6,750.3
ASSESS	1,504	8	6,407.9
DISCUSS	1,602	154	5,923.6
OUTCOMES	1,340	0	5,796.2

### 3.4 The use of corpora on microstructural level

On the *microstructural* level, concordance lines generated from corpora by means of corpus query tools supplement the lexicographer's (native-speaker) intuition. They take him/her to the *heart of actual language usage* by displaying numerous occurrences of the word(s) or phrase(s) in context, allowing the lexicographer to see up to several dozens of cotexts at a time. Compare the following example of how the Sesotho sa Leboa lexicographer could benefit from studying a few corpus lines for the word *bala* 'read, count'

Table 11: Corpus lines for *bala*

šomela pele, o ka ba wa re o kgona go	<b>bala</b>	dikgopolo tša batho ka moka bao ba
, a fihlola ka mae a go gadikwa goba a	<b>bala</b>	dikuranta. Ka morago ga ditumedišo le
ntlhoriša bja go wa. Ruri ge nka re ke	<b>bala</b>	ditšhelete tšeo o senyegetšwego ka
go boela go ba bangwe ba kgale, mme ba	<b>bala</b>	ka menwana ye e atilego. Eupša bjalo
Seroboka - mang le mang mo motseng, ke	<b>bala</b>	le batho ba bagolo ba be ba mmofa. Ge
a ehwa. O ile a re yena o no hlwa a	<b>bala</b>	mo dikuranteng gore mošemane wa
theeditše ka tsebe tše pedi o ka re o	<b>bala</b>	mogopolo wa mokgalabje. Papagwe o be a
e mo tseba goba e mmone. O ile ge a	<b>bala</b>	mphato wa bohloano mmagwe a hloka-fala,
tla hwetša tema yeo o swanetšego go e	<b>bala</b>	pele o etla ka phapošing nako ye e
ka fao ke be ke sa kgone go	<b>bala</b>	sefahlego sa gagwe. "Ga ke tsebe gore

Such corpus lines assist the lexicographer in respect of sense distinction, deciding on translation equivalents, retrieval of typical collocations, pinpointing frequent clusters and in the selection of representative, authentic examples to be included in the dictionary. Let's take a brief look at some of these issues. Without a corpus the lexicographer is always in doubt whether (s)he has covered all the relevant senses of a lemma sign in the definition or in setting up a translation equivalent paradigm.

However, by studying a selection or ‘screenful’ of corpus lines such as these few given for the word **bala** ‘read/count’ the major different senses are easy to detect:

‘read’

... a *fihlola ka mae a go gadikwa goba a bala dikuranta.*  
 ‘... (s)he eats eggs for breakfast or reads newspapers.’

‘count’

*Ruri ge nka re ke bala ditšhelete tšeo o senyegetšwego ka tšona ...*  
 ‘Really if I could count that money you messed with ...’

‘include’

... *mang le mang mo motseng, ke bala le batho ba bagolo ba be ba mmoifa.*  
 ‘... everyone in the village including the adults feared him/her.’

‘do/complete an academic grade/standard’

*O ile ge a bala mphato wa bohloano mmagwe a hlokafala,*  
 ‘When (s)he was doing Standard 5 his/her mother died.’

‘read someone’s thoughts/mind’

... a *theeditše ka tsebe tše pedi o ka re o bala mogopolo wa mokgalabje.*  
 ‘... (s)he listened carefully as if reading the old man’s mind.’

‘read/see the expression of someone’s face’

*O be a mphuraletše ka fao ke be ke sa kgone go bala sefahlego sa gagwe.*  
 ‘(s)he turned away from me and therefore I could not see his/her face.’

The chances of a dictionary compiler gathering all senses and sub-senses of highly polysemous words on the basis of intuition is questionable. Taking corpus lines as one’s point of departure for sense distinctions in order to write definitions (monolingual dictionaries) or find translation equivalents (bilingual dictionaries), is the ideal way to start. The lexicographer should be cautious not to regard each corpus line as a different sense but rather learn to ‘see the senses emerge’ from a digestible number of corpus lines studied. It is normally also not the intention to study thousands of corpus lines for each lemma treated – a few hundred lines sorted in sensible ways, e.g. on the word preceding/following the lemma are usually sufficient.

In the case of examples of usage given in the dictionary, traditional author constructed or so-called ‘made-up’ examples give way to authentic examples taken from the corpus. Fox formulates it as follows:

We read the [corpus] lines to see whether there are any that are suitable for use as examples. (Fox 1987: 147)

The lexicographer should feel free to select or compile examples in whatever way is best in the treatment of a specific lemma. Prinsloo & Gouws (2000:138) give a detailed perspective on authentic (corpus) versus made-up or constructed examples and suggest a continuum of options given in Table 12 that are available to the lexicographer.

**Table 12: Authentic versus constructed examples**

Extreme	Intermediate categories			Extreme
Authentic (corpus examples) taken directly from a corpus without editorial modification	Slightly edited/modified corpus examples	Heavily edited/modified corpus examples	Partially invented, based on a corpus	Constructed examples

Corpora can furthermore assist the lexicographer in finding typical collocations and combinations of words. Consider the top ten clusters for the Sesotho sa Leboa verb *bolela* ‘speak’.

**Table 13: Clusters for *bolela***

Cluster	Freq.	Cluster	Freq.
a bolela a	322		
bolela ka ga	313	o bolela ka	229
ge a bolela	307	go bolela le	199
go bolela ka	299	be a bolela	191
bolela le bona	148	bolela le bona	148

WORD	TOTAL	LEFT	RIGHT	L3	L2	L1	*	R1	R2	R3
BOLELA	9210	66	93	33	25	8	51	36	20	37
A	5226	3424	1802	725	300	99	0	523	411	868
KA	3180	693	2487	306	178	209	0	44	393	350
GO	3037	2367	670	282	237	48	0	77	331	262
O	2646	1812	834	524	241	47	0	102	329	403
LE	2610	623	1987	311	186	126	0	34	228	325
BA	1865	1260	605	387	224	649	0	87	155	363
KE	1622	931	691	301	246	384	0	116	265	310
GE	1416	1023	393	195	828	0	0	76	151	166
RE	1253	794	459	238	227	329	0	38	221	200
GORE	926	232	694	102	129	1	0	520	95	79
E	781	396	385	118	29	249	0	65	128	192
SE	696	406	290	162	52	192	0	62	75	153
GA	689	127	562	110	17	0	0	34	400	128
SA	582	453	129	150	59	244	0	3	58	68
WA	546	336	210	160	75	101	0	27	45	138
BE	510	363	147	44	318	1	0	1	41	105
YA	478	237	241	123	61	53	0	9	76	156
DI	433	335	98	53	13	269	0	13	30	55
BJALO	394	38	356	22	14	2	0	314	24	18

**Figure 2: Collocates of *bolela* generated by WordSmith Tools**

If one instructs the corpus query tool to calculate and list the collocates of the verb *bolela* certain useful conclusions can be drawn such as the frequent use of *bolela gore* ‘say that’ and *bolela bjalo* ‘say so’ in Figure 2.

It is often said that a corpus can only be useful in the study of idioms if it is very large. Prinsloo and De Schryver (2001) however indicate that even relatively small corpora can be successfully used in studying idioms for lexicographic purposes.

Consider Table 14 where respectively a small section of the corpus, two unrelated halves, and finally the entire corpus were queried in respect of the Sesotho sa Leboa idiom *monna ke nku, o llela teng* 'A man is like a sheep, he does not show his feelings (he cries inside)'. The number of occurrences roughly correlates with the size of the sub-section in relation to the full corpus, as indicated by '✓'.

**Table 14: The idiom *monna ke nku, o llela teng* in PSC and sub-sections thereof**

			Phase 1	Half 1	Half 2	Sum
mo dutšego mogolong. O no re	monna ke nku, o llela teng	. A di rumilego bjalo, le tsona tša	✓	✓		✓
! Monna ga se a swanela go lla!	monna ke nku o llela teng	. Ge o ka bona monna a ediša dik	✓	✓		✓
ape go sekhumola. "Tiiša thaka,	monna ke nku o llela teng	." "... ba boletše ... ba boletše, mot	✓	✓		✓
sa mmone. Ee! Baswana ba re	monna ke nku, o llela teng	, fela ge e le Thogorogo yena o il		✓		✓
ba iteile lešepa ka thoka ge ba re	monna ke nku, o llela teng	, gomme a bona ba opile kgomo l		✓		✓
a. Fela ka gore bagologolo ba re	monna ke nku, o llela teng	, o ile a no ikgata pelo a tšwela pe		✓		✓
na ba be ba no šita kgang, ba re	monna ke nku o llela teng	. Ba be ba fetogile difahlegong, b		✓		✓
jo bo rego ke metlae ge go thwe	monna ke nku o llela teng	. Ba ile go felela ka mola mphom		✓		✓
ta. Ee, ke therešo. Sesotho se re	monna ke nku, o llela teng	. Fela le ge se realo, leo morwa' H		✓		✓
megokgo, motho a lebetše gore	monna ke nku o llela teng	. Ga se thaka ya mošemane go go		✓		✓
olo o ile a mo homotša ka go re:	monna ke nku, o llela teng	. Ke ge a be a lemogile gore ga se		✓		✓
ela gore ga se nnete ge go thwe	monna ke nku o llela teng	. Le go llela teng ga nnete go tleg		✓		✓
ela gore ga se nnete ge go thwe	monna ke nku o llela teng	. O ile a bokolela ka pelobohloko		✓		✓
ela bjang, goba ke gona ge ba re	monna ke nku, o llela teng	? Gape taba ke ngwana wa rena w		✓		✓
ore a tle a imologe. Bao ba rego	monna ke nku,	o hwa natšo goba mosadi o fogoh		✓		✓
2l tee fela (bjalo ka seema se,	monna ke nku, o llela teng	, monna le nku di llela teng			✓	✓
šha phefo ganong a lebetše gore	monna ke nku o llela teng	. Ka yeo nako ke ge madira ale a			✓	✓
a ngwana a itiiwe, a lebala gore	monna ke nku o llela teng	. O be a sa itiriše ka gore le go m			✓	✓
tše go di dula ka marago ka gore	monna ke nku.	Aretse, ee, monna ke nku. Mošate			✓	✓
mola lapeng a lebetše gore ba re	monna ke nku.	Basadi bale ba bego ba le moo le			✓	✓

Table 14 (continued)			Phase 1	Half 1	Half 2	Sum
go pitikana fase a lla le ge ba re	monna ke nku	dihlong tša se mo je. O ratharathi			✓	✓
er Verszeile". Šomiša diema tše,	monna ke nku, o llela teng	le Phaga ga e ete, go eta nakedi,			✓	✓
re ke a meletša a kwele ge ba re	monna ke nku,	mafelelong a ba a ntšha phefo gan			✓	✓
oro a šetše molotong, GaMatlala	monna ke nku	o latswa bohloko. E rile e tsena k			✓	✓

Prinsloo and De Schryver (2001:110)

Corpus creation is often followed by *text encoding*, i.e. the raw text is supplemented by a series of so-called standard corpus pre-processing annotations. These processes can consist of any combination of the following: a) word tokenisation, b) part-of-speech tagging, c) lemmatisation, d) syntactic parsing and e) markup and are described in more detail in De Schryver and Prinsloo (2000).

## The user-perspective

---

The user-perspective, so prevalent in modern-day metalexigraphy, compels lexicographers to compile their dictionaries according to the needs and research skills of well-defined target user groups. The dominant role of the user has had a definite effect on the compilation of dictionaries as well as on the evaluation of their quality. Good dictionaries do not only display a linguistically sound treatment of a specific selection of lexical items. Good dictionaries are products that can be used as linguistic instruments by their respective target user groups. The better they can be used, the better dictionaries they are, cf. Haas (1962), Barnhart (1962), Otto (1989) and Prinsloo and Gouws (1996).

A good dictionary is one in which you can find the information you are looking for – preferably in the very first place you look. (Haas 1962:48)

... it is the function of a popular dictionary to answer the questions that the user of the dictionary asks, and dictionaries on the commercial market will be successful in proportion to the extent to which they answer these questions of the buyer. (Barnhart 1962: 161)

The quality of dictionary use, that is the degree of success a user experiences when consulting a dictionary and employing the retrieved information, is determined by a variety of features but one of the most important characteristics of a good dictionary is its accessibility which leads to an unambiguous retrieval of the information presented on both the macro- and microstructural levels. Any theory of lexicography should present strategies to enhance the linguistic quality of dictionaries. However, this should be preceded by strategies to enhance the way in which the target user can identify the data (s)he is looking for in order to retrieve the necessary information and to utilise it in a productive (encoding) and receptive (decoding) way.

A dictionary should not primarily reflect the attitude of the lexicographer but it should rather be aimed at specific needs of a well-defined target user. Modern dictionaries are judged by the success with which the *user* is able to retrieve the sought information and not in the first instance by 'linguistic achievement.'

An analysis of users' needs should precede dictionary design.  
(Hartmann 1989:103)

The lexicographer is in terms of Prinsloo and Gouws (1996: 103) the mediator between linguistics and the everyday dictionary user. The modern trend in lexicography is thus to compile very practical and extremely user-friendly dictionaries. In terms of Barnhart (1962) this means that the lexicographer has to include those words which are most likely to be consulted by the target user and to lemmatise them in a user-friendly way. User-friendliness does not only relate to the central text but also to the front and back matter of the dictionary, especially the user's guide. One could say that

*practicality* and *user-friendliness* virtually determine the methodology and the way in which the access structures of the dictionary are to be designed.

Accessibility and unambiguous retrieval of the information presented on both the macro- and microstructural levels are goals that are not always easily reached by the compilers of dictionaries for especially the African languages. This issue will also be addressed in Chapter 7 where problems regarding lemmatisation are discussed. Consider for now the following example in terms of accessibility and unambiguous retrieval of the information from the perspective of an inexperienced learner of Sesotho sa Leboa. The user wants to look up the word *dikagollišano*. (S)he firstly has to strip the suffixes in order to find the verb stem and then to 'add' the semantic connotations in a cumulative way in order to find the meaning – thus up to 12 steps in total given in Table 1.

**Table 1: Accessibility and information retrieval process for *dikagollišano* in NSDN**

1	dikagollišano	↓	plural deverbative consisting of root + reversive transitive + causative + reciprocal + ending
2	kagollišano	↓	singular deverbative consisting of root + reversive transitive + causative + reciprocal + ending
3	agollišana	↓	verb root + reversive transitive + causative + reciprocal + ending
4	agolliša	↓	verb root + reversive transitive + causative + ending
5	agolla	↓	verb root + reversive transitive + ending
6	<b>aga</b>	↓	<b>verb</b> (stem)
7	build	↓	meaning of the verb
8	break down	↓	reverse or opposite meaning 'un-build'
9	cause to break down	↓	add causative sense of 'let/force'
10	cause each other to break down	↓	add reciprocal sense of 'each other'
11	the process of causing each other to break down	↓	nominalise: 'the process of ...' (singular)
12	the processes of causing each other to break down		change 'the process of ...' to the plural

In step 12 the user concludes that *dikagollišano* means '*the processes of causing each other to break down*' – but it is an artificially constructed meaning and (s)he is still not sure that it is the right conclusion.

The data should be presented in such a way that the user does not make the wrong conclusions. Consider the following examples from isiZulu and Sesotho sa Leboa respectively.

SZD

a. **-nkosi** (i- ama-) (n) king; chief

WAZ

b. **-khosi**, (in-, ama-), b; 1. koning, regent, hoofman

c. **mosegare** dag, middag

For the lemma *-nkosi* the inexperienced user could conclude that the singular for *king* in isiZulu is *inkosi* and the plural is *\*amankosi* or if (s)he has some knowledge of the noun class system, *\*amakosi*. Both conclusions for the plural form are incorrect, it is *amakhosi*. The user's conclusion for the lemma *-khosi* is: *\*inkhosi* 'king', *amakhosi* 'kings', of which the plural form is correct but the singular form is *incorrect* - it must be *inkosi*. This is a very serious mistake, since the dictionary should never guide the user to such incorrect conclusions. These are examples of bad lexicography, violating almost every aim regarding the user-perspective mentioned above. As for the article of *mosegare*, at least justice was done when the lexicographer himself was greeted by one of his students by *\*Mosegare Profesa!* 'Midday Professor!' (instead of *Dumela Profesa!* 'Good day Professor!'). When asked, "which dictionary did you use?" the answer was, "yours".

The dictionary aimed at production should indeed give sufficient guidance so that the users can produce correct text and speech. Once again, for the African languages it could be problematic especially in case of highly complicated structures such as the copulative in Sesotho sa Leboa. Consider the following two articles for the English lemma *is*, aimed at productive and receptive dictionaries respectively.

a. *is* ke [id. cop.], **ke lengwalo, ga se sephuthana** it is a letter, it is not a parcel; **bohodu ke sebe.** theft is a sin. o/e/le...[des. cop.], **aowa, mosadi yo o bohlale, ga a bogale!** No, this woman is clever, she's not cruel!; o/e/le na le [ass. cop.] **Satsope o na le Sara.** Satsope is with Sara. □ **BM: Copulatives**

b. *is* ke, **ke lengwalo** it is a letter; o/e/le..., **o bohlale** she is clever; o/e/le na le **Satsope o na le Sara.** Satsope is with Sara.

For mother-tongue speakers of Sesotho sa Leboa learning English, the b-example provides all the relevant information indicating implicitly by means of examples that all three types of copulatives are expressed by a single strategy in English. For learners of Sesotho sa Leboa, however, the fact that *is* is expressed by three different strategies is explicitly indicated and examples of each type are given with a cross-reference to the back matter where more information especially regarding the semantic relations determining the different types are explained.

Enhancing the quality of dictionaries *en route* to the 'perfect dictionary' should be the mission of the lexicographer. However, even the 'perfect dictionary' in the hands of an unskilled dictionary user has little gain. The ideal is to improve the dictionary skills of the target user with the mission to become an 'ideal user'. Atkins and Varantola formulate it as follows:

There are two direct routes to more effective dictionary use: the first is to radically improve the dictionary: the second is to radically improve the users. (Atkins and Varantola 1998:83)

There is a general belief amongst those concerned with dictionaries that dictionary users do not get the best out of their dictionaries, and, conversely, that dictionaries themselves could be improved so as to serve their users better. (Atkins 1998:1)

However careful the lexicographer, and skilled the user, by its very nature of having to be all things to all users, the print dictionary usually offers either too little or too much information. (Hulstijn and Atkins 1998:11)

The greater the distance between the user's skills on the one hand, and the quality of the dictionary on the other hand, the less effective the dictionary use will be. This is schematically suggested by the following continuum with extreme end points 'bad/useless dictionary or no dictionaries available' as one extreme and 'pre-dictionary culture environment' as the other extreme with stages of relatively low/high in terms of lexicographic achievement and using skills as intermediate stages. The ideal, viewed from both directions and symbolised by the arrows, is aiming at the *perfect dictionary*, consulted by the *ideal user*, in the centre.

**Table 2: Towards the perfect dictionary and the ideal user**

Dictionaries				Users		
▶▶▶▶	▶▶▶▶	▶▶▶▶	▼▼▼	◀◀◀◀	◀◀◀◀	◀◀◀◀
Bad/ useless dictionary or no dictionaries available	Dictionary of relatively low lexicographic achievement	Dictionary of relatively high lexicographic achievement	▼▼▼ <b>Perfect dictionary &amp; Ideal user</b> ▲▲▲	Relatively good dictionary using skills	Relatively poor dictionary using skills	Pre-dictionary culture environment
▶▶▶▶	▶▶▶▶	▶▶▶▶	▲▲▲	◀◀◀◀	◀◀◀◀	◀◀◀◀

This oversimplified schematic illustration is quite representative of the South African situation. For some languages such as English and Afrikaans, the quality of dictionaries is well advanced on the scale from left to right. Likewise, on the right hand of the scale, the average users possess relatively good dictionary skills. For some of the African languages, however, few dictionaries of relatively high lexicographic achievement exist. As for the right hand of the scale, a major problem is that the majority of South Africans find themselves in a pre-dictionary culture environment. Atkins (1998a: 3), after having studied the SA situation remarks as follows:

... the speakers of African languages have not in their formative years had access to dictionaries of the richness and complexity of those currently available for European languages. They have not had the chance to internalize the structure and objectives of a good dictionary, monolingual, bilingual or trilingual.

The task of the newly established NLUs and future compilers of dictionaries for the African languages is immense. On the left of the proposed continuum they have the daunting task of compiling dictionaries of lexicographic achievement. This would include among others intensive study of users' needs. To the right of the continuum they have an obligation or co-responsibility firstly to establish a dictionary culture in instances where it does not exist and secondly to advance the target users in terms of dictionary skills. As a first step a series of dictionary awareness campaigns/initiatives should be embarked upon. Hartmann (1989:102) refers to the "mutual

and often complementary relationship between the compiler and the user". He also points out that different user groups have different needs.

... we must question whether a single type of dictionary can satisfy all the possible reference needs. (Hartmann: 1989: 104)

Users needs are determined by various factors – of these the purpose of the activity is the most important. (Hartmann: 1989: 103)

Hartmann (1992:103) also emphasises the need to teach dictionary skills. Especially in the case of SA where it is presumed that most dictionary users lack a culture of dictionary use, the teaching of such skills especially at primary, secondary and tertiary levels is of vital importance. Well orchestrated initiatives involving education authorities, provincial language councils and NLUs are required to achieve this. These initiatives should not be limited to pupils and students but should be aimed at the broader community as well in the form of community service projects.

Assessment of users' needs, the teaching of dictionary skills and feedback on dictionary use should be done at three different stages. Firstly it should be part of the comprehensive lexicographic process preceding the compilation of a dictionary. These activities should incorporate all the vital aspects discussed in Chapter 2. Secondly it should be continued while the dictionary compilation is still in progress (cf. De Schryver's concept of Simultaneous Feedback (SF) in sources such as De Schryver 1999 and De Schryver and Prinsloo 2000c) and finally done after completion of the dictionary.

Prinsloo and De Schryver (2000) demonstrate the process of Simultaneous Feedback in three feedback phases using two distinct target user groups, beginners and learners on the one hand, and second language and mother-tongue speakers on the other. In a first phase these target user groups were taught how to use the dictionary and were asked to solve exercises interactively. Here especially informal feedback was gathered. In a second phase the two target user groups were instructed to solve exercises using the dictionary, but without guidance from the lexicographers. This type of feedback was already of a more formal type. Finally, in a third phase a questionnaire was distributed to all the users of the dictionary, resulting in very formal feedback.

Table 3 is an example of asking direct questions where the users have the opportunity to express opinions, needs and preferences and Table 4 an example of questions, answers and calculated results determining the knowledge of the users in respect of dictionary conventions.

Table 3: Specific questions on the layout of the dictionary

Layout	
3. What do you think of the layout of the dictionary, that is the entries from A to Z?	<input type="checkbox"/> it is good <input type="checkbox"/> it can be better Do you know a better way? _____ <input type="checkbox"/> it is bad <input type="checkbox"/> other: _____

Prinsloo and De Schryver (2000)

Table 4: Evaluating knowledge of typical dictionary conventions

SeDiPro 1.0 Questionnaire		Results (in %)	
Question	Answer	Target group	
		Learner	Speaker
13. In many dictionaries the headword is replaced by a tilde (~) within an article. This is also done in the dictionary you used. In the dictionary you will find: <b>ntoma</b> bite me; ~ <b>tsêbê</b> tell me a secret a. Which word does the tilde (~) replace here? b. How do you say 'tell me a secret' in Sepêdi?	a. correct word <input type="checkbox"/> wrong or no word <input type="checkbox"/>	86 14	— 100
	b. correct expression <input type="checkbox"/> wrong or no expression <input type="checkbox"/>	71 29	20 80
30. When you see something like this: <b>feela = fêla</b> a. Do you know what you should do? b. Do you know why this was done like this?	a. correct suggestion <input type="checkbox"/> wrong or no expression <input type="checkbox"/>	43 57	10 90
	b. correct suggestion <input type="checkbox"/> wrong or no suggestion <input type="checkbox"/>	43 57	— 100

Prinsloo and De Schryver (2000)

From the results calculated in percentages it is clear in respect of Question 13 that most beginners/learners know how to interpret the tilde '~' as a place-holder symbol but that none of the second language speakers and mother-tongue speakers knew what the function was. This could be vital information for the compilation of dictionaries for this and similar target user groups to rather repeat the full form of the word rather than using a dictionary convention that will impede successful information retrieval for the sake of saving space. Also significant is both target user groups' inability to correctly interpret an implicit cross-reference. This should also be a clear signal to the compilers of dictionaries for these groups to give preference to a more explicit system of cross-referencing, as will be dealt with in Chapter 12.

## Dictionary typology

---

### 5.1 Making realistic choices

An important issue in any lexicographic process is the decision regarding the typological nature of the dictionary to be compiled. Although the compilation of a multivolume comprehensive monolingual dictionary, displaying a comprehensive selection of lexical items to be included as treatment units along with a comprehensive microstructural treatment which allows the inclusion of a wide range of data types, should be the eventual aim and the ultimate goal of all the National Lexicography Units in South Africa, it is not practical to start with such a project. A comprehensive dictionary is a multivolume and multi-decade project. To illustrate this: the WNT was completed after about 148 years, and the work on the WAT has been going on since 1926 with the already published volumes only covering the letters A-Q.

When deciding on the typology of a new dictionary the needs of the target users and the reference situation within a given speech community should play a decisive role. The typical needs of the members of the South African speech communities demand the speedy availability of dictionaries. Consequently the idea and the ideal of a comprehensive dictionary should not dominate the immediate planning of new dictionary projects. It is more important for the users of a speech community to have access to a smaller dictionary than no access to any dictionary at all and only the promise that a comprehensive dictionary is on its way. A small dictionary in the hand is preferred to a comprehensive dictionary in the planning. However, the NLUs should gradually work towards the goal of a comprehensive dictionary by supplying their stakeholders with other dictionaries of a less comprehensive type in order to achieve the necessary communicative empowerment of the speech community and to ensure that the target users do not get frustrated due to a too long period of waiting for a dictionary.

Part of the secondary comprehensive lexicographic process to be performed in each NLU is the drawing up of a typological hierarchy indicating the short, medium and long term priorities of the unit. The first and subsequent dictionaries to be compiled by a unit may not be viewed and evaluated in isolation but should be seen as part of an ongoing endeavour which has the comprehensive monolingual dictionary as an ultimate goal. The first dictionaries to be compiled within the different NLUs may even have to be regarded as limited but functional efforts which can form a basis for subsequent projects.

According to Hausmann (1989:13) the history of lexicography indicates that a lack of harmony can easily exist between lexicography and the general public. Successful dictionary consultation by the intended target user is easily impeded by the sophisticated nature of many modern-day dictionaries. Hausmann (1989:13) refers to this as a conflict between a *dictionary culture* and *user-friendliness* in lexicography. According to Hausmann user-friendliness refers to lexicography adapting to society whereas dictionary culture refers to society adapting to lexicography. The South

African society at large still lacks a dictionary culture and although attempts are being made to promote the introduction of courses focusing on dictionary using skills in schools and tertiary education institutions the average member of the different South African speech communities is still quite naïve when it comes to dictionary use and the choice of specific dictionaries for specific purposes. Consequently the NLUs have to adapt an approach characterised by user-friendliness in the typological choice and eventual compilation of their dictionaries. Failure to do so could easily result in the production of dictionaries that are inaccessible to the target user or dictionaries not directed at the real needs of the target user due to a wrong typological choice on the side of the NLU.

Lexicographers and dictionary users have to realise that no single dictionary can be everything to every person. In lexicography the slogan “horses for courses” is just as applicable as elsewhere. To help to ensure the success of an eventual dictionary as a practical and functional instrument in the hand of a well-identified target user, any new lexicographic process has to negotiate the relation between a dictionary culture and user-friendliness in the speech community of the intended target users. The target users of a general language dictionary are not academics and students but the average members of the speech community who can be empowered by the access to a dictionary. This should be carefully considered when deciding on a specific dictionary type or when debating the functions, structure and contents of a dictionary and the development of a dictionary specific lexicographic process. It is also important to bear in mind that the development of a dictionary culture in a given speech community can run parallel to the development of the dictionary collection. Where the lexicographers in a specific NLU are aware of problems regarding the dictionary culture or dictionary using skills of their intended target users, the typological choice of the dictionary to be compiled has to be determined by these circumstances. Lexicographers could be well-advised to identify and prioritise a given range of dictionary types and to embark on the compilation of an extremely user-friendly product aimed at the most prominent and basic needs of the target users. This could be followed by a more sophisticated product running parallel to a higher degree of lexicographic sophistication among the target users.

## 5.2 Different types of dictionaries

### 5.2.1 No classification is absolute

Dictionary typology has been a favourite topic for discussion among theoretical and practical lexicographers and a wide range of suggestions, classifications and models have been formulated. It has to be emphasised that none of these classifications can be regarded as absolute. The same dictionary can be classified differently by different typological models due to the lack of a general standardised set of terminology to classify dictionary types. Within any given typological model one will find fuzzy borders between the different types of dictionaries. This is due to an overlap between different types of dictionaries and the fact that the same features can be relevant to more than one member of the typological model. No typological model may categorise and classify its different types of dictionaries in such a way that any given dictionary type is placed in a waterproof compartment where it is isolated from all the other types in that model.

The following section contains a brief overview of some of the most common types of general dictionaries. This overview does not rely on any specific model for its classification and terminology but utilises terms from different models. A more comprehensive account of dictionary typology can be found in Zgusta (1971), Hausmann (1989a) and Gouws (1989).

## 5.2.2 A typological classification

### 5.2.2.1 *General versus restricted dictionaries*

A common distinction between broader categories of dictionary types is indicated by the dichotomy *general dictionary* and *restricted dictionary*. The term *general dictionary* is an opposite of the term *restricted dictionary* and refers to dictionaries dealing with a broad selection of lexical items, i.e. not only items taken from one specific field, and it offers a treatment aimed at different linguistic and pragmatic features of the lexical items in question. Restricted dictionaries focus on one field, e.g. a specific semantic field or a specific subject domain (the so-called languages for special-purposes dictionaries, specialised dictionaries or technical dictionaries), or one type of lexical item, e.g. idiom dictionaries, dictionaries of abbreviations, etc., or they limit their treatment to one data category, e.g. pronunciation dictionaries, etymological dictionaries or dictionaries of synonyms. A thesaurus is also a restricted dictionary because it offers a limited treatment of lexical items. Its thematic ordering focuses on semantic relations as the primary data type to be conveyed. The category of restricted dictionaries will not be discussed in this chapter. The South African lexicographic practice is in need of both general and restricted dictionaries. At this stage the terminology section of the Department of Arts and Culture has the assignment to compile restricted dictionaries focusing on languages for special purposes, whereas the NLU's will primarily be involved in the compilation of different types of general dictionaries. However, it may happen that in future NLU's will have the responsibility for the compilation of all the different types of dictionaries. Within specific situations, characterised by a lack of restricted dictionaries and the need among members of the speech community for the lexicographic presentation and treatment of certain technical terms, the choice of lemmata to be included in general dictionaries may exceed the traditional boundaries and venture into the domain of restricted dictionaries. This will imply that lexicographers of general dictionaries will have to deviate from the conventional nature of general dictionaries and perform certain treatment procedures characteristic of restricted dictionaries. The compilation of dictionaries treating languages for special purposes has its own demands and lexicographers working on such dictionaries need to become familiar with these type-specific issues. For a detailed discussion of restricted dictionaries dealing with languages for special purposes, cf. Bergenholtz & Tarp (1995).

Although the NLU's will initially not be compiling dictionaries representing all the typological categories mentioned in this section, it would be good for the lexicographers to take cognisance of the different types in order to decide on the typological nature of their specific project and to situate it within the wider typological domain. In the remainder of this chapter the focus will be on different subtypes within the category of general dictionaries.

### 5.2.2.2 *Encyclopedic and linguistic dictionaries*

One of the first typological distinctions to be made when planning a dictionary is that between encyclopedic and linguistic dictionaries.<sup>3</sup> Encyclopedic dictionaries, often also known as encyclopedias, are directed at the extra-linguistic features of the items to be treated whereas linguistic dictionaries focus on the linguistic and pragmatic aspects. Linguistic dictionaries, especially comprehensive monolingual dictionaries, also contain a limited amount of encyclopedic data but this data will always be subordinate to the linguistic data presented in the specific dictionary. In this regard the data distribution structure of the specific dictionary and the possible choice of synopsis articles and registers included as outer texts, cf. paragraph 6.4, play an important role. A detailed discussion of encyclopedic dictionaries can be found in Hupka (1989).

Linguistic dictionaries can be divided into various types. A first distinction is between *monolingual* and *bilingual* or *multilingual* dictionaries. This distinction usually coincides with a distinction between the main assignment of these dictionary types, i.e. to give an explanation of the meaning of the lemma in monolingual dictionaries and to provide target language translation equivalents for a source language lemma in bilingual and multilingual dictionaries. Within each one of these typological categories a range of subdivisions exists and these subdivisions often run parallel between the two categories. One would for example find a learner's dictionary in the category of monolingual as well as in the category of bilingual dictionaries.

In the following paragraphs the relevant subcategories within the typological categories of monolingual descriptive and bilingual dictionaries will be discussed briefly. The term *bilingual dictionary* will be used to refer to both bilingual and multilingual dictionaries, i.e. all those general dictionaries where the primary treatment is the provision of translation equivalents for the source language item.

### 5.2.2.3 *Monolingual dictionaries*

It has been argued convincingly, cf. Gallardo (1980), Gouws & Ponelis (1992), that the compilation of a bilingual dictionary should be regarded as the first priority when a given language does not have any dictionary or has not yet been fully standardised. These arguments also apply to the South African lexicographic situation. Some NLUs may feel that the existing bilingual dictionaries in which their language functions as one of the treated languages, suffice the immediate needs of the speech community regarding this typological category. In this case the units should be free to continue with the compilation of a monolingual descriptive dictionary. However, these dictionaries should initially be of a restricted nature and should typically belong to the category of desk, standard or learner's dictionaries.

Although an article in a descriptive monolingual dictionary usually contains a variety of data types and although dictionaries are consulted in order to retrieve information from all these different data types, it has been indicated by several investigations that the entry conveying the description or paraphrase of meaning, the so-called lexicographic definition, is the one most frequently looked for by the typical user of a given dictionary. The fact that the semantic data and more

---

<sup>3</sup> This discussion of dictionary typology is an adapted version of a section in Gouws (2001b).

specifically the definition or paraphrase of meaning, cf. Chapter 9, is regarded as the most salient data type can be seen in the name given to this category of dictionaries, i.e. descriptive dictionary.

When compiling a descriptive dictionary the lexicographer has to work within the typological framework applicable to the specific subcategory because the relevant norms and criteria will determine the data distribution, extent and presentation in the articles. The dominating position of semantic data can easily lead to a semantic bias that eschews the need for a whole range of other data types to be included in the articles, cf. Gouws (2000b), and can impede the success of dictionary consultation procedures.

The category of descriptive monolingual dictionaries can be divided into four subcategories, i.e. comprehensive dictionaries, standard dictionaries, desk/college dictionaries and pedagogical dictionaries. When deciding to compile a descriptive monolingual dictionary, the dictionary plan has to give a clear indication of the specific subtypological category of the intended dictionary. Once again it has to be emphasised that these typological categories are not mutually exclusive and often share some features.

#### *Comprehensive dictionaries*

As indicated earlier comprehensive dictionaries typically are multivolume and multi-decade projects. They have an overall-descriptive and informative approach and give an account of the full spectrum of the lexicon, including lexical items from the non-standard varieties. Lexical diversity is covered extensively. To achieve this assignment, a comprehensive dictionary has to rely on a well-established corpus.

This type of dictionary is comprehensive in more than one way and its comprehensive character applies on macro- and microstructural level. On macrostructural level such a dictionary is classified as comprehensive due to the wide-ranging selection of lexical items to be included as lemmata. It endeavours to cover as full a spectrum of the lexical stock of the given language as possible. On microstructural level the comprehensive nature can be seen in two ways. Firstly the dictionary articles include a variety of entries representing a wide range of data types to be used in the treatment of the lemma. The microstructure gives an extensive account of the linguistic features of the lemma signs. This leads to a balanced data density, i.e. the quantitative relation between macro- and microstructural entries, in the dictionary. Secondly each data type is presented in a comprehensive way, e.g. the description of the meaning of a given lemma sign is done in much more detail compared to other types of descriptive dictionaries, and the data on the pronunciation of a lexical item will not only indicate tone or the main stress pattern but will also present the user with a full phonetic transcription.

Comprehensive dictionaries are typically historically oriented and are directed at a lexicographic treatment reflecting the past and the present characteristics of the language. A representative corpus is therefore needed to supply the lexicographers with the necessary data to give a chronological indication of the development of the form and meaning of a given lexical item and to describe its origin and etymology.

It is important to note that the compilation of a comprehensive monolingual dictionary presupposes a fully standardised language, a typological infrastructure

in the given language as well as advanced lexicographic expertise and a sound metalexigraphic basis. This is not the type of dictionary to be compiled as the first project of a NLU. The history of Afrikaans lexicography gives enough evidence of the problems encountered due to the untimely start with the compilation of the comprehensive *Woordeboek van die Afrikaanse Taal*, cf. Gouws & Ponelis (1992), Gouws (1986). The NLUs for the African languages should not make the same mistake. Although it should be the ultimate goal of each NLU to compile a comprehensive dictionary, it also has to be regarded as the lexicographic crown jewel, which is acquired once a network of smaller dictionaries has been completed.

#### *Standard descriptive dictionaries*

Standard descriptive dictionaries can be regarded as products resulting from a well-established lexicographic environment. These dictionaries are the most commonly used monolingual lexicographic instruments and display a wide range of lemmata and microstructural categories. Standard dictionaries usually are single volume products in which a synchronic and normative approach prevails. The macrostructure primarily represents the standard variety of the treated language although a number of high usage frequency items from non-standard varieties, e.g. slang or special fields, may also be included. However, these items from the non-standard varieties should be clearly marked by means of lexicographic labels indicating stylistic, chronolectic, regional or other deviations from the standard variety. Standard dictionaries include a representative selection of macrostructural items and an extensive treatment of these items which presents a variety of data types in each article. These dictionaries consequently have a high data density.

The data distribution structure of these dictionaries has to be well-devised and usually employs a frame structure and the consequent use of outer texts as functional components of the dictionary. Both integrated and non-integrated outer texts are used. (The terms *frame structure*, *outer texts*, *integrated* and *non-integrated outer texts* are explained in Chapter 6.) Although a varied and extensive microstructural treatment is presented, standard dictionaries are characterised by a thorough semantic treatment which includes the use of a variety of definition types and an indication of semantic relations. The definitions contain a limited amount of encyclopedic data but the data distribution structure makes provision for the use of complex articles, cf. paragraph 6.5, to convey extra-linguistic data in the treatment of certain types of lemmata. Standard dictionaries contain medium-sized articles which do not usually display a comprehensive micro-architecture, cf. paragraph 11.3.

In a standard dictionary little attention is given to historical data. The macro- and microstructural presentation should be aimed at the present and future language usage. These dictionaries are compiled for fairly sophisticated users who have acquired a certain level of dictionary culture and who can cope with an access structure leading the user to implicit and explicit data. Their compilation requires a substantial corpus, which gives a valid and representative account of the lexicon and everyday usage of the language. This is an ideal dictionary to be compiled once a speech community has already acquired access to a desk dictionary.

*Desk/college dictionaries*

This category of dictionaries is usually compiled for mother-tongue users and do not display a learner-orientated approach. In comparison with school dictionaries they display an extended macrostructure but a low data density prevails because of a limited microstructural treatment and a restricted article structure. Little cotext assistance is given and the focus is on a brief paraphrase of meaning of the lemma sign. These dictionaries usually contain short articles and do not rely too heavily on a corpus. For new NLUs this could be a good option as a first project.

Compiling a desk dictionary can be regarded as an ideal introduction to the world of dictionary making. It gives lexicographers the opportunity to apply their knowledge and to gain practical experience. A project like this has the additional advantage that it can be completed in a relatively short time to ensure the rapid availability of a dictionary to the relevant speech community. The work done in the compilation process of such a dictionary could form the basis for the next phase in the typological hierarchy, e.g. the compilation of a more comprehensive dictionary like a standard dictionary.

NLUs should pay serious attention to the possibility of focusing their early lexicographic activities on the compilation of a desk dictionary.

*Pedagogical dictionaries*

Dictionaries belonging to this typological subcategory can be divided into two further subcategories, i.e. *school dictionaries* and *learner's dictionaries*.

*School dictionaries* represent a specialised category of lexicographic work, cf. Lombard (1990), the quality of which has in the past too often been neglected due to a false impression that they are easy to compile and merely require a cut and paste approach to extract them from bigger monolingual dictionaries. School dictionaries are aimed at scholars who are mother-tongue speakers of the language treated in the dictionary. Due to the needs of their target users a synchronic approach typifies this dictionary type. The macrostructure of such a dictionary is limited and represents the core vocabulary with which scholars come into contact during typical natural conversations and when working through their study material.

School dictionaries display a low density of data because each article is only allowed a restricted number of microstructural categories. These dictionaries focus on the comment on semantics and more specifically on a brief paraphrase of meaning given as the lexicographic definiens. A limited process of cotextualisation occurs although examples are used to illustrate some typical occurrences of the lexical item functioning as treatment unit. The access structure of a school dictionary should be devised to assist the specific age group, identified as target users of the dictionary, in a functional way. The dictionary will be dominated by the central list and a limited number of texts function as outer texts. It is important that the pedagogical function has to prevail at all times in these dictionaries.

When planning a school dictionary the lexicographers have to take cognisance of the educational and general communication environment of the target users of their dictionary. A school dictionary should empower its users in their attempts to improve their communication skills in their mother-tongue. It should also help them to decode and understand the language they are confronted with on a daily basis.

*Learner's dictionaries* represent one of the fastest growing subtypological categories in modern lexicography, cf. Lemmens & Wekker (1986), Cowie (1987), Otto (1989), Hartmann (1992), Van der Colff (1996), Tarp (2004; 2004a), Tarp & Gouws (2003), Gouws (2004a), Steyn (2004). The existence of these dictionaries is a direct result of the user-driven approach and the consequent attempts to compile dictionaries that respond to the needs of specific target user groups. A learner's dictionary is directed at a user learning a foreign language. Therefore it treats the lemmata in such a way that the learner can have easy access to the presented data in order to achieve an optimal retrieval of information. For the NLUs this typological category should not be of immediate primary interest because these dictionaries are not directed at the needs of the speech communities for whose lexicographic needs the NLUs are responsible. However, many features prevailing in learner's dictionaries can be valuable assets to dictionaries compiled for mother-tongue speakers. NLUs should also be aware of the function these dictionaries can have at a later stage to promote the language of a specific NLU amongst people interested in learning that language.

A learner's dictionary has a macrostructure which represents high usage frequency lexical items. The microstructural treatment includes a variety of data categories, giving the dictionary a high data density, and the treatment is characterised by a very explicit nature. The more explicit the presentation, the easier it is for the foreign language learner to understand the data presented in the article. A typical feature of learner's dictionaries is a limited application of procedures of textual condensation. Textual condensation leads to a more implicit presentation of data which is not the ideal in a learner's dictionary. Where the target users of a dictionary do not have a long tradition of dictionary consultation or where they are confronted with a foreign language, it is wise to abstain from unnecessary procedures of textual condensation, e.g. the use of a tilde to substitute the lemma in an illustrative example or the omission of the stem in an indication of derivations. Learner's dictionaries therefore avoid procedures of textual condensation which typifies dictionaries compiled for mother-tongue speakers. The lexicographer of a learner's dictionary may never rely on the linguistic intuition of the target users.

The comment on semantics in a learner's dictionary also focuses on the paraphrase of meaning and in some learner's dictionaries, e.g. COBUILD and *Basiswoordeboek van Afrikaans* the definitions are given in full sentences. A typical feature of these dictionaries is the prominence of illustrative examples to present the typical context in which the lexical item represented by the lemma sign occurs. The definiens has to be supported by ample examples illustrating the typical occurrence of the word in natural language. In this regard a relation of addressing equivalence, cf. Gouws (2000; 2000a), between context entries and meaning paraphrase entries should prevail. The compilation of a learner's dictionary presupposes the existence of a representative corpus of modern-day language reflecting typical usage.

#### 5.2.2.4 *Bilingual dictionaries*

##### *Subtypological diversity*

Although bilingual dictionaries include a variety of data types in their articles as part of the treatment of the lemma, their primary function is to provide a target language equivalent for a given source language item.

Within a multilingual society bilingual dictionaries have to be regarded as the most commonly used lexicographic sources, cf. Gouws (1996; 2000b). Unfortunately, in spite of some high quality research, metalexigraphers still have not given enough attention to the problems of bilingual dictionaries in multilingual environments. The organisation of the South African lexicographic practice with regard to the planning and compilation of bilingual dictionaries could set an interesting example to the rest of the lexicographic community. A further problem regarding bilingual dictionaries is the fact that in the past too little emphasis has been placed on the different subtypological categories within the category of bilingual dictionaries. More or less the same subtypological distinctions which apply to monolingual dictionaries also apply to bilingual dictionaries, i.e. pedagogical dictionaries (with specific reference to school and learner's dictionaries), desk/college dictionaries and standard dictionaries. The category of bilingual dictionaries, however, does not include a comprehensive bilingual dictionary.

The multilingual and multicultural South African society provides a perfect platform for the introduction of a wide-ranging collection of bilingual dictionaries. As has been indicated earlier in this chapter Gallardo (1980) argues convincingly in favour of the choice of bilingual dictionaries in an environment where the treated language has not yet been fully standardised. This also applies when the treated language lacks good dictionaries. The advantage of a good bilingual dictionary is that it not only treats the source language, the primary object language of the compiler, or, in the South African case, of the specific NLU, but it also gives the target users access to another language and enhances their communicative skills.

The new NLUs in South Africa, especially those NLUs dealing with languages with no or only inferior dictionaries, should consider to embark on the compilation of a bilingual dictionary in which their language is paired with one of the other South African languages relevant to the communication needs of the speech community, cf. also paragraph 5.2.2.5. If a NLU decides to compile a bilingual dictionary, a clear identification of the subtypological classification is of extreme importance. As far as the extent, macrostructural selection, target user and use of a corpus is concerned, a model for bilingual dictionaries should primarily adhere to the same typological principles as applicable to monolingual descriptive dictionaries, cf. paragraph 5.2.2.3. These aspects and the subtypological classifications will not be discussed again. In the following paragraphs attention will be focused on a variety of other aspects relevant to bilingual dictionaries. More detailed information can be found in Gouws (1989;1993;1996), Zgusta (1971;1987), Al-Kasimi (1977), Kromann et al. (1991) and Hausmann (1986).

#### *Some general features of bilingual dictionaries*

The dictionary specific lexicographic process aimed at the compilation of a bilingual dictionary has to give a clear indication of the function of the dictionary because it will have a definite effect on the contents and structure of the dictionary.

#### *Polyfunctional dictionaries*

As indicated in par. 1.5 some theoreticians in the field of dictionary research hold the meaning that for any given language pair at least four and perhaps even eight bilingual dictionaries have to be compiled to meet the diverse needs of the users

coming from both language groups. It was also mentioned that Wiegand (1996:2) pleads for the accommodation of different functions within one dictionary and even one dictionary article and the compilation of only one polyfunctional bilingual dictionary for any given language pair. However, it is extremely important that the data distribution and the data presentation in a polyfunctional dictionary should be devised in such a way that the different functions can be maintained. A polyfunctional dictionary has to adhere to certain structural norms. It should be poly-accessible and should display a well-devised microstructure.

#### Monoscopal dictionaries

A model for bilingual dictionaries should negotiate the principles discussed in Hausmann (1992). He illustrates quite clearly that the typical bilingual dictionary cannot be used with the same degree of success by speakers of both the treated languages. The user is influenced by the positioning of his mother-tongue in a bilingual dictionary with regard to source versus target language status. The distinction between bilingual dictionaries translating to the mother-tongue versus bilingual dictionaries translating from the mother-tongue leads to the demand for dictionaries with a monoscopal lexicographic treatment. In a dictionary with Sesotho sa Leboa and isiXhosa as language pair a monoscopal dictionary will have Sesotho sa Leboa as the only source and isiXhosa as the only target language, cf. Hausmann & Werner (1991:2740) and Wiegand (1996).

Hausmann & Werner (1991:2740) make a terminological distinction between *monoscopal/biscopal*; *monodirectional/bidirectional* and *monofunctional/bifunctional*. This is done in accordance with a more general distinction in bilingual dictionaries between scope, direction and function. According to Hausmann & Werner (1991:2742) *scope* refers to the language direction (*monoscopal* = A>B; *biscopal* = A>B and B>A); *function* refers to the instruction purpose of the dictionary (dictionaries for text production or text reception) and *direction* refers to the mother-tongue of the target users (dictionaries for mother-tongue speakers of the source or the target language or both these languages).

Hausmann (1992:410) regards the compilation of a monoscopal dictionary as the ideal situation. This implies that for a given language pair a polyfunctional dictionary should have two volumes and each volume should be monoscopal. This is a model which can be utilised successfully in the South African context but it represents a dictionary belonging to the subtypological category of standard dictionaries. For a smaller dictionary, e.g. a desk dictionary it should be better to work with a biscopal approach by including both an A>B and a B>A section in the same dictionary. Such a dictionary could later be expanded to the level of two monoscopal volumes.

Hausmann (1992) indicates that commercial publishing houses adapt their bilingual dictionaries to fit the needs of the speakers of both the treated languages. He warns that this leads to a situation where the quality of the dictionary can be severely impeded. According to Hausmann (1992:411) the compilation of bilingual dictionaries may not be seen as only the responsibility of the commercial publishing houses. The state should recognise its language political and culture political responsibilities and should demand that bilingual dictionaries should be compiled in such a way that they can also become instruments of intercultural politics.

For a model aimed at the NLUs Wiegand's preference for a polyfunctional dictionary and Hausmann's preference for a monoscopal dictionary have direct implications. These two points of view are not absolute opposites. The subtypological category of a dictionary should determine which approach needs to be followed. A standard bilingual dictionary can easily be both polyfunctional and monoscopal. However, when compiling a smaller dictionary, lexicographers have to be careful not to try to please the speakers of both languages. In this regard the NLUs have the responsibility to see to the lexicographic treatment of their language. The dictionary specific lexicographic process for a smaller bilingual dictionary, e.g. a desk dictionary, should clearly indicate at which group of mother-tongue speakers the dictionary is directed. Empowering the speech community served by the NLU, the dictionary should primarily be aimed at those speakers and the scope, direction and function should aim to suffice their needs. Although a bilingual desk dictionary may be biscopal, it should primarily focus on the mother-tongue speakers of the language served by that specific NLU. The polyfunctional approach may prevail in more comprehensive bilingual dictionaries, e.g. the standard dictionary.

In a model for the compilation of bilingual dictionaries it is extremely important that the problems of equivalent relations, equivalent discrimination and the difficulties to ensure communicative equivalence should be negotiated sufficiently. The dictionary specific lexicographic process should deal with this in a detailed manner, cf. Chapter 10 as well as Gouws (1989; 1996; 2000; 2000a, 2000b) and Hausmann (1977).

#### 5.2.2.5 *A typological hybrid*

When planning dictionaries lexicographers should not see themselves bound to the existing typological models but should have the freedom to create innovative typological models, cf. Gouws (1999a). Wiegand (1996) has already given an example of a dictionary article containing elements typical of both a monolingual descriptive and a bilingual dictionary. Hartmann (1994) also works with a similar approach when discussing the so-called *bilingualised* dictionaries.

The secondary comprehensive lexicographic process of each NLU could consider the possibility to start their lexicographic endeavours with the compilation of a hybrid dictionary which could satisfy some of the needs for a monolingual descriptive and some of the needs for a bilingual dictionary. A model which would make provision for such an attempt would also allow the expansion and further development of this dictionary into two separate and fully-fledged dictionaries, i.e. a monolingual and a bilingual dictionary.

It has been suggested in a previous paragraph that NLUs could commence their lexicographic endeavours with the compilation of a descriptive monolingual dictionary which belongs to the category desk/college dictionary. The article structure of such a dictionary could easily be adapted to accommodate a translation equivalent in English along with one example to illustrate the typical use of each translation equivalent. This will allow the users of the dictionary to get a paraphrase of meaning of the source language form (the form given in their mother-tongue) as well as the other entries to treat that lemma sign and, in addition, a translation equivalent into a preferred target language, e.g. English.

This hybrid dictionary will be monoscopal in the treatment presented in the central list because only the source language items will be lemmatised. However, an innovative approach could add some biscopal features to the dictionary and make the dictionary poly-accessible. Although desk dictionaries are usually not characterised by too many back matter texts, this dictionary could include an alphabetical equivalent register as a back matter text. Such a register, obtained by means of a computer programme which applies the reversibility principle, cf. Gouws (1989), can contain all the translation equivalents given in the central text, and the treatment of the lemmata in this register can be restricted to the treatment found in reference articles, i.e. giving only an entry indicating the central list lemma where the translation equivalent appears. Users can use this dictionary in a biscopal way by starting their access of the dictionary from the equivalent register. This approach is followed in NWSG, a monolingual Afrikaans dictionary compiled for users with Afrikaans as third or fourth language, cf. Steyn & Gouws (2005).

As a follow-up project the translation equivalents could be extracted along with the lemmata from the central list, and these items could be used as a basis for a more comprehensive bilingual dictionary. The monolingual component of the dictionary could also be expanded to develop into a more comprehensive monolingual descriptive dictionary, e.g. a standard dictionary.

The compilation of a typological hybrid will provide the target users with a functional instrument which can add to their communicative empowerment.

#### *5.2.2.6 A typological hierarchy*

The secondary comprehensive lexicographic process of each NLU should compile a typological hierarchy to guide the editorial staff with regard to the dictionary projects to be attempted. This hierarchy has to negotiate the needs and expectations of the target users and the larger speech community. It is a good approach to supply the users with smaller dictionaries and to make them aware of a continuum of dictionary specific lexicographic processes which will eventually lead to the compilation of a comprehensive monolingual dictionary and the existence of a spectrum of dictionaries. A model for the compilation of dictionaries should avoid the creation of users' frustrations and should endeavour to increase easy and quick access to dictionaries as containers of knowledge, cf. McArthur (1986).

## The frame structure of dictionaries

---

### 6.1 Dictionaries as carriers of text types

Recent work done in the field of metalexigraphy, especially regarding dictionary research, has suggested that dictionaries should be regarded as carriers of text types, cf. Wiegand (1996b). Each dictionary contains a range of different texts which are functional components of the dictionary as a “big” text. The positioning of the texts in a dictionary can be divided into three major areas, i.e. the front matter, the central list and the back matter. This distinction motivates two different approaches to the structure of dictionaries, i.e. the *word book structure* and the *word list structure*, cf. Hausmann & Wiegand (1989). The word list structure only focuses on the central list of a dictionary. The central list, often regarded and referred to as *the dictionary*, is a compulsory component of any dictionary. In general monolingual and bilingual dictionaries as well as dictionaries dealing with languages for special purposes it consists of a series of article stretches, i.e. all the articles included under a specific letter of the alphabet according to the first letter of the lemmata functioning as guiding elements of these articles. The central list typically accommodates article stretches representing the full alphabet but it can also include article stretches representing letters or letter combinations not occurring in the ordinary alphabet but part of the alphabet of a specific language. To illustrate this: Danish has no less than 29 letters in its alphabet. The 26 letters of the Roman alphabet are complemented by the letters *æ*, *ø* and *å*. In dictionaries the article stretch containing lemma signs starting with the letter Z is followed by the article stretches accommodating words starting with *æ*, *ø* and *å* respectively. Each article included in the central list of a dictionary can be regarded as a text in its own right.

The word book structure approach, an approach prevalent in the majority of modern-day dictionaries and strongly promoted by dictionary research, complements the central list by including some additional texts, situated either before and/or after the central list. These texts are collectively referred to as *outer texts* and within the collection of outer texts a distinction can be made between the *front matter texts* and the *back matter texts* as structural components of a dictionary. The front matter contains all the texts preceding the central list and the back matter contains all the texts following the central list. The front and back matter texts constitute the outer texts of a dictionary, and their occurrence establishes a type of dictionary structure known as the *frame structure* of a dictionary, cf. Kammerer & Wiegand (1998).

The central list is a compulsory text in any dictionary, cf. Hausmann & Wiegand (1989:331). The outer texts can usually be regarded as optional texts. However, there is one outer text that has to be included as the second compulsory text in any dictionary, i.e. a text, usually presented in the front matter, which contains the guidelines for the use of the dictionary. No lexicographer may assume that the target user of the specific dictionary will know how to use, to interpret and to understand the full lexicographic presentation. Therefore the structure, contents, presentation

and dictionary specific conventions should be explained to the user. This has to be included as a separate text, i.e. the *user's guidelines*.

## 6.2 Data distribution

When planning a new dictionary lexicographers have to realise that the functionality of the eventual product exceeds the boundaries of the central list. Both the front and the back matter can contain texts, which have a functional role in the presentation of the lexicographic data. A dictionary which exhibits a frame structure increases the options of the lexicographer when planning the lexicographic presentation. Outer texts do not only assist the user to ensure successful dictionary consultation procedures and to obtain an optimal retrieval of information but they also play an important role in the data distribution structure of the dictionary by allowing the lexicographer to accommodate the lexicographic data in more than one text.

Consequently, already as part of the formulation of the dictionary conceptualisation plan, decisions have to be taken regarding the structure of the dictionary and the possible use of outer texts. This is important because the use of outer texts requires a final decision regarding the data to be presented in these outer texts and the relation between the outer texts and the central list. These aspects of the formulation of the dictionary conceptualisation plan, fall within the scope of the *data distribution programme* of the dictionary. This programme organises the distribution of all the lexicographic data between the different texts presented in the dictionary. In a dictionary with a fixed ordering of articles, e.g. according to the alphabetical ordering of the lemmata functioning as guiding elements of the articles, the distribution of data displays a *data distribution structure*, cf. Bergenholtz, Tarp & Wiegand (1999:1779).

Two main types of data distribution structures can be identified, i.e. a *simple data distribution structure* and an *extended data distribution structure*. Where the central list is the only target for the data distribution the dictionary displays a simple data distribution structure. Where outer texts or parts of outer texts are employed to accommodate data as part of the procedure of data distribution, the dictionary displays an extended data distribution structure, cf. Bergenholtz, Tarp & Wiegand (1999:1779).

During the dictionary conceptualisation phase of the dictionary specific lexicographic process the lexicographer has to decide exactly where the different categories of lexicographic data should be accommodated. The data distribution structure does not only determine the distribution of data between outer texts but it also gives clear guidelines regarding the article internal presentation and the different search zones to which data categories are allocated.

## 6.3 Integrated and non-integrated outer texts

Where a dictionary displays a more comprehensive collection of outer texts, these texts, albeit that they are optional components of the dictionary as a big text, often play an important role to enhance the quality of the information transfer to which the dictionary is committed.

Within the collection of outer texts a distinction can be made between integrated and non-integrated outer texts, cf. Kammerer & Wiegand (1998). Some outer texts contain data from which the user can retrieve information about the subject matter of the specific dictionary, e.g. regarding the meaning, grammar, spelling, etc. of lexical items in a general dictionary or regarding technical data in a language for special purposes dictionary. The data distribution in a dictionary could lead to a presentation where the outer texts contain data relevant to the genuine purpose of the dictionary. Where this is the case these outer texts share the feature of presenting data regarding the subject matter of the dictionary in order to accomplish the genuine purpose of the dictionary, with the central list of that dictionary. These outer texts are therefore integrated into the genuine purpose of the dictionary and are called *integrated outer texts*. Contrary to these texts some other outer texts do not contain data from which information regarding the subject matter of the dictionary could be retrieved. Due to the nature of their data these texts are not integrated into the genuine purpose of that dictionary. However, non-integrated texts are functional components of a dictionary and although they do not fall within the genuine purpose of the dictionary they can play a vital role in the information transfer prevailing in a given dictionary and to assist the user with successful dictionary consultation procedures.

When planning the use of a frame structure it is important to make a clear distinction between *integrated* and *non-integrated outer texts*. Non-integrated outer texts function alongside the central list and are not needed to retrieve information presented in the articles of the central list or to contain data relevant to achieving the genuine purpose of the dictionary. Integrated outer texts function in co-ordination with the central list and are needed to ensure an optimal and full retrieval of the data distributed in the dictionary with regard to the subject matter of that dictionary, in order to achieve the genuine purpose. In some dictionaries certain lexical items which could be regarded as within the macrostructural scope of the dictionary are not entered in the central list but only in an outer text – where they receive the full lexicographic treatment. An example of this kind of integrated outer text is the inclusion and treatment of idioms in the translation dictionary with Dutch and English as language pair *Van Dale Groot Woordenboek Nederlands-Engels*. The user retrieves information relevant to the genuine purpose of the dictionary from this outer text. In a similar way lemmata representing abbreviations could be phased out of the central list and could be included in a separate back matter text. Such a text is integrated into the genuine purpose of the dictionary because it provides a presentation and treatment of items selected from the subject matter of the dictionary.

Another type of integrated outer text does not present a lexicographic treatment of lexical items but rather contains data from which information could be retrieved which can be regarded as a prerequisite for the successful interpretation of the treatment offered in the central list. This type of outer text functions in combination with the central list and is also integrated in the genuine purpose of the specific dictionary. As an example of this type of integrated outer text: the grammatical data to be presented in a given dictionary can be allocated to different texts, due to the data distribution programme of the dictionary. One possibility is that each article could include a full treatment of the grammatical aspects of the lemma sign. This implies that grammatical data is only accommodated in the articles of the central

list with each article containing all the data necessary to ensure an optimal retrieval of grammatical information relevant to the specific lemma sign as treatment unit. As an alternative the lexicographer could distribute the grammatical data in a different way by restricting the presentation of grammatical data in the articles of the central list in favour of a comprehensive discussion of grammatical aspects in one of the outer texts. This outer text could focus on systematic aspects like pluralisation, word classes, derivation, etc. Users can be referred from the article to a specific section in the relevant outer text, cf. the *Table of codes* in the *Longman Dictionary of Contemporary English*. This implies that such an outer text does not only function alongside the central list but constitutes a functional part of the lexicographic treatment of the lemmata in the central list and plays an important role in achieving the genuine purpose of the dictionary.

The use of outer texts also allows the lexicographer to include categories of entries, which would not typically appear in a general monolingual descriptive dictionary or a general translation dictionary. Dictionaries like these focus their macrostructural selection on lexical items from the general lexical stock of the given language. Quite often the lexicographers of these dictionaries want to refrain from the inclusion of e.g. proper names like the names of countries, languages, etc. However, due to their frequency of use or other aspects like spelling problems, it is often necessary to include certain proper names. By utilising a frame structure a dictionary can include outer texts consisting of lists of e.g. the names of countries or languages. Those proper names that would have been candidates for inclusion in the central list could be phased out of the central list to this outer text. Items like these do not necessarily need the same treatment given to the lemmata in the central list. A back matter text with these items could include only a limited treatment, focusing on data relevant to the needs of the intended target user. The central list of the Afrikaans monolingual descriptive dictionary *Nasionale Woordeboek* does not include lemmata representing the names of countries, inhabitants or languages. Due to the spelling problems with many of these words the lexicographers have opted for the inclusion of an extensive collection of these forms. The data distribution structure of the dictionary makes provision for a range of data types to be entered in the treatment of each lemma in the central list. The articles of the lemmata in the back matter text dealing with the names of countries, etc. do not contain the same data types as the articles in the central list but receives a restricted treatment limited primarily to orthographical guidance.

Outer texts also offer the lexicographer the opportunity to present other data not necessarily expected in a specific dictionary but still deemed appropriate by the lexicographer for the target users of that dictionary, e.g. cultural data. The typical treatment given in a dictionary article does not allow too much room for a comprehensive presentation of cultural data in the articles of lemmata that represent culturally-bound lexical items. Using a frame structure to accommodate cultural data is discussed in Gouws (2002b).

The dictionary user often consults a dictionary in search of something that falls outside the scope of the specific dictionary. Due to the lack of a dictionary culture people are not always aware of the limitations of a given type of dictionary as a source of reference. In trying to be user-friendly lexicographers will often anticipate some of these non-typical consultation procedures by the inclusion of

outer texts that could assist the target user in a more general reference consultation procedure. In this regard outer texts could be extremely helpful by offering lists of symbols, weights and measures. The inclusion and treatment of these items are not directed at achieving the genuine purpose of the dictionary but rather at providing additional assistance. Such a list can be regarded as a non-integrated outer text, cf. Gouws (2004).

The lemma selection in a desk or standard dictionary is restricted to lexical items with a high usage frequency. Lexical items, which usually qualify for inclusion in an encyclopaedia, will not be the typical candidates for macrostructural selection in a desk or standard dictionary. However, quite often the lexicographer of such a dictionary realises the need of the users to have access to the treatment of such items. The frame structure of the dictionary offers the ideal opportunity to include such items in a variety of addenda, cf. the list of military ranks in the *Longman Dictionary of Contemporary English*. These addenda could also be motivated on thematic grounds by presenting elements from a specific semantic field, cf. the *Animal table* in the *Longman Dictionary of Contemporary English*.

#### 6.4 The planning of outer texts for South African dictionaries

The frame structure creates an ideal environment to expand the data distribution in a dictionary and lexicographers of new dictionaries should be well-aware of this opportunity. Within the South African lexicographic practice the compilers of dictionaries should utilise every opportunity to enhance the transfer of information in terms of the real needs of their target users. General monolingual descriptive or bilingual dictionaries contain a limited selection of lexical items from technical or specialised fields. The prevailing argument is that these terms should be included and treated in dictionaries dealing with languages for special purposes. It is also argued that the expert in a specific field will not use a general dictionary to retrieve information regarding a term from his/her field of expertise but would rather consult a special field dictionary or a technical reference book. These terms are used in the communication between experts in a specific field and they do not qualify for inclusion in a general dictionary.

In contrast to terms only used in communication between experts, there is a whole range of specialised fields with terms also used in the communication between expert and layperson. Ideally these terms should primarily also be treated in a dictionary dealing with the terminology from that special field and laypersons should consult that dictionary when they require any assistance. The South African lexicographic reality confronts both lexicographers and dictionary users with the fact that a lack of dictionaries for languages for special purposes impedes the ideal dictionary consultation procedures. Once again South African lexicographers are compelled to adapt their dictionary models in order to equip their target users with additional information retrieval possibilities.

The data distribution programme of a general dictionary could make provision for the inclusion of a more comprehensive selection of terms from a well-identified range of specialised fields, e.g. from the medical, legal, insurance and computer fields, cf. paragraph 5.2.2.1. The treatment of these terms should not be directed at the needs of the expert in the specific field but rather at the layperson confronted

with those terms in his/her daily communication situations. In this regard outer texts could be used to inform the user of the selection of special fields and the relevant terms accommodated in the dictionary.

The frame structure of such a general dictionary could contain a back matter text entitled *Special fields*. This text could include different lists indicating the relevant special fields represented in the dictionary. The text could start with its own front matter text, a so-called secondary front matter text, cf. Gouws (2001a; 2002; 2004), giving a brief table of contents to indicate the different specialised fields which are represented in the various back matter lists. This secondary front matter text could be followed by a presentation of the different specialised fields with a listing of all the terms from those fields included in the central list of the dictionary. These lists do not have to present a treatment of the terms because the treatment is presented in the central list. Although the relevant terms could be the only entries in these lists, it will give the user a quick indication of the extent of the term collection included in the central list and it would assist the user by grouping the relevant terms together in their specific fields. The outer text *Special fields* could display its own thematic ordering.

The use of back matter texts which contain lists of items that also feature as lemmata in the central list of the dictionary necessarily elevates the dictionary to a *poly-accessible* source because there is more than one position from where a user can find access to a specific lemma. The user can either go to the central list and by following the alphabetical ordering system (s)he can arrive at the required lemma or the user can go to the register in the back matter, find an entry indicating that the specific term has been entered in the dictionary and then continue to look for the lemma in the central list.

## 6.5 The central list

The central list is the most salient component of a dictionary displaying a frame structure. Any dictionary, with or without a frame structure, should contain two compulsory texts, i.e. the central list and the text in the front matter containing the user's guidelines, cf. Hausmann & Wiegand (1989:331). The central list consists of *article stretches* and each article stretch includes a variety of articles which function as texts in their own right. Present-day metalexigraphic research favours a system which displays heterogeneous article structures, i.e. the inclusion of different types of articles in the central list of one dictionary, cf. par. 7.2.3.2 and 7.2.3.3. for a discussion of some of the article types, e.g. nested, niched and cross-reference articles.

The contents of the articles in the central list can also determine their nature. An approach favouring articles with heterogeneous structures does not have a critical influence on the structure of the central list but it does have a critical influence on the nature, extent and contents of the individual articles. These different article structures determine the type and distribution of data presented in the treatment of a given lemma, and the allocation of these data types to these articles should be determined by the data distribution structure of the dictionary. This leads to a further distinction between articles as textual components of the central list, i.e. the distinction between *single articles* and *complex articles*, cf. par. 7.2.3.1.

Single articles display the typical treatment allocated to the average lemma sign and represent the default article structure. Complex articles also display the treatment allocated to single articles but they go one step further and play a value-adding role by allowing the lexicographer to include additional data, relevant to the specific lemma, in the article. Complex articles usually have a stronger encyclopedic approach because the added data often falls outside the scope of the linguistic data categories on offer in single articles. In dictionaries dealing with languages for special purposes Bergenholtz, Tarp & Wiegand (1999) argue in favour of a category of articles which they call *synopsis articles*. In these articles the focus is not only on a treatment of the lemma but the treatment also applies to other lemmata in the dictionary. Synopsis articles are typically found when a more general term has to be treated and the treatment is also directed at more specific terms which are semantically included in the meaning of the lexical item represented by the lemma of the synopsis article. The treatment of the lemma *acid* in a science dictionary will include data also relevant to the lemmata *nitric acid* and *sulphuric acid*. The lemma *acid* will then be the guiding element of a synopsis article. Synopsis articles can be seen as a subtype of complex articles. Typical items to be the treatment units of complex articles in general dictionaries are scientific words used in the general conversation between expert and layperson, e.g. medical and legal terminology as well as culturally-bound lexical items. A dictionary could do well to include a limited number of complex articles in the central list to assist the user with encyclopedic information regarding some issues relevant to the specific speech community, e.g. certain diseases, certain cultural activities, etc. The back matter could also contain a text with a list of all the items which are treatment units in the central list's complex articles. This will also increase the poly-accessible nature of the dictionary.

The use of synopsis articles does not depend on the dictionary typology. Even a bilingual dictionary can include these articles, where the treatment unit is not only given a translation equivalent but also a brief encyclopedic description. Within a multilingual and multicultural society where dictionaries also have a definite nation-building and social responsibility, cf. Wiegand (1997) and Gouws (1999), the secondary comprehensive lexicographic process of a NLU as well as the dictionary specific lexicographic processes should consider the use of synopsis articles in the dictionaries to be compiled.

The central list hosts the most salient structural components of a dictionary. A model for any new dictionary and any dictionary specific lexicographic process will necessarily have to ensure that these structural components are established and used in a functional way.

Research in the field of metalexigraphy has led to the identification of a number of structural components to be negotiated in the central list of a dictionary, e.g. the *macrostructure*, *microstructure*, *access structure*, *addressing structure*, *search area structure* and the *mediostructure*. (There are more structural components but they will not be discussed here.)

- **Macrostructure:** The selection of lexical items to be included in the dictionary as lemma signs. They become the primary treatment units of the lexicographic process.

- **Microstructure:** The selection of data categories given as part of the treatment of the lemma sign. A macrostructural element combined with its microstructural treatment constitute a **dictionary article**.
- **Access structure:** It determines the search route followed by a user to reach a specific lemma sign or data category. A distinction is made between the outer and the inner access structure; resulting in an outer and an inner search route
  - *Outer access structure:* The search route leading the user to the relevant lemma sign.
  - *Inner access structure:* The article internal search route leading the user to the relevant data entry.
- **Search area structure:** The search area structure is constituted by the ordered set of search fields presented in a dictionary article.
- **Addressing structure:** The relation between an entry and the treatment unit at which it is directed is known as the addressing structure. Different addressing procedures can be identified, e.g.
  - *Lemmatic addressing structure:* The lemma is the address of a given entry.
  - *Sublemmatic addressing structure:* A sublemma is the address of a given entry.
  - *Non-lemmatic addressing structure:* Another microstructural element in the article is the address of a given entry.
- **Mediostructure:** The system of cross-referencing which leads a user from a reference position to a reference address.
  - The *article-internal cross-referencing* works within the boundaries of an article.
  - The *article-external cross-referencing* refers a user to an entry in another article or other text in the dictionary.
  - The *dictionary-external cross-referencing* guides the user to an address outside the dictionary.

The above-mentioned structural components are discussed briefly in the next chapters. In theoretical lexicography and in the lexicographic practice dictionary structures are very important. What is more important is to realise that these structures are a means to an end and not the ultimate aim of either metalexigraphic research or the dictionary compilation process. Structures are there to order the data and to function as tools in the hand of the lexicographer. In a user-driven approach to lexicography dictionary structures should be devised and implemented to assist the dictionary user and to help the lexicographer to compile a better dictionary.

## 6.6 Macrostructural issues regarding the central list

Within a frame structure the central list is ordered according to the alphabetical value of the lemmata to be included as macrostructural entries in the different article stretches. A number of aspects regarding the macrostructure have an influence on the nature of the central list of a dictionary.

The work on a dictionary does not start when the first word with the letter *A* is put onto paper and provided with a lexicographic treatment. Once the dictionary conceptualisation is done and criteria for the selection of lemmata have been established, the lexicographic work can commence with the collection and selection

of lemmata to be included as macrostructural elements. The lemmatisation process has its own problems, cf. paragraph 7.1, and especially in African languages issues like stem versus word lemmatisation have to be addressed well in advance, cf. Prinsloo (1994), Prinsloo & Gouws (1996), Gouws & Prinsloo (1997) and Gouws & Prinsloo (2005). The dictionary specific lexicographic processes have to make provision for a clear policy in this regard.

The macrostructure contains a selection of lexical items representing the lexicon of the object language, and this selection may not be made on a random basis. It is important that the compilation of any dictionary must be dominated by well-defined principles regarding the collection and selection of lexical items to be included as lemmata. The lemmata have to be drawn from a representative corpus of the specific language and a vital part of the planning of any lexicographic endeavour is the development of a corpus.

The successful retrieval of information in a dictionary often depends on an unimpeded access to the needed lemma-sign. The macrostructure is an ordering structure, cf. Wiegand (1989a:372). Lexicographic planning includes a clear-cut decision regarding the nature of the macrostructure to be presented in the dictionary. The arrangement of the lemmata is of primary importance and will have definite implications for the central list. The ordering of lemmata in a monolingual dictionary is something too often taken for granted by both lexicographer and dictionary user. The typological characteristics of general dictionaries determine that they should display an alphabetical arrangement. However, lexicographers have to make a distinction between a *straight alphabetical macrostructure* and a macrostructure with a *sinuous lemma file*. Where a sinuous lemma file is presented a further distinction has to be made between *niching* and *nesting* dictionaries.

*Niching* implies a strict alphabetical clustering of lemmata which may or may not be semantically related. *Nesting* implies a clustering which stretches the rules of strict-alphabetical ordering in order to exhibit morpho-semantic relations between words, cf. paragraph 7.4.6 for a more detailed discussion of these issues.

A straight alphabetical ordering demands less from the user. An ordering which allows the inclusion of nested and niched clusters adds to the textual condensation, cf. Hausmann & Wiegand (1989:336), Wolski (1989a). Both nesting and niching can have semantic implications. Clustering is done in order to promote space-saving textual condensation. In a monolingual descriptive dictionary clustering implicates semantic transparency. The question is whether this is always the case. With regard to the so-called *self-explanatory complex items* Philip Gove, editor of the *Webster's Third New International Dictionary* remarked that the *self* in *self-explanatory* should refer to the intended user; not to the lexicographer. Once again a specific lexicographic procedure executed within a component of the frame structure has to be dominated by a user-driven approach.

# Macrostructural aspects

---

## 7.1 Lemmatisation strategies

### 7.1.1 Introduction

In this section the emphasis will be on the lemmatisation of African languages which is believed to be more problematic than say that of Afrikaans or English due to the complicated nominal and verbal derivation systems of African languages.

Lemmatisation can be defined in an over-simplified way as the selection of a specific form from a given paradigm to be used in a dictionary as the starting point for information retrieval. An English lexicographer would for example select *talk* as the lemma to represent the paradigm *talk, talks, talking, talked*, etc. and the Sesotho sa Leboa lexicographer can select *bona* to represent the paradigm *bontšha, bone, bonwa, bonwe*, etc.

Consider the following section of the article of *talk* in the electronic Collins COBUILD (ECC)

**talk**

**1 talk talks talking talked**

When you **talk**, you use spoken language to express your thoughts, ideas, or feelings.

**He was too distressed to talk.**

A teacher reprimanded a girl for talking in class.

The term *lemma* is preferred to *head word* because items smaller, or bigger than words can also constitute a lemma, e.g. in Sesotho sa Leboa:

a. **-go** (*relative suffix*) **monna yo a sepelago**, *the man who is walking*.

b. **ela hloko** note carefully, observe, heed, pay attention

In the a-example a suffix has lemma status and the b-example reflects a multiword lemma.

As far as the lemmatisation of especially *nouns* and *verbs* in African languages is concerned, users often complain that they cannot find the nouns or specific verbal derivations that they are looking for. Compilers of dictionaries for these languages are often blamed for their inability to lemmatise nouns and especially the numerous derivations of verbs satisfactorily within the physical limitations of a printed dictionary, and to render a user-friendly product. Lexicographers also err in including words unlikely to be looked for by the target user at the expense of essential ones. Thus although dictionaries are generally available for African languages, lexicographers agree that these dictionaries generally lack proper lexicographical planning and are not seen as user-friendly products.

Physical limitations on volume, mostly on the number of pages and therefore on the number of entries that can be accommodated in a specific dictionary or sub-dictionary, has a far greater impact on lemmatisation in African languages than one would expect. In simple terms it boils down to the strategy of selection and presentation of words – or the lack thereof. – i.e. which words are to be chosen and how are they to be presented to ensure an optimal utilisation of available space.

... the problem remains as to whether all the lexical units that are likely to be derived from the main entry or the stem should be entered in the dictionary. (Busane 1990:30)

### 7.1.2 Lemmatisation approaches, strategies and traditions

When dealing with lemmatisation in African languages the lexicographer has to negotiate a complex interplay and overlap between (a) lemmatisation approaches, (b) lemmatisation strategies, (c) lexicographic traditions (d) nominal and verbal structures and (e) conjunctiveness versus disjunctiveness. Therefore an amount of repetition should be tolerated in the following presentation, in order to give a clear description and to do justice to all these aspects. Compare the most relevant relations categorically in terms of columns A – E and rows 1 - 5:

**Table 1: Lemmatisation approaches, strategies, traditions, etc.**

	A	B	C	D	E
	lemmatisation approaches	lemmatisation strategies	lexicographic traditions	nominal and verbal structures	conjunctiveness versus disjunctiveness
1	Traditional	Stem	Word	Verbal prefixes	Conjunctive orthography
2	Paradigms	Singular and plural	Stem	Nominal prefixes	Disjunctive orthography
3	Rule orientated	Singular only		Verbal suffixes	
4	Frequency	Left-expanded			
5		First or third letter			

In terms of Table 1 the following complex set of 1-1 relations exists and has to be negotiated in any discussion of the lemmatisation of nouns and verbs in African languages.

A1:B1; A1:C1; A1:D1; A1:E1; A1:B2; A1:C2; A1:D2; A1:E2; A1:B3; A1:D3; A2:B1; A2:C1; A2:D1; A2:E1; A2:B2; A2:C2; A2:D2; A2:E2; A2:B3; A2:D3; A3:B1; A3:C1; A3:D1; A3:E1; A3:B2; A3:C2; A3:D2; A3:E2; A3:B3; A3:D3; A4:B1; A4:C1; A4:D1; A4:E1; A4:B2; A4:C2; A4:D2; A4:E2; A4:B3; A4:D3; B1:C1; B1:D1; B1:E1; B1:C2; B1:D2; B1:E2; B1:D3; B2:C1; B2:D1; B2:E1; B2:C2; B2:D2; B2:E2; B2:D3; B3:C1; B3:D1; B3:E1; B3:C2; B3:D2; B3:E2; B3:D3; B4:C1; B4:D1; B4:E1; B4:C2; B4:D2; B4:E2; B4:D3; B5:C1; B5:D1; B5:E1; B5:C2; B5:D2; B5:E2; B5:D3; C1:D1; C1:E1; C1:D2; C1:E2; C1:D3; C2:D1; C2:E1; C2:D2; C2:E2; C2:D3; D1:E1; D1:E2; D2:E1; D2:E2; D3:E1; D3:E2;

A brief, oversimplified outline of the structure of nouns and verbs is a prerequisite and is given as a point of departure.

Nouns in African languages are grouped into different noun classes. Compare the following table for Sesotho sa Leboa.

**Table 2: Noun classes and examples for Sesotho sa Leboa**

Class	Prefix	Example	Translation
1	mo-	monna	man
2	ba-	banna	men
1a	∅	rrangwane	uncle
2a	bo+	borangwane	uncles
3	mo-	monwana	finger
4	me-	menwana	fingers
5	le-	lesogana	young man
6	ma-	masogana	young men
7	se-	selepe	axe
8	di-	dilepe	axes
9	N-/∅	nku	sheep
10	di+	dinku	sheep
11			
12			
13			
14	bo-	bogobe	porridge
6	ma-	magobe	different kinds of porridge
15	go	go bona	to see
16	fa-	fase	below
17	go-	godimo	above
18	mo-	morago	behind

In the case of verbs numerous derivations of a single verb stem exist, consisting of the root plus one or more prefix(es) and or suffix(es) as is clearly indicated in the following table for the verb stem **reka** 'buy' which is structurally analysed in terms of 18 modules. The complexity of this layout is evident:

**Table 3: Derivations of reka**

1	root + standard modifications	VR	reka
		VRPer	rekile
		VRPas	rekwa
		VRPerPas	rekilwe
02 ANA	root + reciprocal + standard modifications	VRRec	rekana
		VRRecPer	rekane
		VRRecPas	rekanwa
		VRRecPerPas	rekanwe
03 ANTŠHA	root + reciprocal + causative + standard modifications	VRRecCau	rekantšha
		VRRecCauPer	rekantšhitše
		VRRecCauPas	rekantšhwa

		VRRecCauPerPas	rekantšhitšwe
04 ANYA	root + alt. causative + standard modifications	VRAlt-Cau	rekanya
		VRAlt-CauPer	rekantše
		VRAlt-CauPas	rekanywa
		VRAlt-CauPerPas	rekantšwe
05 EGA	root + neutro passive + standard modifications	VRNeu-Pas	rekega
		VRNeu-PasPer	rekegile
		VRPas	
		VRPerPas	
06 ELA	root + applicative + standard modifications	VRApp	rekela
		VRAppPer	reketše
		VRAppPas	rekelwa
		VRAppPerPas	reketšwe
07 ELANA	root + applicative + reciprocal + standard modifications	VRAppRec	rekelana
		VRAppRecPer	rekelane
		VRAppRecPas	rekelanwa
		VRAppRecPerPas	rekelanwe
08 IŠA	root + causative + standard modifications	VRCau	rekiša
		VRCauPer	rekišitše
		VRCauPas	rekišwa
		VRCauPerPas	rekišitšwe
09 IŠANA	root + causative + reciprocal + standard modifications	VRCauRec	rekišana
		VRCauRecPer	rekišane
		VRCauRecPas	rekišanwa
		VRCauRecPerPas	rekišanwe
10 IŠEGA	root + causative + neutro passive + standard modifications	VRCauNpas	rekišega
		VRCauNpasPer	rekišegile
11 IŠETŠA	root + causative + applicative + standard modifications	VRCauApp	rekišetša
		VRCauAppPer	rekišeditše
		VRCauAppPas	rekišetšwa
		VRCauAppPerPas	rekišeditšwe
12 IŠETŠANA	root + causative + applicative + reciprocal + standard modifications	VRCauAppRec	rekišetšana
		VRCauAppRecPer	rekišetšane
		VRCauAppRecPas	rekišetšanwa
		VRCauAppRecPerPas	rekišetšanwe

13 OLLA	root + reversion transitive + standard modifications	VRRevt	rekolla
		VRRevtPer	rekolotše
		VRRevtPas	rekollwa
		VRRevtPerPas	rekolotšwe
14 OLLANA	root + reversion transitive + reciprocal + standard modifications	VRRevtRec	rekollana
		VRRevtRecPer	rekollane
		VRRevtRecPas	rekollanwa
		VRRevtRecPerPas	rekollanwe
15 OLLELA	root + reversion transitive + applicative + standard modifications	VRRevtApp	rekollela
		VRRevtAppPer	rekolletše
		VRRevtAppPas	rekollelwa
		VRRevtAppPerPas	rekolletšwe
16 OLLELANA	root + reversion transitive + applicative + reciprocal + standard modifications	VRRevtApp	rekollelana
		VRRevtAppPer	rekollelane
		VRRevtAppPas	rekollelanwa
		VRRevtAppPerPas	rekollelanwe
17 OLLIŠA	root + reversion transitive + causative + standard modifications	VRRevtCau	rekolliša
		VRRevtCauPer	rekollišitše
		VRRevtCauPerPas	rekollišwa
18 OLLIŠANA	root + reversion transitive + causative + reciprocal + standard modifications	VRRevtCauRec	rekollišana
		VRRevtCauRecPer	rekollišane
		VRRevtCauRecPas	rekollišanwa
		VRRevtCauRecPerPas	rekollišanwe

### 7.1.3 The traditional approach

In the pre-corpus era one could say that words were included in dictionaries ‘as they crossed the compiler’s way’. This is often referred to as the ‘traditional’ approach and is for example characteristic of revisions of bilingual dictionaries bridging Sesotho sa Leboa and English or Afrikaans. With each new revision one could see how more words were merely added to these dictionaries. This approach represents the worst situation where the compiler does not employ any selection strategy and even seems to be unaware of the problem of what to include in and what to omit from the dictionary. In the case of nouns and verbs (s)he conveniently ignores the need to reduce the number of derivations. In most cases it results in the compiler randomly adding words to the dictionary until the publication deadline or the maximum volume or number of pages prescribed by the publisher has been reached.

### 7.1.4 Completing paradigms

In dictionaries such as Ziervogel and Mokgokong's NSDN, compiled in this era, there was also the urge to 'complete paradigms' which is referred to by Prinsloo (1994:97) as an 'enter-them-all' approach. In NSDN it was attempted to enter all nominal and verbal derivations to such an extent that mother-tongue speakers doubt whether many of these derivations are actually and actively used. Compare a section of the article of aga 'build' in NSDN in this regard.

ÁGA	6	ÁGA
<p>and insects); seágo, di- (seagô) impers. dev.; dit wat gebou/opgerig is, boumateriaal // that which is built/constructed, building material; <i>AGÉGA</i> (-agéga, -agégilê) neutr.; boubaar w., kan gebou/opgerig word // b. capable of being built/constructed; <i>AGÉLA</i> (-agéla, -agétse, -agélwa, -agétswé) appl.; bou, ens. vir/by // build, etc. for/at; ~ <i>motho morakó</i> ten gunste van iemand praat // speak in someone's favour; <i>bolwétsi bo</i> ~ <i>mmeleg</i> die siekte word nie beter nie // the illness is not getting better; <i>boágélo</i> (boagélô) lo. dev.; plek waar gebou kan word (bv. waar 'n voënes of hoedernes gemaak is), bouterrein // place where one can build (e.g. where a nest is built), building terrain/area; <i>ikágéla</i> (-ikagéla) refl.; vir jousef bou, ens. // build etc. for oneself; ~ <i>morakó</i> moeilikheid oor jousef bring // invite trouble; <i>kágélo</i>, (n-)/di- (kagélô) man. dev.; <i>moágédi</i>, ba- (moagédi) pers. dev.; <i>moágelwá</i>, ba- (moagélwa) pers. pass. dev.; <i>moágélo</i>, ba- (moagélô) man. dev.; <i>seágélo</i>, di- (seagélô) impers. dev.; heiligdom, tydelike kamp // sanctuary, temporary encampment; <i>ÁGÉLANA</i> (-agélana, -agélane, -agélanwa, -agélanwe) appl. rec.; vir mekaar bou, huise langs mekaar bou // build for one another; build houses next to one another; <i>boágélan</i> (boagélani), cf. <i>kágišano</i>, buurskap // neighbourliness; <i>kágélan</i>, (n-)/di- (kagélanô) man. dev.; <i>moágélan</i>, ba- (moagélani) pers. dev.; buurman, [meerv.] bure wat vir mekaar bou // neighbour, [pl.] neighbours who build for one another; <i>moágélan</i>, me- (moagélanô) man. dev.; <i>ÁGÉLEDISA</i> (-agélédiša, -agélédišé, -agélédišwa, -agélédišwé) caus. &lt; <i>ÁGÉLELA</i>; <i>kágéledišo</i>, (n-)/di- (kagélédišô) man. dev.; <i>moágélediši</i>, mo-/ba- (moagélédišô) pers. dev.; <i>moágéledišo</i>, me- (moagélédišô) man. dev.; <i>ÁGÉLEDISANA</i> (-agélédišana, -agélédišane, -agélédišanwa, -agélédišanwe) caus. rec. &lt; <i>ÁGÉLELA</i> <i>baágéledišani</i> (baagélédišani) pers. dev.; <i>kágéledišano</i>, (n-)/di- (kagélédišanô) man. dev.; <i>moágéledišano</i>, me- (moagélédišanô) man. dev.; <i>ÁGÉLELA</i> (-agéléla, -agéléltse, -agélélwa, -agéléltswé) compl.; toebou in, bou op, tafkamp; fdraad span // enclose within, build on, fence; ~ <i>legora</i> 'n omheining versterk deur meer takke in te steek // strengthen a fence by putting in more branches; ~ <i>motse</i> 'n omheining om 'n stat bou // build a fence around a village; <i>boágélele</i> (boagélélô) lo. dev.; toegeboude/afgekampte plek // enclosed/fenced place; <i>kágélelo</i>, (n-)/di- (kagélélô) man. dev.; opbouing, stigting, edifikasie // edification; <i>moágéledi</i>, ba- (moagélédi) pers. dev.; <i>moágélelo</i>, mo- (moagélélô) man. dev.; fseagélédi, di- (seagélédi) omheiningsmateriaal // fencing material; <i>seágélelo</i>, di- (seagélélô) v. <i>seágélelwa</i>; <i>seágélelwa</i>, di- (seagélélwa) impers. pass. dev.; iets wat</p>	<p>omhein is, ens. // that which is enclosed, etc.; <i>ÁGÉLELANA</i> (-agélélana, -agélélana, -agélélanwa, -agélélanwe) rec. &lt; <i>ÁGÉLELA</i>; <i>baágélelani</i> (baagélélani) pers. dev.; <i>kágélelano</i>, (n-)/di- (kagélélánô) man. dev.; <i>moágélelano</i>, me- (moagélélánô) man. dev.; <i>ÁGÉLELSA</i> (-agélélsa, -agéléltsé, -agéléltsé, -agéléltsé) caus. &lt; <i>ÁGÉLELA</i>; <i>toebou</i>, insluit, beskerm // build around, enclose, protect, shelter; ~ <i>diatla</i> die hande uithou om te ontvang, verdedig // spread out hands to receive, defend; ~ <i>motho</i> iemand verdedig of beskerm // defend or protect a person; ~ <i>motšhidi</i> op tradisionele wyse verloof raak aan 'n meisie, in die bres tree vir iemand wat jy vertrou // become engaged to be married according to custom, step into the breach for someone you trust; <i>kágéletšo</i>, (n-)/di- (kagélétšô) man. dev.; <i>moágéletšo</i>, ba- (moagélétšô) pers. dev.; <i>moágéletšo</i>, me- (moagélétšô) man. dev.; <i>moágéletšwa</i>, ba- (moagélétšwa) pers. pass. dev.; persoon wat beskerm, ens. word, 'n beskermde // person who is protected, etc.; <i>seágéletši</i>, di- (seagélétši) pers. dev.; impers. dev.; persoon wat goed ageletša, ding waarmee ageletša word // person who ageletša well, that by means of which one ageletša; <i>seágéletšo</i>, di- (seagélétšô) impers. dev.; afgekampte plek, omheining, sambree // enclosure, umbrella; <i>seágéletšwa</i>, di- (seagélétšwa) impers. pass. dev.; iets wat beskerm, ens. word // that which is protected, etc.; <i>ÁGÉLELSANA</i> (-agélélsana, -agélélsane, -agélélsanwa, -agélélsanwe) rec. &lt; <i>ÁGÉLELSA</i>; <i>baágéletšani</i> (baagélétšani) man. dev.; <i>kágéletšano</i>, (n-)/di- (kagélétšanô) man. dev.; <i>moágéletšano</i>, me- (moagélétšanô) man. dev.; <i>ÁGÍŠANA</i> (-agišana, -agišitšé, -agišwa, -agišitšwé) caus.; laat/help bou, ens. // cause/help build, etc.; <i>kágišo</i>, (n-)/di- (kagíšô) man. dev.; stigting, vrede, konsiliase // foundation, peace, conciliation; <i>moágiši</i>, ba- pers. dev.; bouassistente, handlanger by bouery // building assistant, building help; <i>moágišo</i>, me- (moagíšô) man. dev.; v. <i>kágišo</i>; <i>moágišwa</i>, ba- pers. pass. dev.; <i>seágiši</i>, di- pers. dev.; <i>ÁGÍŠANA</i> (-agišana, -agišane, -agišanwa, -agišanwe) caus. rec.; mekaar help bou, mekaar laat bou, gelukkig saam woon, goeie bure wees // help/cause one another to build, live happily together, be good neighbours; <i>boágišani</i> v. <i>boágišano</i>; <i>boágišano</i> (boagíšanô) buurskap // neighbourly relations; <i>kágišano</i>, (n-)/di- (kagíšanô) man. dev.; cf. <i>boágélan</i>, buurskap, onderlinge begrip, sosiale omgang/verkeer // neighbourliness, mutual understanding, social intercourse; <i>moágišani</i>, ba-pers. dev.; buurman // neighbour; <i>moágišano</i>, me- (moagíšanô) man. dev.; v. <i>kágišano</i>; <i>ÁGÓLLA</i> (-agolla, -agolotšé, -agolotšé, -agolotšé) rev. tr.; ('n gebou) afbreek, goeie buurskap vernietig // pull down (a building), demolish, destroy neighbourliness; <i>kágólo</i>,</p>	

Figure 1: A section of the article for aga in NSDN

Furthermore, from Figure 1 it is clear that the user often has to struggle through numerous columns of fine print in the dictionary to find the meaning of a word by stripping suffixes and adding suffixal meanings as illustrated for *aga* in Chapter 3 – given here as Table 4:

**Table 4: Accessibility and Information retrieval process for *dikagollišano* in NSDN**

1	dikagollišano	↓	plural deverbative consisting of root + reversional transitive + causative + reciprocal + ending
2	kagollišano	↓	singular deverbative consisting of root + reversional transitive + causative + reciprocal + ending
3	agollišana	↓	verb root + reversional transitive + causative + reciprocal + ending
4	agolliša	↓	verb root + reversional transitive + causative + ending
5	agolla	↓	verb root + reversional transitive + ending
6	<b>aga</b>	↓	<b>verb (stem)</b>
7	build	↓	meaning of the verb
8	break down	↓	reverse or opposite meaning 'un-build'
9	cause to break down	↓	add causative sense of 'let/force'
10	cause each other to break down	↓	add reciprocal sense of 'each other'
11	the process of causing each other to break down	↓	nominalise: 'the process of ...' (singular)
12	the processes of causing each other to break down		change 'the process of ...' to the plural

It is not surprising that this 'enter-them-all-syndrome' resulted in an article such as Figure 2 for *phefa* 'hurl' where the compilers concentrated so hard on completing the modular paradigm that they 'forgot' to add translation equivalents in English or Afrikaans.

**PHÉFA** (-phéfa, -phéfilè, -phéfiša, -phéfilwé)  
(Pb.) v. **PHEMA**; mophéfi, ba- (mophéfi)  
pers. dev.; phéfo, (n-)/di- (phéfò) man. dev.;  
v. phemo; **PHÉFANA** (-phéfana, -phéfanè,  
-phéfanwa, -phéfanwe) rec.; v. phemana;  
baphéfani (baphéfani) pers. dev.; phéfano,  
(n-)/di- (phéfanò) man. dev.; **PHÉFÉGA**  
(-phéféga, -phéfégilè) neutr.; v. phemega;  
**PHÉFÉLA** (-phéféla, -phéfétše, -phéfétwa,  
-phéfétšwe) appl.; mophéfédi, ba- (mophéfé-  
di) pers. dev.; phéfélo, (n-)/di- (phéfélò)  
man. dev.; **PHÉFÉLANA** (-phéfélana, -phé-  
félane, -phéfélanwa, -phéfélanwe) appl. rec.;  
baphéfélani (baphéfélani) pers. dev.; phéfé-  
lano, (n-)/di- (phéfélanò) man. dev.; **PHÉ-  
FÍŠA** (-phéfiša, -phéfišitšè, -phéfišwa, -phé-  
fišitšwé) caus.; mophéfiši, ba- (mophéfiši)  
pers. dev.; phéfišo, (n-)/di- (phéfišò) man.  
dev.; **PHÉFÍŠANA** (-phéfišana, -phéfišane,  
-phéfišanwa, -phéfišanwe) caus. rec.; ba-  
phéfišani (baphéfišani) pers. dev.; phéfišano,  
(n-)/di- (phéfišanò) man. dev.

**Figure 2: The article of *phefa* in NSDN**

From this example it should be clear that it is impossible to do justice, especially in the sense of lexical information, to all derivations of a verb within the physical limitations of even a larger one volume dictionary like NSDN.

### 7.1.5 Rule-orientated approach

Limiting the number of lemmas or sublemmas was consequently attempted by Van Wyk in his revision of Kriel's *Pukuntšu* dictionary (PUKU 2). This was done by lemmatising only singular forms of nouns and only basic verbal stems plus giving sets of rules in the user's guide for the user to strip suffixes and add meaning components, much in the fashion illustrated e.g. in Table 4 above. In principle it still reflects the urge to 'enter-all' and although being quite economical in terms of dictionary space it is user-unfriendly. This strategy, also referred to as the 'regulate-them-in' approach, will be discussed in more detail below for the lemmatisation of nouns.

### 7.1.6 Frequency-based approach

The corpus era introduced for African languages and especially for Sesotho sa Leboa lexicography by Prinsloo (1991), see also Chapter 3, opened new doors for the lemmatisation of nouns and verbs namely lemmatisation based on frequency of use. Using corpus data the lexicographer can ensure that frequently used words are not accidentally omitted and, on the other hand, that precious dictionary space is not taken up by articles of which the lemma is unlikely to be looked-up by the target users. Following this approach the lexicographer could sensibly and even drastically reduce the number of lemmas for a specific verb such as *hlweka* 'clean' based on frequency of use. The best point of departure is a frequency list of actual occurrences of the verb taken from a Sesotho sa Leboa corpus.

*hlweka* (34), *hlwekago* (5), *hlweke* (2), *hlwekege* (1), *hlwekihlweki* (1), *hlwekile* (35), *hlwekilego* (77), *hlwekileng* (1), *hlwekiša* (100), *hlwekišago* (3), *hlwekiše* (2), *hlwekišeditšwego* (1), *hlwekišetša* (1), *hlwekišitše* (4), *hlwekišitšego* (1), *hlwekišitšwe* (3), *hlwekišitšwego* (4), *hlwekišo* (1), *hlwekišwa* (7), *hlwekišwago* (3), *hlwekišwang* (1), *hlwekišwe* (1).

The lexicographer can now lemmatise and reduce this list on the basis of frequency of occurrence in the corpus. (S)he could for example decide not to lemmatise derivations containing the relative suffixes *-go* or *-ng*, and not to include infrequent derivations, thus reducing the list to:

*hlweka*, *hlwekile*, *hlwekiša*, *hlwekišitše*, *hlwekišitšwe*, *hlwekišwa*

Support for the inclusion of highly used regularly derived forms can be found in Zgusta's remark (1989a:300) in reference to agglutinative languages like Sesotho sa Leboa:

... more dictionaries, particularly the modern recent ones will list at least some of the derived forms as separate entries...

There is no need to be hesitant in entering verbal derivations should it be to the benefit of the target user.

### 7.1.7 Word versus stem lemmatisation

These methods or *approaches* manifest in certain lemmatisation *strategies* for nouns and verbs underpinned by two lexicographic *traditions*, namely word versus stem lemmatisation. The ‘enter-them-all’ approach favours the stem tradition for nouns and verbs, with the given stems supplemented by full paradigms of derivations (e.g. NSDN). The ‘regulate-them-in’ approach follows the stem tradition for verbs but a word tradition for nouns, cf. PUKU 2. Consider firstly the ‘regulate-them-in’ approach of lemmatising verbal stems in combination with verbal derivational rules. The lexicographer lemmatises verb stems but not their derivational forms, and provides certain rules/guidelines which should be followed if a word cannot be directly looked up in the dictionary. The target user is expected to interpret (‘reverse’) regularly derived derivations.

PUKU 2 (Kriel and Van Wyk 1989:Preface)

Perfecta:

-dile:	-la,	bv. badile	onder bala
-ditše:	-tša,	bv. bidityše	onder bitša
-etše:	-ela,	bv. rapetše	onder rapela
	-ala,	bv. robetše	onder robala
-itše:	-ša,	bv. bešityše	onder beša
	-tšha,	bv. bontšhitše	onder bontšha
	-sa,	bv. lesityše	onder lesa
	-tswa,	bv. hlatswitše	onder hlatswa

etc.

Applicatives:

-etša:	-ša,	bv. tlošetša	onder tloša
	-tšha,	bv. tsentšhetša	onder tsentšha
	-sa,	bv. lešetša	onder lesa
	-tswa,	bv. hlatswetša	onder hlatswa
	-nya,	bv. senyetša	onder senya
-letša:	-tša,	bv. biletša	onder bitša

For example *hlatswa* ‘wash’ is lemmatised, but not *hlatswetša* ‘wash for’ (applicative) or *hlatswitše* ‘washed’ (perfect) because the user is expected either to know the grammatical derivation rules or if not, to look them up in the user’s guide. Thus the necessary reduction in the number of verbal derivations is achieved in principle by entering only basic forms, e.g. roots and irregular forms. This approach limits redundancy in the sense that many more verbs can be lemmatised in the same number of pages. However, any rule-orientated approach runs into serious difficulty with regard to practicality and user-friendliness. Firstly it has an ‘underlying’ disadvantage. Busane (1990:28) says:

..... many introductory pages [are] usually allocated to grammatical sketches of the language concerned without the knowledge of which it is deemed hazardous to use the dictionary successfully. We believe, however, that these sections and introductory explanations are not sufficient provisions for a user friendly product. Dictionary users are known to allocate little time to the study of these prefatory matters.

It could be argued that the amount of knowledge presupposed from the target user or the utilisation of the directions in the introduction is not unreasonable. However, once they have to deal with more than two suffixes added to the verbal root, chances of successful information retrieval are slim as in the case of *aga* in Table 4 above.

Another weakness of this approach is that extremely highly used verbal derivations are not physically included in the dictionary and that especially the inexperienced user is always in doubt as to whether (s)he has made the right conclusions regarding meaning, circumflexes, tonal pattern, etc.

The compiler of a dictionary based on frequency of use can easily capitalise on the virtues of the rule-orientated approach. In sacrificing a few pages (s)he too can supply guidelines like those given above to guide the user towards the handling of infrequently used words. Take for example *hlwekišetša* (1) listed above. This word will not be included in a pocket-size dictionary compiled on the basis of frequency of use. However, the user can at least be guided towards *hlweka* by means of guidelines.

In the case of nouns these lemmatisation approaches manifest in a number of specific strategies such as lemmatising (a) both singular and plural, (b) only singular forms (c) noun stems, (e) on first or third letter and (e) using left-expanded article structures.

#### 7.1.8 Lemmatising both singular and plural noun forms

Lemmatising both singular and plural noun forms is an extremely user-friendly lemmatisation strategy and very popular among inexperienced users and learners of the language. This is for example the method used by Kriel in all editions of the *Pukuntšu* and *Popular* dictionaries (PUKU 1, POP, etc.). No previous knowledge of the language is required - the user does not even have to know a single word in the language as long as (s)he knows the alphabet. This approach also causes no problem for irregular forms of plural nouns since they are included in terms of the lemmatisation strategy anyway. Unfortunately the *redundancy* factor in terms of dictionary space is almost 80% and has to be weighed up against the advantages in terms of user-friendliness and practicality.

#### 7.1.9 Lemmatising only singular noun forms

Lemmatising only singular noun forms is a sound lexicographic strategy, that is for example followed by Van Wyk, bringing the Kriel tradition of lemmatising both singular and plural forms in the *Pukuntšu* series to an end in 1989. It can indeed in terms of Van Wyk be argued that it is not too much to expect from the target user to know the regular productive rules of the language governing the formation of singular and plural forms. For users who are not so familiar with these rules Van Wyk, in PUKU 2, gives a detailed set of rules in the user's guide as in Table 5.

Once again it has to be pointed out that such rules are not user-friendly and that dictionary users probably do not consult them. However, by lemmatising only singular forms, precious space is saved which can be utilised for other entries. For this dictionary it is estimated that up to 30 pages were saved. The number of pages for words commencing on typical plural forms of the noun, **ba-**, **di-**, **ma-** and **me-** in these editions are compared in Table 6.

Table 5: Rules for looking up plural forms in Pukuntšu (Kriel and Van Wyk: 1989)

Rule		Example	
word starts with	look word up under	word starts with	look word up under
ba-	mo-	basadi	mosadi
bab-	mm-	babetli	mmetli
bo-	(the stem)	bomalome	malome
di-	se-	dilepe	selepe
	(the stem)	dikgomo	kgomo
ma-	le-	maleme	leleme
	bo-	maleke	boleke
mabj-	bj-	mabjang	bjang
mabo-	bo-	mabothata	bothata
me-	mo-	mello	mollo
meb-	mm-	mebutla	mmutla
mef-	mph-	mefoma	mphoma
mengw-	ngw-	mengwaga	ngwaga
nyw-	ngw-	nywako	ngwako

Table 6: Pages utilised for *ba-*, *di-*, *ma-* and *me-* in two editions of Pukuntšu.

PUKU 1 (Singular and plural forms lemmatised)		PUKU 2 (Only singular forms lemmatised)
number of pages		number of pages
7	ba- versus	2
18	di- versus	4
21	ma- versus	16
7	me- versus	2

Unfortunately it is not always that easy to apply the rule since in most cases, from a user's point of view, there is not a straightforward one to one correlation: *di-* (class 8 and class 10) has a one to two correlation, namely either *se-* or noun minus prefix *di-*. In simple terms it means that the user who does not know the meaning of the word *dinku* 'sheep' has to look it up under *\*senku* or *nku*. A one to three correlation exists in the case of *ma-* (class 6) and a one to four correlation in the case of *me-* (class 4).

Say for example the user is confronted with the word *meno* 'teeth.' The rule states that *me-* should be looked up under *mo-*. (S)he finds the word *mono* in the dictionary as 'finger' and consequently concludes that *meno* means 'fingers' whilst it means 'teeth.' The same is true for the word *meetse* 'water.' Taken at face value it is a word in class 4 and according to the rule in Table 5 it should be looked up under *moetse*. Under *moetse* the user finds 'mane, crest' and thus concludes that *meetse* means 'mane, crest' whilst it means 'water.' In fact *meetse* is in class 6, not in class 4, and has no singular form in Sesotho sa Leboa. The lexicographer is forced to lemmatise all such irregular forms separately anyway and consequently does not succeed in avoiding redundancy. Moreover, the inexperienced user has to know whether the particular word that needs to be found is a noun and not something else.

Another weakness of this approach is that extremely highly used plural forms such as e.g. *matšatši* 'days' (395), *basadi* 'women' (387) or *maoto* 'feet' (301) are not

physically included in the dictionary so that the inexperienced user is often in doubt as to whether the right conclusions regarding meaning, circumflexes, tonal pattern, etc. have been made. This is especially true in those cases where the plural forms are *more frequently used* than the corresponding singular forms such as e.g. *diaparo* 'clothes' (141) vs. *seaparo* 'a piece of clothing' (7). It is rather counterproductive to enter the less frequently used singular form just in order to be consistent. This is particularly evident in the case of *badimo* 'ancestral spirits' (296) which is lemmatised under *modimo* 'ancestral spirit' (0).

There is no doubt that such rules successfully combat redundancy but are not user-friendly and sometimes even mislead the user – something that goes against the user-perspective as emphasised in Chapter 4.

### 7.1.10 Lemmatising nouns on the first or the third letter

Lemmatising nouns on the first or the third letter is a method used by Snyman *et al.* in their *Dikišinare ya Setswana English Afrikaans Dictionary Woordeboek* (DS). In principle this is a type of hybrid between word and stem lemmatisation. In DS, irregular forms are automatically accommodated by being entered under their first letter, whilst regular forms are to be found under their third letter.

**kwâlô, le- ma-** dev < **kwala**, letter // brief ; **lo- di-**, book // boek ; **mo- me-**, handwriting, orthography // handskrif, skryfwyse

Redundancy is avoided by not having to compile separate articles for forms such as *lekwalô, makwalô, lokwalô, dikwalô*, etc. A disadvantage of the approach is that there are *always two* options and inexperienced users can get frustrated when selecting the 'wrong' option as a first attempt.

### 7.1.11 Lemmatising stems

Lemmatising nouns and verbs on their stems requires a more elaborate discussion since it goes to the heart of the two conflicting lexicographic traditions followed in the lemmatisation of African languages namely *word* versus *stem*. Central to these traditions stands the issue of *conjunctivism* versus *disjunctivism*, which has to be dealt with first.

Table 7: Conjunctivism versus disjunctivism

Sesotho sa Leboa	ba a mo thuša 'They help him/her'	ba they	a [pres.]	mo him/her	thuša help	
	go be go le motho 'there was a person'	go there	be was	go there	le is	motho a person
isiZulu	bayamsiza 'They help him/her'	ba- they	-ya- [pres.]	-m- him/her	-siza help	
	kwakungumuntu 'there was a person'	kwa there	(be) was	ku there	ng(u) is	umuntu a person

Van Wyk (1995) emphasises that conjunctivism versus disjunctivism is purely a matter of orthographical convention. The stem tradition has mostly been followed for dictionaries for the conjunctively written languages namely isiZulu, isiXhosa, siSwati and isiNdebele, and the word tradition for the disjunctively written ones Sesotho sa Leboa, Setswana, Sesotho, Tshivenda and Xitsonga. The stem tradition is also followed in some dictionaries for the disjunctively written languages and will be discussed in more detail below in respect of stem versus word lemmatisation with regard to nouns. Lemmatisation of verbs according to the stem tradition will be discussed first.

### 7.1.12 Stem versus word tradition in respect of verbs

Van Wyk (1995) says that it is important to note the difference between nouns and verbs when it comes to affixes (prefixes and suffixes). A huge number of prefixes, up to more than 4,000 per verb, combine *freely* and *productively* with verbs, such as subject concords, object concords, negative morphemes, the progressive, the potential, future, etc.

... any verb root can be combined with any subject marker, any modal or aspectual morpheme . . . and any appropriate negative morpheme[.] If it is a transitive root, it can moreover be combined with any object marker (or the reflexive morpheme). The number of combinations possible for a suitable transitive verb stem is, therefore,  $18 \times 19 \times 6 \times 2$ . (Van Wyk 1995: 87).

Compare e.g. for Sesotho sa Leboa and isiZulu:

Sesotho sa Leboa

**Ke/re/o/le/ba** a sepela

(Free combination of any subject concord with the verb stem)

Re tlo **mo/ba/e/di/le/a/** bolaya

(Free combination of any object concord with the verb stem)

Ba **ka/sa/tlo** apara tše ba di ratago kudu

(Free combination of future, progressive, potential with the verb stem)

**Ga** ba thuše/ ba **se** thuše/ ba **sa** thuše/ ba **ka se** thuše

(Free combination of negative morphemes with the verb stem)

IsiZulu

NEG	SUBJ	NEG	MOD	OBJ	STEM
(k)a	ngi	nga	ya	ngi	<b>bona</b>
	u	nge	zo	ku	
	si		yo	si	
	ni		sa	ni	
	u/a		ka	m(u)	
	ba		nga	ba	
	...		...	...	

(Van Wyk 1995:86)

In disjunctively written languages isolating verb stems and verbal derivations and looking it up is non-problematic. Isolating these verbal forms in conjunctively written languages can however be problematic especially for the beginner. Consider the following extract from the isiZulu corpus for the word *hamba* 'walk':

**isiZulu**

nihambe, ukuhamba, kayihambi, ayehamba, sebehamba, ngangilihamba, ngingahamba, hambani, ekuhambeni, ubehambele, ngzihambela, owayehambele, wamhambisa, ayengasahambeli, zihambayo, ngihambile, kabahambanga

It is imperative that the user should be able to isolate **-hamba-** in all of the instances above for successful information retrieval since it is just not possible to enter each verb with all its productive prefixes separately in the dictionary. Redundancy will get totally out of hand if an attempt was to be made to enter a specific verb stem more than once with say different subject concords. Thus it could be said that both the word and the stem traditions are in agreement that verbs should be lemmatised on the verbal stem (and possibly certain verbal derivations), except for the notational device of a hyphen in dictionaries for conjunctively written languages. The entry for 'see' will therefore be found as **bona** in a Sesotho sa Leboa dictionary and as **-bona** in an isiZulu dictionary. For inexperienced users isolating the stem will remain a huge obstacle that would probably only be solved in future electronic dictionaries for the Nguni languages, see Prinsloo (2005).

So the lemmatisation of verbs on the stem is recommended for all SA languages. Some lexicographers refer to this as lemmatising the infinitive form of the verb and would offer for **bona**, for example an infinitive translation equivalent 'to see' as if translating the full infinitive form **go bona**. The opposite is found where the lexicographer takes the trouble to lemmatise the full infinitive form of this verb but translates it as 'see' instead of as 'to see'. Compare the following example from Rycroft's *Concise Siswati dictionary* (CSD):

CSD

(ku)-bona see; understand

Rycroft, in fact gives the full infinitive form of verbs (and nominal forms with their prefixes) but the infinitive (and nominal) prefixes are not considered in the alphabetical ordering.



Gouws and Prinsloo (2005) link this to the strategy of left-expanded article structures and describe it in great detail. Giving the infinitive verbal prefix with the stem has the advantage of the user seeing the full form of the infinitive but it is not expecting too much from the user of a dictionary for a conjunctively written language to isolate the infinitive class prefix (*ku*) from the verb stem. In fact there is no good reason why lexicographers could not break away from this tradition to lemmatise *infinitive* forms and simply lemmatise the verb stem which in most cases have the same form as e.g. the imperative form of the verb. The latter strategy is also not uncommon in the lemmatisation of other African languages on the continent.

### 7.1.13 Stem versus word tradition in respect of nouns

Unlike verbs prefixes do not combine *freely* and *productively* with nouns but the possible combinations are limited to but a few in each case. Consider the number of bold-faced forms versus the asterisked non-existing forms in the following example.

#### Sesotho sa Leboa

**motho** 'human being' (Cl. 1), **batho** 'human beings' (Cl. 2), \*motho (Cl. 3) \*metho (Cl. 4), \*letho (Cl. 5), \*matho (Cl. 6), **setho** 'limb' (Cl. 7), **ditho** 'limbs' (Cl. 8), \*ntho (Cl. 9), \*ditho (Cl. 10), **botho** 'humanity' (Cl. 14), \*gotho (Cl. 16/17) \*motho (Cl. 18).

Van Wyk (1995) pays detailed attention to this misconception and possible other reasons why lexicographers assume that verbs and nouns have to be treated in the same way, namely to lemmatise nouns in conjunctively written and even disjunctively written languages on their stem form. Lexicographers do not have to blindly follow the stem lemmatisation tradition for the sake of tradition, or worse, assume that stem lemmatisation is more 'scientific' than word lemmatisation. Van Wyk (1995) rejects the validity of such an assumption with detailed motivation.

It will be argued that nouns in disjunctively written languages should *not* be lemmatised on their stem forms and that lemmatising nouns with their noun class prefixes in the Nguni languages should be considered by prospective compilers as a viable option due to the many problems involved in stem lemmatisation of which the most crucial ones will be highlighted.

As a first step, consider word versus stem lemmatisation in disjunctively written languages. Lexicographers such as Prinsloo and Sathekge (NSD), Chapole (NSSD), Kriel (PUKU 1), Kriel and Van Wyk (PUKU 2), Wentzel and Muloiwa (TDV) and others lemmatised full orthographic forms of nouns. Ziervogel and Mokgokong (NSDN) opted for lemmatisation on stem forms. Snyman et al. (DS) in a sense use both strategies in the same dictionary. Mabile and Dieterlin (SSED) opted for a hybrid approach by giving the full orthographic form but performing the actual lemmatisation on the stem form, similar to Rycroft in the siSwati examples above. Compare the following series of examples.

#### Sesotho sa Leboa: NSD

**batho** people

**botho** kindness

**motho** person, human being

**setho** human nature  
**sethō** limb, part, member, organ

**Sesotho:** NSSD  
**letho** nothing  
**motho** person  
**ntho** thing object  
**setho** element, member

**Tshivenda:** TDV  
**khangā** (dzi-) tarentaal | guineafowl  
**pfufho** (dzi-) prys, belonging | prize  
**vhurifhi** (ma-) *cf Afr* brief | letter

**Xitsonga :** ETTE  
**lembe** year  
**hosi** chief, king  
**tafula** table

**Sesotho sa Leboa:** NSDN  
**-THO, bo-** ... humanity, goodness, kindness  
**-THO, mo-/ba-** ... human being, person, man (in general) ...  
 ...  
**-THÓ, se-/di-** ... limb, member

**Setswana:** DS  
 a. **tho, bo-**, personality, **mo-**, mankind, **mo- ba-**, person, human being, **se-**, humanely  
 b. **mmútla** pl. **mebútla**, hare // haas

**Sesotho:** SSED

**tho**, ideoph., *ho re -*, to be truly, genuinely human;  
**mōtho** (ba) n., human being; person; - *o ka hana*, one could refuse; - *mo- tonana o kaa ka*  
*tlou*, old description of the Great God;  
**thō**, ideoph., *ho re -*, to drip.  
**lēthō**, n., something; with neg. v., nothing;  
**se thō** (di) n., limb, organ; part of a slaughtered animal;  
**nthō** (di, ma) n., thing, matter; head of cattle;

Stem dictionaries compiled for Sotho languages such as NSDN as well as the stem-based strategy of SSED are regarded as user-unfriendly and rightfully criticised for *introducing* lemmatisation problems such as the urge to identify stems which could easily be avoided in word lemmatisation for these disjunctively written languages.

As a second step consider some major problems in stem lemmatisation for nouns. In terms of Van Wyk (1995) the crucial difficulties in following the stem tradition for the lemmatisation of nouns lie in the fact that in many instances *neither lexicographer*

nor the user of his dictionary can isolate the stem of the noun. He elaborates in great detail on the difficulties in this regard but for the purpose of this discussion a single oversimplified example taken from van Wyk (1995) will suffice.

In Column 1, of Table 8, seven isiZulu nouns containing the class prefix (*n-* or *m-*) of Class 9 is given. For the purpose of compiling a stem dictionary, three possible choices for stem/lemma-sign exist in principle in each case. So, for *impala* for example, **-mpala**, **-pala** and **-phala** are candidates. In some cases if the lexicographer possesses thorough knowledge of synchronic and diachronic grammatical rules the stems can be identified or postulated. But even if the lexicographer can do it, the user often can't. Consider Column 2 for the confusing result of what eventually were isolated and entered as lemma-signs in Doke and Vilakazi (ZED):

**Table 8: Problematic cases in Class 9 (Van Wyk 1995:90)**

NOUN	LEMMA
<i>impala</i> 'impala'	<b>-mpala (impala)</b>
<i>impilo</i> 'health'	<b>-philo (impilo)</b> [<phila]
<i>intaba</i> 'mountain'	<b>-ntaba (intaba)</b>
<i>intombi</i> 'girl'	<b>-thombi (intombi)</b> [<thomba]
<i>ubuntombi</i> 'maidenhood'	<b>-ntombi</b> [<intombi]
<i>inkosi</i> 'king'	<b>-khosi (inkosi amakhosi)</b>
<i>inkabi</i> 'ox'	<b>-nkabi (inkabi)</b>

This problematic situation results in different lemmatisation strategies in the same dictionary or dictionaries, having different 'stem'-lemmas for the same word often abandoning the stem principle as is the case with **-mpala**, **-ntaba**, **-ntombi**, **-nkabi** in Table 8 and in the following examples.

IsiZulu: CZD

**-nkosi** (i- ama-) (n) king; chief

IsiZulu: WAZ

**-khosi**, (in-, ama-), b; 1. koning, regent, hoofman

The nasal in lemmas such as **-nkosi**, **-nkabi**, **-ntombi**, etc. is included as the prefix of Class 9, thus no longer reflecting true stem lemmatisation. For the user it simply means 'keep on trying' to find the lemma – a severe impediment on accessibility. The latter aspect as well as the numerous incorrect conclusions (s)he could come to in examples such as **-nkosi** and **-khosi** has been discussed in Chapter 4. This abandoning of the stem principle in a stem dictionary is even more severe in instances where the user will not suspect that the stem principle has been abandoned. Consider the treatment of *ukuthi* in the following isiZulu dictionaries:

CZD

**-thi** (v) (*ukuthi*), to say.

EZD

**-thi** defect. V. 1. say. 2. mean, intend; ...

**ukuthi** conj. 1. (foll. by indic.) that. 2. (foll. by subjunct.) so that, in order that.

In CZD *ukuthi* is lemmatised strictly according to the stem tradition based on the infinitive, on its stem form *-thi* and the full infinitive form given in brackets. No treatment, either as a separate lemma or within the given article is given for the use of *ukuthi* as a conjunctive, which represents more than 90% of its use in the corpus. In the first example from EZD the stem tradition is followed for the meaning of *ukuthi* in the sense of 'say' but for the conjunctive meaning of *ukuthu* the lexicographer totally abandons stem lemmatisation for word lemmatisation by entering *ukuthi* in the alphabetical stretch *U*. This is done without cross-referencing from either *-thi* to *ukuthi* or vice versa. It is unlikely that a user of a stem dictionary would try to look up *ukuthi* under *U*, especially if no guidance in this regard is given in the user guide to the dictionary.

A typical argument against word lemmatisation for conjunctively written languages is that most nouns will fall in the alphabetical categories *I*, *A* and *U*, i.e. the first letters of the class prefixes of classes 1–14, and that all verbs will be under *U* for isiZulu and *K* for siSwati for example. For verbs the 'problem' will be solved if they are lemmatised on their stem forms as suggested above. As for nouns, firstly it is not uncommon in dictionaries of many languages that certain alphabetical stretches are much bigger than others. In Afrikaans for example, the category *S* contains 12% of the lemmas and for Sesotho sa Leboa the category *M* contains 17% respectively in comparison with categories such as *C*, *Q*, *X*, *Y*, *Z* with less than 1% for Afrikaans.

Nguni lexicographers could for instance put the *i*, *a* and *u* in italics to make the search easier on the eye in long stretches of *I*, *A* and *U*. Consider the following example as a subsection of such a presumed long alphabetical stretch for *I* and note that the problems in respect of e.g. Table 8 are solved for the lexicographer and the user.

<i>impala</i>	impala
<i>impilo</i>	health
<i>inkabi</i>	ox
<i>inkosi</i>	king
<i>intaba</i>	mountain
<i>intombi</i>	girl

Using stem lemmatisation for verbs and word lemmatisation for nouns in future dictionaries for the conjunctively as well as the disjunctively written African languages will mean a single lemmatisation strategy for, say, a bidirectional isiZulu ← → Sesotho dictionary.

The discussion of lemmatisation strategies brings us to the question whether it is possible to formulate a lemmatisation strategy for nouns which avoids the shortcomings and pitfalls and which exploits the virtues of those very same approaches at the same time. The recommended option is to lemmatise full forms of both singular and plural forms and to give full treatment at the member of the pair that is more frequently used.

**meriri** n. cl. 3/4 LHL hair (on the head) (plural), *meriri ye mešweu ke lehumo* grey hair is a treasure; ~ **wa titsana** soft hair, **motho wa** ~ a reliable person

**moriri** n. cl. 3/4 LHL (one) hair | see **meriri**

In this way one exploits the *virtues of lemmatising nouns on both singular and plural forms* such as (a) very user-friendly, (b) no previous knowledge of the grammar required, (c) in most cases not necessary to consult the guidelines to the dictionary, and (d) solving the problems relating to irregular forms as well as instances where the plural form is more likely to be looked up than the singular form. At the same time one could avoid the *major shortcoming* of this approach, namely redundancy, by giving only essential information in a smaller font size at the member of the pair less likely to be consulted, with a cross reference to the other one where a full treatment is offered.

## 7.2 Different types of lemmata and articles

### 7.2.1 Focusing on a specific environment

Users consult dictionaries to get some information about items from the vocabulary of a specific language. Research in dictionary use has shown that the average dictionary user does not regularly use the user's guidelines text when consulting a dictionary. They usually move directly to the central list where they try to solve the problem that motivated their search. Too often the search does not lead them to the required destination because the user is not familiar with the system of the dictionary, with the positioning of a specific microstructural item in a dictionary article, or does not know how to access the item they are looking for via the lemma sign functioning as guiding element of the relevant article. In the planning, compilation but also evaluation of dictionaries ample attention should be given to the different types of lemmata included in the macrostructure of a dictionary as well as to the different types of articles presented in the central list to accommodate the lexicographic data. A discussion of the nature and structure of dictionary articles and the different types of lemmata has to be done within the broader discussion of the structure of dictionaries and the way in which the macrostructure gives a representative account of the lexicon of the language treated in a given dictionary.

For the lexicographer it remains ever so important to be familiar with the reference needs and reference skills of the intended users of the dictionary, the function of the dictionary and the different types of usage situations where the specific dictionary will be consulted as a source of knowledge. Decisions regarding the different types of lemmata and articles need to be made with this user-perspective and usage-perspective in mind.

In the field of theoretical lexicography sophisticated research has already focused on the types of lemmata and dictionary articles, cf. Wolski (1989a), Wiegand (1989a; 2002, 2002a, 2003a). This section will not attempt to present a detailed or comprehensive discussion of the different lemma and article types. The focus will rather be on the issues relevant to the South African lexicographic environment.

## 7.2.2 Types of lemmata

### 7.2.2.1 *Moving from the lexicon to the dictionary*

The lemma selection of a dictionary should be done in accordance with the functions and the type of that dictionary. If a general dictionary has a text reception function the user should be able to find the words encountered in the day to day general language usage in that dictionary. The lemma selection of a dictionary should be done in such a way that the section of the lexicon falling within the scope of the dictionary type is adequately reflected in the dictionary.

The lexicon of any language consists of a collection of lexical items. These lexical items do not all share the same formal characteristics. Looking at the lexicon of many languages a threefold distinction can be made to provide a broad classification for the different types of lexical items. *Words* constitute the majority of lexical items in the lexicon of a language. The lexicon also includes *items smaller than words* that play an important role in word-formation processes. These items, usually *stems* and *affixes*, are fully-fledged lexical items and should be considered for inclusion in the lemma candidate list of a dictionary. The language-specific nature and features of the language treated in the specific dictionary will determine the decisions regarding the inclusion of stems and affixes in the dictionary. Lexical items can also consist of *more than one word*. These multiword items include idioms, group prepositions and language borrowings but also lexical items, determined by, among others, whether the language has a conjunctive or a disjunctive writing system.

Dictionaries have often been characterised and dominated by a word-bias, cf. Gouws (1989; 1991). This has led to a situation where the macrostructural selection has only focused on words and not on lexical items smaller than words or lexical items consisting of more than one word. A lexical-based approach to the macrostructure emphasises the need to lemmatise all the different types of lexical items. This implies that the macrostructure should contain words, entered as lexical lemmata, stems and affixes, entered as sublexical lemmata, and multiword units, entered as multilexical lemmata. In this way the macrostructure will reflect the lexicon of the target language of the dictionary. It is also important that the presentation of these different types of lemmata should indicate their equal status as treatment units of the dictionary.

In the following discussion the focus on words and stems as lemma candidates does not refer to the lemmatisation issue in some African languages where decisions need to be taken regarding stem or word lemmatisation. This lemmatisation problem is dealt with in par. 7.1. The present discussion merely focuses on the fact that in certain circumstances the stem occurrence of a lexical item does qualify for separate inclusion as a lemma.

### 7.2.2.2 *Lexical lemmata*

Words form the default lemma type and these lexical items are included in a dictionary as lexical lemmata. Yet again, the word-formation system of a particular language determines the types of words in that language but a typical distinction within the category of words is that between simplex and complex words. Both these types of words should be considered for inclusion in the lemma candidate list of a given dictionary. For English words like *honey*, *bee* and *keep* would be examples

of simplex words whereas words like *honeybee* and *keeper* would be complex words. Complex words can be divided into two categories, i.e. compounds and derivations. A compound typically consists of more than one stem, e.g. *honeybee*, whereas a derivation consists of at least one stem and at least one affix, i.e. a prefix, suffix or infix, e.g. *keeper*.

### 7.2.2.3 Sublexical lemmata

In general language use stems and affixes do not function independently but as parts of complex words. These complex words are potential candidates for inclusion in the lemma candidate list of a dictionary. Some affixes play a productive role in the expansion of the lexicon because they are frequently used to form new complex words. These affixes are fully-fledged lexical items and dictionary users have every right to find them as treatment units in a dictionary. In Afrikaans the prefix *ont-*, frequently used in the formation of verbs, is a polysemous lexical item. One of its senses is “to take away” and this sense is activated in words like *onthoof* (decapitate), *ontwater* (dehydrate), *ontmasker* (unmask). A second sense of this word “to loosen” prevails in a word like *ontkoppel* (disconnect). In words like *ontbrand* (ignite) and *ontsteek* (infect) the prefix *ont-* has the sense “to begin to”. This is part of the information a user needs to retrieve from a dictionary and a prefix like *ont-* should be regarded by the lexicographer on an equal basis with words when it comes to the selection of lexical items for the lemma candidate list.

In languages like English and Afrikaans lexical items often have the ability to function either as a word or as a stem. The lexical items *honey* and *bee* function as words in an example like *the bee produces honey* but they function as stems in the compound *honeybee* in an example like *the honeybee lives in a hive*. In the compound *honeybee* the stems *honey* and *bee* have the same meaning as the words *honey* and *bee* in the sentence *the bee produces honey*. Stem and word occurrence are variant uses of the same lexical item. If a dictionary includes the words *honey*, *bee* and *honeybee* as lemma signs there is no need to include the stems *honey-* and *bee-* as lemmata because little will be added to the information transfer already achieved. The Afrikaans lexical item *grond* (soil) is polysemous. This lexical item is frequently employed as word and as stem and in many of these occurrences the stem functions in one of the senses which is part of the polysemous paradigm of the word. In compounds like *grondbeginsel* (basic principle) and *grondwet* (constitution) the stem *grond-* displays a sense which the word *grond* never has, i.e. “first or most important”. It is one of the polysemous senses of the lexical item *grond* which is only activated in some of the occurrences of this lexical item as a stem. In order to present this sense of the lexical item *grond* it is necessary to include a separate lemma, i.e. the sublexical lemma *grond-*, in the macrostructure of Afrikaans dictionaries. Where the stem variants of lexical items display a sense not activated in the occurrence of that lexical item as a word, the lexicographer will do well to include the stem as a sublexical lemma in the dictionary.

In technical jargon one is often confronted by a whole range of complex lexical items with the same stem or combining form as first component, e.g. complexes like *aerodynamics*, *aerodrome*, *aeronautics* which have the form *aero-* as first component. Many of these techno forms have a productive occurrence in a specific technical language and they should be entered as sublexical lemmata with a treatment which

offers a paraphrase of meaning that will enable the user to use the given form in a productive way.

#### 7.2.2.4 *Multilexical lemmata*

The structure of a lexical item should not determine whether it qualifies for inclusion as lemma in a dictionary or not. Whether a lexical item is a word, an item smaller than a word or a combination of words it should be considered in terms of the general selection criteria of a given dictionary for inclusion in the lemma candidate list. The fact that lexical items consist of more than one word is often a result of the orthographic system of a given language. For many years the Afrikaans lexical item *weer eens* (yet again) had to be written as two words. Because it is a single lexical item it had to be included in Afrikaans dictionaries notwithstanding the fact that it is written as two words. Changes in the Afrikaans orthography have led to a situation where this item can now be written as either one or two words. The change in spelling has not affected its status as lexical item. It has been and will remain a single lexical item and should still be included in a dictionary as a lemma. The fact that *water bird* is written as two words and *waterfowl* as one word should not disqualify *water bird* from being included as a lemma whilst *waterfowl* is included. Lexical item status and not word status should be one of the main criteria for lemma selection. This will necessarily lead to the inclusion of different types of lemmata if the lexicographer tries to reflect the lexicon of a given language in the dictionary.

In spite of the orthographical system of a given language one often finds that the process of lexical borrowing results in words entered into the lexicon but maintaining the orthography from their language of origin. These borrowed forms become integral parts of the lexicon and qualify for inclusion in dictionaries. Afrikaans has a number of borrowings from e.g. Latin and Greek where a word group in these languages becomes a single unified lexical item in Afrikaans, although it is still written as more than one word. These items need to be included in a dictionary as multilexical lemmata, cf. examples like the following borrowings from the classical languages: *carpe diem*, *nolens volens*, *ex vires*, *ex post facto*, *sine qua non*.

Idioms are a part of the lexicon of a language and an idiom is to be regarded as a single lexical item. On semantic level it is quite clear because the meaning of an idiom is not the product of the meanings of the different words in the idiom but the idiom has a single meaning which is not related to the individual words. The meaning of the Sesotho sa Leboa idiom translated into English as *to catch the chicken by its beak* has nothing to do with the meaning of the individual words in this idiom but indicates that one rises early. Likewise an English idiom like *blood is thicker than water* has nothing to do with blood or water but indicates that you will do more for your relatives than for your friends. Idioms qualify for inclusion in dictionaries. However, for practical reasons it is often difficult to include idioms in the alphabetical ordering of lemmata. Therefore lexicographers do have to find a way of including the idioms as treatment units in their dictionaries. A typical procedure is to include an idiom in the article of a lemma sign that represents a key word from the idiom. In such an article a special slot is reserved for idioms, cf. the following example from the HAT where the entry "UITDR:" introduces the article slot occupied by idioms and a number of idioms in which the word *baard* has

been identified as a key word are grouped together in this article slot where they are presented as multiword lexical items and treatment units within this dictionary article.

**baard** s.nw. (-e) **1** Hare om die mond, op wange en kin, veral by 'n volwasse man: *Jou baard laat groei. Baard kweek.* ∇*Manuel, . . . / met die hol gesig en die bietjie baard* (D.J. Opperman). **2** lets wat aan ('n) baard herinner, bv. baardagtige aanhangsel by diere of plante, punt van 'n ploegskaar, blad van 'n sleutel, oneffenheid aan die rand van voorwerpe, soos drukletters. UITDR.: **Bietjie baard** maar klipsteenhard – gesê van 'n ervare persoon. *In jou baard brom*, mompel – meestal om ontevredenheid te kenne te gee. *Die baard groei deur sy keel* – gesê van 'n seun wie se stem begin breek. *Al baard kry*, al man word. **Praat jy wat baard het** – uitnodiging aan 'n ouer persoon om sy opinie te gee; soms effens spottend gerig tot 'n jonger persoon (asof hy oor baie kennis en insig beskik). **baard**: ~**agtig**, ~**draer**, ~**hawer**, ~**koring**, ~**kweker**, ~**loos**, ~**siekte**, ~**skurfte**, ~**stoppels**.

Following an approach like this one would mean that an idiom like the English *someone's ears must be burning* could be included in a search field in the article of the lemma sign *ear*. Although they are entered as items in an article introduced by another lemma sign as guiding element, the idioms remain treatment units and can be regarded as a special type of sublemma, cf. Gouws (to appear).

#### 7.2.2.5 Main lemmata and sublemmata

In par. 7.3. the emphasis is on different macrostructural ordering methods and a distinction is made between lemmata adhering to a straight alphabetical ordering and those ordered in a sinuous lemma file. This leads to a distinction between lemmata that are ordered vertically and those that are ordered horizontally, cf. the following example:

**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**´ance d picture, gebalanseerde (ewewigtige) voorstelling;  
**bal**´ance r, 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.

In this partial article stretch the lemmata display a straight alphabetical, i.e. a vertical ordering, cf. par. 7.3.6. The following example takes the same partial article stretch and changes the vertical ordering to a sinuous lemma file by moving all the complex words with *balance* as initial element to a horizontally ordered cluster, attached to the article of the main lemma *balance*. Horizontal ordering is primarily done for space-saving reasons and in this example a further space saving procedure by means of textual condensation has omitted the element *balance* in all the horizontally-ordered

lemmata, resulting in a cluster of partial lemmata in which the element *balance* has been substituted by a place-keeping symbol.

**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)

Vertically ordered macrostructural elements are main lemmata and within the category of horizontally ordered lemmata a distinction is made between main lemmata and sublemmata. Compare the lemmata in the previous example with those in the next example:

**bag**a´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)

In this example the horizontally-ordered lemmata have not been subjected to the process of textual condensation that omits their mutual initial element. The horizontally-ordered lemmata in this partial article stretch can be reached by following the alphabetical arrangement of lemmata. They are presented in their full form, i.e. not reduced to partial lemmata by means of a process of textual condensation substituting a stem by a place-keeping symbol, and function, like the vertically-ordered lemmata, as main lemmata. The lexical item *balance sheet*, lemmatised as ~ *sheet* in the cluster attached to the article of the lemma *balance* can only be reached via the main lemma *balance*. The lemma *balance* functions as a lemma in the niche entrance and gives access to the horizontally-ordered lemmata. These horizontally-ordered lemmata that can only be reached via the preceding vertically-ordered lemmata are known as sublemmata, cf. Wiegand (2002; 2002a; 2003a), Gouws (to appear).

All the vertically-ordered lemmata presented in full are main lemmata. Horizontally-ordered lemmata can be either main or sublemmata. Those horizontally-ordered lemmata that are given in full, i.e. without a lemma part omitted through a process of textual condensation, e.g. lemmata like *bagasiebewys*, *bagasieburo* and *bagasiewa* in the previous example, are main lemmata. Those horizontally-ordered lemmata presented as partial lemmata due to the application of a process of textual condensation that has led to the omission of a lemma part, e.g. lemmata like ~*bob*, ~*bridge* and ~*wheel* in the cluster attached to the article of the lemma sign *balance*, are regarded as sublemmata.

When planning and compiling a dictionary, lexicographers should not merely follow existing macrostructural patterns, like the use of main and sublemmata, without a critical look at the purpose and function of the specific dictionary and the reference skills of the target users. The use of two systems of ordering the lemmata, i.e. the vertical and the horizontal approach, and the distinction between two different types of lemmata, i.e. main and sublemmata, when using a horizontal ordering, makes the presentation of macrostructural elements and consequently the access to these elements and to the data presented as part of their treatment more complex.

Although the use of a sinuous lemma file, i.e. horizontal ordering, and the use of sublemmata, can play an important role in the space-saving endeavours of the lexicographer, the decision should be made from the perspective of the dictionary user and the typical situation of dictionary use. For the sake of the average dictionary user it is better to restrict the ordering to one system, i.e. a vertical ordering with all the lemmata presented in their full form as main lemmata.

### 7.2.3 Different types of articles

#### 7.2.3.1 *Single and complex articles*

In par. 8.3 the focus is on the article structure and the data categories in general monolingual and bilingual dictionaries, and it is indicated that all articles do not have a homogeneous structure or contents. The dictionary type but also the data relevant to a specific lexical item will have an influence on the treatment and consequently on the nature of the data accommodated in a given article. The data distribution structure of a dictionary also plays a decisive role when decisions regarding the article structure are taken. This leads to the distinction between *single articles* and *complex articles*.

The majority of articles in a dictionary will accommodate a set of search fields to which specific data types have been allocated. These search fields include e.g. a zone for items giving the pronunciation, for items giving morphological data, for items giving a paraphrase of meaning or the translation equivalents, etc. The data distribution structure will make provision for these search fields that have to accommodate the data types that occur typically in all the articles. These search fields need to be presented in a systematic and consistent way to lead to a search area structure, cf. Chapter 11. A user who is familiar with the contents of the user's guidelines text of a dictionary should be able to know what kind of data to expect in a dictionary article and also in which search field a given type of data will be found. Displaying a standardised structure and microstructural data categories that is characteristic of the default article of a given dictionary leads to these articles to be classified as *single articles*. A single article will always display at least an obligatory microstructure, cf. par. 8.5, but it can also display an extended obligatory microstructure.

The type of dictionary and the data distribution structure will determine the freedom lexicographers have to include more than the default data presentation in a given dictionary article. Ideally the data distribution should make it possible for lexicographers to include data over and beyond the required minimum or the treatment on offer in all default articles where they feel the user needs to get additional guidance or help. Yet again, lexical item specific features should determine whether the lemma in a specific article needs a more comprehensive treatment. Where this is needed, it will lead to a change in the article structure from a single to a complex article. A complex article is characterised by the inclusion of additional data categories or search fields or by a more comprehensive treatment of a specific aspect of the lexical item represented by the lemma sign. In this regard lexicographers may use article-internal inserted inner texts, cf. par. 8.5, or other procedures to convey the additional data. In the POD usage notes are often included and this convention leads to a complex article:

**aggravate** v. (-ting) 1 make worse or more serious. 2 annoy. □ **aggravation** n.  
[Latin *gravis* heavy]

**Usage** The use of *aggravate* in sense 2 is regarded by some people as incorrect, but it is common in informal use.

In a similar way *Basiswoordeboek* also uses usage notes, cf. the following note added to the article of the lemma *sign afbrand*:

**Afbrand** word meestal ten opsigte van geboue gebruik. 'n Motor wat deur vuur vernietig word, sal uitbrand.

(**Afbrand** (to burn down) is generally used to refer to buildings. If a car is destroyed by fire it will burn out.)

In NWSG the usage notes, given in a text block, are also marked by means of a small drawing of a bee – the same marker that is used in the accompanying text book to focus the attention of learners on something special:

<p><b>grootte</b> naamwoord ■</p> <p>Die grootte van iets is hoe groot dit is.          ⇒ <i>Wat is die grootte [size] van jou skoene?</i>          ⇒ <i>Sy weet nie wat die grootte [size] van die swembad is nie.</i></p> <div style="border: 1px solid black; padding: 5px; margin: 10px 0;">  <p>Pasop vir <b>grootte</b> en <b>grote</b>.          ⇒ <i>Daardie vis is 'n grote [big one].</i> ⇒ <i>Watter grootte [size] broek dra jy?</i></p> </div>
--

Complex articles can play an important role in giving a lexicographer the opportunity to present more data in an article than the prescribed minimum accommodated in default articles.

One of the types of lexical items that ideally qualify for a treatment in a complex article is the category of culture-bound items. The typical treatment on offer in the comment on semantics, cf. par. 8.3.1.2, makes provision for among others a paraphrase of meaning presented as a lexicographic definition. However, when it comes to culturally-bound items a more comprehensive treatment than a mere definition is often necessary. In this case the lexicographer may add a cultural note as an additional macrostructural category or the explanation of meaning may be expanded to include data on the cultural value of the given lexical item. In the *Longman Dictionary of English Language and Culture* the articles of the lemma signs *beer* and *nursing home* have the following respective cultural notes attached to it:

**CULTURAL NOTE** In the UK, people usually say 'beer' when they are talking about a brown or dark brown form of the drink, such as BITTER or STOUT. People in the US usually call this darker type of beer ALE or 'dark beer'. The clear, pale, yellow

CARBONATED (= with gas) form of the drink, which is simply called 'beer' in the US and in most other countries, is usually called LAGER in the UK.

CULTURAL NOTE In the US and UK it is common for people to put their parents in a nursing home when their parents are too old to take care of themselves.

As discussed in par. 6.5 Bergenholtz, Tarp & Wiegand (1999) make a distinction between single articles and *synopsis articles*. A synopsis article includes the typical data presented in a single article but it goes further by also presenting additional data, often of a more encyclopedic or general nature. According to Bergenholtz, Tarp & Wiegand (1999:1780) synopsis articles with encyclopedic data in dictionaries dealing with languages for special purposes do not only include data relevant to the lexicographic treatment of the lexical item represented by the lemma sign of the specific article but they also include data relevant to lemma signs of some single articles in the specific dictionary.

In general dictionaries synopsis articles often present the treatment of a lemma sign representing a lexical item functioning as superordinate of a semantic field, with the different hyponyms of the semantic field represented by lemma signs of other single articles in the dictionary. The synopsis article includes data, which is also relevant to the single articles, and these single articles contain a cross-reference entry guiding the user to the relevant synopsis article. In a dictionary which contains synopsis articles an optimal awareness of and access to these articles could be achieved by employing an outer text which contains an alphabetical list of all those lemmata featuring as guiding elements of synopsis articles. Such a list will assist the user to utilise the data distribution of the dictionary to the full. In the comment on semantics of a synopsis article one typically finds a paraphrase of meaning that has a general character and includes the meaning of a number of other lexical items, cf. the treatment offered in the articles of the lemma signs *acid* and *sulphuric acid* in the POD:

**acid** —*n.* 1 a any of a class of substances that liberate hydrogen ions in water, are usu. sour and corrosive, turn litmus red, and have a pH of less than 7...

**sulphuric acid** *n.* dense oily highly corrosive acid.

A number of features given in the paraphrase of meaning of the first sense of *acid*, e.g. *liberate hydrogen ions in water, are usu. sour and corrosive, turn litmus red, and have a pH of less than 7*, also apply to *sulphuric acid* and have not been repeated there.

Synopsis articles are frequently used in dictionaries dealing with languages for special purposes but they are also relevant for general dictionaries. Due to the additional data on offer in a synopsis article they can be regarded as a subtype of complex articles.

### 7.2.3.2 Cross-reference articles

Single articles display at least an obligatory microstructure, cf. par. 8.5. In the following example from the POD the lemma sign represents the orthography, the part of speech is indicated along with a brief paraphrase of meaning:

**lampshade** *n.* translucent cover for a lamp.

Dictionary articles often contain a restricted treatment with less data categories on offer than the minimum required to qualify the article as displaying an obligatory microstructure, cf. the following example from the POD;

**woodman** *n.* forester.

Compared to other articles in this dictionary the article introduced by the lemma sign *woodman* offers a restricted treatment of the lemma sign, cf. Gouws (2003a). This kind of limited treatment occurs when the lemma is a lesser used member of a synonym group and the treatment is primarily directed at a cross-reference entry, guiding the user to the lemma which represents the synonym with a higher usage frequency. It also occurs e.g. where the user is cross-referred to a lemma representing a spelling variant or a plural/female/ form of the lexical item represented by the guiding element of the article with the limited treatment, cf. the following examples from the *POD*:

**pollock** var. of \*pollack.

**women** pl. of \*woman.

**godmother** *n.* female godparent.

These articles displaying a restricted treatment with a cross-reference entry often being the most salient entry in the article are known as *cross-reference articles*. Derivations are often included as lemmata and their treatment primarily consists of items giving the specific morphological data and a cross-reference entry guiding the user to an article where a comprehensive treatment of the simplex is provided, cf. the following examples from the *NSDN*:

**thekollano**, (n-)/di- v. REKA

**thekollelano**, (n-)/di- v. REKA (NSDN)

Some cross-reference articles contain only the lemma sign and the cross-reference entry, cf. the following example from the *NEN*:

**molelo**, see **mollo**. (NEN)

In the following example from *NeW* a subarticle, cf. par. 7.2.3.4, with the lemma part ~ *ring* as guiding element functions as a cross-reference article with only a place keeping symbol, a lemma part and the cross-reference entry in the article:

**growth** *n.*: ~ *ring* → **annual ring**.

Various aspects of cross-referencing are discussed in Chapter 12.

### 7.2.3.3 *Niched and nested articles*

The classification of dictionary articles as *single articles*, *complex articles*, *synopsis articles* or *cross-reference articles* is done on the basis of microstructural criteria. The nature and extent of the microstructure of an article determines its being classified as one of the above-mentioned article types. Article typology can also be determined on a macrostructural basis and this leads to the distinction between *niched articles*

and *nested articles*. In the discussion of the ordering of lemmata in a dictionary, par. 7.4, a distinction is made between a straight alphabetical ordering where the lemmata are presented in a vertical ordering and a sinuous lemma file where the lemmata are presented in a horizontal ordering, cf. the following examples, also given in par. 7.2.2.5, that illustrate these two types of ordering:

**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`ance d picture**, gebalanseerde (ewewigtige) voorstelling;  
**bal`ance r**, 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.

**bagas`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)

The lemmata *balalaika* – *balancing-pole* are ordered vertically whereas the lemmata *bagasiebewys* – *bagasiewa* are ordered horizontally. Each dictionary article contains at least a lemma sign. Each lemma, whether ordered vertically or horizontally, is the guiding element of an article. Horizontally-ordered lemmata imply horizontally ordered articles and these clustered articles can be divided into two types, i.e. *niched articles* and *nested articles*. A niched article functions within a cluster of niched articles, i.e. articles with niched lemmata as guiding elements. Although these articles are presented in a horizontal ordering their ordering is done along strict alphabetical lines, as seen in the above-mentioned article where all the horizontally-ordered articles, introduced by the lemmata *bagasiebewys* – *bagasiewa*, display a strict alphabetical ordering.

The following cluster of horizontally-ordered articles deviates from a strict alphabetical ordering:

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

Within this cluster of horizontally-ordered articles different subgroupings can be identified. Within each subgrouping an alphabetical ordering prevails but the cluster as such does not display a strict alphabetical ordering. The articles in this cluster are *nested articles*. A nested article is an article functioning within a cluster of nested articles, i.e. a cluster which displays a deviation from the strict alphabetical ordering, cf. par. 7.4 for a more detailed discussion of niching and nesting.

#### 7.2.3.4 *Main articles and subarticles*

In par. 7.2.2.5 a distinction is made between main lemmata and sublemmata. This distinction has direct implications for the classification of articles. An article with a main lemma as a guiding element is known as a *main article* whereas an article with a sublemma as guiding element is known as a *subarticle*. Although both main and sublemmata are macrostructural elements that deserve an equal microstructural treatment, many dictionaries offer an inferior or at least a restricted treatment of sublemmata. As a result subarticles often display a less comprehensive lexicographic treatment compared to articles with a main lemma as guiding element.

#### 7.2.4 **A well-informed decision**

Lexicographers have to be well-aware of the different types of lemmata and the different types of articles. When planning a dictionary, ample consideration should be given to the types of lemmata and types of articles that will suit a specific dictionary the best in terms of its typology, its target user and its situations of usage. These decisions may not be taken in a random way and should also be influenced by the functions identified for the specific dictionary. Easy access to the lemmata and an optimal retrieval of information from the articles still have to be guiding principles in the decision-making process.

### 7.3 **The ordering of lemmata**

#### 7.3.1 **General remarks**

Users of dictionaries often have an uncritical approach towards the ordering of the lemmata in a dictionary. They are satisfied to find a lemma in the section of the dictionary where they are looking for it and where they expect to find it. From a metalexigraphic perspective the ordering of lemmata is not as self-evident as regarded by the dictionary user. During the planning of a dictionary and in the formulation of a dictionary conceptualisation plan the lexicographers have to pay attention to this matter and they have to decide on a number of relevant issues. In the following paragraphs some of these issues will be discussed quite briefly.

#### 7.3.2 **Alphabetical or thematic ordering**

The history of lexicography gives evidence of two major ordering principles that have been employed in dictionaries, i.e. alphabetical and thematic ordering. Many of the early dictionaries displayed a thematic ordering whilst alphabetical ordering predominantly prevails in modern-day dictionaries. Once again the target user and the typology of the intended dictionary play a decisive role in determining the type of ordering to be employed in a specific lexicographic product. Today procedures of thematic ordering are still applicable in certain dictionary types, e.g. in thesauri. The use of an alphabetical ordering is the preferred procedure in general descriptive and bilingual dictionaries. However, even in dictionaries with an alphabetical ordering, the educational value of thematic ordering may never be underestimated.

The frame structure of dictionaries, cf. Chapter 6, allows the inclusion of an unlimited number of outer texts in the front and back matter sections of a dictionary. Especially in the back matter, registers are often included as outer texts. These registers can present a selection of lexical items from various specialised or

other fields. Quite often lexical items from a specific semantic field are grouped together in a register included in the back matter component. These registers do not necessarily have to display the same ordering principle as the central list of the dictionary. The descriptive monolingual German learner's dictionary *Wörterbuch Deutsch als Fremdsprache* (WDF) has an alphabetical ordering in the central list. One of the back matter texts with the title *Wortfelder* (word fields) presents a semantic classification of the lemmata included in the central list by identifying a range of semantic fields and ordering the lemmata from the central list in these semantic fields. This thematic ordering in a dictionary with a predominant alphabetical ordering has a functional value. Depending on the needs of the target users of a dictionary and the data distribution structure lexicographers should consider using the thematic ordering in certain outer texts of their dictionaries.

The remainder of this section will primarily focus on different aspects of the ordering of lemmata in the central list of alphabetically-ordered dictionaries.

### 7.3.3 Strict alphabetical ordering in an access alphabet

The most typical ordering system in modern-day dictionaries is known as the strict alphabetical ordering. This implies that lemmata are ordered according to their first and subsequent letters, i.e. a full alphabetical ordering. When employing a strict alphabetical ordering a lexicographer has to take cognisance of a number of issues that could have an effect on the positioning of certain lemmata in the relevant article stretches.

Bergenholtz (1990) presents a lexicographic "instruction book" in which he deals with a whole range of practical problems confronting the lexicographer. These are problems that need to be addressed in the planning phase of a dictionary and the decisions have to be included in the dictionary conceptualisation plan. One of the issues discussed by Bergenholtz regards the ordering of lemmata in a dictionary with a strict alphabetical ordering system. Bergenholtz identifies a number of problematic areas for which the lexicographer has to find solutions. Some of the issues to be negotiated by the lexicographers are the following: the influence of diacritic signs on alphabetisation, the ordering of multiword lemmata, the influence of a hyphen (in a lemma where the hyphen is a place-keeping symbol, indicating the status of the lexical item as a stem or an affix) and the ordering of lemmata differing only in terms of capital letters versus lower case letters in their initial positions. The ordering of lemmata with a symbol or number as component should also be clarified during the dictionary-planning phase.

Nielsen (1995:190) makes some very valid remarks with regard to the alphabetical ordering. According to him a distinction has to be made between the alphabet (as it is used in everyday language, referring to a set of letters arranged in a particular order) and the access alphabet, "which is the alphabetic arrangement principle used in a particular dictionary". This access alphabet is not necessarily identical with *alphabet* as it is used in its everyday sense. Nielsen argues that the access alphabet may contain additional letters from other alphabets, hyphens, numbers, etc. All graphemes used in writing are potential lemma constituents and the arrangement or ordering in a dictionary has to make provision for that. If a lexicographer decides to employ an ordering system based on the use of an access alphabet which differs from the traditional alphabet it does imply certain deviations from a traditional

strict alphabetical ordering. The term *strict alphabetical ordering* still applies but it has to be interpreted in terms of the access alphabet and not in terms of the traditional alphabet.

Lexicographers should be aware of the fact that Nielsen's use of the term *access alphabet* supports that of Wiegand (1989a) but differs from the way in which Wiegand (1991) uses it. In the following example from TFW the capital G introduces the beginning of a new article stretch.

G

**gab** *have the gift of the gab* § glad wees met die bek★/mond, 'n gladdie bek★/  
mond hē.

**gad**<sup>1</sup> [v.] *gad about/around* rondjakkter, rondrits, rinkink.

Wiegand (1991) regards this alphabet letter placed at the beginning of a new article stretch to introduce that article stretch, as an article external non-typographic structural indicator which is an immediate text constituent of the central list. According to him the set of all the structural indicators of this kind represents the access alphabet of the dictionary.

#### 7.3.4 Guidelines for an access alphabet

Following the idea of an access alphabet, as propagated by Nielsen (1995), and the criteria given in Bergenholtz (1990), lexicographers should formulate guidelines for the arrangement of lemmata in their specific dictionary according to a strict alphabetical ordering on the basis of the access alphabet. Although an access alphabet should be devised for each dictionary according to the needs of the object language, lexicographers could try to follow some general guidelines that are applicable to all dictionaries. These guidelines will have to be stated in the front matter texts of the dictionary to assist the user in interpreting the ordering of lemmata in that dictionary in an unproblematic way.

A typical point of departure in the formulation of rules for an access alphabet could be that the unmarked form should always precede the marked form. This would imply, e.g., that a letter without a diacritic sign precedes a letter with a diacritic sign. Such an approach is in line with Bergenholtz (1990) as well as Wiegand (1989a:377). It would also imply that lower case letters precede capital letters and that an unhyphenated form precedes a hyphenated form.

Lower case letters are regarded as the unmarked form and they precede the marked capital letters. Consequently the English lemma sign *jersey* (piece of clothing) should precede the form *Jersey* (light brown cow). The specific ordering of these lemmata is not that important. Unmarked may just as well follow marked. What is important is that the application of this ordering may not be done in an arbitrary way but has to be done in a consistent way. Therefore it is necessary for every dictionary plan to include a section in which the ordering practice is described.

Where a lemma has a hyphen functioning as a place-keeping symbol, cf. Wolski (1989a:365), to indicate that the lemma sign is a stem or an affix, this hyphen does not have an own value within the access alphabet. Where, however, such a hyphen is the only difference between two lemma signs, the hyphenated lemma follows the

unmarked form, cf. the following example from HAT where the lexical lemma *binne* precedes the sublexical lemma *binne-*:

**bin<sup>1</sup>ne<sup>4</sup>** vs.

**1** Aan of na die binnekant van: *Binne die skool, dorp.* **2** Korter of nie langer nie as die genoemde tyd: *Binne 'n dag, uur.* b.nw. en bw. Deur grense ingesluit: *Die kinders is al binne. Die boot is binne (die hawe).* UITDR.: *Jou iets te binne bring, roep, in die gedagte, gemoed. Van binne en buite ken, deur en deur. Dit val, skiet my te binne, ek onthou dit.*

**bin<sup>1</sup>ne<sup>2</sup>**

As eerste lid van wwe.: *binnegaan, -kom*; of van s.nwe.: *binnesak, -huis, -kant.*

Lemmata with a hyphen in the final position, the so-called post-hyphenated lemmata, precede lemmata with a hyphen in the initial position, the so-called pre-hyphenated lemmata, cf. Gouws (1989). In HAT this principle is not applied consistently but in the following partial article stretch the ordering goes from unmarked ( $a^2$ ) to marked ( $\grave{a}^3$ ) to post-hyphenated ( $a^{-4}$ ) and to a pre-hyphenated ( $-a^5$ ) lemma

**a<sup>2</sup> tw.**

**1** Uitroep van vreugde, genot: *A, nou verstaan ek alles, nou het ek jou!* **2** Uitroep van ergernis, ontkenning, gebruik met *nee* of *ag nee*: *A nee a, nou word ek kwaad. Ag nee a, nou maak jy my vies.*

**\grave{a}<sup>3</sup> vs.**

**1** Tot: *8 à 10 m.* **2** Teen: *R500 à 7 persent belê, 10 cm à 20 sent koop.*

**a<sup>-4</sup> voorv.** (*an-* voor vokale)

Aanduiding van neutraliteit, houdingloosheid, en by uitbreiding van die teendeel of afwesigheid; anti-, on-, bv. in *agodsdienstig, anasionaal, amoreel, asosiaal, anorganies.*

**-a<sup>5</sup> agterv.**

Meestal by verkorte manlike persoonsname, ter aanduiding van vertroulikheid, e.d., soos in *Benna, Pieta, boeta.*

The general use of hyphens should not disturb the ordering system of a dictionary. An English word like *chicken-hearted* should be ordered according to the alphabetical value of its components and placed between say *chickenfeed* and *chickenpox*. In a case like this the hyphen should not be regarded as a separate access alphabet sign. Once again if the hyphen is the only difference between two words, the form without a hyphen should precede the hyphenated form.

Multiword lemmata are treated in a similar way as single word lemmata. In English the lemma *livery company*, consisting of two words, is ordered between the two single word lemmata *livery* and *liveryman* because the two-word lemma fits in this alphabetical slot. Where the space between two words in a lemma sign has a differentiating value, the multiword lemma should follow the single word lemma. Where there is doubt, the space can be regarded as the first sign of the access alphabet.

The access alphabet makes provision for an ordering of lemmata where the lemma sign is combined with a numerical marker to identify a lemma as representing a lexical item that is a member of a homonym group. The strict alphabetical ordering is not enough and additional ordering criteria are needed. Homonyms are not ordered in an arbitrary way but in a way clearly formulated in the dictionary conceptualisation plan. In the field of lexical semantics homonyms are regarded as lexical items with the same form, spelling and pronunciation but with unrelated meanings. English has two lexical items *arm* which both belong to the same part of speech category (noun) but with unrelated meanings. They are included in the POD as homonyms and presented as separate lemmata, distinguished by the use of a numerical superscript marker:

**arm<sup>1</sup>** *n.* 1 upper limb of the human body from shoulder to hand. 2 forelimb or tentacle of an animal. 3 a sleeve of a garment. b arm support of a chair etc. c thing branching from a main stem (*an arm of the sea*). d control, means of reaching (*arm of the law*). ... [Old English] (POD)

**arm<sup>2</sup>** *—n.* 1 (usu. in *pl.*) weapon. 2 (in *pl.*) military profession. 3 branch of the military (e.g. infantry, cavalry). 4 (in *pl.*) heraldic devices (*coat of arms*). *—v.* 1 supply, or equip oneself, with weapons etc., esp. in preparation for war. 2 make (a bomb etc.) ready. ... [Latin *arma* arms] (POD)

The relative ordering of the members of a homonym pair depends on various principles dominating the ordering in the specific dictionary. In a dictionary which is based on historical principles the ordering of homonyms should display the historical ordering and the lemma sign representing the oldest lexical item will be presented as the first homonym. In general synchronic dictionaries that focus on the language as it is currently used, the ordering is usually determined by the usage frequency of the words. The word with the highest usage frequency will be presented as the first homonym.

Some dictionaries work with an alternative approach to the classification of homonyms, i.e. the principle of grammatical homonyms. Differences in the grammatical category of a single lexical item, i.e. where one lexical item can be used in more than one part of speech function, motivates the presentation of homonyms. A typical pattern then is that a noun should precede a verb, cf. the following examples from the VAW. The Afrikaans word *skil* (peel) functions as a noun (peel) and as a verb (to peel) and these different functions motivate the inclusion of two homonym lemmata:

**skil<sup>1</sup>**, (s), **-le.** 1. Omkleedsel, bas. 2. Buitenste laag van 'n vrug. 3. Vlies oor die oog, dop; *die skille het van sy oë geval*, 'n skilbedekking het van die oë geval; hy het begin verstaan. (VAW)

**skil<sup>2</sup>**, (w), **ge-** 1. Die omkleedsel of die bas verwyder. 2. Die skil van 'n vrug met 'n mes wegsny. (VAW)

A problem arises for the user when a dictionary uses this approach but also follows an approach to classify different lexical items with the same form and spelling but unrelated meanings albeit that they belong to the same grammatical category as

homonyms. This happens in the VAW, as is seen in the following examples where the Afrikaans homonyms *koraal* which are both nouns are presented as separate lemmata:

**koraal**<sup>1</sup>, **..rale**. **1.** Geraamte van in kolonies lewende seepoliepe wat kalk afskei. **2.** Rooi kalk van seepoliepe. **3.** Glasballetjie met 'n gaatjie deur om aan 'n toutjie te ryg, soort kraal; **koraalvormend**; **koraalvormig**. (VAW)

**koraal**<sup>2</sup>. **1.** Liturgiese kerkgesang. **2.** Kerklike koorsang. **3.** Psalm of gesang, eenstemmig gesing met meerstemmige begeleiding. **4.** Die melodie daarvoor. **5.** Koorsanger, koorknaap; **..ra**<sup>1</sup>lies. (VAW)

The front matter text presenting the user guidelines should give a clear indication of the criteria according to which homonyms are identified and the order in which they are lemmatised.

Different criteria are used in the presentation of lemmata as homonym pairs and this is to the detriment of the user. Further confusion for the user follows where both types of the classification of homonyms are used with regard to one set of lemmata presented as homonyms, as seen in the following partial article stretch from the VAW where the word *grou* is lemmatised as five homonyms. The first and third lemmata are semantically related, being different part of speech functions of one lexical item, as is the case with the second and fourth lemmata. The fifth lemma, a verb, is a true semantic homonym. The user does not know whether this lemma participates in a relation of semantic or grammatical homonymy with the preceding lemmata and whether its fifth place in the paradigm of homonyms is merely determined by its grammatical category:

**grou**<sup>1</sup>, (s). **1.** Grou kleur, grouheid. **2.** Skemering. **3.** Laere volksmassa.

**grou**<sup>2</sup>, (s). Bitsige, onvriendelike, skerp woord.

**grou**<sup>3</sup>, (b), **-er**, **-ste**. Grys(kleurig), askleurig; **grouerig**; **grouheid**.

**grou**<sup>4</sup>, (w), **ge-**. Knor, brom, grom.

**grou**<sup>5</sup>, (w), **ge-**, (gewest.). Grawe. (VAW)

The ordering of lemmata within a paradigm of homonyms should be done according to one classification principle and this principle needs to be applied in a consistent way.

### 7.3.5 Deviating from a strict alphabetical ordering in an access alphabet

In a Sesotho sa Leboa dictionary the application of a rule stating that the unmarked form precedes the marked form would result in an ordering where the letter *s* precedes the *š*. The rigid application of a strict alphabetical ordering (on the basis of an access alphabet) could still cause some problems for the user in his/her attempts to find a given lemma sign. In this regard the treatment of *s* versus *š* in Sesotho sa Leboa dictionaries seems to be quite problematic. Ziervogel and Mokgokong (NSDN) handle *s* versus *š* as separate categories by entering *s* first, working through the alphabet on the letters following *s* i.e. *sa*, ... *se*, ... *si* ... before commencing with *ša*, ... *še*, ... *ši*. The way in which NSDN arranges these lemmata can be justified because a rigid application of a strict alphabetical ordering is employed on the basis of the relevant

access alphabet. However, from the perspective of a user looking for a given lemma sign this approach could be quite frustrating.

A less rigid application of the strict alphabetical ordering results in a slightly different arrangement where all lemmata starting with *s* do not precede all lemmata with *š*. Nothing only precedes something when the lemmata in which the marked and unmarked letters occur are identical in every other respect. In Sesotho sa Leboa it would mean that *sa* will precede *ša* and *si* will precede *ši*. But it does not imply that *ša* will only be entered after all the lemmata starting with *s*, e.g. *sa*, *se*, *si*. The ruling only applies where the two forms, i.e. the one without and the one with the diacritic sign, are in direct opposition.

In the descriptive monolingual Afrikaans dictionary HAT the same, less rigid principle applies. The Afrikaans lemmata *se*, *sê* and *se-* follow each other in this particular order. The motivation is that the unmarked form (*se*) precedes the marked forms (*sê* and *se-*) and in an opposition of marked forms the form with the diacritic sign (*sê*) precedes the form with a hyphen (*se-*).

**se**<sup>1</sup> vnw.

1 Van, behorende aan: *Ma se jas. ...*

**sê**<sup>2</sup> ww. (gesê)

1 In woorde uitspreek, mondelings te kenne gee: *Hy sê nee. Mōre sê ...*

**se**<sup>-3</sup> (L.)

Voorvoegsel met die bet. "apart, sonder"

It is important to note that the article of the lemma sign *sê* is not ordered after all the articles of lemma signs starting with *se*, like *sebra*, *sedan*, *see*, etc. Within the strict alphabetical order provision is made for the occurrence of marked forms.

The use of an access alphabet demands a consistent application of the formulated principles. As mentioned earlier the compilers of NSDN work through the alphabet on the letters following *s* i.e. *sa*, ... *se*, ... *si* ... before commencing with *ša*, ... *še*, ... *ši* ... As a result of such an approach *se* is listed almost three pages apart from *še*. The compilers of this dictionary, however, enter *ê* alphabetically directly following *e*, and *ô* directly after *o*, etc. From a lexicographic perspective, there is no justification for such a differentiation between the handling of *e*, *ê*, *o* and *ô* on the one hand and *s* versus *š* on the other. The lexicographer should treat *s* versus *š* exactly like *o* versus *ô* and *e* versus *ê*. Such an approach will contribute towards user-friendliness in the sense that *še* will directly follow *se* in the dictionary. Van Wyk rightfully disregarded this separation in his revision of the 1989 edition of *Pukuntšu*.

For Sesotho sa Leboa it means that clarity in respect of alphabetical ordering for uppercase and lowercase *S/s* versus *Š/š* in combination with *ê/e* and *ô/o* should be clearly determined. The recommended ordering is as follows for this paradigm:

*se* → *Se* → *sê* → *Sê* → *še* → *Še* → *šê* → *Šê*

Rule 1: *lowercase* precedes *uppercase*;

<b>s</b> → <b>S</b>	and	<b>š</b> → <b>Š</b>		
<i>se</i> , <i>Se</i>	<i>sê</i> , <i>Sê</i>	<i>še</i> , <i>Še</i>	<i>šê</i> , <i>Šê</i>	Rule 1

Rule 2: *without circumflex precedes with circumflex*

e → ê	and	š → Š		
se, Se	sê, Sê	še, Še	šê, Šê	Rule 2

The lexicographer has to be careful not to stretch the use of an access alphabet which differs from the ordinary alphabet too far. In their access alphabet Ziervogel and Mokgokong deviate from an ordinary alphabetical sorting of the entries by utilising a phonemic one, namely: *A, B, BĴ, D, E, F, FS, FŠ, G, H, HL, I, J, K, KG, KH, L, M, N, NG, NX, NY, O, P, PH*, etc., because this is in their opinion 'more scientific.' Detailed criticism of this extremely user-unfriendly arrangement in which, for example, in the article stretch of the letter *B* an entry like *bu* 'alphabetically' precedes *bj*, will not be given here. In principle the golden rule should be to keep the access alphabet as close to the ordinary alphabet as possible.

In any language certain candidates for the lemma list of a dictionary confront the lexicographer with lemmatisation and ordering problems. This is due to the fact that their orthographical form includes signs which are not part of the ordinary alphabet and for which an access alphabet will have to make special provision. Typical examples are lexical items like the following: *4x4 vehicle, α-particles, 3-D pictures, .303 rifle, @* (in an e-mail address), etc. If the dictionary is to include the items as lemma signs a system has to be devised to ensure a systematic and predictable positioning of this type of lemma. Once again lexicographers should be aware of this problem and solutions have to be included in the dictionary conceptualisation plan. Fortunately one can look at the way in which existing dictionaries order these items in their macrostructure. One convention which is applied quite often is that a number is given the value it would have had if it was presented in a written form, e.g. *4x4* is interpreted and ordered as if it is *four by four*, cf. Nielsen (1995:191). This would imply that *4x4 vehicle*, in this form, would be entered in the article stretch where the lemma *four* occurs. In an example like *.303 rifle* the period (.) is ignored by the access alphabet and the lemma is entered as *.303 rifle* in the article stretch where the lexical item *three* appears. This kind of lemmatisation represents a deviation from the system of a strict alphabetical ordering. However, if the lexicographer conveys this to the users in the text containing the user's guidelines and applies it consistently, it will enhance the user-friendliness of the dictionary.

The successful retrieval of information in a dictionary often depends on an unimpeded access to the needed lemma-sign. Consequently lexicographic planning should include a clear-cut decision regarding the macrostructural ordering and presentation in the dictionary. The arrangement of the lemmata is of primary importance. The ordering of lemmata in both descriptive monolingual and bilingual dictionaries is something too often taken for granted by both lexicographer and dictionary user.

### 7.3.6 Employing a sinuous lemma file

General descriptive and bilingual dictionaries typically display a strict alphabetical ordering in which the use of the traditional alphabet is complemented by the use of an access alphabet which may allow certain deviations from the strict alphabetical order. The use of a strict alphabetical ordering, also when the scope is extended by means of a more comprehensive access alphabet, does not necessarily imply the use

of a straight alphabetical ordering. A straight alphabetical ordering results in what is generally known as a *vertical lemma file*. The lemmata are all ordered in a vertical line with each lemma sign appearing as guiding element of a new article on a new line, as seen in the following example from the POD:

**him** *pron.* 1 objective case of \*he (I saw him). 2 *colloq.* he (it's him again; taller than him). [Old English, dative of \*he]  
**himself** *pron.* 1 a *emphat.* form of \*he or \*him (he himself will do it). b refl. form of \*him (he has hurt himself). 2 in his normal state of body or mind (does not feel quite himself today). □ be himself see \*oneself. by himself see by oneself. [Old English: related to \*him, \*self]  
**hind**<sup>1</sup> *adj.* at the back (hind leg). [Old English *hindan* from behind]  
**hind**<sup>2</sup> *n.* female (esp. red) deer, esp. in and after the third year. [Old English]  
**him** *pron.* 1 objective case of \*he (I saw him). 2 *colloq.* he (it's him again; taller than him). [Old English, dative of \*he]  
**hinder**<sup>1</sup> *v.* impede; delay. [Old English]  
**hinder**<sup>2</sup> *adj.* rear, hind (the hinder part). [Old English]  
**Hindi** *n.* 1 group of spoken dialects of N. India. 2 literary form of Hindustani, an official language of India. [Urdu *Hind India*]  
**hindmost** *adj.* furthest behind.  
**hindquarters** *n.pl.* hind legs and rump of a quadruped.  
**hindrance** *n.* 1 hindering; being hindered. 2 thing that hinders.  
**hindsight** *n.* wisdom after the event.  
**Hindu** *—n.* (*pl.* -s) follower of Hinduism. —*adj.* of Hindus or Hinduism. [Urdu *Hind India*]  
**Hinduism** *n.* main religious and social system of India, including the belief in reincarnation, several gods, and a caste system. *v.* impede; delay. [Old English]

This vertical ordering of lemmata is a typical application of straight alphabetical ordering and is predominantly used in dictionaries of a lesser macrostructural extent, e.g. school and desk dictionaries. It is a user-friendly way of presenting the lemmata, it does not constitute a high degree of textual condensation and it does not confront the user with demanding dictionary consultation procedures, cf. Wiegand (1989a).

In dictionaries with a bigger macrostructural selection lexicographers are trying to perform different space saving procedures in order to accommodate as many macrostructural entries as possible in the dictionary. One of the most commonly used strategies is directed at the macrostructure with the lexicographer endeavouring to combine the vertical ordering with a horizontal ordering by also employing a sinuous lemma file, cf. the following examples from GW and VAW respectively:

**Eas**ter, Pase, Paasfees; Paasvakansie; ~ **Day**, Paasdag; ~ **egg**, Paaseier; ~ **holidays**, Paasvakansie; ~ **lily**, Maartlelie, misryblom; belladonnalelie. (GW)

**lift**ing, hef= ~ **bracket**, hefarm; ~ **bridge**, hefbrug; ~ **crane**, hyskraan; ~-**hook**, hefhaak; ~ **jack**, domkrag; ~ **power**, hefvermoë; stygkrag; ~ **pump**, hefpomp; ~ **screw**, uitnemer; ~ **truck**, hefwa; ~ **wire**, hefhaak. (GW)

**ei**er, -s. **1.** Wat gelê word deur 'n hoender, voël, ens. **2.** Vroulike saadseel. **3.** Eievormige voorwerp. **4.** Nul, geen punte; *op 'n eier BROEI*, besig wees om 'n plan

te beraam; *altyd 'n eiertjie* BYLÊ, altyd in die rede val om iets te sê; 'n HALWE eier is beter as 'n leë dop, iets gerings is beter as totaal niks; *moenie al jou eiers onder een HEN sit nie*, moenie al jou geld in een onderneming belê nie; *op eiers LOOP*, baie stadig loop; *die eier wil slimmer wees as die HEN*, maak of hy baie slimmer is as die persoon wat meer geleentheid gehad het om ondervinding op te doen; **eierdop; eiergeel; eierkissie; eierklitser; eierlaai; eierlegging; eiermandjie; eiersel; eiersnyer; eiersous; eieruitstoting; eierversameling; eierwit.** (VAW)

The examples starting with the lemma signs *Easter* and *lifting* are characterised by the results of a procedure of textual condensation which is a commonly used space-saving procedure, cf. Hausmann & Wiegand (1989:336) and Wolski (1989; 1991), but the use of this procedure has additional implications. Textual condensation in these examples can be seen in various ways. It leads to the lemma signs following *Easter* and *lifting* to be ordered horizontally. It also reduces the form of the lemma by omitting the part represented by the preceding main lemma, i.e. *Easter* and *lifting* in these examples. The position where this part of the lemma has been omitted is filled by a substitute, the place-keeping symbol, a tilde (~). This results in a cluster of partial lemmata. In the examples from GW the main lemma, i.e. the lemma positioned in the vertical ordering and given in its full form, functions as an entrance element to the cluster containing the horizontally-ordered lemmata given in part and therefore functioning as sublemmata. These sublemmata can only be accessed via the main lemma. There is no way to find the macrostructural element *lifting-hook*, presented in this dictionary as *~-hook*, without going via the main lemma *lifting* to enter the cluster of horizontally-ordered lemmata, presented as a stretch of partial lemmata. Access to these partial lemmata and the retrieval of the relevant information depends on the co-occurrence of the preceding main lemma, functioning as entrance lemma. Although these sublemmata need the main lemma for their macrostructural realisation they remain fully-fledged macrostructural elements of the dictionary and they are the guiding elements of their own articles, albeit that these articles are subarticles. It is important to note that these horizontally ordered lemmata are not part of the article of the lemma signs *Easter* and *lifting*. Each lemma sign introduces its own article, a horizontally ordered article.

The example taken from VAW displays a lower degree of textual condensation because the horizontally ordered lemmata have not been reduced in form to render partial lemmata consisting of a place-keeping symbol and a lemma part. The horizontally ordered articles have a simplified article structure, consisting only of the lemma sign, but these lemmata are presented as full lemmata.

The use of horizontally ordered lemmata leads to the clustering together of a number of articles into an article cluster or text block. Such a text block contains a cluster of lemmata. The procedure of horizontal ordering is primarily employed as a space-saving device. The lexicographer has to make sure that this procedure does not impede or diminish the intended information transfer of the given dictionary. Once again this method of ordering may only be employed if the dictionary using skills of the target users are sophisticated enough to ensure a successful retrieval of all the necessary information. The dictionary typology should also play a decisive role. Pedagogical and desk dictionaries should refrain from using this method because the degree of textual condensation is too high.

The examples from the bilingual dictionary GW and the descriptive monolingual dictionary VAW illustrate different applications of the procedures of horizontal ordering. In GW the lemma signs remain the treatment units at which translation equivalents are addressed. In VAW no other entries but the lemma signs are included in the articles. This is typical of the use of a sinuous lemma file in many monolingual descriptive dictionaries. In descriptive monolingual dictionaries sinuous lemma files are often used to present complex lexical items, i.e. compounds and derivations. The lexical item represented by the vertically ordered lemma sign immediately preceding the first horizontally ordered lemma, in the quoted example from VAW it would be the lemma sign *eier*, functions as a stem, usually the first stem, in all the horizontally ordered lemmata, e.g. *eierdop*; *eiergeel*; *eierkissie*; *eierklitser*; *eierlaai*, etc. As a space-saving strategy the lexicographer decides to include a selection of complex items as part of a sinuous lemma file. These items are included as so-called self-explanatory or unexplained lemmata. This clustering implies that the lexicographer regards these lemmata as being semantically transparent and therefore not in need of an explanation. Semantic transparency implies that the meaning of the complex can be unambiguously deduced from its components. The use of a procedure of the horizontal ordering of lemmata in a monolingual dictionary where the lemma signs are the only elements in the clustered articles is only valid if they are indeed semantically transparent. Quite often dictionaries include complex lexical items as unexplained lemmata but these lemmata do not display semantic transparency. With regard to these so-called *self-explanatory complex items* Philip Gove, editor of the *Webster's Third New International Dictionary*, remarked that the *self* in *self-explanatory* should refer to the intended interpreter of the word and not to the word itself (nor the lexicographer), cf. Gove (1966:184).

However, the unexplained compounds in the above-mentioned partial article stretch are not the only complex lexical items with *eier* as a first stem to be included in the macrostructure of this dictionary, cf. the following example:

- ei'er, -s.** 1. Wat gelê word deur 'n hoender, voël, ens. 2. Vroulike saadseel. 3. Eievormige voorwerp..... **eierdop**; **eiergeel**; **eierkissie**; **eierklitser**; **eierlaai**; **eierlegging**; **eiermandjie**; **eiersel**; **eiersnyer**; **eiersous**; **eieruitstoting**; **eiersameling**; **eierwit**.
- ei'erboer, -e.** lem. wat veral vir die eieropbrengs met hoenders boer.
- ei'erboor, ..bore.** Werktuig van bepaalde insekte om 'n gaatjie, bv. in vrugte, te maak, of sprinkane in die grond, vir die lê van hulle eiers.
- ei'erbrandewyn.** Brandewyn waarin 'n eier geklits is.
- ei'erdans, -e.** 1. Dans tussen eiers deur wat op die grond lê. 2. Poging om jou uit 'n netelige posisie te red.
- ei'erdooiër, -s.** Die geel van 'n eier.
- ei'ereër, -s.** lem. (iets) wat eiers eet; soort slang.
- ei'ergang.** Eierleier.
- ei'ergeld.** 1. Geld wat vir die verkoop van eiers ontvang is. 2. Geld wat vir die koop van eiers bedoel is.
- ei'ergereg, -te.** iets te ete wat hoofsaaklik uit eier bestaan.
- ei'erglans.** Die besondere glans van 'n eierdop.
- ei'erhouer, -s.** Toevoubare houër van karton met kelkievormige vakkies waarin eiers vir die handel verpak word. (VAW)

Following the cluster of unexplained compounds the same dictionary includes a variety of compounds with *eier-* as first stem as main lemmata in the vertical ordering of the dictionary. This is usually done because the lexicographer regards these items as semantically opaque. A comparison of the compounds given as unexplained and explained lemmata respectively does not substantiate the presumed differences in terms of semantic transparency and semantic opaqueness. Some complex lexical items are included as fully explained lemmata in spite of their unambiguous semantic transparency. Once again the dictionary specific lexicographic process has to be dominated by a user-driven approach. A user should be able to know where a specific compound will be lemmatised. Lexicographers should not expect users to be able to distinguish between semantic opaque and semantic transparent compounds or to know when a given compound, e.g. *eieruitstoting*, will be lemmatised as an unexplained compound in a horizontal ordering whereas e.g. *eierdooier* will be included as an explained compound in the vertical ordering. For a more detailed discussion of unexplained lemmata, cf. Gouws (1989:77).

It is extremely important that lexicographers have to be well aware of the problems users can experience with the horizontal ordering of lemmata. The decision to use a sinuous lemma file may never be seen as the mere continuation of a specific lexicographic tradition. In the planning of every dictionary the lexicographer needs to take cognisance of the potential target user of the dictionary as well as the typical usage situations. This should determine, among others, the decision to use or not to use a sinuous lemma file. The functions of the specific dictionary should also play a role in this regard. The inclusion of unexplained compounds in a sinuous lemma file may suffice in a dictionary with text reception as primary function because the knowledgeable user will be able to interpret the unexplained horizontally-ordered lemma as a self-explanatory form, and for the comprehension of a text in which the user is confronted with this lexical item the unexplained presentation could still lead to a successful dictionary consultation procedure. However, where text production is the prevailing function the dictionary needs to present a treatment of these lemmata, e.g. by means of entries giving the paraphrase of meaning or entries giving typical usage examples, that will enable the user to utilise them in the production of new texts. Some of the problematic issues relevant to the lexicographer in this regard are discussed in Gouws (to appear).

Lexicographers can apply the procedure of the horizontal ordering of lemmata in different ways and can utilise it to convey specific information. In this regard it is important to identify different types of sinuous lemma files. A sinuous lemma file contains clusters of lemmata which display a horizontal ordering and are attached to the article of a vertically ordered main lemma. This lemma, e.g. *eier* in the quoted example, functions as an entrance lemma to the cluster. The lemma is the guiding element of the article block which accommodates the lemma cluster. These lemma clusters can be divided into two distinct groups, i.e. niched and nested lemmata, cf. Hausmann & Wiegand (1989), Wolski (1989a). The entrance lemma functions either as nest or as niche entrance lemma. A niche is formed through a process of *niching* and is characterised by a strict alphabetical clustering of lemmata. The lemmata in a niche may or may not be semantically related. A nest is formed through a process of *nesting* and is characterised by a clustering which stretches the rules of the strict alphabetical ordering in order to exhibit morphosemantic relations between

words. These two types of horizontal ordering will be discussed in the following paragraphs.

### 7.3.6.1 *Niching*

The space-saving function can be regarded as the most dominant motivation for the application of procedures of niching. Niched lemmata adhere to an alphabetical ordering with respect to both the horizontal and the vertical ordering. The lemmata entered within the niche display an internal alphabetical ordering and they also precede the next vertically ordered main lemma alphabetically. This type of cluster merely illustrates a deviation of the direction of macrostructural ordering, i.e. horizontal in stead of vertical, but does not imply any deviation from the prevailing strict initial alphabetical ordering, i.e. an ordering in which the alphabetical value of each letter in the lemma co-determines its macrostructural arrangement. The lemmata included in a lemma niche, that is the cluster of niched lemmata, can display semantic relations and this is often the case due to the fact that these lemmata are compounds with the same first component. However, this semantic relation is no prerequisite for the lexicographic procedure of niching. Even when a lemma niche does display semantic relations it is merely coincidental if it also displays morphosemantic relations between the niched lemmata. No formal distinction is made between compounds and derivations. The space saving function of niched lemmata can be seen as the most important motivation for this macrostructural procedure. The following example from HAT illustrates this type of horizontal ordering:

**krieˆket** (< E.)

Opelugspel op 'n grasbaan met harde bal, kolf en paaltjies, gespeel deur twee spanne van elf spelers elk waarby die twee spanne om die beurt boul na die ander wat kolf om soveel lopies as moontlik aan te teken. **krieket**: ~baan, ~bal, ~kolf, ~paaltjies, ~seisoen, ~speler, ~veld, ~wedstryd.

**kriel** (w.g.) ww. (gekriel)

In this text block the lemma cluster displays an internal alphabetical ordering and all the lemmata in the cluster precede the following main lemma (*kriel*) alphabetically. The treatment of the main lemma *krieket* includes only one paraphrase of meaning and consequently all the niched lemmata do display a semantic relation.

The following example from HAT also illustrates an application of niching:

**krimp** ww. (gekrimp) **1** Saamtrek; kleiner, dunner, smaller word; teenoor *uitsit*: *Klere wat krimp in die was. 'n Gesig soos 'n gekrimpte suurlemoen.* **2** Krom trek, lett. of fig., as gevolg van 'n onaangename aandoening: *Krimp van die pyn.* **3** Kleiner, minder word: *Die dae krimp nou dat die winter kom.* **4** Laat saamtrek: *Die materiaal moet eers gekrimp word.* **krimp**: ~baar, ~bestand, ~erig, ~ing, ~verlies.

Due to the polysemous nature of the lexical item *krimp* (to shrink) the comment on semantics, cf. paragraph 8.3.1.2, contains four subcomments on semantics with a paraphrase of meaning presented in each subcomment on semantics. The ordering in the lemma cluster is determined by a strict initial alphabetical principle and no indication is given to indicate which specific sense of the main lemma applies in a given sublemma. The sublemmata presented in the niche belong to different

morphological categories: *krimpbaar*, *krimperig* and *krimping* are derivations whilst *krimpbestand* and *krimpvoerlies* are compounds. The user cannot deduce this from the way in which the niched lemmata are presented.

In some niches the alphabetical ordering is so dominant that it eschews semantic differences. In GW a number of examples can be found where the niche in an article block with a member of a homonym pair as niche entrance lemma sign contains compounds that actually belong in a text block introduced by the other member of the homonym pair. The treatment of the homonym pair *link*<sup>1</sup> and *link*<sup>2</sup> illustrates this:

**link**<sup>1</sup>, (n) toorts, fakkel.

**link**<sup>2</sup>, (n) skakel; skakelman (*mil.*); string; binding (*skeik.*); mouskakel, mansjetknoop; (v) (aaneen-) skakel, verbind, aaneensnoer; aaneenkoppel, vaskoppel; inhaak; ~ *HANDS with*, aansluit by; saamhang met; ~ *ON*, aanhaak; aansluit; ~ *TOGETHER*, aaneensluit; verbind; ~ *UP*, verbind; ~ *UP with*, aansluit by; saamhang met; ~**age**, verbinding, aaneenskakeling; ~**chain**, skakelketting; ~**committee**, skakelkomitee; ~**ed**, gekoppel; verbonde; ~*ed ARMS*, arm-in-arm, ingehaak; ~*ed HORSES*, gekoppelde perde; ~**ing**, aaneenhegting; ~**ing-up**, vereniging, samesmelting; ~**man**, fakkeldraer; ~**pin**, kettingbout; ~**plate**, skakelverbinding; ~ *road*, skakelpad.

**links**, (n, *pl.*), gholffbaan.

The text block attached to the second member of the homonym pair includes a typical niche which adheres to an internal and external strict alphabetical ordering. The alphabetical principle is so dominant that the lexical item *linkman*, which links semantically with the first member of the homonym pair is ordered in the niche which follows the second member of the homonym pair. Judged from a semantic point of view this lexical item must have been lemmatised as a sublemma in a text block with *link*<sup>1</sup> as niche entrance lemma. Such a strict adherence to the alphabetical ordering principle impedes successful dictionary consultation procedures because the user does not expect to find the niched lemma in a semantically unfamiliar environment.

### 7.3.6.2 Nesting

The macrostructural procedure aimed at the inclusion of nested lemmata makes provision for two distinct types of article clusters. The one type, *first level nesting*, has a limited lexicographic function whereas the second type, *second level nesting*, has to be regarded as a more sophisticated lexicographic tool. First level nesting actually lies between niching and second level nesting. As is the case with niching, the ordering within a cluster of first level nesting is not determined by morphosemantic relations although first level nesting often contains semantically related lemmata. Semantic relatedness is not a prerequisite for first level nesting. First level nesting shares a further feature with niching, i.e. that the cluster internally also displays a strict initial alphabetical ordering. However, it differs from niching in one important respect. Where a lemma niche fits perfectly in the alphabetical ordering of the preceding and following main lemmata the first level nest deviates from this ordering because the alphabetical sequence between the preceding and the following vertically ordered main lemmata is interrupted by the lemma nest. Although the lemmata included in the nest follow the preceding vertically ordered lemma alphabetically the nest includes lemmata which do not precede the following vertically ordered lemma

alphabetically. A deviation from the strict initial alphabetical ordering is the most characteristic feature of the procedure of nesting, shared by both first and second level nesting. Compare the following examples from NeW and HAT respectively:

**ba-by** *-bies*, *n.*: ~ **batterer** babaslaner. ~ **battering** baba-, kindermishandeling. ~ **blues** nageboortelike depressie, nageboortede depressie, bababedruktheid. ~ **boom** geboortegolf, baba-, geboorteontploffing. ~ **boomer** naoorlogse baba/kind. ~ **bouncer** huppel-, wiptuig. ~ **boy** (baba)seuntjie. ~ **buggy** stootwaentjie, -karretjie, baba-, kindervaentjie. ~ **bust** afname/daling in geboortes. ~ **doll** babapop; (*mooi meisie/vrou*) pop(lap), poppie. ~ **face** babagesig(gie). ~ **girl** (baba)dogtertjie. **B~gro** *-gros*, (*handelsnaam*), ~ **grow** *-grows* groeipakkie. ~ **like** babaagtig, baba-agtig. ~ **minder** babawagter, -oppasser. ~ **snatcher** babadief; (*infml.:* *ouerige man met 'n jong nooi/vrou*) wiegie-, kuikendief, ou bok met 'n groen/jong blaar; (*infml.:* *ouerige vrou met 'n jong kêrel/man*) wiegie-, kuikendief, ou blaar met 'n jong bok. ~ **tooth** melktand. ~ **walker** loopring.

**Baby-lon** *n.*, (*hist.*) Babilon; verdorwe/dekadente plek/stad/ens. (NeW)

**koe`ël** □ *s.nw.* (-s) **1** Langwerpige, silindervormige projektiel met effens spits punt waarmee uit skietwerktuie geskiet word: *lemand 'n koeël deur die kop jaag. Getref deur 'n verdwaalde koeël. Die outydse koeëls was rond.* ▽ *Die koeëls blits dat die klippe so brand* (Toon van den Heever). Vgl. PATROON. **2** (*meer D.*) Ronde balletjie soos in 'n koeëllaer. UITDR.: *Die koeël is deur die kerk*, die saak is beslis, beklink. □ *ww.* (gekoeël) (*w.g.*) Skiet na; by uitbr., gooi na: *Hulle het my met akkers gekoeël.*

**koeël:** ~ **gat**, ~ **tjie**, ~ **vormig**, ~ **wond**.

**koe`ëlas** As wat op koeëls (bet. 2) loop. (HAT)

In both these examples the nest displays an internal alphabetical ordering but the strict alphabetical ordering with regard to the following main lemma in the vertical order (*Babylon* and *koeëlas* respectively) is interrupted because a strict initial alphabetical ordering would have ordered these two main lemmata before some of the sublemmata included in the nests.

The example from HAT illustrates a similarity with niching because no indication is given of the applicable sense of *koeël* in the complex lexical items and no distinction is made between the derivation *koeëltjie* and the compounds *koeëlgat*, .. *koeëlwond*. This is typical of first level nesting.

Although second level nesting is also characterised by a deviation from the strict initial alphabetical ordering, it shows significant differences from first level nesting with regard to the relations holding between the members of the lemma nest and the possible degree of deviation from the strict-initial alphabetical ordering. Second level nesting typically displays a higher density of data compared to first level nesting. This usually also leads to a higher degree of textual condensation in the lemma nests. Compare the following example from the descriptive monolingual dictionary NW:

**broei** (**ge-**) *ww.* **1.** *op eiers sit en hulle warm hou om hulle te laat uitkom.* **2.** *voortkom, ontspruit.* Daaruit sal onheil -. **3.** *peins, planne maak.* Oor iets -. **4.** *ontwikkel, in wording wees.* Daar is iets aan die -. **5.** *hitte ontwikkel, warm word.* Die mis, lug -. **6.** *warm word deur gisting.* Die hooi -. **7.** *deur spesiale verwarming vroeër laat bloei of ryp word.* **8.** *warm kry.* In die son sit en -. **9.** *kleintjies voortbring.* Die jakkalse - in die lente. '**broeiery**, **broeiing**; **broei-eend**, **-eiers**,

-gans, -hen, -hok, -kamer, -kolonie, -paar, -proses, -sak, -tent (by 1); -mis (by 5);  
 -aarde, -bed (by 6); -bak, -glas, -huise (by 7).  
 'broeiend (-e; -er, -ste) b.nw. 1. drukkend, ...

As is the case with first level nesting the nest interrupts the alphabetical ordering with regard to the following main lemma in the vertical ordering (*broeiend*). However, in this example the text block internal ordering also deviates from a strict alphabetical ordering. The alphabetical ordering is interrupted several times. This is due to morphosemantic motivations. Within the lemma nest a distinction is made between derivations and compounds. The first two nested lemmata (*broeiery* and *broeiing*) are derivations and are therefore grouped together and separated from the compounds by means of a semi-colon. The first group of compounds, i.e. *broei-eend, -eiers, -gans, -hen, -hok, -kamer, -kolonie, -paar, -proses, -sak, -tent*, display an internal alphabetical ordering. This partial article stretch represents the application of one specific polysemous value of the lexical item *broei*, i.e. the first polysemous sense indicated in the comment on semantics of the article of the lemma sign *broei*. This is clearly indicated by the entry (*by 1*). In a similar way the remainder of the lemma nest consists of partial article stretches which contain one or more than one lemma sign, displaying an internal alphabetical ordering where possible, and complemented by items indicating the applicable polysemous sense of the lexical item *broei*. This is a typical example of second level nesting where the nest, although also used for space-saving purposes, displays morphosemantic data.

When planning the ordering of lemmata for a given dictionary one has to work on a model that allows a clear and unambiguous access to the different macrostructural elements. Decisions regarding the ordering of lemmata should always aim to assist the identified target user of the given dictionary in the kind of dictionary usage situation (s)he needs and, consequently, to achieve a successful dictionary consultation procedure that leads to a successful retrieval of the kind of information that prompted the search.

#### 7.4 Balancing article stretches: multidimensional lexicographic rulers

Nothing is more difficult to predict or control than a dictionary begun from scratch. (Landau 2001:398)

This remark is equally applicable to dictionaries that were compiled without the availability of a corpus as discussed in Chapter 3.

Many unfortunate examples of dictionary compilation in the African languages exist where the lexicographers enthusiastically start with *A* and compile lengthy articles for every possible lemma starting with *a-* that they can think of but by the time they approach the end of the alphabet articles tend to become shorter and lemmas that should have been included, are missing. Consider the following example of a Sesotho sa Leboa dictionary where the discrepancy between a random section taken from the alphabetical stretch *A* compared to a sample taken from *T* is apparent even to the naked eye, without any help from measuring instruments.

aka	2	ala
<p>deel vir, vonnis vir (of) tot; uitspraak gee vir, – <i>lehu</i>, ter dood veroordeel. <i>ahlotlwa</i>, gevonniss woord, veroordeel word, geoordeel word. <i>ahloteng</i>, julle moet oordeel; <i>se</i> –e, moenie oordeel nie. <i>ahloté</i>, mag/kan oordeel. <i>ahlotšwe</i>, gevonniss wees, uitspraak is gegee. <i>moahlodi</i>, regter, beoordelaar. <i>baahloti</i>, regters, beoordelaars.</p> <p><b>aka, a.ka.</b> (-ile, -etše), lieg, leuens vertel, jok, onwaarheid spreek (dial. kyk: <i>aketsa</i>).</p> <p><b>aka, a.ka.</b> inhaak, vashaak, haak, aanhaak, soen, omarm, lieg, liefkoos; <i>akwa</i>, gehaak/ingehaak word; <i>akéla</i>, haak vir; <i>akelana</i>, mekaar liefkoos, vriendskaplik verkeer; <i>akelwa</i>, ingehaak word vir; <i>akiwa</i>, ingehaak word; <i>ake, ga, sa</i>, nie (in)haak nie; <i>aké</i>, mag/moet haak of inhaak; <i>moaki</i>, haker; <i>baaki</i>, hakkers.</p> <p><b>akalala, a ka la.la.</b> sweef, hang oor, oorhang; <i>akalaléla</i>, sweef vir/oor; <i>akalatsa</i>, laat sweef, vlerke oopsprei om te sweef; <i>akaladitše</i>, het laat sweef; <i>se bone nong go</i> –, <i>go wa fase ke ga lona</i>, hoogmoed kom tot 'n val; <i>akalatswa</i>, genoodsaak om te sweef; <i>akalatswa</i>, gesweef word; <i>akaléla</i>, hang/sweef oor, wydsbeen staan oor; <i>akalétsé</i>, het gesweef oor; <i>moakaladi</i>, persoon wat sweef.</p>	<p><b>akere, 'a ké, 'ré.</b> akker.</p> <p><b>aketsa, a ke.tša.</b> leuen vertel, lieg, jok; <i>akeditše</i>, het (gelieg) 'n leuen vertel. <i>sa aketše</i>, nie lieg nie.</p> <p><b>akga, a.kga.</b> werp, gooi, slinger, swaai, beweeg. <i>akgaakga</i>, heen en weer beweeg (soos branders), slinger, skommel; <i>akgaakgwa</i>, heen en weer geslinger word; – <i>diatla</i>, arms swaai, met leë hande loop. – <i>dinao</i>, voet in die wind slaan; <i>akgwa</i>, beweeg/geslinger word; –<i>akgéga</i>, skommel, swaai; –<i>akgéla</i>, slinger, swaai, werp. <i>akgéla</i>, slinger na/vir, tou om die horings gooi, met 'n vangtou vang, uitkrap, soos kole uit 'n vuur. <i>akgelwa</i>, geslinger word, gevang word met 'n tou. – <i>dikobo</i>, klere uitpluk.</p> <p><b>akgamala, a kga.ma.la.</b> verwonder, verbaas wees.</p> <p><b>akgatha, a kga.tha.</b> o/d skouer dra.</p> <p><b>akgofa, a kgo.fa.</b> haas, yl, snel, roer, gou maak, gou wees; <i>akgoféla</i>, haas na of vir; <i>akgofisa</i>, haastig maak, versnel, aanjaag, bespoedig, aanroer; <i>akgofela</i>, haas na of vir; <i>akgofago</i>, wat haas, haastige, vinnige, kits, gou; <i>akgofe, ga/ša</i> –, nie haas nie; <i>akgofé</i>, kan/mag haas.</p> <p><b>akgolia, a kgo.la.</b> opvang, vlugtig vang, harmonieer, onderskep; <i>akgodisa</i>, kous, laat opvang, vlugtig laat vang.</p>	

tao	292	tebetebe
<p><b>tao, ta.ó.</b> lêplek, (van 'n roofdier).</p> <p><b>taodisaserafe, 'ta o 'di šá se ra. fe.</b> volksbeskrywing, etnografie.</p> <p><b>taoditšeto, ta o di še. tšo.</b> biografie, verduideliking (van) aan, beskrywing (vir) van.</p> <p><b>taodiso, 'ta o di šó.</b> beskrywing, opstel, verhaal, vertelling, komposisie, stel-oefening, bewering, stelwerk, stelling.</p> <p><b>toidisophelo, 'ta o di šó 'phe.ló.</b> biografie, lewensbeskrywing; <i>tabataba ya</i> –, biografiese tema.</p> <p><b>taodiswana, dim., artikeltjie.</b></p> <p><b>taola, ta o.lá.</b> dolos, waarsêmiddel.</p> <p><b>taolamesifa, ta o la me šifa,</b> spierbeheer.</p> <p><b>taolelo, 'ta o lé.ló.</b> mandaat, reglement, ordonnansie, predestinasie; <i>naga</i>-, mandaatgebied; <i>maatla a</i> –, magte kragtens mandaat.</p> <p><b>taolá, ta o lá,</b> bevel, outoriteit, magtiging, kontrole, heerskappy, regering, gesag, beheer; – <i>ya (therafike) dina-melwa</i> verkeersreëling.</p>	<p>'n kind leer om te loop; <i>tataišo</i>, oefening.</p> <p><b>tatakgope, ta ta'kgo.pé,</b> koringsprinkaan.</p> <p><b>tatomola, 'ta ta mo.la.</b> uitstrek, los-torring.</p> <p><b>tatomolebo, ta 'ta mo le.bó.</b> rank.</p> <p><b>tatomologa, 'ta 'ta mo lo.ga,</b> losgaan, lostrek, losgetorring word, afrol, uitrek.</p> <p><b>tatomoloko, 'ta 'ta mo lo.kó,</b> geboortemerk, moesie, vlek.</p> <p><b>tatampana, 'ta 'ta m.pa.na,</b> kleuter.</p> <p><b>tatampua, ta ta m.pua,</b> blaas, blaar.</p> <p><b>tatasela, 'ta ta šé.la,</b> waggel, wankel.</p> <p><b>tate,</b> kyt: <i>tata</i>, vader, my vader.</p> <p><b>tatela, ta 'té.la,</b> opwen, opdraai, oprol, omdraai, toedraai, ombind.</p> <p><b>tatelano, 'ta té la.nó,</b> volgorde. opeenvolging, reeks; <i>ka</i> –, respektiewelik; – <i>e kwagalang</i>, logiese volgorde; –<i>melila</i>, chronologiese volgorde; <i>ka</i>-, chronologies.</p> <p><b>tatelo, 'ta té.ló,</b> navolging.</p>	

In order to address such inconsistencies on the macrostructural level, Prinsloo and De Schryver in sources such as Prinsloo and De Schryver (2002, 2005) studied the balance between alphabetical categories for English, Afrikaans and the African languages in existing dictionaries and electronic corpora of these languages. They designed practical instruments of measurement (and even prediction) for the relative length of alphabetical stretches in alphabetically-ordered dictionaries according to the generally accepted principle that alphabetical categories in any given language do not contain an equal number of words. A single glance at a few popular English dictionaries reveal that the alphabetical categories or article stretches for *A*, *B*, *P* and especially *C* and *S*, contain huge numbers of lemmas occupying almost 40% of the dictionary, while categories such as *J*, *K*, *Q*, *U*, *V*, *X*, *Y* and *Z* are relatively small, and consequently fill only a few pages in these dictionaries. For a dictionary such as the *Macmillan English Dictionary* (MED), where the alphabetical categories are marked with coloured thumb tags, one does not even have to open the dictionary in order to appreciate this breakdown which can also literally be measured by putting an ordinary ruler against the dictionary to roughly measure the 'thickness' of each article stretch in millimetres.

An alphabetical list of types generated from the Sesotho sa Leboa corpus show that roughly 17% of all words in this language fall under the single category *M* while categories such as *C, J, Q, U, V, W, X, Y* and *Z* are virtually empty.

The question is thus whether a specific balance, preferably one that could accurately be measured, exists between the different categories in a given language. In depth and exhaustive research for a number of languages by Prinsloo and De Schryver proved that this is indeed possible.

The full set of rulers for the 11 official languages of South Africa is given by Prinsloo and De Schryver (2005:116-8).

Consider Figure 1 as an example of a measurement ruler for Setswana (Prinsloo: 2004):

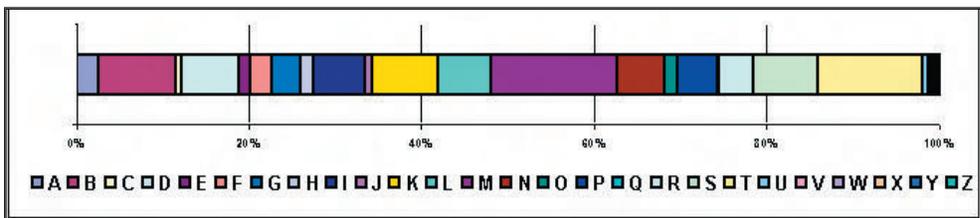


Figure 1: A Ruler for Setswana

For the revision of SESD as described in Prinsloo (2004), the focus is shifted from an alphabetical breakdown in the sense of the balance between the 26 letters of the alphabet on which the rulers are based (a to z) into a *percentage* breakdown in the form, referred to as a ‘Block System’ in Table 9.

Table 9: A Block System for Setswana

1	ALAF	21	FELE	41	KOUS	61	MOTL	81	SELE
2	AROG	22	FOLO	42	LAEL	62	MPHE	82	SERA
3	BADI	23	GAGW	43	LEBO	63	NATE	83	SETO
4	BANN	24	GATS	44	LEKI	64	NGWA	84	SIMO
5	BATW	25	GOLO	45	LERI	65	NKUK	85	SUAS
6	BIRO	26	GWET	46	LETS	66	NTEM	86	TALE
7	BOGA	27	HUBE	47	LOKO	67	NTSH	87	THAA
8	BOLA	28	IJES	48	MAAD	68	NYOR	88	THIB
9	BONK	29	IKGO	49	MAHA	69	OOMA	89	THWE
10	BORU	30	INOL	50	MALE	70	PANT	90	TLAM
11	BOUT	31	IPUS	51	MARA	71	PHAK	91	TLHA
12	DAAM	32	ITIS	52	MATL	72	PHIM	92	TLHO
13	DIFA	33	ITSH	53	MEFA	73	PITL	93	TLWA
14	DIKG	34	JOKO	54	MESU	74	PUDU	94	TSAP
15	DINK	35	KANY	55	MMAL	75	RAMO	95	TSHE
16	DIRA	36	KERO	56	MMOL	76	RENG	96	TSHW
17	DITH	37	KGAR	57	MOFI	77	ROKG	97	TSUN
18	DITU	38	KGOM	58	MOKG	78	RURU	98	UBAU
19	EGEP	39	KHAN	59	MONG	79	SEBA	99	WABO
20	ETLH	40	KODU	60	MORW	80	SEHI	100	ZIMB

Although based upon the same statistics, the Block System opens the door to a number of very practical applications and multi-dimensional utilisation in the compilation of a new dictionary or the revision of an existing dictionary. For the lexicographers and editors it gives a clear guidance in terms of page allocation, average length of the articles, progress in terms of time and even remuneration intervals for the part-time compilers.

# Microstructural aspects

---

## 8.1 Introduction

A dictionary article typically consists of a macrostructural element, functioning as lemma and as guiding element of the article, and an indefinite number of microstructural entries, primarily presented as part of the treatment of the lemma. Different types of microstructural entries can be distinguished and will be discussed in par. 8.5. A user-friendly dictionary demands that the microstructural entries should not be presented in an arbitrary way but rather in a systematic order. The knowledgeable user who is familiar with the system of a dictionary, with its data distribution structure and with the way in which the microstructural entries are presented, should be able to predict what data types can be found in a given article and also where to find a specific data type. Different types of microstructures have been developed and lexicographers should apply the microstructural type they opt for in a meticulous and consistent way. Some of these types of microstructures are discussed in paragraph 8.5. The publication of a user-friendly dictionary compels the lexicographer to pay ample attention during the planning phase of the dictionary to the article structure. The article structure and the data distribution structure will determine the types of data categories to be represented in the microstructure as well as the way in which they will be presented. The article structure and the different data categories are discussed in par. 8.3.

## 8.2 Categories of entries

### 8.2.1 Different types of entries in a dictionary article

When looking at the microstructure of a dictionary or when planning the microstructure of a dictionary it is important that one should be well-aware of the different types of entries to be included as microstructural elements. It is also important to realise exactly what the term *entry* refers to. This term is used in different ways by different scholars. One of the frequent uses of the term *entry* refers to the dictionary article. This is not how the term will be used in this publication. The term *entry* is rather used to refer to each and every constituent of a dictionary article, cf. the following article from the POD:

**chair** —*n.* **1** seat for one person usu. with a back. **2** professorship. **3 a** chairperson.  
**b** seat or office of a chairperson ...(POD)

In this article *seat for one person usu. with a back* is a single entry. It is a paraphrase of meaning giving one of the polysemous senses of the word *chair*. The *n* indicating the part of speech is also an entry. Entries can also be smaller constituents. In the following article from GW the comma between *chair* and *seat* as well as the semicolon between *stool* and *see* are entries because they are textual elements marking a specific relation within the translation equivalent paradigm of this article (cf. par. 10.4).

**stoel**, (s) (-e), chair, seat, stool; see (of bishop); pedestal; stool (plant); ... (GW)

Following the proposals made in Wiegand (1989c:427) the class of dictionary entries can be divided into two distinct types, i.e. *items* and *indicators*, usually referred to as *structural indicators*.

*Items* refer to those entries from which the dictionary user can retrieve some information regarding the subject matter of the specific dictionary. In a monolingual explanatory dictionary the general lexicon of the given language will be the subject matter and the treatment allocated to the lemmata will include a variety of data types, e.g. definitions, pronunciation, morphology, etc. Each entry presenting data that represents such a category will be seen as an item. In the above-mentioned article from POD both the definition *for one person usu. with a back*, and the part of speech marker *n* will be items because the user can retrieve some information regarding the word *chair* from these entries. In the article from GW an entry like “(-e)” is an item giving the plural form of the noun represented by the lemma. Likewise entries like the different translation equivalents, e.g. *chair, seat, stool*, are items. This also goes for entries indicating the context in which a given translation equivalent should be used, e.g. the entries (*of bishop*) and (*plant*) in this article. From each one of these items the knowledgeable dictionary user can retrieve information regarding the subject matter of the specific dictionary. Different types of microstructural data categories are discussed in par. 8.3

*Structural indicators* are not entries from which the user can retrieve information regarding the subject matter of the dictionary but they are those entries that mark a specific item or indicate a specific search field in a dictionary article. Structural indicators are the entries assisting the dictionary user to identify the different types of items, data categories and search fields in a dictionary article. Two kinds of structural indicators are used, i.e. *typographical* and *non-typographical structural indicators*. *Typographical structural indicators* are the different typefaces, e.g. bold, italic, roman, and the use of capitals, small caps, etc. in a dictionary. The function of these indicators is to mark specific search fields or data categories. In a monolingual dictionary one often finds the lemma sign to be presented in bold, the paraphrase of meaning in roman and illustrative examples in italics. The following example from the HAT illustrates the use of typographical structural indicators:

**bak**<sup>1</sup> □ s.nw. (-ke) **1** (Oop) kis waarin iets bewaar kan word: 'n *Bak vir meel, hout, steenkool*. **2** Deel van wa, kar, motor, ens. wat op die onderstel rus. **3** Houer vir vloeistof: 'n *Bak met water. Drink-, suip-, doopbak*. **4** Hoeveelheid wat in 'n bak gaan: 'n *Bak water, meel*. **5** Hol kant van 'n kromming: *Die plate staan met 'n bak*. UITDR.: *In die bak raak*, agter raak. *Iemand in die bak sit* (nie alg.), hom oortref. □ b.nw. en bw. Soos 'n bak: *Sy bene, ore staan bak. Die hande bak hou*.

**bak**<sup>2</sup> ww. (bakkende; gebak) **1** Gaarmaak deur hitte: *Eiers in die pan bak. Gebakte eiers*. **2** Gaar word; warm kry: *In die son sit en bak*. **3** Hard (laat) word deur hitte: *Stene bak*. **4** Hitte afgee: *Die son bak op die stoep. In die bakkende son*. ∇ *Die son bak blink* (M.M. Walters). UITDR.: *Mooi broodjies bak* - sien onder BROOD. *Bak en brou*, knoei. **bak**: ~dag, ~hoender, ~huis, ~kis, ~mengsel, ~skottel, ~trog.

**bak**<sup>3</sup> b.nw. en bw. (geselst.) Piekfyn, uitstekend: 'n *Bak kêrel. Dit gaan vandag net bak*. Vgl. BAKGAT. (HAT)

In this example the main and the sublemmata are given in bold, e.g. **bak**<sup>1</sup> and **bak: ~dag, ~hoender**, the definitions in roman, e.g. “(Oop) kis waarin iets bewaar kan word”, and the illustrative examples in italics, e.g.: ‘*n Bak met water* and *Die plate staan met ‘n bak*. Labels are given in brackets and in italics, e.g. (*geselst.*). A carefully devised and consistent use of typographical indicators plays an important role to enhance successful dictionary use and to help the user finding the desired data as quickly as possible. One of the advantages of the use of typographical structural indicators is the way in which it isolates the search field of a specific data category from other search fields. A dictionary user who is consulting the dictionary only for the sake of e.g. the illustrative examples presented in a specific article does not have to work through the article to find these illustrative examples. If the user is familiar with the system of structural indicators, as explained in the user’s guidelines text in the front matter of the dictionary, (s)he can have a rapid access to the article slot where this data category is accommodated by merely looking for the use of italics.

Capital letters can also be used as typographical structural indicators. In HAT idioms are included in the central list of the dictionary and they are presented as treatment units in a text block within the article of a lemma representing a keyword from the idiom. The text block for idioms is preceded by an entry “UITDR.” (an abbreviation for “uitdrukking” = expression). The entry is given in capital letters and this enhances the rapid access to the text block entrance because being the only entry given in capital letters it becomes much more noticeable, cf. the following article taken from the HAT:

**maag** (mae) **1** Belangrikste spysverteringsorgaan van die mens (en sommige diere), waar kos ‘n tydlang gehou word: ‘*n Sterk, swak maag hê. Met ‘n vol maag gaan slaap. Moenie jou maag oorlaai nie, te veel eet.* **2** Onderste deel van die romp; buik: *Daardie ou het ‘n groot maag.* UITDR.: *Dit sit my **dwars** in die maag,* ek is daarmee verleë, opgeskeep. *Sy **oë** is groter as sy maag,* hy wil meer hê as wat hy kan opeet (behartig). *Jy kan dit op jou maag **skryf** en met jou hemp afvee ..., daar sal niks van kom nie. Jou maag **vashou** van die lag,* onbedaarlik lag. ‘*n Maag soos ‘n **volstruis** hê,* alles kan eet. *Van jou maag ‘n **wolsak** (afgod) maak,* vraatsugtig wees. **maag:** ~aandoening, ~bloeding, ~druppels, ~kanker, ~kramp, ~kwaal, ~lyer, ~operasie, ~siekte, ~streek, ~wand, ~wond. (HAT)

Within the reserved text blocks the idioms are given in italics, with one of the words in the idiom printed in bold italics, cf. the words ***dwars**, **oë**, **skryf**, **vashou**, **volstruis*** and ***wolsak*** in the different idioms presented in the relevant text block of the article of the lemma sign **maag**. Where a text block contains more than one idiom these idioms are ordered alphabetically according to the alphabetical value of the words given in bold italics. This is clear from the ordering in the article given above.

*Non-typographical structural indicators* are symbols and signs used to mark the beginning of a certain search field or data category and they play an important role in the inner access structure of a dictionary. Dictionaries employ different types of non-typographical structural markers, e.g. diamonds, triangles, squares, brief headings, etc. In the article of the lemma sign **bak**<sup>1</sup> from HAT a square is used to indicate the different part of speech functions of a given word. The word *bak* presented by the lemma *bak*<sup>1</sup> can be used as a noun and as an adjective and adverb. Its classification as noun is preceded by the non-typographical indicator “□”, cf. the entries “□ *s.nw. (-ke)*” given immediately after the lemma sign. This square tells

the user that the word can also be used in a different function and the user merely has to look for the next occurrence of the square because it marks the treatment of another part of speech function of the word *bak*. Such a search leads the user to the entries “ □ *b.nw. en bw.*” where the lemma is treated in terms of the word *bak* in its part of speech function as an adjective or adverb.

In HAT a distinction is made between two types of illustrative examples, i.e. citations, taken from printed texts, and those examples constructed by the lexicographer. This distinction may be relevant to users, especially if they are interested in the way a given word has been used in the printed medium. Consequently a non-typographical structural indicator, an upside down triangle, “▽”, is employed and it precedes all the citations presented in the search area of illustrative examples. This is seen in the above-mentioned article of the lemma sign *bak*<sup>2</sup> with the inclusion of the following entry: ▽*Die son bak blink* (M.M. Walters). The citation is followed by an entry giving the name of the author from whose work the citation comes.

In the *Woordeboek: Nederlands-Afrikaans/Afrikaans-Nederlands* (ANNA), a bilingual Dutch-Afrikaans dictionary currently being compiled, different indicators are used to mark different search fields and data categories, cf. Gouws (2001); Martin & Gouws (2002). In the treatment of examples in this dictionary a distinction is made between those examples where the Dutch and Afrikaans forms are non-contrastive and those examples where they are contrastive. Non-contrastive examples are preceded by the indicator “-” and the Dutch example gets no Afrikaans equivalent. Contrastive examples are preceded by the indicator “•” and the Dutch example gets an Afrikaans translation. Idioms are preceded by the indicator “♦”. The following excerpt from an article of this dictionary illustrates the use of these indicators:

**bril**  
**1** [om te kijken] #  
 - een bril hebben/dragen; hij heeft zijn bril niet op; ...  
 • <inf.> een bril moeten 'n bril moet kry ...  
 ♦ elk ziet door zijn eigen bril *elkeen kyk deur sy eie bril*  
 door 'n roze bril kijken *deur 'n rooskleurige bril* kyk  
 ► iemand 'n bril op die neus sit iemand te grazen nemen

When planning a new dictionary the focus should not only be on the inclusion of items presenting the users with opportunities to retrieve information regarding the subject matter of the specific dictionary. Ample attention should also be given to the use of structural indicators that would make it easier for users to find the different items they may be looking for.

## 8.3 The article structure and the data categories in general monolingual and bilingual dictionaries

### 8.3.1 The article structure

As a part of the dictionary conceptualisation plan the lexicographer has to formulate a microstructural programme. This programme will determine the nature and extent of the microstructure, the article structure and the way in which the different slots in the article will be filled with data types. Once the microstructural programme has been formulated and the macrostructural selection has been completed the

lexicographers are in a position to pursue the construction of the dictionary articles as texts in the central list of the dictionary. The lemma functions as guiding element of each dictionary article and the microstructural programme orders the entries included as part of the treatment of the lemma in such a way that the article displays a definite structure. The article structure can be divided into two major article components, i.e. the *comment on form* and the *comment on semantics*. Every data category included in the microstructural programme belongs to one of these components. The distinction between the comment on form and the comment on semantics applies to all general bilingual and monolingual dictionaries.

### 8.3.1.1 *The comment on form*

#### *Orthography*

The *comment on form* is the search field accommodating those data types that reflect on the form of the lemma sign, i.e. the morphological, phonetic and orthographic form. The lemma sign is a part of the comment on form because it conveys data regarding the spelling of the treatment unit. People often need orthographic guidance and their dictionary consultation procedure only goes as far as finding the lemma and retrieving the necessary spelling information from the lemma sign. The comment on form can also accommodate additional spelling guidance if the lexical item included as lemma has spelling variants, cf. the following articles from the POD and HAT

**disfavor** (*Brit. disfavour*) –*n.* 1 disapproval or dislike. 2 being disliked. –*v.* regard or treat with disfavour. (POD)

**gentleman's agreement** *n.* (also **gentlemen's agreement**) agreement binding in honour but not enforceable. (POD)

**we<sup>ˆ</sup>dergebore**, ook **weergebore** b.nw. Herbore, opnuut gebore; bekeerd: *Die wedergebore sondaar. we<sup>ˆ</sup>dergeborene.* (HAT)

**re-stou-rant<sup>ˆ</sup>** (-e, -s) (< *F.*), ook **restaurant** Openbare inrigting van behoorlike gehalte waar voorbereide etes en drank bedien word: *In 'n restaurant die middagmaal geniet.* (HAT)

#### *Pronunciation*

Users also consult dictionaries for information regarding the pronunciation of words. This also falls within the comment on form because pronunciation has to do with the sound form of lexical items. Pronunciation can be presented in various ways and dictionaries differ in terms of the amount of pronunciation guidance on offer in a dictionary article. A typical treatment of the sound form of a word focuses on its phonetic representation and its stress pattern. Yet again dictionaries use different methods in the treatment of pronunciation. Some dictionaries would give a comprehensive phonetic transcription, using the symbols from the International Phonetic Alphabet (the IPA), whereas other dictionaries would give a partial transcription of the word or only an orthographic transcription, trying to capture the pronunciation of the word in the ordinary writing system. A minimalist approach in the treatment of pronunciation is to restrict this form of guidance to a mere marking of the stress pattern or by only indicating the main stress for a given

word. An indication of the syllable division forms part of the presentation of the sound form of a word, cf. the following example:

**cha-pe-ro<sup>ˈ</sup>ne** (uitspr. *sja-pe-roo<sup>ˈ</sup>ne*) (-s) **1** Ouer dame wat as begeleier van 'n ongetroude jong dame optree om fatsoenlikheid te waarborg. .... (HAT)

In this example the lemma sign does not only convey the orthography of the word but the lemma sign also accommodates further data belonging to the comment on form. The different syllables are separated by means of dots (·) and the syllable with the main stress is immediately followed by a superscript stress indicator (ˈ). The lemma sign is followed by an entry given in parenthesis and introduced by the pronunciation marker *uitspr.* (an abbreviated form of *uitspraak* = pronunciation). The pronunciation marker is followed by an orthographic transcription of the lemma in which the general alphabetical system is used to indicate the pronunciation. This orthographic transcription also contains its own syllable dividers and main stress indicator.

Yet again, the dictionary function(s) should play a determining role in the decisions regarding the nature, extent and presentation of phonetic data. Where text reception or text production are the dominant functions and the dictionary has to assist users in oral communication, a thorough treatment of the sound aspects of a given word is needed. Then the lexicographer will do well to give a full transcription with an indication of the stress pattern and the syllable divisions. The usage situation and the reference skills of the user need to be considered when decisions are taken with regard to the treatment of pronunciation. Although the use of the IPA leads to a precise account of the pronunciation, many dictionary users will not be familiar with this form of transcription and they will not be able to interpret the pronunciation data effectively. One possibility, found in many dictionaries, is to present the IPA in a front matter text along with a transcription of typical examples taken from the language treated in the dictionary. Some dictionaries also repeat the key to the IPA (or whatever pronunciation system is used) as a footer on every page of the dictionary. Thorough consultation should precede decisions regarding the presentation of pronunciation data.

For the African languages guidance in respect of *tone* is crucial. Louwrens (1994) in his *Dictionary of Northern Sotho Grammatical Terms* (DGT) describes tone as follows:

DGT

**tone** (*segalô, toon*)

Tone can be defined as *pitch variations* which affect the meaning and function of words. *Tone* is one of the distinctive features of the Bantu language family (see *Bantu languages*), and in these languages differences in *tone* between words which have exactly the same shape, result in a difference in meaning. Two basic tones (also called *tonemes*) are usually distinguished, namely a *high tone* and a *low tone*, although more detailed distinctions are often drawn between, for example *rising* and *falling tones*, *mid*, *mid-high* and *mid-low* tones, etc. A *tone* (or *toneme*) is always associated with a particular *syllable*, i.e. there are as many *tones* in a word as there are *syllables* since *tones* realise on *vowels*. This is one of the reasons why *vowels* are often referred to as *syllable nuclei*. (See: *nucleus*.) ...

In the following examples the correct pronunciation in respect of high versus low tone is crucial to distinguish between three unrelated meanings, *dassie* 'rock rabbit', *xilofoon* 'xilophone' and *haastigheid* 'haste':

PUKU 2

pela<sup>1</sup> ... HL: *dassie* ...

pela<sup>2</sup> ... LL: musiekinstrument van die xilofoon-tipe ...

pela<sup>3</sup> ... LH: vinnigheid, haastigheid; ...

A more user-friendly way to indicate tone is on the syllables themselves indicating high tones only or both as follows:

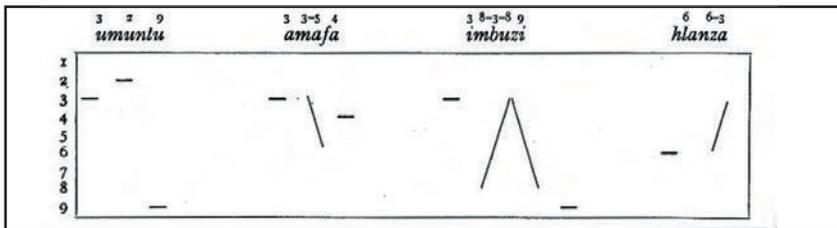
pèlà<sup>1</sup> ... HL: *dassie* ...

pèlà<sup>2</sup> ... LL: musiekinstrument van die xilofoon-tipe ...

pèlà<sup>3</sup> ... LH: vinnigheid, haastigheid; ...

Exact tone distinction in African languages is a complex issue as the following extract from ZED clearly illustrates.

The Zulu speaker employs a nine-tone system; that is to say, his range of tones in speech covers nine different pitches. These nine tone points cannot be indicated in musical notation, for they depend upon relative and not absolute height. The intervals between the notes are the important things. The whole range is generally slightly above an octave, with a man much lower in the scale than with a woman. No satisfactory method of recording the tones of Zulu words has yet been devised, ... 1 to 9 have been used to indicate the tone heights of the various syllables.



(ZED: xi)

### *Morphological data*

The most typical other entries accommodated in the comment on form are items conveying data regarding the morphology of the lemma as well as certain grammatical features. In the treatment of a lemma representing a noun the comment on form may include, where applicable, entries indicating morphological data like the plural and diminutive forms. These entries should always be presented in such a way that the target user can have an unproblematic retrieval of the relevant information. Where the data is presented in a condensed version the lexicographer has to make sure that this textual condensation does not confuse the target users. In the following examples the enlarged entries illustrate the plural forms of the nouns represented by the respective lemmata presented in the comment on form of various articles in different dictionaries:

**bab**´y, (**babies**), kind, wig, baba, babatjie; ... (GW)  
**maag**<sup>1</sup>, **mae**. Vernaamste spysverteringsorgaan; (VAW)  
**profeet**<sup>1</sup>, ..**fete**. **1.** lem. wat die heilige roeping van God ontvang om te vermaan en tot inkeer te probeer bring. (VAW)  
**cha-ris**´ma (**-ta**) **1** Buitengewone geestelike gawe of krag deur God geskenk. (HAT)  
**termyn**<sup>1</sup>, **-e**. **1.** Tydruimte van ´n bepaalde, gesette tyd. **2.** ...

In the first two articles the orthography of the stem changes when the noun is used in the plural form. Consequently the dictionary has to give the plural form in full so that the user can get the necessary guidance. In the third example the plural form *profete* also undergoes a spelling change but the lexicographer of this dictionary has opted not to give the full form but only the part of the word in which the change occurs. Such an entry demands more from the user than an entry where the full form, i.e. *profete* would have been entered in the slot for morphological data. The last two articles give a treatment of words where the plural is done in a systematic way by a mere addition of a plural suffix. Only the relevant suffix is entered in the relevant slot in the comment on form. In a dictionary where the users are familiar with the system such a condensed entry could be sufficient. In a dictionary for learners or users with limited dictionary experience it should not be regarded as the preferred treatment. The following example from NWSG illustrates a better option:

**aanval** naamwoord ■  
 (aanvalle, aanvalletjie) ...(NWSG)

This dictionary is compiled for primary school learners who are not mother-tongue speakers of Afrikaans. Where applicable the plural and diminutive forms of nouns are given and these entries are presented in full (*aanvalle*, *aanvalletjie*) so that the user is not confronted with a highly condensed version.

The comment on form should contain the relevant morphological data for words belonging to all the different parts of speech. The following examples illustrate the treatment of verbs:

**braai** (**het gebraai**) ww. **1** ´n Mens braai kos ... (BASIS)  
**drei**´neer´ ww. (gedreineer) **1** Drooglê; van oortollige water ontlas: ... (HAT)  
**aan**´bied, **-ge-**. **1.** Vrywillig tot beskikking stel ... (VAW)  
**demp**, **ge-**. **1.** Bedwing, onderdruk (oproer) ... (VAW)  
**beoor**´deel, (**-**), judge, criticise, adjudicate (merits); evaluate, assess; review (books); rate, value. (GW)

The first example gives the best guidance as the past tense form of the verb is given in full along with the auxiliary verb. The second example also gives the verb in full but without the auxiliary verb which is a compulsory component to indicate past tense in Afrikaans. Seeing that this auxiliary functions systematically the lexicographer can assume that the user who is more or less familiar with the grammatical system of Afrikaans should know that the past tense form of the word must always be accompanied by this auxiliary verb. The third and fourth examples are highly condensed entries. In the third example the past tense of the verb *aanbied* is *aangebied*. In this example the past tense affix *ge* is flanked by two hyphens, signalling that the suffix becomes an infix to function between the two

components of the particle verb *aanbied*. Yet again the lexicographer relies on the linguistic knowledge of the user. The typical past tense in Afrikaans is formed by means of a prefix *ge-*. The fourth example merely gives this prefix to indicate that the past tense of the word *demp* is formed in a systematic way. The last example has a tilde in parenthesis and this indicates that the word *beoordeel* does not get a prefix *ge-* or any other change in form when it is used in the past tense.

Dictionaries are also consulted to retrieve information regarding the morphological features of adjectives. The following examples illustrate typical entries to be found in the treatment of Afrikaans adjectives:

**heel**<sup>3</sup>, (b) (**hele**), whole, entire, complete, unbroken, undamaged; sound (GW)  
**psigede'lies**, **-e**. **1.** Wat die uitwerking het om die sintuie sterker, skerper te laat voorkom as in die werklikheid, waardeur vreemde, opgewerkte sensasies van krag, (VAW)  
**bit'ter**<sup>2</sup>, (b), **-der**, **-ste**. **1.** Wat ,n skerp, onaangename smaak veroorsaak. **2.** Skerp, griewend, bytend. **3.** In hoë mate; ~ *MIN*, uiters min; ~ *in die MOND maak die maag gesond*, medisyne wat sleg smaak is baie goed; ~**agtig**; ~**amandel**; ~**heid**. (VAW)  
**privaat**<sup>2</sup>, (b, bw), (**..vate**); **..vater** of **meer private**, ~**ste** of **mees private**. **1.** Wat op ,n persoon, ,n individu self en alleen betrekking het of aan hom behoort, (VAW)

Some Afrikaans adjectives get a derivational *-e* added on as a suffix when they are used attributively. The first two examples illustrate this with the first example giving the full derivational form *hele* whereas the second example only gives a condensed form indicating that an *-e* should be added to the adjective *psigedelies*. Degrees of comparison need to be indicated as in example 3 where the condensed version indicates that the degrees of comparison of the adjective *bitter* are *bitterder* and *bitterste*. For some adjectives degrees of comparison are morphologically marked by means of the grading suffixes *-er* and *-ste* whereas some adjectives have a lexical marking by means of the words *meer* and *mees* preceding them to mark the comparison. Some adjectives have both these ways of taking degrees of comparison and the lexicographer needs to assist the user in this regard. This is done in the fourth example (*privaat*) where a comprehensive morphological treatment is presented, albeit in a condensed form. These entries indicate that the adjective is subjected to derivation in attributive position (*private*) and the degrees of comparison lead to the possibilities *privater* or *meer privaat* and *privaatste* or *mees private*.

The way in which the morphology of adjectives is entered should always be determined by the dictionary using skills and the dictionary culture of the intended target users. Lexicographers often employ a system of textual condensation in the comment on form in an attempt to save space. This leads to a presentation characterised by the use of place keeping symbols, complex abbreviated entries and markers to indicate the non-occurrence of derivation. To illustrate this: the Afrikaans adjective *lui* can be used in attributive function with or without the suffix *-e*. The lexicographical treatment of this lexical item should include this variation in the comment on form along with the suffixes used to form the comparative and superlative forms of this adjective when degrees of comparison are required. The typical way in which this is done in a dictionary leads to the following entries:

**lui**, *adj.* (- of **-e**; **-er**, **-ste**)

This is a typical example of textual condensation rendering a condensed version of the full version:

*lui*, *adjektief* (*lui of luië; luiër, luiëte*)

In the condensed version the lemma sign is substituted by the place keeping symbol “-”. To interpret this version correctly demands a certain degree of dictionary using skills on the side of the target user. The dictionary plan should not only prescribe the data types to be included in the comment on form but it should also give clear and unambiguous guidance regarding the form and presentation of these entries.

Looking at the African languages an extensive range of morphological entries could be included in the comment on form of dictionary articles, cf. the following examples:

NSDN

**BÉNE**, -/di- (bêhê) (< Eng.) afleweringswa // van, delivery van ...

-bêni, se-/di- v. BÉNA

-bentšhi, ba- pl. <mmentšhi v. BENYA

SESD

**dipiriti** N. CL.8 di-, PL. OF sepiriti [FOR.], methylated spirits.

**dipitsa** N. CL.10 di-, PL. OF pitsa, pots.

**khurumêlaka** V. S. EXT., put covers, or lids on vessels, ...

**khurumêlêga** V. S. NEUT., be coverable, as a vessel ...

**khurumetsa** V. S. CAUS. < khurumêla, cover something up; overshadow.

### *Part of speech*

Dictionaries are often consulted for the verification of the part of speech of the word represented by the lemma sign. This data type is also presented as part of the comment on form. In the planning of their dictionaries lexicographers need to give a clear indication of the extent of their presentation of part of speech data. The type of dictionary, the data distribution structure, the functions, the users and situations of use but also the nature of the language treated in the dictionary should have a determining influence. In many dictionaries the part of speech guidance is restricted to an item giving the main part of speech class to which the word belongs, e.g. verb, noun, and adjective, cf. the following examples:

**chair** —*n.* **1** seat for one person usu. with a back. **2** professorship. **3** a chairperson.  
**b** seat or office of a chairperson ...(POD)

**conventional** *adj.* **1** depending on or according with convention. **2** (of a person) bound by social conventions. **3** usual; of agreed significance. **4** not spontaneous or sincere or original. **5** (of weapons etc.) non-nuclear. ...(POD)

This treatment will often be sufficient but a specific dictionary, due to its functions and situations of usage, may need a more detailed account of the part of speech. The POD is primarily compiled for mother-tongue speakers of English and classifications like noun and adjective could be quite sufficient, especially when the dictionary is used in a text reception function. For text production a more detailed treatment may be necessary, e.g. by indicating whether an adjective is used attributively or predicatively or whether a verb is transitive or intransitive. The

COBUILD dictionary, compiled for learners of English, uses a special column for part of speech data and gives a variety of entries. It is indicated whether nouns are countable or uncountable, whether a verb takes an object or not and whether the verb also requires an adjunct with the object. To the learner of English who uses the dictionary in a text production function this data is very relevant.

Language-specific features will have a definite influence on the part of speech data and lexicographers should plan their dictionaries accordingly. In some of the African languages it may be imperative to state explicitly not only that a word is e.g. a noun or a verb but also to which noun class it belongs or which verbal suffix(es) a verb stem includes.

SESD

**ngaka** N. CL. 9 N-SING. OF *dingaka*, ...

**thabuégêla** V.S. APPL. OF *thabuêga*, ...

Lexicographers often use abbreviations to mark the part of speech, e.g. markers like *n* (noun), *v* (verb), *adj.* (adjective), etc. Dictionaries should refrain from using their own domestic abbreviations as part of speech indicators where there are official abbreviations for the specific part of speech markers in the language. By using the established abbreviations the users have a better chance of immediately interpreting the abbreviation correctly. Where the dictionary is compiled for learners it is even better to give the part of speech marker in an unabbreviated form, cf. the following examples from NWSG where the entries *werkwoord* (verb) and *naamwoord* (noun) give the part of speech:

<b>aanval</b> (het aangeval) ... (NWSG)	werkwoord ■
<b>aar</b> (are, aartjie) (NWSG)	naamwoord ■

To ensure a systematic retrieval of the information presented in the comment on form it is essential that the dictionary plan should prescribe a fixed ordering of the data types and subtypes. These entries should be given in such a way that the user who is familiar with the system followed in the dictionary will know exactly what the status and function of each one of these entries are. Although a general model for the compilation of dictionaries can identify the comment on form as a component of the article structure, the contents of this component has to be determined along language-specific lines. Each dictionary project should decide on the entries to occupy this search field in an article.

### 8.3.1.2 *The comment on semantics*

*Subcomments on semantics, definitions and translation equivalents*

The *comment on semantics* is the search area accommodating those data types that reflect on the semantic and pragmatic features of the lexical item represented by the lemma. This component typically displays a range of data types. The nature and extent of the comment on semantics are also determined by the type of dictionary, the dictionary user and the situations of usage.

Dictionary research has indicated that the type of data most commonly looked for in dictionaries is semantic data. Monolingual dictionaries are frequently consulted for the explanation of meaning, also known in lexicography as the paraphrase of meaning, which is presented by means of lexicographic definitions, cf. Chapter 9 and Zgusta (1971), Gouws (1989) for a discussion of different types of definitions. Bilingual dictionaries are frequently consulted for another type of semantic guidance, i.e. the translation equivalents, presented as target language items for a given source language form. When it comes to the explanation of meaning in a monolingual dictionary or the presentation of translation equivalents in a bilingual dictionary the lexicographer has to make sure that the treatment reflects the complete spectrum of semantic values relevant for the specific dictionary. Where a given lemma sign represents a monosemous word the comment on semantics should include a treatment of that single meaning of the word. One definition, in a monolingual dictionary, or one translation equivalent, with possible synonym equivalents where applicable, will constitute the dominant entries in the respective comments on semantics. However, where a lemma sign represents a polysemous word all the polysemous senses should fall within the scope of the treatment. Negotiating the different polysemous senses should be done in such a way that the user gets a clear indication of the variety of senses a given word has. Consequently the comment on semantics is then divided into as many subcomments on semantics as needed to accommodate each polysemous sense in its own subcomment on semantics. If a given lexical item has e.g. four polysemous senses the comment on semantics will include four subcomments on semantics and, in a monolingual dictionary, each one of these subcomments on semantics will accommodate the definition given to represent one of the polysemous senses of the lexical item represented by the lemma sign. The ordering of the different subcomments on semantics within the comment on semantics may not be done in an arbitrary way. Looking at the polysemous senses of a lexical item, the lexicographer should apply clearly devised criteria to present these senses in a dictionary. The type of dictionary plays an important role in the article-internal ordering of the subcomments on semantics. A dictionary based on historical principles will typically order the senses from the oldest to the youngest. In general synchronic dictionaries one usually finds the ordering determined by the usage frequency of the senses. The sense with the highest usage frequency will be given as the first sense. For a discussion of other approaches to the ordering of senses, cf. Gouws (1989).

It is the task of the lexicographer to present the different subcomments on semantics in such a way that the user can realise they represent different senses of the relevant lexical item. Entries functioning as polysemy markers are used to guide the user in this regard. In monolingual dictionaries the different subcomments on semantics are usually indicated by means of a number preceding the subcomment on semantics, cf. the following example from the POD of the article of the lemma sign *aberration*, representing a word that has four polysemous senses (the numbers functioning as polysemy markers have been enlarged):

**aberration** *n.* **1** aberrant behaviour; moral or mental lapse. **2** *Biol.* deviation from a normal type. **3** distortion of an image because of a defect in a lens or mirror. **4** *Astron.* apparent displacement of a celestial body. (POD)

In bilingual dictionaries numbers are also used but very often the lexicographer utilises a semi-colon to separate the different subcomments on semantics, cf. the following example from GW where the article of the lemma sign *daintily* accommodates translation equivalents to represent four polysemous senses of the word, as indicated by the semi-colons separating the different translation equivalents:

**dain`tily**, kraaksindelik; fyntjies; kieskeurig; keurig. (GW)

In an article like the one given from GW the semi-colons do not function as mere punctuation markers but rather as functional dictionary entries. This use of semi-colons is discussed in more detail in Chapter 10.

In a monolingual dictionary the lexicographic definition is the item dominating the comment or subcomment on semantics and the definition is the one entry that will appear in every default article of such a dictionary. In many articles the comment on semantics will include only a single entry, i.e. the paraphrase of meaning presented as the lexicographic definition, cf. the following example from the POD:

**thoroughfare** *n.* road or path open at both ends, esp. for traffic. (POD)

Here the comment on form consists of the lemma sign and the part of speech marker whereas the comment on semantics only includes the entry *road or path open at both ends, esp. for traffic*, functioning as the lexicographic definition.

In a bilingual dictionary the presentation of translation equivalents dominates the comment on semantics and in some articles the comment on semantics will include a single translation equivalent as its only entry, cf. the following example from NeW:

**bim-boy** *n.* mooiseun. (NeW)

#### *Context and cotext entries*

In a dictionary compiled for text reception exclusively it is an acceptable procedure to limit the comment on semantics in many articles to the mere presentation of a paraphrase of meaning or a translation equivalent in monolingual and bilingual dictionaries respectively. If, however, text production is a function of the dictionary the lexicographer has to assist the user to use the words presented by the lemma sign and the translation equivalents in active communication. This compels the lexicographer to add some complementing entries in the comment on semantics that will help the user to interpret the words presented by the lemma sign and the translation equivalents not only in isolation but as part of the relevant language system. This could be done, among others, by means of entries giving the relevant context or cotext for the lemma and the translation equivalents. The *context* of a given word can be regarded as the pragmatic environment in which it is typically used. The context is usually indicated by means of glosses, i.e. a single word indicating something about the usage of the word, or by means of lexicographic labels. The *cotext* refers to the syntactic environment in which it is typically used. This is usually indicated by means of illustrative example material like collocations and example phrases and sentences. Context and cotext entries play an important role in both monolingual and bilingual dictionaries, cf. the following examples:

**conventional** *adj.* 1 depending on or according with convention. 2 (of a person) bound by social conventions. 3 usual; of agreed significance. 4 not spontaneous or sincere or original. 5 (of weapons etc.) non-nuclear. ...(POD)

**mis**<sup>4</sup>, (w) (**ge-**), miss (train); do without (drink); lose (boat); lack (experience); sy *DOEL* ~, miss the mark; be ineffective; *kan jy die GELD* ~, can you spare the money?; *dit* ~ *NIE*, it never fails; *ek* ~ *twee STOELE*, two chairs are missing; ....(GW)

**dry** —*adj.* (**drier**; **driest**) 1 free from moisture, esp.: a with moisture having evaporated, drained away, etc. (*clothes are not dry yet*). b (of eyes) free from tears. c (of a climate etc.) with insufficient rain; not rainy (*dry spell*). d (of a river, well, etc.) dried up. e using or producing no moisture (*dry shampoo*; *dry cough*). f (of a shave) with an electric razor. 2 (of wine) not sweet (*dry sherry*). 3 a plain, unelaborated (*dry facts*). b uninteresting (*dry book*). 4 (of a sense of humour) subtle, ironic, understated. 5 prohibiting the sale of alcohol (*a dry State*). 6 (of bread) without butter etc. 7 (of provisions etc.) solid, not liquid. 8 impassive. 9 (of a cow) not yielding milk. 10 *colloq.* thirsty (feel dry). ...(POD)

The lemma sign *conventional* is treated as being polysemous with five different polysemous senses. Each sense is treated in a separate subcomment on semantics and these subcomments on semantics are marked by means of numbers (1-5), functioning as polysemy markers. The lexicographer regards it necessary to indicate in two of these subcomments on semantics the typical context in which the word *conventional* is used to activate that specific sense. The entries, given in parenthesis, (*of a person*) and (*of weapons etc.*) give the relevant context of the word. It can, however, rightly be asked why the lexicographer did not give similar contextual guidance in the other three subcomments on semantics.

The word *mis*, presented as the lemma sign *mis*<sup>4</sup> – being the fourth member of a group of homonyms, is treated as a word with four polysemous senses. Following the system used in GW a translation equivalent is provided for each sense of the lemma and the different subcomments on semantics are separated by means of semi-colons. In this article each sense of the lemma only receives one translation equivalent with no target language synonyms accompanying these equivalents, cf. Chapter 10. Each translation equivalent is immediately followed by a single word, given in parenthesis, to indicate the typical pragmatic environment in which the word *mis* should be translated with the specific translation equivalent. These context entries are known as glosses, brief explanatory comments, added to the translation equivalent or the paraphrase of meaning, to guide the user in a successful interpretation of the treated words.

In the article of the lemma sign *dry* the different subcomments on semantics, indicated by the numbers 1-10, allow a further subdivision within the different senses to make provision for usage differences in a given sense of the word, indicated by alphabetical markers, e.g. *a-f* in the first subcomment on semantics. Although the same sense prevails in these different uses of the word, the pragmatic environment differs and this is indicated by means of context indicators like *of eyes*, *of climate*, *of a river*, etc.

Context entries like these assist the user in both text reception and text production but for text production purposes it is even better when a dictionary also offers context guidance by giving typical illustrative collocations, phrases and example sentences

to illustrate the way in which the specific word functions within the linguistic system of the language. In the article of the lemma sign *dry* entries like *clothes are not dry yet*, *dry shampoo*; *dry cough* and *feel dry* assist the users in their active language usage. In the article of the lemma sign *mis* entries like *sy DOEL ~, miss the mark*; *kan jy die GELD ~, can you spare the money?* and *ek ~ twee STOELE, two chairs are missing* illustrate the use of the word represented by the lemma sign along with the English translations for these examples.

The supporting examples play an important role in both bilingual and monolingual dictionaries and they have to be selected in such a way that they complement the meaning paraphrase or translation equivalent. This can only be achieved in a consistent way if a representative lexicographic corpus is used. From a linguistic perspective examples play an important role because they ensure that the word represented by the lemma sign is not only seen in isolation but as part of the language system. The careful selection of examples should be done in such a way that typical combinations (collocations) are presented but also the typical grammatical constructions in which the word represented by the lemma sign functions in real language use. This implies, e.g., that the treatment of a given noun should indicate typical combinations with adjectives, verbs and prepositions. The examples given in the treatment of a verb should show whether the verb functions as transitive or intransitive verb. Cotext entries should be used as instruments to convey desperately needed grammatical information. Various issues regarding the use of examples are discussed in Prinsloo and Gouws (2000).

It is very important that the cotext and context entries should not be given in an arbitrary way but rather in accordance with criteria formulated in the dictionary conceptualisation plan. Too often in bilingual dictionaries one finds that only some of the subcomments on semantics include context and/or cotext entries. Procedures like these leave certain translation equivalents stranded in terms of supporting entries. This impedes the text production function of the dictionary. Problems associated with insufficient context and cotext assistance are discussed in Gouws (2002a).

### *Lexicographic labels*

Lexicographic labels are frequently employed in the comment on semantics to give explicit contextual guidance. As pragmatic markers labels are used to relate an item in a dictionary to the world outside the dictionary and they can be used to mark either a macro- or a microstructural item in a dictionary article. When a label is used to mark a lemma sign it implies that all the senses of the word represented by the lemma sign fall within the scope of the label. A label could also be used to mark a specific microstructural item, e.g. a specific item giving the pronunciation or one sense in the treatment of a polysemous item. This implies that only that single item falls within the scope of the label.

Labels are used to mark deviations from the neutral or default value prevailing in a dictionary. In a special field dictionary dealing with terms from the field of chemistry the lemmata will not be labeled as belonging to the field of chemistry because this is the default value of the items selected for inclusion in the macrostructure of this specific dictionary. However, when a term from the field of chemistry is selected for inclusion in the macrostructure of a general dictionary it will typically be labeled

to indicate that it deviates from the default value of the dictionary, i.e. the general lexicon, cf. the following example from the POD, a dictionary presenting the general lexicon of English:

**sulphuric** *adj.* (US **sulfuric**) *Chem.* containing sulphur with a valency of six. (POD)

The word *sulphuric* is labeled as belonging to the field of chemistry because it deviates from the default subject matter of the POD.

Although dictionaries use their own domestic labels the majority of these labels can be categorised in a few well-established classes of lexicographic labels. Three major classes of labels that are commonly used in dictionaries are *subject field labels*, *stylistic labels* and *chronolectic labels*, cf. Gouws (1988; 1989).

*Subject field labels* are used to indicate that an item belongs to a specific specialised field which is not the section of the lexicon primarily targeted in the specific dictionary. When dealing with subject field labels the term *subject field* should not be interpreted in a too narrow sense. It does not only refer to academic, scientific or professional fields but also to hobbies and sport. In treating the word *cantabile* the POD labels it as belonging to the subject field of music:

**cantabile** *Mus. –adv. & adj.* in smooth flowing style. –*n.* cantabile passage or movement. [Italian, = singable] (POD)

This label is directed at the lemma sign. In contrast the lemma *gully* is not labeled because it is part of the general lexicon of English. However, the word *gully* is also used as a cricket term and the subcomment on semantics accommodating this specific sense, subcomment on semantics 3, has been marked accordingly with the label *cricket*:

**gully** *n.* (pl. -ies) 1 water-worn ravine. 2 gutter or drain. 3 *Cricket* fielding position between point and slips. [French *goulet*: related to \**gullet*] (POD)

*Stylistic labels* are widely used in general dictionaries to mark deviations from the standard variety and neutral register and style of everyday language use. Labels like *formal*, *colloquial*, *figurative* and *slang* are often encountered, cf. the following examples:

**job** –*n.* 1 piece of work to be done; task. 2 position in, or piece of, paid employment. 3 *colloq.* difficult task (had a job to find it). 4 *slang* crime, esp. a robbery. 5 state of affairs etc. (*bad job*). (POD)

**footsie** *n. colloq.* amorous play with the feet (POD)

**cruelhie** *n.*, (*infml.*: *gemene opmerking/ens.*) 'n lae hou. (NeW)

**haar'klowery** (-e) (*fig.*) 1 Die maak van uiters fyn onderskeidings; vitterige kritiseerdery; kleingeestige stryery. 2 Nietige onderskeiding; gestry, getwis oor 'n nietige verskilletjie. (HAT)

**laag**<sup>2</sup> (lae) (*veroud.*; *digt.*) Wat gereed gelê is om iemand te oorval; hinderlaag, valstrik: ... (HAT)

In the last example from HAT the word is labeled as both *veroud* (= *verouderd* = archaic) and *digt.* (= *digterlik* = poetic). The second label is a stylistic label whereas the first one is a chronolectic label.

Curses and other taboo words often have a high usage frequency in the general language and should therefore be included in the lemma selection of a dictionary. They should, however, be labeled to warn the user not to use them in a general conversation. All taboo forms do not have the same degree of taboo and a dictionary should make provision for labeling items in terms of their position in a taboo hierarchy. The use of labels like *vulgar*, *obscene*, *derogatory* and *coarse* and their relative taboo value should be explained in the user's guidelines text. The following articles contain labels indicating various forms of taboo:

**bloody** —*adj.* (-ier, -iest) **1** of, like, running with, or smeared with blood. **2 a** involving bloodshed. **b** bloodthirsty, cruel. **3** *coarse slang* expressing annoyance or antipathy, or as an intensifier (*bloody fool*; *a bloody sight better*). **4** red. (POD)  
**shit** *tw.*, (*taboesl.*) fok, kak, bliksem, dēm, demmit, verdomp; *oh ~ my* (goeie) fok, fokkit. (NeW).

In a bilingual dictionary it is important that labels should be used to mark both the lemma and its translation equivalents to indicate whether the same deviation applies in both source and target language. In the above-mentioned example from NeW some of the equivalents given for the lemma *shit* share the same value, i.e. taboo slang, as indicated with regard to the lemma. However, some of the equivalents display a lesser degree of taboo, e.g. the word *verdomp*. This should be indicated clearly. In bilingual dictionaries translation equivalents often go unlabeled, cf. the following example from GW:

**vom** it, ... (v) braak, vermeer, vomeer, kots, opgooi, jongosse inspan; ... (GW)

The source language item is a neutral form and should be unlabeled. The equivalents given for the verb *vomit* are not equal in terms of their use in Afrikaans. *Braak* is a neutral form and should be unlabeled, *vermeer* is an archaic form, *vomeer* lies between neutral and formal, *kots* is a coarse form, *opgooi* an informal word and *jongosse inspan* an informal expression. Unfortunately no labels have been included in the comment on semantics to mark the translation equivalents. This is detrimental to the lexicographer's attempts to ensure communicative equivalence, cf. Chapter 10, and impedes the user when employing the dictionary in a text production function.

*Chronolectic labels* mark a word or one of its senses or uses as deviating in terms of its typical time of use, i.e. being archaic and outdated or being a very new form. This is illustrated in the following examples where the labels *archaic* and *neol* (= neologism) mark the chronolectic deviation of the given words:

**jackanapes** *n. archaic* rascal. [earlier Jack Napes, supposed to refer to the Duke of Suffolk] (POD)

**haal** baar *b.nw. (neol.)* Wat bereik, verwesenlik kan word: *Dit lyk asof ons doelwit na al die harde werk uiteindelik haalbaar is. Haalbare resultate, winste.* (HAT).

Archaic forms are more likely to be labeled in dictionaries because neologisms rapidly become part of the standard variety and then no longer need to be labeled.

*Etymological data*

In spite of its name the comment on semantics does not only include semantic data. All the data in a dictionary article not related to the form of the lemma falls within the comment on semantics. Many dictionaries contain entries giving some guidance regarding the origin of the word under treatment. These etymological entries also form part of the comment on semantics. Especially in comprehensive and historical dictionaries users can expect to find etymological and other historical data. This type of data is not needed for text production purposes and even in text reception it plays a minor role. However, because dictionaries are regarded as containers of linguistic data, users often consult these dictionaries to retrieve some form of etymological guidance. It is one of the types of linguistic data that the average member of a speech community finds interesting and people often consult a dictionary merely as a matter of interest to inquire about the etymology of a given word. In general synchronic monolingual and bilingual dictionaries etymological data can be regarded as a bonus and not as a compulsory component of the comment on semantics. Consequently one would find that many dictionaries do not reserve a slot in the comment on semantics for this data type. Where etymological data has been included in the comment on semantics it should be regarded as part of the knowledge-orientated functions of the dictionary and not as part of its communication-orientated functions, cf. the following example:

**acerbic** *adj.* harsh and sharp, esp. in speech or manner. □ **acerbity** *n.* (*pl.* -ies).  
[Latin *acerbus* sour] (POD)

In this example taken from the POD the horizontally ordered lemma *acerbity*, attached to the article of the lemma *acerbic*, has a very limited treatment. The comment on form contains a part of speech marker and an entry giving the condensed plural form of the word. The comment on semantics includes only one data type, i.e. the etymological data. This complex entry consists of an entry indicating the language from which the word *acerbity* originates, i.e. Latin, along with the Latin form (*acerbus*) and an English equivalent of this Latin word. Due to space-saving reasons etymological data in standard and desk dictionaries are often presented as condensed entries, cf. the example taken from HAT:

**chri`so-fil** Geel kleurstof in blare. [G. *chrysos* goud + *phullou* blaar]

The condensed complex entry giving etymology consists of an abbreviated indication of the language from which the word originates (G = Greek) as well as the two Greek words *chrysos* and *phullou* from which the Afrikaans form *chrisofil* has been derived, along with the present-day Afrikaans equivalents (*goud* = gold and *blaar* = leaf) of these two Greek words

In historical dictionaries and in restricted dictionaries focussing on etymology a much more comprehensive account is given with a detailed account of the language development in terms of changes in the form and meaning of a given word. The comprehensive twenty volume historical dictionary, *The Oxford English Dictionary* is a good example of a dictionary giving ample etymological guidance. For Afrikaans the restricted dictionary EWA, the *Etimologiese woordeboek van Afrikaans*, focuses exclusively on etymology as a data type and gives a much more detailed discussion

of etymology than one would find in a dictionary where etymology is only one of many data types to be included.

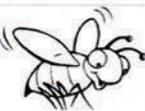
#### Other data

The paraphrase of meaning and the translation equivalents are not the only types of semantic data that can be presented in the comment on semantics. In the planning of the data distribution structure of a dictionary the lexicographer may decide also to include an indication of some relevant semantic relations. This will lead to a presentation where e.g. items giving antonyms or synonyms also occur as part of the treatment of a given lexical item. This type of data should be entered in the comment on semantics and provision should be made for a specific slot for these items.

The comment on semantics is also the venue for other data types not always presented in the default article of a given dictionary. *Inserted inner texts*, inserted into an article, are often used to focus the attention of the user on a specific aspect of the use or the meaning of a given word. These inserted texts, often presented in so-called commentary boxes, are not given in all articles but where they do occur, they usually function within the comment on semantics. In TAW these text boxes are put to good use to convey relevant data which falls outside the scope of the default categories presented in the normal search fields of the article. The article of the lemma *meat* includes the following inserted inner text as part of the comment on semantics:

The meat from some animals has a different name from the animal itself: the meat from a cow is called **beef** and that from a pig **pork**, but the meat from a lamb is called **lamb**. For fish and for birds such as chicken or duck the same word is used for both meat and animal.

In NWSG the lexicographers also use article-internal inserted text blocks to make the users aware of issues that are not treated in the default presentation of the data categories, cf. the following examples:

<p><b>verveeld</b> byvoeglike naamwoord ■  <b>(verveelde; verveelder, die verveeldste)</b>          As 'n mens verveeld is, stel jy nie in iets belang nie. ⇒ <i>As die onderwyser maklike werk doen, is die meeste kinders verveeld [bored].</i> ⇒ <i>Die onderwyser sukkel om die verveelde [bored] leerders wakker te hou.</i>          kyk by <b>vervelig</b></p> <p><b>vervelig</b> byvoeglike naamwoord ■  <b>(vervelige; verveliger, die verveligste)</b>          As iets vervelig is, is dit nie interessant nie en stel jy nie daarin belang nie. ⇒ <i>Die les was so vervelig [boring] dat ek amper aan die slaap geraak het.</i> ⇒ <i>Die lew op ons dorpie is nooit vervelig [boring] nie. Daar gebeur elke dag iets opwinden is.</i></p>	 <p>Pasop vir die verskil tussen <b>verveeld</b> en <b>vervelig</b>: iemand is <b>verveeld</b>, iets is <b>vervelig</b>. 'n Mens gebruik <b>verveeld</b> vir iemand wat nie in iets belangstel nie, byvoorbeeld: ⇒ <i>Die slim student is verveeld [bored], want die werk is oninteressant.</i> 'n Mens gebruik <b>vervelig</b> vir iets waarin nie belangstel nie en wat jou nie boei nie, byvoorbeeld: ⇒ <i>Die slim student hou nie van vervelige [boring] lesse nie.</i></p>
--	---



categories and search fields. All microstructural items should be included as part of the treatment offered in the specific dictionary and lexicographers and users should know exactly at which treatment unit a specific item is directed. Although the lemmata are the typical treatment units in a dictionary it is important for the lexicographer and the dictionary user to realise that some items, albeit not lemmata, need additional supporting entries to assist the user in an unambiguous retrieval of information. With the lemmata functioning as first level or primary treatment units the microstructural items in need of supporting entries are elevated to treatment units and become second level or secondary treatment units.

The lemmata are included in the macrostructure of a dictionary as guiding elements of an article. As the primary or first level treatment units they are the targets of the majority of items presented as part of the lexicographic treatment. In lexicographic terms a lemma is regarded as the *address* of all the entries directed at the lemma as treatment unit. Each microstructural item is part of the treatment of either the lemma of a given article or of another microstructural item in the same article. An item can also function as part of the treatment of a lemma or other item in an article elsewhere in the same dictionary. Microstructural items are directed or addressed at specific targets. The addressing structure of a dictionary is the system according to which these procedures of one item being directed at another is employed. The application of an addressing structure leads to a situation where every microstructural item is directed at another item, the address of the given microstructural item. The lemma is the most typical address in an article but other items may also function as addresses. This leads to a distinction between two major types of addressing, i.e. *lemmatic* and *non-lemmatic addressing*.

*Lematic addressing* is a procedure where a main lemma is the address of an entry. In a dictionary, which adheres to a straight initial alphabetical ordering, all the lemmata would be arranged vertically and each lemma will be the guiding element of an article. Lemmatic addressing will always have one of these lemmata as the address of a given entry. In a dictionary where a sinuous lemma file prevails, cf. Chapter 7, niched and nested articles can have sublemmata as their guiding elements and primary treatment units. These sublemmata are addressed by items in the subarticles and such a sublemma functions therefore as the address of an item in a subarticle. This addressing procedure is known as *sublemmatic addressing*. The sublemmata remain part of the macrostructure of the dictionary; therefore sublemmatic addressing is a type of lemmatic addressing, cf. the following example from GW:

**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.

This is a condensed niche with *early*, the main lemma, in the remote niche entrance position. The niched articles are headed by sublemmata and the partial article stretch

starts with the article of the sublemma *early-bearing*, presented in a condensed form as the partial lemma *~bearing* and ends with the article of the sublemma *early-warning system*, presented in a condensed form as the partial lemma *~warning system*. These sublemmata are the guiding elements of subarticles and these articles contain translation equivalents given for the sublemmata. These translation equivalents are addressed at the sublemmata and this constitutes a procedure of sublemmatic addressing.

If all the microstructural items in an article are addressed at the lemma, a situation of *full lemmatic addressing* prevails, cf. Hausmann & Wiegand (1989:349).

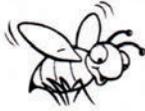
*Non-lemmatic addressing* is an addressing procedure in which the lexicographic treatment is directed at an item not functioning as lemma. This item becomes an address in the given procedure. Whilst lemmatic addressing is directed at macrostructural items, non-lemmatic addressing is directed at microstructural items. The address is the topic of the specific treatment procedure. The use of non-lemmatic addressing implies a system of topic switching within the dictionary article because each non-lemmatic address introduces a new treatment unit as topic, cf. Hausmann & Wiegand (1989:329). Although topic switching increases the level of textual condensation in a given dictionary article it also leads to a more comprehensive treatment procedure where the lemma is complemented by other items as treatment units. The user does not only retrieve information regarding the lemma but also of other items in the article. This is of extreme importance in the articles of bilingual dictionaries, cf. Chapter 10, where procedures of non-lemmatic addressing elevates the translation equivalents to treatment units.

Especially bilingual dictionaries have traditionally been dominated by a lemmatic addressing bias. This has created a situation where the translation equivalents are merely functioning as elements given as part of the treatment of the lemma. A lack of procedures of non-lemmatic addressing has impeded attempts to add data addressed at the translation equivalents in order to help the user with the choice of a correct translation equivalent for a given context. Translation equivalents should also function as secondary treatment units in a dictionary article. This will lead to the application of procedures of non-lemmatic addressing which will enhance the establishment of a relation of communicative equivalence between source and target language forms because the entries directed at the translation equivalents will ensure successful procedures of equivalent discrimination, cf. the following two articles from GW where the one article displays procedures of non-lemmatic addressing whereas the other article only has an application of lemmatic addressing:

**leng**, (ge-), shed (blood); spill (liquor); offer (libation); ... (GW)  
**pre kerig**, (-e), preachy; drawing; predicatory; ... (GW)

Each one of the context words, cf. Chapter 10, following the translation equivalents in the first example has the preceding translation equivalent as address and this non-lemmatic addressing helps the users in their choice of a translation equivalent for a given occurrence of the source language item. This help is not available in the second example which is characterised by an exclusive lemmatic addressing procedure.

Although addressing procedures usually function on an article-internal level the address of a given item could also be an item in another article. This implies that addressing procedures can exceed the limits of an article. This type of article-external addressing is also known as remote addressing. The following examples, taken from NWSG illustrate one of the uses of remote addressing:

<p><b>gesin</b>                                      naamwoord ■  <b>(gesinne, gesinnetjie)</b>  'n Gesin bestaan gewoonlik uit 'n pa, 'n ma en kinders. ⇒ <i>Ons gesin [family] hou altyd saam vakansie.</i> ⇒ <i>Hulle is 'n baie groot gesin [family].</i></p>		<p>Pasop vir <b>gesin</b> en <b>familie</b>.  <b>Gesin</b> verwys net na 'n pa, ma en kinders, terwyl <b>familie</b> ook verwys na jou ooms, tantes, oumas, oupas, niggies en neefs (die hele familie).</p>
<p><b>familie</b>                                      naamwoord ■  <b>(families, familietjie)</b>  ① Jou familie is jou ooms, tannies, neefs, niggies en die mense met wie hulle getroud is. ⇒ <i>My ma en pa was albei enkelkinders; daarom het ons nie so 'n groot familie [family] nie.</i> ⇒ <i>My vriendin kom uit 'n baie ryk familie [family].</i>  kyk <b>gesin</b></p>		

In these examples the inserted inner text has been included in the article of the lemma sign *gesin*. The pragmatic guidance given in this text block is not only addressed at the lemma *gesin* but it is just as applicable to the lemma *familie*. This inserted inner text is addressed at the lemma *gesin* by means of a typical procedure of article-internal lemmatic addressing. A further form of lemmatic addressing prevails with the lemma *familie* being the article-external address of the inserted inner text. This is a form of remote lemmatic addressing. The inner text is addressed at two different lemmata and this represents a form of double addressing, cf. Gouws (1996a), Louw & Gouws (1996). In the examples taken from NWSG the article of the lemma sign *familie* contains a cross-reference entry that guides the user to the article of the lemma sign *gesin* where the relevant inserted inner text can be found. When using procedures of remote addressing it is important that the article in which the remote address occurs should contain a cross-reference to the article from where it has been addressed. The absence of a cross-reference to *gesin* in the article of the lemma sign *familie* would have decreased the value of the inserted text given in the article of the lemma sign *gesin* because users would not have been able to access the data in this inserted text, relevant for the use of the word *familie*, from the article with *familie* as lemma sign. Remote addressing is also a space-saving procedure because, in a case like the given examples, the same inserted text does not have to appear in two different articles. The optimal success of this type of remote addressing co-depends on a well-implemented system of cross-referencing.

A theoretically founded model for the compilation of dictionaries should emphasise the need for a well-devised addressing structure, which allows the application of both lemmatic and non-lemmatic addressing procedures.

## 8.5 Types of microstructures

Although the microstructure of a dictionary can informally be described as the set of entries in a dictionary article accompanying the lemma and presented as treatment of the lemma, the term *microstructure* demands a much more precise interpretation, and this interpretation should be influenced by a number of other features. When planning the microstructure of a given dictionary one should negotiate the results of dictionary research that have led to the identification of different types of microstructures, cf. Wiegand (1989b; 1989c; 1996).

The dictionary-specific lexicographic process of each project has to instruct the lexicographers with regard to the type of microstructure to be employed in the dictionary. This decision may not be taken in an arbitrary way or be isolated from other decisions regarding the specific dictionary. As one of the most important ordering structures of a dictionary the microstructure should be seen as an instrument to help achieve the genuine purpose of the dictionary. Yet again, the typological classification of the dictionary, the users, their needs and reference skills, the situation of dictionary use and the functions of the dictionary should influence the decisions. The microstructure of any given dictionary should be a factor of all these features.

Although metalexigraphic research has formulated models for various types of microstructures, cf. Wiegand (1996), all these types are not as relevant for general dictionaries. However, in the planning of any new dictionary lexicographers should decide between at least two major types of microstructures, i.e. an *integrated* and a *non-integrated microstructure*. Within these categories, and applying to both these types of microstructures, Gouws (1999a, 2003) makes provision for a subdivision between *obligatory microstructures* and *extended obligatory microstructures*. The choice between these two subdivisions depends on the data distribution structure of a given dictionary. The distinction between the two major types of microstructures, i.e. integrated and non-integrated microstructures, is made on the grounds of the proximity and the directness of the relation between each entry representing a paraphrase of meaning (in a monolingual dictionary) or each entry representing a translation equivalent (in a bilingual dictionary) and the supporting cotext entries representing illustrative examples in the specific article. The relative positioning, i.e. the proximity, between the cotext entries and the respective core entry of each subcomment on semantics, i.e. the paraphrase of meaning/translation equivalent given for each polysemous sense of the lemma sign, cf. par. 8.3.1.2, determines the type of microstructure.

An *integrated microstructure* displays a close proximity between a cotext or context entry and the relevant paraphrase of meaning/translation equivalent, cf. the following articles from TAW and HAT respectively:

**dom** ❶ *slow* *He is clever at languages but a little slow at maths.* Hy is slim in tale maar 'n bietjie **dom** in wiskunde. ❷ *foolish, silly, stupid* *It is foolish/silly/stupid*

to go too deep into the sea if you can't swim. Dit is **dom** om te diep in die see in te gaan as jy nie kan swem nie. ... (TAW)

**da'nig** b.nw. en bw. **1** (bw.) Baie, erg; in hoë mate: *Danig verlief wees. Dit is nie so danig ver na hulle toe nie. Danig in jou skik wees met iets.* **2** Oorvriendelik: *Danig wees met die nuwe landdros. Danig met mekaar wees, duidelik verlief.* **3** (ietwat neerhalend) Van groot belang; waffers: *Jy hou jou verniet so danig. Gaan leen by jou danige vriende.* (HAT)

In both these articles the treatment is directed at a lemma representing a polysemous lexical item. Each sense of the polysemous item is treated in its own slot, i.e. its own subcomment on semantics. In the article from the bilingual dictionary TAW the source language item *dom* gets a translation equivalent *slow* for its first sense and three target language synonyms (*foolish, silly, stupid*) as translation equivalents for the second sense. Illustrative examples, presented as cotext entries, are provided to illustrate the typical use of these translation equivalents. The cotext entries are presented as immediate neighbours of the translation equivalents within the same subcomment on semantics. The treatment in the monolingual dictionary HAT shows a similar pattern. The lemma *danig* represents a polysemous lexical item with three different senses. Each one of these senses is presented by means of a paraphrase of meaning, either as a synonym or an explanation, in its own slot, i.e. its own subcomment on semantics. Within each subcomment on semantics the paraphrase of meaning is immediately followed by an illustrative example, presented as a cotext entry, to illustrate the typical use of the word represented by the lemma sign in that particular sense. In both TAW and HAT the translation equivalent/paraphrase of meaning and the respective cotext entries function within the same subcomment on semantics. This postulates a relation of close proximity between the translation equivalent/paraphrase of meaning and the respective cotext entries and makes it easy for the user to achieve a rapid retrieval of information regarding the translation equivalent or a specific sense of the lemma as presented in the paraphrase of meaning and the accompanying cotext entry. In these articles the translation equivalent/paraphrase of meaning and the respective cotext entries appear in a single search field of the article, the same subcomment on semantics. Such a search field is also known as an integrate. Where the translation equivalent/paraphrase of meaning and the respective cotext entries occur in the same *integrate* the dictionary article displays an *integrated microstructure*.

The fact that no other occurrences of paraphrases of meaning/translation equivalents come between a given paraphrase of meaning/translation equivalent and its cotext entry decreases the textual condensation and makes it easier for a user to interpret the contents of the subcomment on semantics correctly. Especially in the treatment of a lexical item with many polysemous senses, the direct relation between cotext entry and paraphrase of meaning/translation equivalent ensures an optimal retrieval of information. This is a type of microstructure ideally suited for monolingual and bilingual pedagogical, desk/college and standard dictionaries.

A non-integrated microstructure does not have the cotext or context entries in the same integrate as the respective translation equivalents or paraphrases of meaning at which they are addressed. Such a dictionary typically includes all the different subcomments on semantics in a text block and follows this with a separate text block which contains cotext entries. Although cotext entry and translation equivalent/

paraphrase of meaning are not presented in the same integrate, resulting in a non-integrated microstructure, the presentation and ordering of the cotext entries in the relevant text block are done in such a way that the relation between a given cotext entry and the item at which it is addressed is clearly marked, and the skillful user should have little difficulty in pairing the cotext entries with their relevant addresses, cf. the following constructed example from Gouws (2003):

**bak** ww. (het gebak)

**1** Gaarmaak deur hitte. **2** Gaar word; warm kry. **3** Hard laat word deur hitte. **4** Hitte afgee.

*Eiers in die pan bak (1). In die son sit en bak (2). Stene bak (3). Die son bak op die stoep (4).*

In this article the treatment of the polysemous word *bak* makes provision for four paraphrases of meaning. This text block with the different subcomments on semantics is followed by a next text block that contains all the cotext entries. Each cotext entry is followed by a number which relates it to one of the paraphrases of meaning presented in the preceding text block. Each cotext entry is addressed at a paraphrase of meaning which represents one of the senses of the word represented by the lemma sign. Although the numbers following the cotext entries explicate a relation to a specific subcomment on semantics this relation is not as clear as the relation prevailing in an integrated microstructure. From an addressing perspective one could call this a procedure of article-internal remote addressing because a given cotext entry and the item at which it is addressed are not in such close proximity as is the case in integrated microstructures. Because a non-integrated microstructure demands more dictionary using skills from the user, the lexicographer should only employ this type of a microstructure where it will not impede the dictionary consultation procedures of the intended target user.

A non-integrated microstructure is a specific type of microstructure and not merely a microstructure that is not integrated, cf. Gouws (2003). The following examples come from GW:

**bot** heid, bluntness; stupidity; obtuseness; lethargy; leaden-heartedness. (GW)

**ei** e, own, private; personal; natural, native; peculiar, specific, distinct, friendly, familiar, intimate; ~ **AAN**, peculiar to; *uit* ~ **BEWEGING**, of one's own free will (one's own accord); *jou* ~ **DING doen**, do your own thing, do what you want to do; be independent; ~ **DOEN** ~ *geen skade nie*, dog does not eat dog; *die* ~ **EK**, the self; ego; *vir* ~ **GEBRUIK**, for one's private use; *in* ~ **GELD**, in the local currency; *IETS wat hom* ~ **is**, something characteristic of him; *in* ~ **KRING**, in one's own circle, privately; *jou* ~ **MAAK**, master something; get the better of something; make something one's own; ~ *wees* **MET iem.**, be intimate with someone; *op* ~ **NAAM**, in one's own name; *hulle is* ~ **NIGGIES (neefs)**, they are first cousins; *in* ~ **PERSOON**, personally, in person; *'n* ~ **SAAK begin**, set up business on one's own account; ~ **STABILITEIT**, inherent stability; *sy was 'n nooi N. VAN haar* ~, her maiden name was N.; ~ **WEES aan iets**, be characteristic of something; *op jou* ~ **WERK**, work on your own (alone); *sy* ~ **WOORDE**, his very words; (GW)

The first example contains no less than five subcomments on semantics with a single translation equivalent in each one of these subcomments on semantics, indicating that the word represented by the lemma sign has five polysemous senses. This

article does not display an integrated microstructure because none of the integrations contain any cotext entries. This does not, however, imply that the article has a non-integrated microstructure. A non-integrated microstructure presupposes the occurrence of cotext entries presented in a well-systematised way in a separate text block. This example from GW demonstrates a rudimentary microstructure, i.e. a microstructure that contains less than the minimum entries to make the retrieval of information successful. Such a rudimentary microstructure forms the basis for the formation of either an integrated or a non-integrated microstructure.

The second example contains a text block in which a number of equivalents are given for the polysemous word *ie*. A separate text block contains cotext and other entries. This presentation of a text block with the translation equivalents followed by a text block with the cotext entries does not qualify this example for a non-integrated microstructure. One of the criteria for qualification as a non-integrated microstructure is that the cotext entries must be marked to show a direct relation with a specific subcomment on semantics. No such marking is seen in this article and the user gets no assistance to relate a cotext entry to a specific translation equivalent. This article leads to further confusion because the text block containing the cotext entries give a mixed presentation of cotext entries and fixed expressions. This presentation can be characterised as a primitive microstructure, cf. Gouws (2003), Gouws & Wiegand (to appear). Lexicographers should avoid this confusing type of microstructure because it offers limited guidance to the user. The microstructure of a dictionary needs to be planned with precision and care in order to offer the users an optimal chance of successful dictionary consultation procedures.

The distinction between *obligatory microstructures* and *extended obligatory microstructures* has to do with the data distribution structure and the extent of the data categories to be included in an article. When planning a dictionary the lexicographer should decide on the minimum contents every article should contain. This establishes the microstructural contents of the default article. The *obligatory microstructure* refers to the microstructural items that will be found in each and every article. In a monolingual dictionary the obligatory microstructure may contain e.g. the lemma sign, the item giving the part of speech, a paraphrase of meaning presented as a lexicographic definition for each one of the polysemous senses of the word represented by the lemma sign, and an illustrative example presented as a cotext entry to illustrate the typical use of the word. These data categories will prevail in each and every article that displays the obligatory microstructure. All the single articles in a dictionary, cf. par. 7.2.3.1, will display at least an obligatory microstructure.

The data distribution structure should make provision for additional items and data categories that might be extremely important in the treatment of certain lexical items. All words do not take morphology but in the treatment of a noun it is important to have an article slot where e.g. the plural form of the noun can be indicated and in the treatment of a verb it may be necessary to have an article slot where the past tense of the verb is indicated. In the comment on semantics the inclusion of synonyms and antonyms will not be relevant to all words because all words do not have synonyms and antonyms. However, where a word does enter into a semantic relation of synonymy with other words the dictionary user may want to retrieve that kind of information from the dictionary. Consequently an article slot should be

available to accommodate this type of data category. These additional items or data categories are not included in all articles but are included where they are relevant to a given word and in these articles they could be regarded as obligatory. The articles in which these additional data is accommodated still displays the obligatory microstructure of the default articles because they have to include all those data types. However, the microstructural presentation in these articles has been extended to include more than the obligatory categories of the default articles. Where the microstructural presentation in an article includes more than the default categories such an article displays an *extended obligatory microstructure*. In many articles an extended obligatory microstructure merely leads to a presentation of all the data categories needed to ensure an adequate treatment of the lexical item represented by the lemma sign in terms of the minimum requirements of the specific dictionary. The microstructure in the article of a lemma sign representing a preposition will differ from the microstructure in the article of a lemma sign representing a noun. The data types included in the article of the preposition will also be found in the article of the noun. The article of the noun will most probably include a few data types necessary for the minimum treatment of the noun but not relevant to the treatment of the preposition. The article of the preposition will display the obligatory microstructure whereas the article of the noun will display an extended obligatory microstructure. Although the obligatory microstructure is identical in all articles the extended obligatory microstructure, where it prevails, is not the same for all articles. The part of speech of the lexical item represented by the lemma sign in the specific article but also the network of semantic relations of a given word will determine the extent of the extended obligatory microstructure. A comparison of the treatment allocated to two nouns may show that both articles display an extended obligatory microstructure because of, e.g., the occurrence of a search field for items giving the plural form. Such a comparison may also show that the treatment differs because no synonyms or antonyms may have been entered in the one article whereas the other article includes items giving synonyms and antonyms. The nature and extent of an extended obligatory microstructure relies on the nature of the lexical item represented as primary treatment unit in the article.

Where the data transfer in an extended obligatory microstructure still falls within the typical treatment allocated to a given type of lexical item, e.g. that the treatment in the article of a lemma sign representing a noun should include data on the possible plural form of that noun, or that the treatment of a lexical item that has synonyms or antonyms should include entries giving these synonyms or antonyms, the extended obligatory microstructure is still accommodated in a single article, cf. par. 7.2.3.1.

An extended obligatory microstructure does not necessarily imply a complex article but all complex articles display an extended obligatory microstructure because the extended obligatory microstructure makes provision for the inclusion of those additional data categories, items or more comprehensive treatment within a given search area that differentiates a complex article from a single article. This addition can be done by means of e.g. article-internal inserted inner texts, boxes with lexicographic commentary or a treatment characterised by a stronger encyclopedic approach in the comment on semantics to assist users in specific situations of dictionary use.

## 9.1 Introduction

An appropriate way of starting a chapter on lexicographic definitions, also referred to as the paraphrase of meaning, might be a brief look at the definitions of *define*, *definition*, *paraphrase* and *meaning* given in dictionaries.

MED

**define** ... to explain the meaning of a word: Defining the word 'love' can be very difficult. Manual work is broadly defined as work that you do with your hands. ...

OALD

**definition** ... a statement giving the exact meaning of a word or phrase: *Definitions should not be more difficult to understand than the words they define.*

COBUILD 2

**paraphrase** ... If you **paraphrase** someone or **paraphrase** something that they have said or written, you express what they have said or written in a different way.

**meaning** ... The **meaning** of a word, expression, or gesture is the thing or idea that it refers to or represents and which can be explained using other words.

Lexicographers constantly strive to enhance the quality of definitions in monolingual dictionaries to best suit the needs and level of their target users. Landau (2001:162) emphasises that the definition must define and not just talk about the word or its usage. It must answer the question “what is it” that is defined, and it must answer it directly and immediately. As the given example in the article of *definition* above suggests, the lexicographer should strive to describe the meaning of a given lemma in easier language than the lemma itself. This issue is somewhat controversial, i.e. whether it is always possible to honour such a principle, but the basic aim not to use words that are more difficult from the viewpoint of the target user than the lemma to be defined, is sound and should always be kept in mind.

The MED uses a defining vocabulary of 2,500 words, presented in the back matter, which are regarded as the most common words in English. The selection is based on frequency in large corpora consulted by the compilers. Consider the following extract from the alphabetical stretch D in MED’s defining vocabulary.

damage, damaged, dance, dancer, danger, dangerous, dark, darkness, date, daughter, day, dead, deaf, deafness, deal, deal with, death, debate, decay, December, decide, decision, ...

It is clearly stated in MED that it is not possible for a dictionary to restrict its defining vocabulary to such a limited set and it is sometimes necessary to use additional words as well. Landau, while admitting the virtues of Zgusta’s rule which states

that, words which are more difficult than the word defined, should not be used, also indicates that it is often impossible to apply this rule.

What should be avoided, say for the average user of a general dictionary, are definitions such as the one for *feather* from *Webster's Ninth New Collegiate Dictionary* (W9):

any of the light epidermal outgrowths that form the external covering of the body of birds and that consist of a shaft bearing on each side a series of barbs which bear barbules which in turn bear barbicles commonly ending in hooked hamuli and interlocking with the barbules of an adjacent barb to link the barbs into a continuous vane.

Here a common word is defined by means of words of which many will probably be unknown to the user of a general dictionary such as *epidermal*, *barbs*, *barbules*, *barbicles*, *hamuli* and *vane*. The user will have to look up the meanings of each of these words in the hope to eventually find out what *feather* means, but they in turn will probably be defined using new words which are also incomprehensible to the user at his/her level. This clearly illustrates that the lexicographer can easily err in his or her compilation of a dictionary in many ways, which results in definitions that are either too difficult, like the one mentioned above, or definitions which are too basic for the target user.

Consider MED's definition of *feather*.

**feather**<sub>1</sub> / . . . / A noun [count] \*  
one of the narrow tubes with thin soft hairs on each side that cover a bird's body

The word must be defined in such a way that the user will get all the answers to the question that made him or her consult the dictionary. Laufer (1992: 71) calls it the objective *knowing* a word:

Knowing a word would ideally imply familiarity with all its properties ... When a person "knows" a word, (s)he knows the following: the word's pronunciation, its spelling, its morphological components, if any, the words that are morphologically related to it, the word's syntactic behaviour in a sentence, the full range of the word's meaning, the appropriate situations for using the word, its collocational restrictions, its distribution and the relation between the word and other words within a lexical set ... The foreign language learner knows a much smaller number of words ... In many cases word knowledge is only partial, i.e. the learner may have mastered some of the word's properties but not the others.

In COBUILD it is stated that, "users expect more and more from their dictionaries, and in particular want to gain confidence in using a word". The latter remark also underlines the responsibility of the lexicographer to supply enough productive (encoding) information to the user, and even more important, that this information should be on the level of the user. In Collins Cobuild (xviii-xix) it is firstly emphasised that typical collocates of a word should also be shown in the definition. The example of *savoury* is cited typically co-occurring with food rather than with something else.

Secondly it is claimed that their definitions give information about the grammatical structures that a word is used in and thirdly information about context and usage.

## 9.2 Basic types of definitions

### 9.2.1 Genus and differentiae definition

In such definitions it is firstly attempted to distinguish a semantic category/class or superordinate to which the lemma belongs (the genus) and then to list a number of specific characteristics of the lemma that differ from other members of this category or class (the differentiae). Consider the following example:

**apple** ... An **apple** is a round fruit with smooth green, yellow, or red skin and firm white flesh. ...

**avocado** ... are green pear-shaped tropical fruit. They have hard skins and contain large stones. (COBUILD 2)

In this example the genus *fruit* is identified for both *apple* and *avocado* but they differ in terms of the one being *smooth, green, firm white flesh, etc.* and the other *pear shaped* and *tropical*... Hartmann (1983:90) says:

... the methodology of the lexicographer's semantic analysis is basically contrastive. In order to isolate, it is essential first to compare, to set up a range of lexical items that across a spectrum of semantic features match exactly in many particulars, so that areas of contrast are thrown into relief.

### 9.2.2 Synonym definition

A synonym, i.e. a word that has the same meaning as another word, is given instead of a full explanation.

**sidewalk** pathway

The comment on semantics for the lemma *sidewalk* is limited to giving the synonym *pathway*. In such cases the meaning of the lemma is not explained but a cross-reference is given to a lemma that will be defined more comprehensively. This means that if the user knows the meaning of the synonym, successful information retrieval has taken place and if not, the required information can be found in the article of the synonym, *pathway* in this case. The lexicographer, as in the case of all other instances of cross-referencing, has to ensure that the latter lemma is indeed included and treated in the dictionary.

### 9.2.3 Circular definition

In the case of circular definitions the lemma is partially defined in terms of itself which can result in definition failure if not handled correctly.

MED

a. **archaeologist** ... someone who studies archaeology

b. **geologist** ... a scientist who studies geology

In the a-example *archaeologist* is simply linked to *archaeology*. That will be of no use to the user of the dictionary unless (s)he knows the meaning of *archaeology* or consult the article of the latter. Once again it is imperative that a full treatment of *archaeology* should be given in the appropriate alphabetical stretch. In the b-example the inclusion of *scientist* in the comment on semantics at least puts *geologist* in a semantic class but the circularity still lies in *geology* which needs to be explained.

The choice of definition type is largely arbitrary in the sense that no strict guidelines can be laid down, say lemmas which have X-characteristics should be defined by means of genus and differentiae definitions and those with typical Y-characteristics by means of say, circular definitions. Taken at face value it can be expected that *-ologists* such as *archaeologist*, *geologist*, *ecologist*, etc. would be defined using the same strategy. However MED used different types of definitions.

**ecologist ...**

- 1 a scientist who studies the environment and the way plants, animals, and humans live together and affect each other
- 2 someone who believes that protecting the environment is important

In the case of closely related sets such as the days of the weeks and the months of the year it is advisable that a specific defining strategy should be applied consistently.

**MED**

**Monday ...** the day after Sunday and before Tuesday ...

on Mondays (=every Monday): The bank is open later on Mondays and Fridays.

that Monday morning feeling used for saying that you wish that it was still the weekend and that you did not have to go to work

**Tuesday ...** the day after Monday and before Wednesday ...

on Tuesdays (=every Tuesday): We close early on Tuesdays.

**Wednesday ...** the day after Tuesday and before Thursday ...

on Wednesdays (=every Wednesday): I never work on Wednesdays.

For each day of the week one fixed positioning phrase *the day after ... and before ...* is used as well as a fixed strategy for the plural *on ... days (=every ... day)* supplemented by a number of typical examples of usage. Where a specific use/connotation in relation to a day is applicable, it is included, e.g. *that Monday morning feeling*. This brings some kind of consistency in the defining strategy.

In MED's back matter it is clearly stated that only the most basic and central meanings are used in the definitions. They cite for example *cold* for which the sense 'unfriendly' was not considered.

Pictorial illustrations, or ostensive definitions can also be utilised fruitfully in the defining process especially in cases where it is difficult to clearly explain the meaning of a lemma by means of the descriptive definition.

**MED**

**skip1 ... 1** [intransitive] to move forwards by jumping first on one foot and then the other: Julie skipped along the pavement.

1a. to jump over a rope that you or two other people swing above your head and then under your feet:

This definition is fine but a user who is totally unfamiliar with the concept will benefit much from the added pictorial illustration:



Figure 1: Pictorial illustration of *skip* in MED

### 9.3 Enhancing the quality of definitions

Lombard (1991:166) identifies a number of defining criteria that would result in good definitions namely *completeness, clarity, accuracy, consistency, independency, objectivity* and *neutrality*. Consistency has been illustrated by means of definitions for the days of the week above but the other criteria would be elaborated on in more detail.

As far as *completeness* is concerned one could say that all details, features and characteristics necessary to understand what the word means should be given. Lombard (1991) warns against incompleteness on the one hand and over-completeness and over-specifying on the other. A definition in terms of Lombard is incomplete if it does not give enough information about the lemma. A typical example is where a specific tree or bird is simply defined as ‘a kind of tree’ or ‘a kind of bird’ respectively as in the following example.

**Motshwere** mohuta wa setlhare se se dikgong di thata (THAN)

Defining *Motshwere* simply as ‘a kind of tree with hard wood’ might go some way but is unlikely to satisfy the user’s needs. It does not distinguish this tree from many others that have hard wood. The flipside of incompleteness is over-completeness where too much extra-linguistic information is given.

**tennis** ... a game in which two or four people use RACKETS to hit a ball across a net. A player scores a point when their opponent cannot hit the ball back. Tennis is played on a tennis court. (MED)

**tennis** ... Tennis is a game played by two or four players on a rectangular court. The players use rackets to hit a ball over a net which is placed across the middle of the court. (COBUILD 2)

**tennis** s.nw. (< F.) Spel vir twee of vier spelers wat 'n bal met 'n raket heen en weer oor 'n net slaan. **tennis**: ~bal, ~klub, ~net, ~raket, ~skoel, ~spel, ~speler, ~toernooi, ~wedstryd. (HAT)

**tennis** ... (also **lawn tennis**) ... a game in which two or four players hit a ball with racket ... backwards and forwards across a net on a specially marked playing area ... □ picture (OALD)

A comparison of these four articles reveal that core elements of the definition of *tennis* are *game, ball, racket(s), players, playing* and *area/court*. The fact that the court is rectangular, the way points are scored, that the net is placed across the middle of the court and that the court is a specially marked area are extra-linguistic information which could be regarded as not essential to the definition of *tennis*. Detailed specification of the size of the court, the height of the net, the position of the umpire's chair, etc. could be regarded as over-completeness. It has, however, to be kept in mind that the modern lexicographer tends to be more tolerant towards the inclusion of such encyclopedic or extra linguistic information. However, it should be limited to what is really functional in each case. OALD included a pictorial illustration at the article of tennis indicating a lot of such extra-linguistic information.

As far as *over-specifying* is concerned, it is true that the definition can be too specific. For example, Lombard (1991:171) cites *kruk* 'crutch', being defined as 'made of wood' and rightly says it is too specific and can easily become dated since most crutches are currently made of metal rather than wood.

*Clarity*, simply means that a word can be easily understood, thus the question can the definition be easily understood? Lombard (1971:172) says respondents' feedback suggest that definitions should be less complicated, less detailed, and not repetitive or fragmented.

As for *accuracy* the lexicographer should ensure that the definitions are factually correct. Lombard (1991: 177) cites a number of interesting examples of incorrect definitions where it is not the case such as the definitions of *lion* which incorrectly states that it is the biggest catlike animal or that *malaria* is caused by a mosquito.

*Independency* of a definition means that it will not be necessary for the user to consult more than one definition to obtain the meaning of a specific lemma or sense. As explained in Chapter 12 the purpose of following up on cross-references is to obtain more information about the lemma and not to accumulate enough information in order to understand what the lemma means.

Finally, as for *objectivity* and *neutrality* Lombard (1991: 180) sums up this complicated issue by stating that definitions should be without personal preferences and judgment by the lexicographer, and should not have ideological, racist, religious or sexist connotations.

These guidelines could be very useful to the lexicographer as measures and counter-measures to employ and to note in the compilation of lexicographic definitions.

## 9.4 Electronic corpora as the key to writing better definitions

The availability of corpora and the possibility of studying every lemma sign in context prior to the compilation of a definition revolutionized dictionary compilation. Firstly, utilisation of a main or 'general' corpus such as the Pretoria Sesotho sa Leboa Corpus can help the lexicographer to write definitions for the typical general user of the dictionary. However, in addition to the main corpus, dedicated sub-corpora

comprising of a representative sample of reference works used by different target user groups will give a clear indication of the required level of compilation for such target users. This means that the different sub-corpora and in particular, the corpus lines studied, will reflect the level on which the definition should be compiled right from the start.

Concordance lines generated from corpora constitute a useful if not essential tool for the lexicographer. When dealing with a word, often a single glance at a number of concordance lines helps the lexicographer to determine different senses of the word and goes a long way in supplying useful information for the defining process. Consider the following two examples for English and Sesotho sa Leboa.

**Table 1: Concordance lines for *bank* in Pretoria English Corpus PEC**

ye. On one occasion the canoe neared a	bank	on which a large flock was sitting.
So you don't want to be a	bank	robber?
r enemies; but in your canoe, behind a	bank	of reed, nothing can harm you." Th
The moon is disappearing behind a	bank	of gray clouds. EXT.
She stops. Walks to him, framed by a	bank	of t.v. monitors.
ned. I received a letter, containing a	bank	draft for £2, from a friend to who
y meals, and even get real money from a	bank	. *The absolute best method is to
Problem. She doesn't have a	bank	account.
ness. "Why, yes; I was employed in a	bank	at one time. I think I told you t
t a white light seemingly from a cloud-	bank	far away in front of us. It dispens
the Wachovia Loan and Trust Company's	bank	. As we did so, John said:
s, which had been thrust out from each	bank	into the stream, had been driven fi
of bluffs on the opposite side, or east	bank	, of the river. As soon as the ene
of Thebes. On the right, the eastern	bank	of the Nile, rose the buildings of
nd his desk watching the battle on his	bank	of monitors.

Such concordance lines can be used for sense distinction and the writing of a definition for each sense and the selection of appropriate examples to complement the definition in each case.

**Bank ... 1** a financial institution that people or businesses can keep their money in or borrow money from. The main banks used by ordinary people are called high-street banks:

Marge works for the Royal Bank of Scotland.

a New York investment bank

1a. an office of a bank:

I need to go to the bank this morning.

1b. [only before noun] belonging to or connected with a bank:

a bank manager/loan/robbery

2 a raised area of land along the side of a river:

A man was fishing on the opposite bank.

bank of: The village lies on the east bank of the river Derwent.

2a. a long area of land with sloping sides:

There was a steep bank looking onto the football ground.

2b. a long pile of earth, snow, or sand:

bank of: Great banks of snow and ice lined the roads.

2c. a large mass of cloud or FOG

- 3 a large number of things in a row, especially pieces of equipment:  
 bank of: a bank of TV monitors  
 3a. a large collection, especially of information or ideas:  
 bank of: The library has a valuable bank of old documents.  
 an impressive data bank  
 3b. a store of something that is available for use when it is needed:  
 a blood/gene/sperm/organ bank (MED)

Consider also a similar strategy for the Sesotho sa Leboa word *gare*.

**Table 2: Concordance lines for *gare***

Suzan a bego a dutše. Ka <i>where Suzan was sitting. In</i>	gare	ga seswaramelora, sekerete se Suzan a <i>of the ashtray, the cigarette Suzan was</i>
goba a ka yo botšiša mang <i>Or who can he ask in</i>	gare	ga bošego bjoo. <i>of that night</i>
lešata. Lekota a swaba ge a bona <i>a noise. Lekota was disappointed when he saw that</i>	gare	ga bona go na le ba babedi ba go <i>them were two who</i>
Diphapano tše bohlokwa <i>Important differences</i>	gare	ga Sepedi le Setswana ... Diphapano tša medumo <i>Sepedi and Setswana. Differences of sound</i>
Pharologano ya bobedi <i>The second difference</i>	gare	ga seema le seka <i>a proverb and an idiom</i>
Hlaloša phapano <i>Explain the difference</i>	gare	ga kgalapatšo le tengwafatšo. [4] <i>palatalization and velarisation</i>
Lesibana a thabela go hwetša Tholo <i>Lesibana was glad to find Tholo</i>	gare	ga banna. A re, "Aowa, Tholo, o monna. <i>the men. He says, "No, Tholo, you are a man</i>
Mahlo a yona a be a gadima ka <i>His eyes were glittering</i>	gare	ga leswiswi ge e be e boela go <i>the dark when he returned to</i>
ka 1977-03-31. Gona bjale sekolo se <i>in 1977-03-01. Now the school is</i>	gare	ga mehlare, matšoba a mehutahuta <i>the trees, different kinds of flowers</i>

*gare* [1] (*lehlathi*) **1** lehlathafelo leo le hlathago lefelo leo le lego makgatheng a dilo tše pedi goba go feta : **Lesibana a ema gare ga banna ka dihlong**; **2** lehlathafelo leo le hlathago lefelo leo le lego ka go se sengwe : **Ka gare ga moago wa mohuta woo go dula mahodu le babolai**; **3** lehlathafelo leo le hlathago tiragalalo le seemo seo se sa tšwelago pele : **Ke hlabile mampša gare ga lesolo** (PUKUTLH)

There is no doubt that studying concordance lines could be very useful in the writing of definitions. Compilers should however be cautioned against regarding every second concordance line as 'a new sense'. Well chosen examples of usage are crucial in the treatment of a lemma and should complement the definition by actually extending it. The examples given for the lemma *gare* taken from the corpus, illustrate this point well because they further clarify and contextualise the descriptive definitions for senses 1-3.

# Equivalent relations in bilingual dictionaries

---

## 10.1 Introduction

Bilingual dictionaries can be regarded as one of the typological categories most frequently used by the average member of a speech community. This is especially true in a multilingual environment. When discussing different types of dictionaries it is important to realise that in spite of the category-specific features of any given dictionary, many features, both in terms of structure and contents, are mutually shared by dictionaries belonging to different typological categories. In general monolingual and bilingual dictionaries the article structure is one of the typical features shared by both these dictionary types. The lemma sign functions as guiding element and main treatment unit of the article and all the data entries in the article have been positioned in either the comment on form or the comment on semantics.

That bilingual dictionaries share many features with monolingual dictionaries becomes clear when one looks at the data types on offer in these two types of dictionaries and at the way in which they are presented. In the comment on form of both these dictionary types one would typically find entries like items giving the pronunciation of a word and those presenting morphological data, e.g. an indication of affixes of pluralisation. These dictionaries are often consulted to find the correct spelling as indicated by the lemma sign. The main difference between these two typological categories lies in the comment on semantics where one can identify the core focus of the respective articles as texts. Although the comment on semantics in both these dictionary types makes provision for an article slot accommodating illustrative examples the main focus in the comment on semantics is on two different types of data categories. In a general monolingual dictionary the focus is on the paraphrase of meaning, presented as a lexicographic definition. In a general bilingual dictionary the focus is on the items presenting translation equivalents for the word represented by the lemma sign. In a monolingual dictionary it is often necessary to give more than one paraphrase of meaning in a single article in order to provide for the different polysemous senses of a word represented by a given lemma sign, cf. the following example from the HAT:

**effektief** <sup>ˆ</sup> b.nw. en bw.

**1** Wat die gewenste effek, uitwerking het; doeltreffend: *Effektiewe maatreëls toepas. Effektiewe medisyne.* **2** Treffend: *'n Effektiewe beskrywing.* **3** Werklik: *Die effektiewe sterkte van 'n leër.*

In this article no less than three paraphrases of meaning are given to account for the polysemous nature of the Afrikaans word *effektief*. The comment on semantics includes no less than three subcomments on semantics to make provision for the presentation of these different senses of the treated lexical item. The comment on semantics in a bilingual dictionary should also make provision for a treatment procedure which has all the polysemous senses of a lexical item in its scope. Polysemy is a word specific feature which implies that for a polysemous word

in the source language one will not necessarily find a target language translation equivalent with exactly the same polysemous senses. In lexicographic practice it boils down to a situation where a lexicographer often has to provide a separate translation equivalent for each one of the polysemous senses of a lemma. The Afrikaans word *bak* is polysemous and the GW makes provision for the treatment of the different polysemous senses by including different translation equivalents:

**bak**<sup>4</sup>, (s) (-ke), basin, bowl; vat, trough; body (of a carriage); (snake's) hood; dustbin; ash bucket; parterre, pit (theatre); forecastle (fo'c's'le); hutch (ore);

In this article the comment on semantics accommodates a whole range of translation equivalents. The collection of translation equivalents, whether one or more than one, presented in the comment on semantics of a single article is referred to as the *translation equivalent paradigm* of the given article. It is the duty of the lexicographer to make sure that the target user of a given dictionary can achieve a successful retrieval of information from a translation equivalent paradigm. Where a translation equivalent paradigm consists of more than one member the lexicographer may not rely on the intuition of the target user to make the correct choice of equivalents.

## 10.2 Translation equivalent versus paraphrase of meaning

When users consult a bilingual dictionary they often work with the assumption that this type of dictionary offers them data to be regarded as the meaning of the source language item, represented by the lemma. Dictionary users typically see the translation equivalents as "the meaning of the source language word in the other language", i.e. the translation equivalent represents to them an indication of the meaning of the source language form (the lemma) given in the target language by means of a brief paraphrase of meaning. They seldom realise that the data on offer is not a paraphrase of meaning or even a statement about meaning but it is rather a list of translation equivalents, cf. Gouws (1996), Louw (1985:53, 54). Compare the following articles from the POD, HAT and GW:

**chair** —*n.* **1** seat for one person usu. with a back. **2** professorship. **3** a chairperson. **b** seat or office of a chairperson ...(POD)

**stoel** □ s.nw. (-e; -tjie)

**1** Meubel vir een persoon om op te sit, gewoonlik met 'n rug en vier pote en dikwels met twee armléunings: *Neem 'n stoel en gaan sit. 'n Harde, sagte, beklede stoel. 'n Bank en twee stoele koop vir die sitkamer. Arm-, kinder-, leun-, leuning-, russtoel.* **2** Vasstaande setel by 'n amp gebruik; vandaar ook, die amp self of die persoon wat die amp beklee: *Die regter spreek na hom af uit sy stoel. Die stoel aanspreek, die voorsitter. Vrae uit, aan die stoel stel. 'n Stoel (in chemie, geografie, Afrikaans-Nederlands) beklee, professor in die vak. Die pouslike stoel, die stoel van Rome, die Heilige Stoel, die pouslike gesag. Bieg-, preek-, regterstoel.* **3** Onderstel waarin of waarop iets rus: *Spoorstawe word op stoele van gegote yster gelê. Die stoel van 'n balans, van 'n globe, e.d.* **4** Gedeelte van 'n plant met stronk en stingels, gewoonlik net bogronds; krop; pol: ... (HAT)

**chair**, (n) stoel, gestoelte; setel; voorsitter(stoel); leerstoel, professoraat; plat (spoor); kagebint; stoelbed (vir tappe); ... (GW)

**stoel**, (s) (-e), chair, seat, stool; see (of bishop); pedestal; stool (plant); ... (GW)

In the first article *chair* is treated as a polysemous lexical item with three distinct senses. For each sense a lexicographic definition presents the relevant paraphrase of meaning. In a similar way the second article treats the Afrikaans word *stoel* as a polysemous lexical item. It is, however, interesting to note that although *chair* and *stoel* share some of the polysemous senses, they also show some differences. In their comment on semantics the third and fourth articles contain translation equivalent paradigms which accommodate a range of translation equivalents. Some of these equivalents represent target language synonyms whereas some represent target language forms for the different polysemous senses of the source language form. Comparing these articles from a bilingual dictionary with the articles from the two monolingual dictionaries, it is clear that there is no one to one relation of polysemy holding between the Afrikaans word *stoel* and the English word *chair*. The word *stoel* has translation equivalents like *chair*, *see*, *stool* and *pedestal* to represent some of its polysemous senses. From a semantic perspective it would be wrong to argue that anyone of these equivalents can be regarded as the meaning of the word *stoel*. The Afrikaans word *stoel* does not mean *pedestal* but in a specific environment it could be translated with the word *pedestal*.

In the teaching of dictionary using skills and the nature, function and contents of different types of dictionaries, users should be made aware of the fact that the nature of the core entries in a bilingual dictionary differs from that of the entries in a monolingual dictionary. Translation equivalents may not be regarded as entries giving the meaning of the lemma but they should be seen as target language lexical items that may be used to substitute the source language item in a specific situation. However, it should be emphasised that the use of these translation equivalents will be determined by the context and the cotext of the source language item. Consequently it is of extreme importance that the lexicographic treatment presented in a bilingual dictionary may not leave the translation equivalents isolated from their typical contexts and cotexts. Where these supporting entries are not given as part of the lexicographic treatment, the users are at a loss in their attempts to retrieve information that can lead to a successful use of the target language forms.

### 10.3 Relations between lemma and translation equivalents

Translation equivalents are presented as part of the treatment of the lemma, functioning as guiding element of a particular dictionary article. In the article of a bilingual dictionary the dominant addressing procedure is one of lemmatic addressing. The lemma is the primary target of the treatment with the majority of microstructural entries directed at the lemma. Procedures of non-lemmatic addressing also prevail but to a far lesser extent than the procedures of lemmatic addressing. This will be discussed in a following section. Between the lemma as address and the translation equivalent paradigm as a collection of entries addressed at the lemma a relation of translation equivalence exists. The direction of this relation is from the source language form to the target language forms. This relation of translation equivalence holds between the lemma and all the translation equivalents in the translation equivalent paradigm. Different types of translation equivalent relations can be found and within one single article the relation between the lemma and the members of the translation equivalent paradigm can represent

different types of equivalent relations. These different types will be discussed in the next section.

## 10.4 Different types of equivalent relations

Aspects regarding the different types of equivalent relations have been discussed extensively in various publications, e.g. Gouws (1989; 1996; 2000; 2002a), Trautmann (1993). One golden thread going through the discussions is the fact that lexicographers have an obligation towards their users in ensuring a presentation and treatment of translation equivalents that will enable an unambiguous retrieval of information from the data on offer in the comment on semantics of a bilingual dictionary. The proper presentation and treatment of translation equivalents prerrequisite a clear understanding of the different types of equivalent relations.

The attempts of the lexicographer to co-ordinate source language items with target language items that can be used to substitute the source language forms in a translation, leads to the recognition of three major types of equivalent relations, i.e. full equivalence, partial equivalence and zero equivalence. These different relations of equivalence confront lexicographers with different challenges to ensure that the users will be able to achieve an optimal retrieval of information from a given dictionary article. Due to the fact that so many dictionary articles do not display a one to one relation between the source and target language items, lexicographers are compelled to include additional entries as supporting material in order to assist the user to make an informed choice when selecting the appropriate translation equivalent for a given occurrence of the source language item.

### 10.4.1 Full equivalence

Full equivalence prevails where a source language item, represented by the lemma sign, is co-ordinated with a single target language item, represented by a translation equivalent, and this one to one relation exists on both a lexical and a semantic level. This type of full equivalence is also known as *congruence*. The source language item and the target language item have exactly the same meaning, function on the same stylistic level and represent the same register. This implies that the target language item can be used as a translation equivalent of the source language item without any restrictions. The following example from NeW illustrates such a relation of full equivalence:

**bon-i-ness** *n.* *benerigheid*.

According to this presentation the English word *boniness* can be translated with the Afrikaans form *benerigheid* in all its occurrences.

Due to the fact that source and target language items have exactly the same meaning the lexicographer does not have to add too many entries to assist the user in selecting the proper target language item for a given source language item. Depending on the function(s) of a dictionary the existence of full equivalence could lead to the presentation of the translation equivalent without any contextual or cotextual entries in the comment on semantics. This is especially true in a dictionary with text reception as its primary function. The absence of any supporting entries would imply that the source and target language forms are equal in all aspects.

Full equivalence does not always demand a one to one relation between source and target language. The target language may have two absolute synonyms as equivalents for a given source language item. This will still be a form of full equivalence although no longer a relation of congruence but rather a subtype of divergence, cf. paragraph 10.4.2 for a discussion of the relation of divergence.

#### 10.4.2 Partial equivalence

Partial equivalence prevails where the source and target language items do not display a one to one relation. This type of equivalence manifests itself in different ways and confronts lexicographers with a challenge in their attempt to assist dictionary users in achieving an unambiguous interpretation of the relation between the relevant source and target language items. Partial equivalence typically occurs where there is not a one to one relation between source and target language items. This can be on the lexical or the semantic level or on both lexical and semantic levels. However, partial equivalence may also prevail where a one to one relation exists between the source and target language but it only applies on the lexical level and not on a semantic level. On a lexical level a one to one relation may prevail between source and target language when a given target language item is the only possible candidate to function as translation equivalent of the source language item but where the two items do not share the exact meaning. Consequently the lack of a one to one relation on semantic level leads to a relation of partial equivalence. This lack of a one to one relation on semantic level can be the result of the source language item being monosemous whereas the target language item may be polysemous and functions as equivalent for the source language item because one of its senses overlaps with the only sense of the source language item. In GW the English lemma *debilitate* only has one translation equivalent, i.e. the Afrikaans word *verswak*.

**debil`itate, verswak.**

*Debilitate* is a monosemous item and *verswak* is the Afrikaans form that conveys the meaning of *debilitate*. The Afrikaans form *verswak* is not monosemous. This is clear from the treatment of this word in GW:

**verswak`**, (-), weaken (eyes); enervate, attenuate, devitalise, enfeeble, debilitate; run down (battery); **~te kragte**, weakened strength; ..

Although the treatment of *debilitate* is done by means of a one to one relation it does not represent a form of congruence but rather partial equivalence. To assist the user, the lexicographer should have provided additional entries, e.g. illustrative examples, to give a clear indication of which sense of *verswak* is activated when it is presented as translation equivalent for *debilitate*. The unsatisfactory treatment given for the lemma *verswak* will not be discussed in this section. This type of problematic treatment is discussed in more detail in paragraph 10.5.

Another situation of partial equivalence occurs where a one to one relation on the lexical level does not have a parallel on the semantic level because the source and target language items do have the same meaning but do not fall within the same register, cf. the following example from the TFW:

**lepel ..... so lank as die lepel in die pappot staan** \* never say die

According to the system of this dictionary, as explained in the user's guidelines text, the star following the Afrikaans idiom indicates that it is an informal form. The English equivalent goes unmarked, implying that it is a neutral form. Although a one to one relation exists it represents partial equivalence because the source language form cannot substitute the target language form in every occurrence. The dictionary does well to inform the users of this difference in register.

A frequent type of partial equivalence sees the establishment of a one to more than one relation between source and target language. Such a type of partial equivalence is also known as an equivalent relation of *divergence*. A one to more than one relation, i.e. divergence, does not necessarily imply partial equivalence. One source language item can have two target language items as translation equivalents where these two translation equivalents are synonyms in the target language, cf. the discussion in paragraph 10.4.1. This one to more than one relation could represent a subcategory of full equivalence. However, in the majority of one to more than one relations of equivalence the relation of divergence represents a form of partial equivalence.

#### 10.4.2.1 Divergence

Divergence is characterised by a one to more than one relation between source and target language forms. For a given lemma the translation equivalent paradigm will contain more than one translation equivalent. Divergence is typically a relation which falls within the domain of partial equivalence. However, as stated in paragraph 10.4.1, a one to more than one relation can also represent full equivalence if the target language forms are absolute synonyms. In an article displaying an equivalent relation of divergence different subtypes can be distinguished, i.e. *lexical divergence* and *semantic divergence*.

##### *Lexical divergence*

*Lexical divergence* prevails where a monosemous lexical item, functioning as lemma sign, has more than one translation equivalent. These equivalents are usually partial synonyms in the target language and therefore constitute a relation of partial equivalence. Where these equivalents are absolute synonyms they constitute a relation of full equivalence, as indicated in the preceding paragraph. Lexical divergence is typically indicated by means of a comma, used as non-typographical structural marker, to separate these equivalents, e.g. in GW, the Afrikaans lemma *gulheid* has two translation equivalents, i.e. *generosity* and *liberality*.

**gulheid**, generosity, liberality.

The comma separating these two translation equivalents indicates a relation of lexical divergence, i.e. that the translation equivalents are target language synonyms.

On a semantic level there is a one to one relation between source and target language but on a lexical level there is a one to more than one relation, i.e. a relation of divergence. Where lexical divergence prevails, the lexicographer has to ascertain whether the translation equivalents are full or partial synonyms. If there is a relation of complete synonymy, which is very seldom the case, the lexicographic treatment can be of a similar nature as in the case of congruence. A lack of context and context items will indicate that the lemma and all the equivalents share the same semantic value. More often than not an equivalent relation of lexical divergence displays

equivalents which are partial synonyms. The user should not only be made aware of the fact that these partial synonyms can substitute the source language form but they should also be cautioned that the target language forms cannot substitute one another in all environments. Consequently the lexicographer should enter some kind of either contextual or cotextual guidance to indicate the typical environment where the common semantic value of the translation equivalents is activated. In the above-mentioned example a cotext entry like the following could suffice:

*We appreciate the generosity/liberality with which he has helped the poor people.*

The lexicographer may also consider using a structural marker to indicate to the user that the translation equivalents are only partial synonyms. Be it as it may, lexical divergence does not demand a sophisticated system to ensure an optimal retrieval of information but it does require a consistent application of a well-devised model.

### *Semantic divergence*

*Semantic divergence* can be regarded as the most typical occurrence of partial equivalence. It prevails where the lemma sign represents a polysemous lexical item. Polysemy is a language specific phenomenon and the chances are minimal that a single target language item will have the same semantic load as the polysemous source language item. To solve this problem the comment on semantics contains a subcomment on semantics for each one of the polysemous senses and a translation equivalent has to be entered for each polysemous sense of the source language form. When comparing a monolingual dictionary like the POD with a bilingual dictionary like the GW one notices some telling differences in the article layout, cf. the following examples as an illustration of some of the salient differences:

**chair** —*n.* **1** seat for one person usu. with a back. **2** professorship. **3** a chairperson.  
**b** seat or office of a chairperson ...(POD)  
**chair**, (n) stoel, gestoelte; setel; voorsitter(stoel); leerstoel, professoraat; plat (spoor); kagebint; stoelbed (vir tappe); ... (GW)

Although both these articles present the treatment of a polysemous lexical item, they use different ways to mark the different polysemous senses. In the POD a system is used where the paraphrase of meaning given for each polysemous sense is preceded by an item giving a polysemy marker, i.e. the entries 1, 2, and 3. These entries form part of the rapid access structure, cf. par. 11.2.2, and helps the user on his/her search route to the desired entry. The use of numbers as items giving a polysemy marker is a well-established and standardised lexicographic convention which is not only used in monolingual dictionaries but also in some bilingual dictionaries. As far as bilingual dictionaries are concerned an alternative but equally established convention is to use a structural marker like the semi-colon to mark a relation of semantic divergence. In the above-mentioned article from GW commas and semi-colons are utilised within the translation equivalent paradigm. These commas and semi-colons are not mere punctuation markers but rather functional lexicographic text elements, signalling something regarding the translation equivalents they precede.

As indicated in the previous section the comma between two translation equivalents marks a relation of lexical divergence. In the translation equivalent paradigm of

a bilingual dictionary a semi-colon is used to separate translation equivalents representing different polysemous senses of the lexical item represented by the lemma sign. Compare the following example from GW:

**eenmalig, (-e)**, single; unique; non-recurrent.

This presentation implies that the Afrikaans word *eenmalig* is polysemous and that it has three senses, translated into English by means of the equivalents *single*, *unique* and *non-recurrent*.

In planning a dictionary a lexicographer should endeavour to employ a system which would be the most beneficial for the intended target user. Using numbers to mark the relation of semantic divergence brings a bilingual dictionary in line with the system with which the users are familiar in their confrontation with monolingual dictionaries and in some bilingual dictionaries. This is also a much more explicit way of marking polysemy, compared to the use of semi-colons, cf. the following example from the TAW:

**maak** 1. make [a] *My mother can make clothes. My ma kan klere maak ...* 2. go  
*Ducks go "quack". Eende maak "kwaak". ...* 6. put *She tried to put the baby to sleep. Sy het die baba aan die slaap probeer maak.*

From a user-perspective it could be argued convincingly that the use of numbers should be seen as the preferred way to mark the subcomments on semantics accommodating translation equivalents for the different polysemous senses of the lexical item represented by the lemma sign. The value of these numbers to ensure a rapid access to the relevant entries could be enhanced by a proper article structure and micro-architecture, cf. par. 11.3.

No lexicographer may assume that the users of the dictionary will intuitively know which translation equivalent to choose for a given situation. Consequently the lexicographer is compelled to utilise additional strategies to ensure an optimal retrieval of information. These strategies include procedures of contextualisation and cotextualisation and will be discussed in paragraph 10.5.

### 10.4.3 Zero equivalence

Zero equivalence prevails where the target language has no item to be co-ordinated as a translation equivalent with a lemma representing a source language item. The lexicon of a language does not necessarily develop parallel to the lexicon of any other language. When one language acquires a word for a given concept it does not imply that the next language will also acquire a word for that concept. Language reflects the communication needs of the speakers of that language and finds words to assist the users to express themselves. Although the majority of languages will have lexical items for a core of basic concepts there are and always will be noticeable differences in the lexical stock of different languages. When comparing any two languages one soon becomes aware of the absence of words in a given language. In linguistics this phenomenon is known as lexical gaps.

Lexical gaps are language specific and occur where a language lacks a word for a given concept. Dagut (1981) identifies different types of semantic gaps in the transfer of a text from one language to another. He distinguishes between gaps

due to linguistic and gaps due to extra-linguistic factors. These two categories are called linguistic and referential gaps respectively, cf. Dagut (1981), Gouws (1989). When comparing two languages, e.g. in the compilation of a bilingual dictionary, a linguistic gap is identified where the speakers of both languages are familiar with a certain concept but where the one language does not have a word to refer to it, whereas the other language does have such a word. In South Africa the speakers of both Afrikaans and English are familiar with the concept of food which you take along for the road when you are on a long journey, e.g. by car. Afrikaans has a word for this concept, i.e. *padkos*. English has no single word for it. This is a linguistic gap. A referential gap can be postulated where a lexical item from language A has no translation equivalent in language B because the speakers of language B do not know the referent of the lexical item from language A. The Nguni languages have the word *lobola* and the speakers of these languages are familiar with the word and with the specific cultural concept. In e.g. the Netherlands the speakers of Dutch do not have a word that can be regarded as a proper equivalent for *lobola* because they are not familiar with this concept. Referential gaps typically occur when a source language form is a culturally bound lexical item and the speakers of the target language do not share in that culture.

The lack of translation equivalents where lexical gaps exist, establishes a relation of zero equivalence. The existence of zero equivalence adds yet again to the challenges faced by the lexicographer because these items may not be presented in a dictionary without a treatment that will enable the user a proper understanding of the source language item. Zero equivalence often leads to the inclusion of surrogate equivalents, i.e. a target language entry substituting a translation equivalent. In an Afrikaans-English bilingual dictionary the word *padkos* should be entered as an Afrikaans lemma. No exact translation equivalent exists and the lexicographer will have to use a surrogate, e.g. a brief description like "food for the road". Partial equivalents are sometimes offered in the treatment of a relation of zero equivalence but they do not ensure a proper semantic co-ordination between source and target language and consequently additional entries, belonging to the domain of surrogate equivalents, should be included. In a dictionary with English and one of the Nguni languages as treated language pair, the word *lobola* will have to be included as a lemma. English has a lexical gap in this regard and the lexicographer will have to use a surrogate equivalent in the treatment of the lemma *lobola*. English does have a partial equivalent that can be offered in the treatment of this lemma, i.e. the equivalent *bride's price*. However, this English word does not convey the cultural value of the word *lobola* and the lexicographer should preferably give some additional entries as surrogate equivalents to complement the partial equivalent and to ensure a proper explanation of the meaning of the word *lobola*.

The existence of zero equivalence compels lexicographers to use innovative approaches to ensure that the users of the dictionary receive sufficient guidance. The nature of the lexical gap in the target language will determine the extent of the guidance given by the lexicographer in treating the source language item. In the case of a linguistic gap the lexicographer knows that the speakers of the target language are familiar with the specific concept. Consequently a brief description of meaning will suffice. Where the lexicographer is dealing with a referential gap a more comprehensive treatment is needed because the speakers of the target language are

not familiar with the concept represented by the source language lemma. Where dictionaries use pictorial illustrations as a type of microstructural entry, articles where zero equivalence prevails could be ideal candidates to accommodate pictorial illustrations. The old saying that a picture says more than a thousand words is still true and in order to achieve a proper transfer of meaning of a source language word representing a concept unknown to the speakers of the target language the use of a pictorial illustration as a surrogate equivalent may be beneficial to the target users of the dictionary, cf. Al-Kasimi (1977).

#### 10.4.4 Polidivergence

Divergence is a one to more than one relation between the lemma and the translation equivalent paradigm whereas zero equivalence stems from a one to zero relation and this zero is substituted by a surrogate equivalent, which often is a brief description of the meaning of the source language item. Equivalent relations in bilingual dictionaries are constituted by the relations between a given lemma and the equivalents presented in the translation equivalent paradigm. A given dictionary article can present a single equivalent relation. This is the case in articles where exclusive relations of e.g. congruence, lexical divergence, semantic divergence (with each polysemous sense of the source language item having only one translation equivalent) or surrogate equivalence prevails. These relations of exclusive equivalence are illustrated in the following examples:

##### *Congruence*

**bon-i-ness n.** *benerigheid.*

##### *Lexical divergence*

**gul´heid,** *generosity, liberality.*

##### *Semantic divergence*

**een´malig, (-e),** *single; unique; non-recurrent.*

##### *Surrogate equivalence*

**padkos,** *provisions (food) for a journey.*

In each one of these examples a relation of mono-equivalence exists, i.e. there is only one type of translation equivalence prevailing in the relation between the lemma and the translation equivalent paradigm. However, dictionary articles very often do not display such a relation of exclusive equivalence but frequently one finds more than one type of equivalent relation prevailing in a single article. Compare the following examples:

##### *Lexical and semantic divergence*

**level ...** *waterpas, paslood; peil; stand; standaard, hoogte, niveau; verdieping; vlak, plan; laag; gelykte; ...*

*Semantic divergence and surrogate equivalence*

**roei**<sup>4</sup>, (s) (-e), tail of a comet; lattice; mullion.

*Lexical divergence and surrogate equivalence*

**hard** <sup>o</sup>p, aloud, in a clear voice (GW)

*Lexical divergence, semantic divergence and surrogate equivalence*

**gat**, (-e), hole, opening, gap, pit; socket; eye; hole of a place, unattractive place, room, etc.

In all these articles one sees relations of inclusive equivalence, i.e. the equivalent relation includes more than one subtype of equivalence. These examples have a one to more than one relation between source and target language and therefore fall within the broader category of divergence. Where a relation of divergence displays inclusive equivalence the equivalent relation is known as *polydivergence*. Polydivergence is the prevailing equivalent relation in many dictionary articles and, yet again, confronts the user with a whole range of problems in finding the proper translation equivalent for a given occurrence of a source language item. It compels the lexicographer to negotiate the use of additional entries to ensure sufficient guidance for the target user of the dictionary.

## 10.5 Communicative equivalence

In the articles of bilingual dictionaries one or more relations of equivalence exist between a lemma and the translation equivalent paradigm. Translation equivalence, the first aim of the lexicographer of a bilingual dictionary, implies a semantic co-ordination between a lemma and its translation equivalent paradigm. In the treatment of a given lemma the lexicographer will provide translation equivalents for all the different senses of that lemma. The function of the specific dictionary, e.g. a text reception or a text production function, should have an influence on the nature and the extent of the additional entries given to complement and support the translation equivalents. The majority of bilingual dictionaries are compiled as bifunctional products, i.e. to assist their users with at least the above-mentioned functions of text reception and text production. However, the data presented in the comment on semantics in bilingual dictionaries too often gives the user very little chance of choosing the correct equivalent. Unfortunately lexicographers too often restrict their attempts to ensure translation equivalence to the mere listing of a number of target language items. Although these items are equivalents that collectively represent the full semantic value of the lemma in the target language, and although the relation between the lemma and the translation equivalent paradigm is one of semantic equivalence, the dictionary users receive no guidance to assist them in the choice of equivalents or in using the target language items in a proper way. From a text reception point of view it may suffice but from a text production point of view the lexicographic treatment is extremely inferior, cf. the following examples from TW and GW respectively:

**level** ... waterpas, paslood; peil; stand; standaard, hoogte, niveau; verdieping; vlak, plan; laag; gelykte; ... (TW)

**op'sigter (-s)** custodian, caretaker; overseer; clerk of works; conservator; checker; green-ranger; keeper; groundsman; banksman; gaffer; ganger. (GW)

In both these articles an equivalent relation of polydivergence prevails with a combination of lexical and semantic divergence. Semantic divergence implies that the source language item is polysemous and that translation equivalents are supplied for the different polysemous senses of the word represented by the lemma. Although this treatment may have secured semantic equivalence between the lemmata and their respective translation equivalent paradigms, communicative equivalence cannot be achieved by this presentation. Communicative equivalence can only be achieved if the treatment is not restricted to a listing of equivalents but if these equivalents are complemented by context and cotext entries that can help the user to choose the correct equivalent for a given occurrence of the source language item and to use this equivalent in a proper way.

The addressing structure of bilingual dictionaries makes provision for procedures of lemmatic and non-lemmatic addressing. The above-mentioned examples only give evidence of lemmatic addressing. This is in order if text reception is the only function of a dictionary. However, if the lexicographer aims to assist the user with text production the procedures of lemmatic addressing need to be complemented by procedures of non-lemmatic addressing. This implies that the lemma should not be the only treatment unit in a dictionary article but that the translation equivalents as the target language items should be elevated to receive treatment unit status. Such a procedure of non-lemmatic addressing will not lead to a comprehensive treatment of the members of the translation equivalent paradigm but some additional entries will be included and these entries will be directed at the translation equivalents so that the user can have a better idea of how to use them and when a given translation equivalent should be used to replace the source language item in a specific occurrence. This would help the user to achieve communicative equivalence. The following example from TAW illustrates how the use of examples can help to achieve communicative equivalence:

**taak 1** task *It is my **task** to lay the table for dinner in the evening.* Dit is my **taak** om saans die tafel vir ete te dek. **2** duty *He works in a bank and his main **duty** is to receive and pay out money.* Hy werk in 'n bank en sy belangrikste **taak** is om geld te ontvang en uit te betaal. ... (TAW)

The Afrikaans word *taak* is polysemous and in the treatment of this word the different subcomments on semantics, containing the treatment of the different polysemous senses of the source language item, are marked, using numbers as structural indicators, to explicitly show that the lemma represents a polysemous lexical item. Due to the implementation of an integrated microstructure each translation equivalent is immediately followed by an illustrative example, addressed at the translation equivalent, indicating how this target language item is used to represent the specific polysemous sense of the lemma. Procedures of non-lemmatic addressing to help with communicative equivalence can also be achieved by means of context entries, cf. the following example from GW:

**berei'**, (-), prepare (meal); dress (leather, tools); preserve (foodstuffs); cure (tobacco); concoct. (GW)

An equivalent relation of divergence prevails in this article. Semi-colons are used to separate translation equivalents representing different polysemous senses of the lemma and they indicate that no less than five polysemous senses have been allocated to the Afrikaans word *berei*. Although the dictionary has not used any illustrative examples, the user should be able to make a correct choice of equivalents due to the entries indicating the typical context in which a specific target language item should be used as translation equivalent of the lexical item represented by the lemma. For text production some additional entries, especially illustrative examples, may be needed but for text reception these context entries sufficiently assist the user.

# The access and search area structure

---

## 11.1 General remarks

The data on offer in a dictionary should be presented in such a way that the knowledgeable target user can access it in order to retrieve the kind of information that motivated a given dictionary consultation procedure. To ensure the success of a dictionary consultation procedure in a typical usage situation the lexicographer has to embark on a dictionary plan that makes provision for a well-devised dictionary structure. The dictionary conceptualisation plan needs to focus on all the structural components of a dictionary with the necessary attention to the access structure and the article structure. In a typical dictionary consultation procedure where a user needs to retrieve information from a dictionary article the access structure, i.e. the search route the user follows to reach the desired entry in a dictionary, should lead the user to the specific article but also into the article and to the relevant data category. The success of finding the data without delay depends on the access structure but also on the article structure, and more specifically the way in which the data categories in the article have been arranged and presented as different search fields. The search area structure complements the access structure and these structures function as help structures to guide the user to the required data.

## 11.2 The access structure

The access structure determines the search route a user follows to reach an entry in a dictionary. The search of a typical user often starts with the choice of a given dictionary on account of the title of that dictionary as it is presented on the spine of the cover or the front outside cover page. From there the user proceeds to the inside of the dictionary and finally reaches the article. The search route leads the user into the article to a specific microstructural entry. The procedure of accessing a dictionary and following a search route must be devised as part of the dictionary plan in the dictionary specific lexicographic process. Within the access structure a distinction is made between an *outer* and an *inner access structure*, cf. Hausmann & Wiegand (1989:337), resulting in an outer and an inner search route.

### 11.2.1 The outer access structure

The *outer access structure* determines the part of the search route which leads the user from the entries on the cover of the dictionary to the lemma sign presented as guiding element of a given article. Lexicographers do not always realise the importance of the entries on the cover of a dictionary. These entries are not only there for promotional purposes or as a form of lexicographic cosmetics to enhance the looks of the dictionary. When not chosen in an arbitrary way these are functional entries and they form an integral part of the lexicographic presentation. The spine and front cover of a dictionary often represent the first encounter a user has with the dictionary. Consulting a dictionary is motivated by specific needs and users

have to select the proper type of dictionary to solve the specific problem. The entries on the cover of a dictionary should inform the user what to expect in the specific dictionary. A user may wish to consult a dictionary which is known to him/her by the title of the dictionary or the names of the authors or even by the publishing house. Dictionaries are often referred to by means of an abbreviated form and such an abbreviated form becomes a household name in a specific speech community. However, for people not familiar with the dictionary this abbreviated title has no meaning and they are not able to choose that dictionary, even if it is the one they need for a specific purpose, by merely looking at the abbreviated form. Therefore it is important that the abbreviated form, with which the speech community is familiar, as well as the full form, to draw new users or to give new users a good idea of the dictionary type, should be printed as entries on the cover and on the title page. HAT, the *Verklarende handwoordeboek van die Afrikaanse taal*, is best known within the Afrikaans speech community by its abbreviated form. However, both the abbreviated form and the full form appear on the cover and on the title page:

#### HAT

VERKLARENDE  
HANDWOORDEBOEK  
VAN DIE  
AFRIKAANSE TAAL

In the past dictionaries were too often evaluated merely in terms of the contents of the central list. Wiegand (1996b) argues that dictionaries should be regarded as carriers of text types, with the central list being only one of the texts in the dictionary. The frame structure, cf. Chapter 6, shifts the focus from a central list bias to an approach which also acknowledges the importance and function of the outer texts, i.e. the texts in the front matter and in the back matter of the dictionary. From an access perspective the frame structure can be pivotal in guiding users to the required data. The data distribution structure determines where a specific type of data should be accommodated in a dictionary. The use of a frame structure offers the lexicographer a wide range of possibilities, and data types that do not fit into the central list can be included in the outer text section of a dictionary. The selection of outer texts is done on a dictionary-specific basis and users not familiar with a specific dictionary has no idea regarding the nature of the data to be included in the outer texts. Access to these texts may not be left to the intuition of the user or a random consultation procedure. The outer access structure should be the instrument to guide a user to those outer texts that could provide solutions to the problems that motivated the specific dictionary consultation procedure. In this regard a text presenting the table of contents of a dictionary has an important role as part of the outer access structure of the dictionary.

Gouws (2002) indicates that the table of contents is a functional part of a dictionary as a compound of texts, or “big text”, cf. Kammerer & Wiegand (1998). It should guide the user over textual boundaries to different parts of the dictionary. The purpose of a table of contents should not only be to give an overview of the contents of the dictionary but also to increase the access of the dictionary as a big text by means of an indication of page numbers ensuring a rapid progress to the different texts constituting the big text. In this regard the table of contents puts the user on the

dictionary internal search route. The table of contents does not offer the user access to the different articles or even the article stretches but to the central list as well as to other texts of the dictionary. Contrary to the general access structure the table of contents does not lead a user to the different lemma signs functioning as guiding elements of the articles but rather to the beginning of the central list and a selection of outer texts functioning within the compound of texts. Because the average user of a general dictionary is not that used to outer texts co-accommodating the lexicographic data, the lexicographer has the responsibility to make the users aware of these texts and of their contents. In this regard the table of contents plays an important role and in the table of contents the lexicographer should give a clear indication of the type of outer text to which the user is guided.

The table of contents is a special part of the outer access structure and when planning dictionaries in which a frame structure prevails, lexicographers will do well to include a table of contents as one of the front matter texts.

The search route in the central list of a dictionary has to lead the user to the required lemma sign but for many users this route goes via different lexicographic road signs that form part of the outer access structure. One of these road signs is the alphabetical letter indicating the beginning of a new article stretch, cf. the following examples from TFW and NeW respectively:

## B

**babe** *a babe in arms* 'n suigeling, 'n kind op die skoot; *be a babe in the woods* § (soos) 'n groot kind wees «, 'n naïweling wees, naïef wees (TFW).

## B

**B** *B's, Bs, n.:* ~ **flat** *a., (mus.)* B-mol. ~ **sharp** *a., (mus.)* B-kruis.

**B<sup>1</sup>-B bomber, B<sup>2</sup> bomber** *n, (mil.)* → **stealth** (NeW).

Many dictionaries also have a thumb index on the open outside of the dictionary. This index gives an indication where the various article stretches start. By putting a finger on the specific letter and opening the dictionary right there, the desired article stretch is reached and the user can continue with the search within the relevant article stretch. These thumb index markers form part of the rapid outer access structure of the dictionary. In a dictionary with a strict alphabetical ordering the user looking for a lemma starting with e.g. the letter *B* could also page through the dictionary until the marker “*B*”, indicating the beginning of the article stretch of lemmata starting with the letter *B* is reached. The thumb index marker gives the user a quick orientation within the central list regarding the relative position of the article stretch within which the target lemma falls.

Within the central list the access structure is also realised by means of the search words presented as headers on each page to indicate the first and the last lemma sign featuring on the specific page. In some dictionaries two search words are given on each page, with the left hand word indicating the first lemma sign and the right hand word indicating the last lemma sign on that page. Other dictionaries have one

search word on the left hand page, indicating the first lemma sign to be entered on that page, and one search word on the right hand page, indicating the last lemma sign on that page. When searching a specific lemma in a dictionary, knowledgeable dictionary users will closely watch the search words presented as headers to help with a rapid access to the required lemma sign.

The lemma sign is the final destination of the outer access structure. As macrostructural items they head their respective articles and, although they stand in the most salient position within an article, i.e. right at the beginning where, as guiding elements, they introduce the article, dictionaries still use a typographical structural marker, bold letters, to let the lemma sign stand out from the other entries in the article. Both vertically-ordered lemmata and horizontally-ordered lemmata are macrostructural elements and are presented in bold. This typographical structural marker forms part of the access structure of the dictionary, cf. the following example from the VAW:

**bakeliet** *˘*. Sintetiese hars of plastiese materiaal wat vir die vervaardiging van plastiese voorwerpe gebruik word.

**ba** *˘***ken**, **-s**. **1.** Grensmerk van 'n plaas. **2.** Merkteken in of by die see vir skepe. **3.** Radiosender wat as gids spesiale tekens vir bote en vliegtuie uitsaai; *baken* MAAK (steek), van 'n perd afval; 'n plaas koop; broek losmaak; *die bakens* VERSIT, jou by die omstandighede aanpas; **bakenlig**; **bakenpunt**; **bakenstasie**; **bakenstoring**; **bakenvuur**.

**ba** *˘***kenlanding**, **-s**. Landing van 'n vliegtuig met behulp van 'n radiobaken.

**ba** *˘***kenry**, **-ge**-. Praktiese metode deur die Voortrekkers gebruik om te perd plase af te meet deur 'n bepaalde tyd lank in verskillende rigtings te ry. (VAW)

Where the article does not fit into a single line, cf. the article of the lemma sign *baken*, the layout of the article makes provision for a hanging paragraph that enhances the optical salience of the lemma sign. This is also clear from the following examples from the VAW:

**bak** *˘***kiespomp**, **-e**. Groot wiel met bakkies om water te skep en dit na 'n hoër plek te laat loop.

**bak** *˘***kop**, **-pe**, **bak** *˘***kopslang**, **-e**. Koperkapel of rinkhalsslang wat sy nek breed trek wanneer hy kwaad is en 'n straaltjie gif uitspuit. (VAW)

This procedure to make the lemma sign more salient is also followed in the GW, cf. the following example:

**bait**, (n) lokaas, aas; aanloksel; verversing; (v) aanlê; ververs, afsaal, voer gee; lok; aas aansit; aanhits; terg; aanval; **~boy**, aasjong; **~er**, terger, koggelaar; **~hook**, aasstok; **~ing**, tergery, koggeling; **~money**, lokgeld. (GW).

In these articles the layout has a specific role as part of the lexicographic process. When planning and compiling a dictionary, lexicographers always have to take cognisance of the importance of the dictionary layout. The form of the lemma sign also plays an important role in the outer access structure. The decision to employ a horizontal ordering has implications for the easy access of a user to the lemma

sign because the user has to deviate from the vertical to the horizontal search route. Where the horizontal search route still presents the macrostructural items as main lemmata, i.e. in their full form, the access is not that strenuous a procedure, cf. the example from the VAW:

**ba´ken, -s.** 1. Grensmerk van 'n plaas. 2. Merkteken in of by die see vir skepe. 3. Radiosender wat as gids spesiale tekens vir bote en vliegtuie uitsaai; **baken** **MAAK** (steek), van 'n perd afval; 'n plaas koop; broek losmaak; **die bakens** **VERSIT**, jou by die omstandighede aanpas; **bakenlig**; **bakenpunt**; **bakenstasie**; **bakenstoring**; **bakenvuur**.

**ba´kenlanding, -s.** Landing van 'n vliegtuig met behulp van 'n radiobaken.

The presentation of *bakenlig*; *bakenpunt*; *bakenstasie*; *bakenstoring*; *bakenvuur* as horizontally ordered main lemmata makes it relatively easy for the user to find them and it does not complicate the access structure of the dictionary too much. However, with the use of sublemmata where the horizontal ordering presents a partial article stretch with partial lemmata functioning as guiding elements and the only access to a lemma sign being via the preceding main lemma, the access structure becomes complicated and user-unfriendly. This is clear from the following example from GW where a lemma like *bait-hook* is presented by means of a place-keeping symbol, the tilde, and the partial lemma *~hook* functioning as sublemma. Access to this lemma is only possible via the preceding main lemma *bait*:

**bait**, (n) lokaas, aas; aanloksele; verversing; (v) aanlê; ververs, afsaal, voer gee; lok; aas aansit; aanhits; terg; aanval; **~boy**, aasjong; **~er**, terger, koggelaar; **~hook**, aasstok; **~ing**, terger, koggeling; **~money**, lokgeld. (GW).

When opting for the use of sublemmata, presented in an horizontal ordering as partial lemmata, the access could be improved if the preceding main lemma is not the only entrance to the niched or nested cluster but if the mutual first component of the sublemmata could be presented as a nest- or niche-external lemma part functioning in the nest or niche entrance position, compare the following examples from the HAT and the GW respectively:

**fa´bel (-s)**

**1** Verdigte vertelling met opvoedkundige doel: *Die fabels van Esopus*. **2** Versinsel, leuen, verdigsel: *Dis somer pure fabels. Dit behoort tot die ryk van die fabels*, is 'n versinsel. [L. *fabula*] **fa´bel**: **~dier**, **~kunde**, **~kundige**, **~land**, **~leer**, **~literatuur**, **~wêreld**. (HAT)

**fa´bel**, (s) (-s), fable, legend, fiction; myth; *dis maar ~s*, that is all stories; **~aar**, fabulist, fibber; **~agtig**, (-e), fabulous; incredible; mythic(al), mythologic(al); **~agtigheid**, fabulosity, fabulousness; **~boek**, book of fables; **~dier**, fabulous animal; **~digter**, fabulist, writer of fables; **~kunde**, **~leer**, mythology; study of fables. (GW)

The HAT example repeats the mutual element immediately before the niche whereas the user of the GW has to link the main lemma with the sublemmata. The HAT gives a more direct link between the partial lemmata and the element that has been omitted in the lemma niche. This displays a higher degree of user-friendliness and increases the quality of the outer access structure. The occupation by a lemma part of the nest or niche entrance position becomes crucial where the complex lexical

items, presented as sublemmata, contain a variant of the lexical item represented by the main lemma that includes a combining form which does not feature in the presentation of the main lemma, cf. the following example from HAT where the lexical item *kind* has the variant *kinder-* in complex words:

**kind** (kinders, (*veroud.*; *deftig*) kindere; -jje)

**1** Jong, onvolwasse menslike wese; veral: **1 (a)** Nog ongeboore mens: 'n *Kind afdrywe, verwek. (b) Baba: Die kind huil. Kinders oppas, bad, aantrek, soog, voed. Die kind(jie) Jesus.* ∇ 'n *Kind se gestalte is soepel en teer (Elisabeth Eybers). Sag soos 'n soentjie op kindjie se wangetjie (C.J. Langenhoven).* (c) *Jong mens; onryp menslike wese: Kinders op skool hê. Die kind word nou groot. 'n Soet, stout, dom, slim, dierbare kind. Die kinders van vandag.* ∇ *Ek wens dat kindjie groter was om al die diertjies op te pas D.J.Opperman (D.J. Opperman).* **2** Volwassene wat aan 'n kind herinner: *Hy sal altyd kind bly. Vergeleke met my is hy 'n kind. Hy is nog 'n kind in hierdie dinge, onervare.* ∇ *Jy bly in jou volwassenheid tog maar 'n kind D.J.Opperman (D.J. Opperman).* ... . **kinder:** ~bad, ~bediende, ~begrip, ~biblioteek, ~broek, ~drank, ~droom, ~fantasie, ~gehuil, ~gelag, ~geluk, ~hospitaal, ~kliniek, ~kwaal, ~lawaaï, ~leed, ~lektuur, ~leiding(s)kliniek, ~liedjie, ~luier, ~misdaad, ~natuur, ~opvoeding, ~pret, ~roman, ~roof, ~siel, ~skool, ~smart, ~speelgoed, ~sport, ~tal, ~tehuis, ~tug, ~verdriet, ~wieg. (HAT)

Lexicographers need to plan the ordering of lemmata and the possible use of sublemmata along with main lemmata carefully and they have to consider the implications these decisions have for the access structure of a dictionary and the eventual success of the dictionary consultation procedures in typical situations of dictionary use.

### 11.2.2 The inner access structure

The inner access structure determines the search route a user follows within the dictionary article to reach the specific item or data category (s)he is looking for. It is the task of the lexicographer to devise the article structure in such a way that the data presentation makes it easy for a user to find any data category in the quickest possible way. In this regard it is important that the lexicographer should realise that users very seldom want to read through an article. The typical dictionary consultation procedure is aimed at finding one specific item or data type. This required item may not necessarily be the one which has been allocated the first or a prominent position in the article. However, the presentation of data types should be done in such a way that easy and rapid access to any category is possible for the target user during a typical situation of dictionary use. A well-devised inner access structure relies on the search area structure, cf. par. 11.3, and the use of structural indicators.

Articles should be structured in such a way that the user can clearly distinguish the different data categories. One way of assisting the user in this regard is the use of structural indicators as microstructural entries, cf. par. 8.2.1. Structural indicators, both typographical and non-typographical structural indicators, are entries that identify a specific item or data category, cf. the following examples from the Dutch-Afrikaans dictionary, currently being compiled, and the HAT respectively:

**bril****1** [om te kijken] #- een bril hebben/dragen; hij heeft zijn bril niet op; ...• <inf.> een bril moeten 'n bril moet kry ...♦ elk ziet door zijn eigen bril elkeen kyk deur sy eie brildoor 'n roze bril kijken deur 'n rooskleurige bril kyk▶ iemand 'n bril op die neus sit iemand te grazen nemen (ANNA)

In this article the non-typographical structural indicators “-”, “•” “♦” and “▶” are employed to guide the user to specific types of items. This system is explained in detail in the user’s guidelines text in the front matter of the dictionary. As indicated in par. 8.2.1, where the same example is discussed, the indicators “-” and “•” mark specific types of illustrative examples. The user interested in examples that show a contrast between Dutch and Afrikaans merely looks for the “•”, the structural indicator marking this type of an example. Rapid access to non-contrastive examples goes via the structural indicator “-”. The presentation of Dutch idioms with their Afrikaans equivalents is preceded by the marker “♦” and the user only interested in this type of data does not have to read through the whole article but once the outer access structure has led him/her to the desired lemma, (s)he can immediately proceed to the search field marked by the “♦” to find the type of data that motivated the specific dictionary consultation procedure. The indicator “▶” is used to mark the presentation of Afrikaans idioms as source language items, co-ordinated with their Dutch equivalents.

The HAT also employs a variety of structural indicators, as illustrated in the next example:

**kalf** □ s.nw. (kalwers)

**1** Jong bees; die kleintjie van 'n koei: *Die kalf suip aan die koei. Hou die kalwers weg voor hul die koeie uitsuip.* ∇ (D)ie *kalfie* (wat) kort-kort om sy moeder wals op nog onvaste horrelbeen (Totius). **2(a)** Die kleintjie van sommige ander - gewoonlik groot - diere van dieselfde familie as dié waartoe die beste behoort, nl. die *Bovidae*: *Die kalf van 'n eland.* **(b)** Die kleintjie van sommige ander groot soogdiere: *Die kalf van 'n olifant, van 'n renoster, van 'n walvis.* **3** Dommerige, goedaardige persoon: So 'n *kalf kan jy enigiets wysmaak.* UITDR.: *Die gemeste* (vetgemaakte) *kalf slag*, (a) uit vreugde die beste en kosbaarste aan iemand as welkomswewys voorsit, hom rojal onthaal; (b) 'n gebeurtenis feestelik vier; Vgl. Luk. 15:23. *Die goue kalf aanbied*, (a) rykdom najaag; (b) kruiperige hulde aan ryk mense bewys; Vgl. Ex. 32:4. *Nie al jou kalwers in die hok hê nie*, nie heeltemal reg wees nie. .... □ ww. (gekalf) Geboorte skenk aan 'n kalf; 'n kalf werp: *Ons koei het vannag gekalf.* **kalfs**: ~**been**, ~**gebraad**, ~**gehak**, ~**huid**, ~**karmenaadjie**, ~**kotelet**, ~**lewer**, ~**nier**, ~**vet**, ~**wors**. Vgl. ook ss. met **kalwer**-. (HAT)

The non-typographical structural indicator “□” indicates that a word is used in more than one grammatical function and it precedes each one of the items giving the part of speech. By seeing this marker in the beginning of a dictionary article the user is immediately made aware of the fact that the word functions as more than one part of speech. If the user is not looking for data regarding the first grammatical function it is not necessary to read through the whole article but the user merely proceeds to the next occurrence of the indicator “□”, where the following grammatical category is introduced. In the presentation of illustrative examples this dictionary

distinguishes between made-up or constructed examples and citations. All citations are preceded by an upside down triangle “▽” as structural indicator. This entry also allows rapid access to the specific data type and guides the user on the article-internal search route.

In this article from HAT typographical structural indicators also play an important role as route markers on the inner access structure. The different typefaces, e.g. bold, italic and roman, indicate specific search fields or data categories. The lemma signs (both the main lemma and the sublemmata) are given in bold, the paraphrase of meaning in roman and the illustrative examples in italics. A user who is only interested in finding the examples can have a rapid access to the article slot where this data category is accommodated by merely looking for the use of italics. Italics are also used for the idioms but these entries are clearly separated from the preceding subcomment on semantics by the structural marker UITDR. (abbreviation for “uitdrukking” = expression). Within the article slot where idioms are treated the rapid access structure uses bold italics to mark a keyword in the idioms according to which the idiom has been ordered in the listing of idioms. By looking for these words in bold italics the user can quickly move from one idiom to another.

A dictionary plan has to make provision for an access structure to assist the target users in their dictionary consultation attempts and the access structure has to make it possible for them to reach the required data with as little trouble as possible. Once again the lexicographers have to be aware of the dictionary using skills of the intended target user to ensure a functional access structure.

### 11.3 The search area structure

The success of a dictionary consultation process does not only rely on whether the dictionary contains the relevant data a user is looking for and whether the user manages to find this data. The quicker and easier the access to a specific item or data type the higher is the level of appreciation the user has for the dictionary and the better are the chances of successful dictionary consultation procedures.

When planning the data distribution structure of a dictionary the lexicographers should focus on where specific data should be presented but also on how it should be presented. In this regard the access structure and the search area structure are of vital importance. The lexicographer decides on the different data categories to be included in the default article of the dictionary. Prior to the compilation process a clear understanding should be reached with regard to the article structure. Bergenholtz, Tarp & Wiegand (1999:1770) make a distinction between articles that display a micro-architecture and those that do not. A dictionary article that displays a micro-architecture is characterised by definite text topological relations, i.e. top to bottom and left to right relations. The following examples illustrate the difference between an article without and one with a micro-architecture:

**editor** *n* -s. 1. a person who is in charge of a magazine or newspaper and who is responsible for its policy and organization. *The new editor of the magazine has many innovative ideas to improve its quality.* 2. someone who does the proofreading and editing of a text. *As editor it is her responsibility to read all the texts and prepare them for publication.*

**editor n -s.**

1. a person who is in charge of a magazine or newspaper and who is responsible for its policy and organization. *The new editor of the magazine has many innovative ideas to improve its quality.*
2. someone who does the proofreading and editing of a text. *As editor it is her responsibility to read all the texts and prepare them for publication.*

In contrast to the first example the second example makes provision for a clear distinction between the comment on form and the comment on semantics and treats the different subcomments on semantics in different text blocks. The division between the text blocks becomes even more clear when a white line is used to separate them:

**editor n -s.**

1. a person who is in charge of a magazine or newspaper and who is responsible for its policy and organization. *The new editor of the magazine has many innovative ideas to improve its quality.*
2. someone who does the proofreading and editing of a text. *As editor it is her responsibility to read all the texts and prepare them for publication.*

Within the comment on semantics of the article with a micro-architecture a top-down relation prevails. The user has a far better and quicker access to these different text blocks compared to the access in the first example. Especially where an article has many or lengthy subcomments on semantics or a varied presentation of data category types, the use of a micro-architecture enhances the accessibility and the user-friendliness of the dictionary. For space-saving reasons some dictionaries only introduce a partial micro-architecture with e.g. the different subcomments on semantics presented in one text block but separate text blocks for other data categories. This is illustrated by the following example from the TAW:

**only**<sup>1</sup> ① *al Rachel was al kind in die klas wat vol punte in die toets gekry het. Rachel was the only child in the class to get full marks in the test. . ② enigste John is die enigste seun met skoene aan; al die ander is kaalvoet. John is the only boy with shoes on; all the others are barefoot.*

◆ **only one** *enigste Rachel was die enigste wat vol punte in die toets gekry het. Rachel was the only one who got full marks in the test.*

□ **only adjective**

**Only** should be positioned as near as possible to the word it refers to in order to make the meaning clear: *I have an egg for breakfast only on Sundays* (not on other days). *I have only an egg* (nothing else) *for breakfast on Sundays.*  
**Only** *I have an egg for breakfast on Sundays* (no one else)

This presentation displays a user-friendly partial micro-architecture with its different text blocks. A full micro-architecture would have resulted in the subcomments on semantics also displaying a top to bottom ordering which would have increased the accessibility, albeit at the cost of space:

only<sup>1</sup> ❶. al *Rachel was al kind in die klas wat vol punte in die toets gekry het.* Rachel was the **only** child in the class to get full marks in the test.

❷. enigste *John is die enigste seun met skoene aan; al die ander is kaalvoet.* John is the **only** boy with shoes on; all the others are barefoot.

◆ **only one** enigste *Rachel was die enigste wat vol punte in die toets gekry het.* Rachel was the **only one** who got full marks in the test.

□ **on • ly adjective**

**Only** should be positioned as near as possible to the word it refers to in order to make the meaning clear: *I have an egg for breakfast **only** on Sundays* (not on other days). *I have **only** an egg* (nothing else) *for breakfast on Sundays.*  
**Only** *I have an egg for breakfast on Sundays* (no one else)

The following examples from NWSG also display a micro-architecture:

**ontmoet** werkwoord ■  
**(het ontmoet)**  
 ❶ Wanneer jy iemand ontmoet, dan leer jy die persoon vir die eerste keer ken. ⇒ *Daar is iemand wat ek graag wil hê jy moet ontmoet [meet].* ⇒ *Het jy al vir Johannes Coetzee ontmoet [met]?*  
 ❷ Wanneer jy iemand ontmoet, dan reël jy om die persoon op 'n sekere tyd en plek te kry. ⇒ *Ek ontmoet [meet] hom môreand op die lughawe.* ⇒ *Jeanne wil hê jy moet haar vandag voor die biblioteek ontmoet [meet].*

The micro-architecture in this dictionary also makes provision for the different part of speech functions to be treated in separate text blocks, clearly marked by the structural indicator “■”:

**sny** naamwoord ■  
**(snye, snytjie)**  
 ❶ 'n Sny is 'n plek waar 'n mens jou gesny het. ⇒ *Ek het 'n sny [cut] aan my vinger waar ek my met die mes raakgesny het.* ⇒ *Die seun het 'n sny [cut] aan sy voorkop omdat hy van sy fiets afgeval het.*  
 ❷ 'n Sny is 'n plat stuk wat van iets anders afgesny is. ⇒ *My ma sit elke dag vir my twee snye [slices] brood in vir pouse.* ⇒ *Kan ek asseblief 'n dik sny [slice] koek kry?*

The use of a micro-architecture leads to the formation of text blocks which accommodate specific data types. According to Bergenholtz, Tarp & Wiegand (1999)

text blocks play an important role in the reception of texts. They improve the access confidence of the user and diminish the access time. The article-internal structure which accommodates these text blocks in a consistent way throughout the dictionary defines the search area structure of a dictionary, i.e. the systematic ordering of article-internal text blocks in a top to bottom relation. A search area structure can be seen as the order-structure that presents the different text blocks and article slots of a dictionary as search fields, ordered according to fixed criteria. Each search area will contain a number of search fields presented in an order devised for the specific dictionary. Users who are familiar with the system of a specific dictionary know where to find a specific data category and they approach these text blocks as search fields and the targets of very specific search procedures. In the TAW the comment on semantics is followed by a text block, marked with a diamond (◆) which contains frequently-used word combinations. This text block is followed by the one containing the grammatical data, marked with the structural indicator "□". The user interested in grammatical data utilises the search area structure and immediately proceeds to the text block marked with the "□". A well-devised search area structure establishes a good inner access structure.

The planning of a dictionary with a micro-architecture and a search area structure compels the lexicographer to give ample attention to the layout of the dictionary. Layout used to be seen as an activity for the publishing house or printer. The implications that the layout of a dictionary have for the access structure, the search area structure and the micro-architecture demand a different opinion regarding layout within the lexicographic process. The layout of a dictionary is essential to its function. The layout helps the user to achieve better and easier text reception, to link the meaning of a word to its relevant cotext entries for text production and to present the data in such a way that the knowledge-oriented function of the dictionary as a source of reference can prevail.

According to Nielsen (2003), when applying a lexicographic approach to dictionaries and not a linguistic approach, the three distinguishing features of a dictionary are its functions, its structure and the lexicographic data it contains. Nielsen (2003a) indicates that a dictionary has at least these three significant features, which together describe and explain the concept. When dealing with e.g. the access structure and the search area structure it becomes clear that these structures are not merely abstract suggestions from the theoretical lexicographer but they are tools in the hand of the practical lexicographer to enhance the quality of the dictionary as a practical instrument. This helps to establish an improvement in the cultural activity of dictionary use and helps to create a dictionary culture. Yet again, structures are not devised for the sake of structures but rather to order the data and to guide the user to the data on offer so that the relevant lexicographic functions and the genuine purpose of the dictionary can be achieved.

# Cross-referencing as a Lexicographic Device

---

## 12.1 Introduction

The quality of *dictionary use*, that is the degree of success a *user* experiences when consulting a dictionary and employing the retrieved information, is determined by a variety of features but one of the most important characteristics of a good dictionary is its *accessibility* which leads to an unambiguous retrieval of the information presented on both the macro- and microstructural levels. Any theory of lexicography should present strategies to enhance the linguistic quality of dictionaries. However, this should be preceded by strategies to enhance the way in which the target user can identify the data (s)he is looking for in order to retrieve the necessary information and to utilise it in a receptive or productive way.

The system of cross-referencing, that is the mediostructure, is a lexicographic device that can be used to establish relations between different components of a dictionary. According to Wiegand (1996c:11) it interconnects the knowledge elements represented in different sectors of the dictionary on several levels of lexicographic description to form a network. Working with a dictionary as a carrier of texts, the mediostructural entries can guide the user between different texts, e.g. between the central text and any text in the front or back matter or between various articles functioning as subtexts in the central word list.

Dictionary research has led to the establishment of different structures of printed dictionaries, e.g. the *macrostructure*, *microstructure*, *access structure* as well as the *mediostructure*.

This chapter focuses on different mediostructural strategies and their practical application in general synchronic dictionaries. From a metalexicographic perspective the structure of dictionaries is discussed in order to explain the application domain of a system of cross-referencing. It is shown how textual coherence, achieved by the interaction of the various structural components, is promoted by the use of a system of cross-referencing and improved by an innovative approach towards a mediostructure-orientated lexicography. Although the mediostructure of dictionaries is a central topic of this chapter, it has to be stressed that references to the theory of mediostructures will only cover a tiny segment of this structural component. A detailed discussion can be found in Wiegand (1996c).

## 12.2 Some basic terms relating to a theory of mediostructures

Wiegand (1996c) gives an exposition of the fundamental terms employed in a theory of mediostructures. According to his theory of mediostructures a lexicographer cross-refers the dictionary user from a *cross-reference position* to a *cross-reference address*. This is usually done by means of a *cross-reference entry* in which a *cross-reference marker* is used and gives the user access to additional relevant lexicographic data. A cross-reference relation is established between the cross-reference entry and

the cross-reference address. In the *Webster's Ninth New Collegiate Dictionary* (W9) the article of the lemma sign *frog* contains the following entries:

1: any of various smooth-skinned web-footed largely aquatic tailless agile leaping amphibians ... - compare TOAD

In this excerpt this specific slot in the article of the lemma sign *frog* is the *cross-reference position* and the lemma sign *toad*, the separate macrostructural entry to which the user is cross-referred, is the *cross-reference address*. In this example the *cross-reference entry* consists of two separate textsegments, i.e. the entry marking the cross-reference relation (*compare*), henceforth referred to as the *cross-reference marker*, and the entry indicating the cross-reference address (*toad*). A cross-reference does not necessarily link one entire article with another entire article. It often happens that a specific entry in an article is linked with a specific entry in another article.

A variety of *cross-reference markers* is used in different dictionaries and often also in one dictionary, e.g. textsegments like *see*, *compare*, *cf.*, *→*, *⇒*, etc. In the English-Dutch translation dictionary *Van Dale Groot Woordenboek Engels-Nederlands* (VDGW) a single arrow is used as one of the cross-reference markers. In the article of the lemma sign **track system** the cross-reference entry “→ tracking” consists of the cross-reference marker “→” and the entry *tracking*, indicating the cross-reference address. In addition to implicit cross-referencing, two types of explicit cross-referencing are used in *Thanodi ya Setswana* (THAN) namely:

| = what follows is an alternative to the defined lemma.

BONA (“SEE”) = the following is related to the explained.

THAN

**paraganya** ... taboga ka go tloatlola ga diphôlôgôlô fa di itumetse

BONA mokaragana

Here the user is cross-referred by means of the cross-referent marker BONA ‘see’ to the lemma sign **mokaragana**.

ZED

**jiya** ... Become thick, stiff;... [cf. shuḡa.]

**shuḡa** ... 2. Become thick, firm, set (of food, as porridge; of cement, plaster of Paris). [cf. jiya, shuḡa.]

TDV

**mutete** (*adj*) *cf -tete*

The user who looks up **mutete** finds the lemma sign **mutete** and the part of speech indicator ‘adjective’ as well as a cross-reference entry *cf -tete* but no further treatment is given. By means of the cross-reference marker, **cf**, (s)he is guided to the lemma sign representing the root **-tete** where an exhaustive treatment of this lexical item is given.

**-tete** (*adj*) sag(te) | soft **eg muroho mutete** sagte groente | soft vegetables

A rather interesting example of cross-reference is found in the Dictionary of Lexicography (DL) in its handling of the entries **Circular reference** and **Reference circularity**:

**Circular reference**

⇒ Reference circularity

and finds

**Reference circularity**

⇒ Circular Reference

After having felt victim to this cunning technique by which the user is put into an unending loop (s)he will exactly understand what the basic idea conveyed by 'circular reference' is! However, the application of circular reference diminishes the explicit transfer of information.

The lexicographer has to deal with three important types of cross-reference addresses namely the *internal*, *external* and *dictionary external* cross-reference addresses.

### 12.3 The internal cross-reference address

The first category is that of the internal cross-reference address. An article-internal mediostructural relation assists the user to relate various microstructural entries employed in the same article. With an internal cross-reference address the mediostructural relation *does not exceed the boundaries of the article*. This type of cross-referencing is used to ascertain coherence between different microstructural entries in one article. The *Woordeboek van die Afrikaanse Taal* (WAT) contains the following entries in the article of the lemma sign *kroon* (crown):

**kroon.** I ... **1.a.** Hoofsieraad ...

**b.** [Simboliese] voorstelling of afbeelding van 'n kroon (bet. I, 1 a), ...

**3. a. i.** Ornamentele kopbedekking wat herinner aan, of 'n namaaksel, voorstelling is van 'n kroon (bet I, 1 a) ...

In the treatment of sense 1 b and sense 3 a i cross-references are made to sense I 1 a, i.e. to a cross-reference address within the article.

### 12.4 The external cross-reference address

A second type of cross-reference address is the external cross-reference address. The cross-reference *exceeds the boundaries of the article*. Two search domains can be identified for external cross-reference addresses. Dictionary articles are texts but they also function as subtexts of the central list which is the dominating lexicographic text. The external address can be located either elsewhere in the central list, e.g. another lemma sign or a specified microstructural element in another article, or in a separate text outside the central list. Compare the articles of *gyro* and *stow* in the *Collins Dictionary of the English Language* (CDE) and the *Longman Dictionary of Contemporary English* (LDOCE) respectively.

**gy-ro** ... *n.*, *pl.* **ros**. **1.** See **gyrocompass**. **2.** See **gyroscope**.

**stow** ... *v* [X9 (AWAY)] **1** to put away or pack, esp. for some time: to stow goods (away) in boxes **2 stow it!** *s!* Be quiet!

The lexical item *gyro* is polysemous and has two different senses. The article of this lemma sign displays no meaning paraphrase for either of the polysemous senses but cross-refers the user instead to the treatment presented for two other lemma signs, i.e. *gyrocompass* and *gyroscope*. These lemma signs are the external cross-reference addresses located elsewhere in the central list. In the article of the lemma sign *stow* LDOCE includes the textsegment “[X9(AWAY)]”. The X9 cross-refers the user to a text in the back matter of the dictionary which contains a table of codes indicating a variety of grammatical values. X9 is explained in this table as a verb with one object as well as an additional descriptive word or phrase, e.g. *put + it + in the box*. The textelement *away* in the quoted textsegment is the additional word to be used with the verb *stow*. In this example the textsegment X9 is a cross-reference entry indicating an external address located in another text of the dictionary.

Quite often a combination of external and internal cross-reference addresses are given in one reference entry. In *The Concise Oxford Dictionary* (COD) the article of the lemma sign *ghosting* contains the following entries:

the appearance of a ‘ghost’ (see GHOST *n.* 4) or secondary image in a television picture.

In this excerpt the meaning paraphrase of the lemma sign *ghosting* is the cross-reference position containing a triple address which consists, as the main address, of an external cross-reference address located in the central list, i.e. the lemma sign *ghost*, as well as two additional internal addresses, i.e. a secondary address, the nominal function of this lexical item, and a tertiary address, the fourth polysemous sense of this item. The last two cross-reference addresses identify textsegments in the article of the lemma sign *ghost*.

## 12.5 The dictionary external cross-reference address

The third category of cross-reference addresses is the dictionary external cross-reference address. This mediostructural procedure *links a textsegment in a dictionary to a source outside the dictionary*. In his *A Dictionary of Language Planning Terms* Cluver (1993) puts the strategy of dictionary external cross-referencing to good use. The back matter of the dictionary contains a bibliography of sources in which more information regarding the terminology treated in the dictionary can be found. Many articles contain condensed bibliographical references which lead the user to the bibliography in the back matter, which is the cross-reference position from where the user is guided by means of a complete reference to the specific source. The condensed bibliographical references in the articles are clearly indicated by the cross-reference marker “**Bibl.**”. In the article of the lemma sign *primary language* the following textsegment is found: “**Bibl.** Mühlhäusler 1986:9”. The bibliography gives the full reference, i.e. “Mühlhäusler, P. 1986. *Pidgin and creole linguistics*. Oxford: Basil Blackwell.” By means of the dictionary external cross-reference address the lemma sign is linked to this external source. Compare also *Klein Noord-Sotho woordeboek* (KNS):

KNS

-i ... is the reflexive prefix. ... N.B. verbs commencing with the reflexive prefix should be looked up under the first sound of the basic verb; remember the sound changes effected by the i-; cf. the Handboek

A variety of other reference addresses can also be identified but they are not relevant for the present discussion.

## 12.6 Combatting decontextualisation

The use of a mediostructural strategy of external cross-reference addresses endeavours to enhance the functionality of a dictionary as a source reflecting aspects of the linguistic reality. One of the real problems experienced by the users of alphabetically ordered dictionaries is the decontextualisation of lexical items. Bolinger (1985:69) maintains that lexicography is an unnatural occupation. "It consists in tearing words from their mother context and setting them in rows - carrots and onions and beetroot and salsify next to one another - with roots shorn like those of celery to make them fit side by side, in an order determined not by nature but by some obscure Phoenician sailors who traded with Greeks in the long ago." He continues, arguing that half the lexicographer's labour "is spent repairing this damage to an infinitude of natural connections that every word in any language contracts with every other word, in a complex neural web knit densely at the centre but ever more diffusely as it spreads outward." According to Bolinger a "bit of context, a synonym, a grammatical category, ... and a cross-reference or two" are the additives that accomplish the repair. Compare Figure 1 as an excellent example of the lexicographer's effort to 're-unite' the different types of vegetables.

Thus from both a semantic and a pragmatic perspective the lexicon has to be regarded as an ordered set of lexical entries. However, the *alphabetical ordering* of a dictionary defies the network of semantic relations existing between this set of lexical entries. The mediostructure of a dictionary is a powerful mechanism to re-establish some of the lexical relations. Dictionaries employ the mediostructure to cross-refer the user to external addresses which are linked with the lemma sign of the cross-reference position article in relations such as synonymy, oppositeness of meaning, hyponymy, dialectal, stylistic, chronolectic and other forms of variation, etc. For the language learner as well as the seasoned native speaker of any given language these cross-references represent an added value which assists them in improving their communicative potential. South African dictionaries should employ external cross-reference addresses in a more general and consistent way. However, it is of extreme importance that these strategies be explained comprehensively in the front matter of the dictionary.

For the African languages, apart from the defiance of semantic relations, *alphabetical ordering* has serious detrimental consequences for *grammatical relations*. Many traditional compilers, although following an alphabetical ordering in principle, regard the importance of combined semantic and grammatical coherence as too important to break and revert to complicated lumping strategies as in Figure 2.

This view implies that in the case of African languages the mediostructure is incapable of re-establishing the most relevant lexical relations. In most dictionaries

this results in a hybrid approach where different derivations, sometimes a hundred or more, of a single word are treated within the article of a nominal or especially verbal stem in a complex article with numerous niched articles, in addition to these lexical items being entered as separate lemmas in their appropriate alphabetical positions. Compare the following section from the article of reka 'buy' in NSDN.

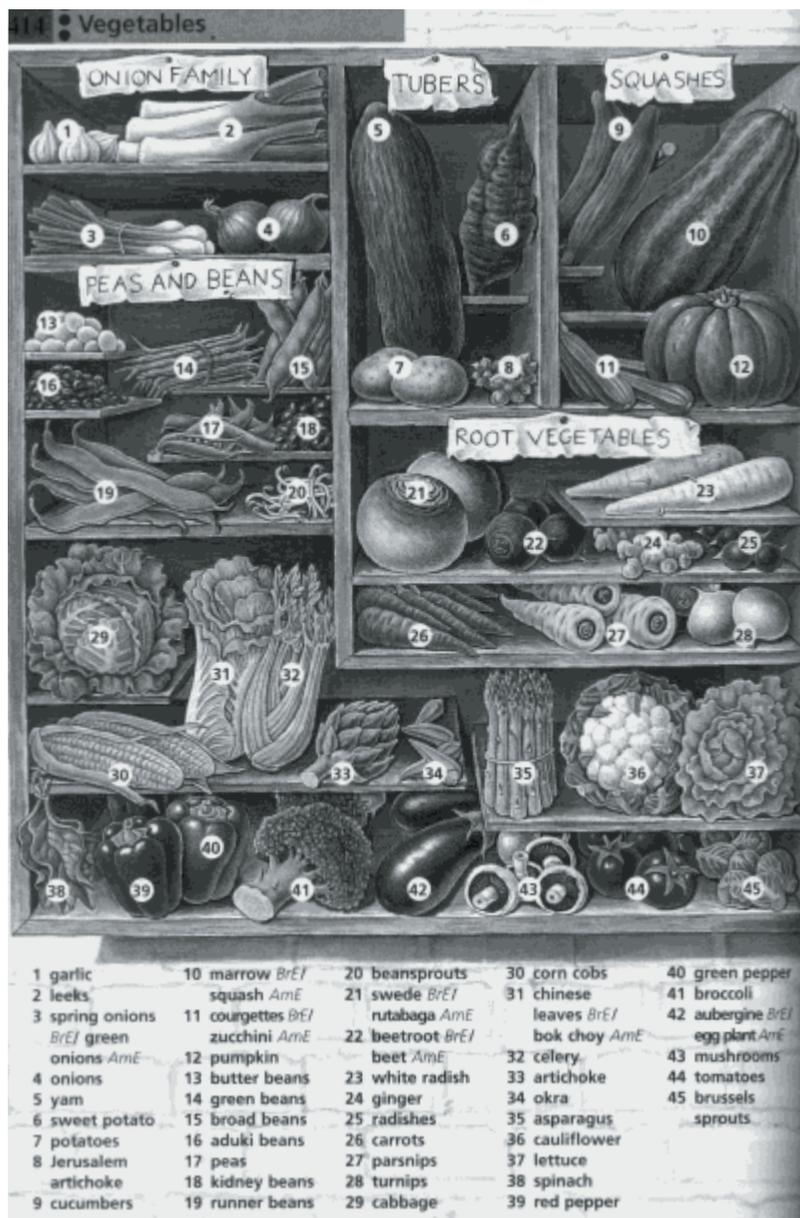


Figure 1: Colour plate for vegetables from LDOCE

**RÉKA** (-rēka, -rēkilē, -rēkwa, -rēkilwē) koop, aankoop, ruil // buy, purchase, barter; ~ *polasa* in weelde lewe // live in comfort/luxury; ~ *o lebélētše godimo* kat in die sak koop // buy a pig in a poke; *nku e rēkwa mosela* 'n mooi geboude dame is 'n aantrekkingskrag vir jongmans // a lady with a good figure easily attracts young men; **dirékárekáne** (**dirékarékane**) verskeidenheid gekoopte goedere // variety of things bought; **leréko**, **ma-** (**lerékó**) gewoonte/neiging om te koop // habit of buying, inclination to buy; **moréki**, **ba-** (**moréki**) pers. dev.; koper // buyer, purchaser; **seréki**, **di-** (**serékó**) pers. dev.; lustige koper // keen buyer; **seréko**, **di-** (**serékó**) impers. dev.; wat gekoop word, aankope // purchase(s); **théko**, (**n-**)/**di-** (**thékó**) man. dev.; koopwyse, prys // manner of buying, price; **RÉKANA** (-rēkana, -rēkane, -rēkanwa, -rēkanwe) rec.; ruil met mekaar // exchange with one another; *a re rēkané, wēna o mphē hēmpē éla, nna ke go fē diēta isē* laat ons met mekaar ruil, jy gee my daardie hemp en ek gee jou hierdie skoene // let us exchange, you give me that shirt, I will give you these shoes; **barékáni** (**barékani**) pers. dev.; **thékáno**, (**n-**)/**di-** (**thékano**) man. dev.; **RÉKANTSHA** (-rēkantsha, -rēkantshītšē, -rēkantshwa, -rēkantshītšwē) caus. < **RÉKANA**; (om)ruil, wissel (geld), inruil // exchange, barter, trade in, swop; **morékántshi**, **ba-** (**morékantshi**) pers. dev.; **serékántshwá**, **di-** (**serékantshwa**) impers. pass. dev.; **thékántsho**, (**n-**)/**di-** (**thékantshó**) man. dev.; omruiling, inruiling, wisseling // exchange, bartering, swopping; **RÉKANYA** (-rēkanya, -rēkantšē, -rēkanywa, -rēkantšwē) caus. < **RÉKANA**; (om)ruil, wissel (geld) // exchange, barter, swop; **morékányi**, **ba-** (**morékanyi**) pers. dev.; **serékanywá**, **di-** (**serékanywa**) impers. pass. dev.; **thékányo**, (**n-**)/**di-** (**thékanyó**) man. dev.; v. **thékántsho**; **RÉKÉGA** (-rēkēga, -rēkēgilē) neutr.; koopbaar w. // b. purchasable; **RÉKÉLA** (-rēkēla, -rēkētšē, -rēkēlwa, -rēkētšwē) appl.; koop vir // buy for; ~ *kolobé kgetsing* (< Afr.) kat in die sak koop // buy a pig in a poke; **borékélo** (**borékéló**) lo. dev.; koopplek // place where things are bought; **morékédi**, **ba-** (**morékédi**) pers. dev.; **morékélwá**, **ba-** (**morékélwa**) pers. pass. dev.; **serékélo**, **di-** (**serékéló**) impers. dev.; iets waarin jy koop // that into which one buys; **thékélo**, (**n-**)/**di-** (**thékéló**) man. dev.; maat, skaal (waarin bv. bier gekoop word) // measurement, bowl (one used for buying beer); **RÉKÉLANA** (-rēkélana, -rēkélane, -rēkélana, -rēkélane) appl. rec.; **barékélani** (**barékélani**) pers. dev.; **thékélano**, (**n-**)/**di-** (**thékélanó**) man. dev.; **RÉKISA** (-rēkiša, -rēkišitšē, -rēkišwa, -rēkišitšwē) caus.; laat/help koop, verkoop, van die hand sit // cause/help buy, sell; ~ *ka leleme* kul, mislei, verdraai // deceive, mislead, pervert;

~ *leleme* praatsiek w., skinder // gossip, b. loquacious, b. garrulous; ~ *motho a sa phela* iemand kul // deceive someone; ~ *motho lebake* iemand kul, 'n tevergeefse belofte maak, iemand verag weens sy slegte gedrag // deceive someone, give a vain promise, despise someone because of his bad conduct; ~ *segae* iets aan iemand so verkoop dat hy 'n goeie slag slaan omdat jy sy vriend of familielid is, afslag gee // sell to someone at bargain price because he is your friend/relative, give discount; **morékiši**, **ba-** (**morékiši**) pers. dev.; verkoper, verkoopsman, winkelier // seller, salesman, storekeeper; **serékišwá**, **di-** (**serékišwa**) impers. pass. dev.; **thékišo**, (**n-**)/**di-** (**thékišó**) man. dev.; verkoping, uitverkoping, afset, bemerking // sale, selling, market, marketing; **RÉKISANA** (-rēkišana, -rēkišane, -rēkišana, -rēkišanwe) caus. rec.; ruil met mekaar // exchange with one another; **barékišani** (**barékišani**) pers. dev.; **thékišano**, (**n-**)/**di-** (**thékišanó**) man. dev.; **RÉKIŠEGA** (-rēkišēga, -rēkišēgilē) neutr. < **RÉKISA**; verkoopbaar w. // b. sellable; **RÉKIŠETSÁ** (-rēkišētša, -rēkišēditšē, -rēkišētšwa, -rēkišēditšwē) caus. appl.; verkoop vir // sell for; **borékišetšo** (**borékišetšó**) lo. dev.; verkoopplek // selling place; **morékišetši**, **ba-** (**morékišetši**) pers. dev.; tagent // †(business) agent; **thékišetšo**, (**n-**)/**di-** (**thékišetšó**) man. dev.; **RÉKIŠETSANA** (-rēkišētšana, -rēkišētšane, -rēkišētšana, -rēkišētšanwe) caus. appl. rec.; sake verrig // transact business; **barékišetšani** (**barékišetšani**) pers. dev.; **thékišetšano**, (**n-**)/**di-** (**thékišetšanó**) man. dev.; besigheidstransaksie // business transaction; **RÉKÓLLA** (-rēkólla, -rēkólōtšē, -rēkóllwa, -rēkólōtšwē) rev. tr.; terugkoop, terugruil, geld terugva, los // buy back, exchange back, ask for a refund, redem; **morékólli**, **ba-** (**morékólli**) pers. dev.; **serékóllwá**, **di-** (**serékóllwa**) impers. pass. dev.; 'n ding wat teruggekoop word // that which is bought back; **thékólló**, (**n-**)/**di-** (**thékólló**) man. dev.; (Bl.) lossing // (Bl.) redemption; **RÉKÓLLANA** (-rēkóllana, -rēkóllane, -rēkóllana, -rēkóllanwe) rev. rec.; **barékóllani** (**barékóllani**) pers. dev.; **thékóllano**, (**n-**)/**di-** (**thékóllanó**) man. dev.; **RÉKÓLLELA** (-rēkóllēla, -rēkóllētšē, -rēkóllēlwa, -rēkóllētšwē) rev. appl.; **morékóllēdi**, **ba-** (**morékóllēdi**) pers. dev.; **thékóllēlo**, (**n-**)/**di-** (**thékóllēló**) man. dev.; **RÉKÓLLELANA** (-rēkóllēlana, -rēkóllēlane, -rēkóllēlana, -rēkóllēlanwe) rev. appl. rec.; **barékóllēlani** (**barékóllēlani**) pers. dev.; **thékóllēlano**, (**n-**)/**di-** (**thékóllēlanó**) man. dev.; **RÉKÓLLISA** (-rēkólliša, -rēkóllišitšē, -rēkóllišwa, -rēkóllišitšwē) rev. caus.; **morékólliši**, **ba-** (**morékólliši**) pers. dev.; **thékóllišo**, (**n-**)/**di-** (**thékóllišó**) man. dev.; **RÉKÓLLISANA** (-rēkóllišana, -rēkóllišane, -rēkóllišana, -rēkóllišanwe) rev. caus. rec.; **barékóllišani** (**barékóllišani**) pers. dev.; **thékóllišano**, (**n-**)/**di-** (**thékóllišanó**) man. dev.

Figure 2: Treatment of *reka* in NSDN

In dictionaries such as NSDN word stems and their derivations are clustered together in one huge article with the noun or verbal root as the lemma often containing up to

eighteen levels of niched lemmata. Where derivations are entered separately in their appropriate alphabetical positions as lemma sign in (NSDN), the articles are hardly more than cross-reference articles containing only minimal grammatical information along with a cross-reference back to the lemma sign heading the cluster.

NSDN

**thekollano**, (n-)/di- v. REKA

**thekollelano**, (n-)/di- v. REKA

**thekollelo**, (n-)/di- v. REKA

**thekollišano**, (n-)/di- v. REKA

**thekollišo**, (n-)/di- v. REKA

**thekollo**, (n-)/di- v. REKA

In this way mediostuctural procedures are exhausted/overused for the sole purpose of maintaining structural links. Little or no realisation of the mediostucture as a powerful access procedure is achieved. Once cross-referred back to the main cluster, its unlikely that the user will be able to work out the meaning, especially for those cases which lie relatively deep in the modular structure as in the case of **dithekollišano**. The user has to look up this word under the singular **thekollišano** and is then cross-referred to **reka** in Figure 2, and eventually, after having struggled through this lengthy article, (s)he finds **thekollišano** at the end with no translation equivalents given, cf. Prinsloo (1994) for similar examples and a detailed discussion on problematic aspects of the lemmatisation of verbs.

This obsession of keeping together what in their view semantically and grammatically “belong together” thus results in extremely user-unfriendly articles in which successful retrieval of information virtually becomes impossible.

It was stated in the introduction that one factor for the evaluation of a dictionary is the extent to which it is useful to the user. Dictionaries such as these fail on this main criterion. Students consequently opt for less sophisticated dictionaries with less data categories, less exhaustively treated lemmas, i.e. a lower density of data.

## 12.7 Cross-references aimed at additional information retrieval

It is of the utmost importance that the user should find more data at the cross-reference address. Otherwise the value of cross-referencing is devaluated. A cross-reference, or more specific, the position of a cross-reference entry, simply indicates to the user that this is only the starting point in the process of information retrieval. It initiates a new dictionary consultation procedure. The usage frequency of the item which stands in the cross-reference position is lower than that of the cross-reference address.

Compare the articles of **kgarebê** and **lekgarebê** in THAN and Thanodi ya Setswana ya Dikole (THAND).

THAN

**kgarebê** TTT |lekgarebê *In./9.* ma-. mosetsana yo o godileng mme a ise a nyalwe

**lekgarebê** TTTT *In./9.* ma-. 1. mosetsana yo o lekaneng go nyalwa 2. mosetsana yo o itlkhôkômêlang a apara sentlê

Explicitly cross-referring the user from **kgarebê** to **lekgarebê** makes good sense since apart from the meaning “girl who can be married”, which is similar to that given in sense 1. of **lekgarebe**, an *extended* meaning “a neatly, well-dressed girl” is given as sense 2. The fact that no explicit cross-reference from **lekgarebe** to **kgarebe** is given, is also quite acceptable since the user who looked up **lekgarebê** will not find any new data in looking up **kgarebê**. However **kgarebe** must be given as a synonym directly following the sense 1 definition. THAN’s handling of **kgarebê** can also be improved in respect of the *position* allocated to the cross-reference entry. The explicit cross-reference |**lekgarebê** should not be given in the focus position of the article. It can be regarded as a non-necessary or even non-functional cross-reference which is interfering with the user’s information retrieval process. Formulated differently, the information primarily needed by the user who looks up **kgarebê** is that given in the *definition*. Once given the definition, (s)he might be interested to consult the cross-reference address for additional information.

Consider now THAND’s treatment of the same words which are both lemmatised:

**kgarebê (ma)** mosetsana yo o godileng mme a ise a tsewe (nyalwe).  
**lekgarebê (ma)** kgarebê; mosetsana yo o ka tšewang.

In the case of **kgarebê** only a definition is offered while a synonym as well as a definition is given for **lekgarebê**. No explicit cross-reference is given. Since no cross-reference is given from **kgarebê** to **lekgarebê** it suggests that **kgarebe** is the entry with the higher usage frequency. However the user gets more information from the treatment of **lekgarebê**, namely a synonym as well as a definition, than from **kgarebê**. This is confusing. The article of the lemma sign **lekgarebe** is a cross-reference position of the cross-reference entry **kgarebe**. Normally, for economical reasons, the same definition is not given in two places. Two definitions and the lack of a cross-reference have a negative effect on coherence. Here the user cannot derive/conclude which one is the more frequently used. The more frequently used word is the one likely to be treated more comprehensively; thus in itself, an indication of higher frequency of use. It would thus be better to enter **kgarebe** with a definition adding **lekgarebe** as a synonym. It is normal practice to give a list of synonyms after the definition since they comply to the criteria to be lemmatised themselves. Such synonyms can be listed in order of frequency of use if such criteria are available or otherwise alphabetically. Thus, since all synonyms have to be entered as lemmas, **lekgarebe** will be entered as a lemma sign but only with a cross-reference to **kgarebe**.

It is also not clear why in both THAN and THAND the definitions differ in respect of the concept “grown up”. In the case of **lekgarebe** “a girl who can be married” and in the case of **kgarebe** “a grown up girl, one who is not yet taken/married”. In comparison the user can get the incorrect impression that **kgarebe** implies an adult and **lekgarebe** not.

## 12.8 Dead cross-references

One of the basic errors sometimes made by lexicographers is to give a cross-reference entry cross-referring the user to a cross-reference address that *does not exist*. Consider the following examples:

## BASIS

**uitgee** ... 1. Wanneer 'n mens geld uitgee op iets, gebruik jy geld daarvoor. ... Kyk BESTEE. 2. Wanneer 'n boek uitgegee word, word hy gedruk en te koop aangebied. ... Kyk PUBLISEER.

**ontydig** ... iets wat ontydig gebeur, gebeur nie op die regte tyd nie. ... Sien TYDIG

All three cross-references to *bestee* (spend), *publiseer* (publish) and to *tydig* (timely) are dead references since these words are not entered as lemma signs in the dictionary.

In the guidelines to *A Learners Chichewa and English Dictionary* (LCE) the compilers explain the policy not to lemmatise derived forms when the meaning is readily ascertainable from the root + suffix combination. In support of this far reaching decision for lemmatisation of an African language, they include cross-references to the central text: "Thus, both **-mva** 'hear understand' and its derived form **-mvana** 'get along together' are listed" That's fine, but the *very examples* that they quote to illustrate their policy were not treated as such: **-mva** is listed but not **-mvana**. The damage done by the dead cross-reference to **-mvana** is far reaching since the user is now in doubt about the treatment policy not only in respect of a single entry but a whole category of entries. Exactly the same thing occurs in the following sentence: "the derived verb **-mverana** 'listen to each other' is not listed because its meaning is readily determined from the root **-mvera** 'listen to' plus an affix". However, once again the root **-mvera** is not listed, clearly violating the claim "verbs are entered according to their root forms...". Cross-references from the front matter texts, especially the user's guidelines, to the central text, are crucial to the user for successful or optimal retrieval of information. Dead references, especially in the guidelines text of a dictionary are serious mistakes which undermine the trust of the user in the dictionary as a trustworthy source of data, and in the value of the cross-referencing system as a whole. Such dead references often do not effect only one cross-reference address, but a key to a whole section can be lost.

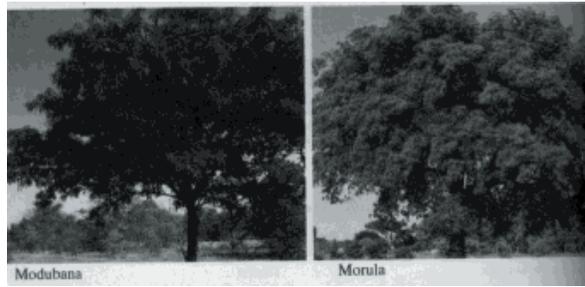
## 12.9 Failure to utilise cross-referencing where needed

The lexicographer should not miss out on golden opportunities to utilise a system of cross-referencing, especially in those cases where an excellent potential cross-reference address exists.

## THAN

**morula** ... setlhare se se tswang dikungwa tse di kgolokwe tse di jewang e bile di kgona go dira bojalwa

**modubana** ... mofuta wa setlhare se dikungwa di kima



In this dictionary both **morula** and **modubana** are entered and treated in the central word list and full colour pictures are given in the back matter. Unfortunately no cross-references are given from **morula** or **modubana** to the back matter nor from the back matter to the central list.

### 12.10 Cross-references to the wrong cross-reference address

The lexicographer should make sure that the user is referred to the correct cross-reference address especially in cases where homonyms or closely related polysemous meanings are given.

Cross-referencing has not been employed to its full potential in dictionaries for most African languages. Typical errors and shortcomings will be briefly outlined below:

Consider *New English Northern Sotho Dictionary* (NEN)'s handling of **molelo** versus **mollo**:

**molelo**, see **mollo**.

'**mol'lo**, n., fire, witch-weed, principal wife;...

**mol'lo**, n., cry, (manner of) crying.

The correct cross-reference address to which the user should be cross-referred from the entry for **molelo** is '**mol'lo** and not **mol'lo**. The latter has no relation to **molelo** whatsoever. The convention " ' " is not explained in the dictionary anyway. Since '**mollo** and **mollo** represent a fairly rare situation where two words in Sesotho sa Leboa can neither be distinguished phonetically (both **mollô**) and having an identical *tonal pattern* (Low-Low-Low), the compiler should therefore distinguish by means of homonym numbers, i.e. **mollo**<sup>1</sup> and **mollo**<sup>2</sup>. The decision of the compiler to refer the user who looks up **molelo** to **mollo** without treating **molelo** is however acceptable in terms of frequency of use criteria, since **mollo** is frequently used and **molelo** not. Thus no cross-reference from **mollo** to the less frequently used **molelo** is necessary or appropriate because the target user of this dictionary is looking for translation equivalents in the target language and is not interested in (more) information in the source language. However, within the article of **mollo**<sup>1</sup>, a cross-reference should be given to **molelo** but then labeled as *dialectical* or treated by means of an *inserted text*. Compare the additional data given by means of inserted text in RD in the case of **rekenaar** versus **komper**:

**kom'per** =s computer; *vid.* **Rekenaar**, **rekenoutomaat**.

## WORDS IN ACTION

**Komper, rekenaar, rekenoutomaat**

Komper (computer) is the word used by some speakers and writers in the Western Cape.

Other people there and most people elsewhere in South Africa use rekenaar. Rekenaar is also the word preferred by people in the computer profession. ...

Cross-references not guiding the user to an unambiguous cross-reference address have a negative effect on the target user. Once disappointed, it will discourage him/her from following up on other cross-reference entries since (s)he is unable to distinguish functional and valid cross-references from non-functional and invalid cross-references in this dictionary.

### 12.11 Cross-references that misguide the user in respect of information retrieval

Consider the following articles in the *Northern Sotho Terminology and Orthography* (NSTO)

<b>complainant</b> (see plaintiff)	klaer	mmelaedi, molli, mmegi
<b>plaintiff</b>	eiser, klaer	mmelaedi, molli, mmegi, motlalei

In the article of the lemma **plaintiff** the addition of the translation equivalents **eiser** in the Afrikaans column and **motlalei** in the Sesotho sa Leboa column, raise a few questions. Firstly, it implies that **eiser** and **motlalei** are suitable equivalents for **plaintiff** but not for **complainant**. Secondly, giving **eiser** as the first translation equivalent for **plaintiff**, suggests that it is the best option. However, in the case of **motlalei**, although it is added to the translation equivalent paradigm for the sake of **eiser**, it is given at the end. Thus the entire relationship between **complainant** and **plaintiff** is unclear. The user cannot determine in which relation they stand to each other.

It is unclear why no cross-reference from **plaintiff** to **complainant** is given. Such a cross-reference is necessary because equivalents in both target languages are given under **complainant**. A central list internal cross-reference should strengthen the coherence, as was correctly done in the case of **molelo** versus **mollo** above. In the case of **complainant** versus **plaintiff** this coherence is actually broken off. The user who wants to find translation equivalents in Afrikaans and Sesotho sa Leboa is cross-referred to another word where the same treatment is given for no reason.

Consider now NSTO's handling of **bracket** versus **brackets**:

NSTO		
a. <b>bracket</b> (symbol) (see: brackets)	hakie	lešakana
b. <b>brackets</b>	vierkantige hakies	mašakanakhutlwana

In the a-example translation equivalents in Afrikaans and Sesotho sa Leboa are given for **bracket**. The cross-reference to **brackets** is quite appropriate since the latter is more frequently used. Also, due to frequency of use considerations, no

cross-reference from **brackets** to **bracket** is necessary. However, in looking up **brackets**, the user does not get any additional information, e.g. in respect of types and use of brackets. To the contrary, (s)he is misguided by the additional information given at the cross-reference address namely that the lemma **brackets** is translated in Afrikaans and Sesotho sa Leboa as necessarily *square*. Thus the plural form, in contrast to the singular, **bracket**, excludes other types of brackets .

### 12.12 Using cross-referencing to avoid a full treatment of the lemma

Lexical items most likely to be looked up by the target user should be treated, rather than unnecessarily having a cross-reference to another entry. The lexicographer may never utilise the system of cross-referencing simply because (s)he is too lazy to give proper treatment to the items in question. If it is in the interest of the target user that a specific lemma should be entered and treated, it has to be done. The non-treatment of the word **syllable** in *Dictionary of Northern Sotho Grammatical Terms* (DGT) can be offered as a typical example:

**syllable** (*noko, sillabe/lettergreep*)      See *nucleus*.

Firstly **syllable** deserves full treatment, especially in a dictionary of grammatical terms. Apart from translation equivalents in Sesotho sa Leboa and Afrikaans no definition is given, only an explicit cross-reference to **nucleus**. In the article of **nucleus**, many references are once again made to **syllable**, such as “[a nucleus] is used to characterise the nature of a *syllable* ... As far as the *syllable* is concerned, it is maintained that vowels form the nuclei of syllables...”, etc. However, **syllable** itself remains undefined.

### 12.13 Unidirectional versus bi-directional cross-referencing

Compare the following dictionary articles of Sesotho sa Leboa:

**bagolo** ... | -golo  
**legolo**... | -golo  
**magolo**... | -golo  
**megolo**... | -golo  
**mogolo**... | -golo  
**segolo**... | -golo

-golo *adj.* big, large; HL; *lepokisi le legolo a big box*

In this example the unidirectional (one way) reference is quite appropriate because the inexperienced user who looks up the full form of the adjective (class prefix plus stem) is correctly guided by means of a cross-reference to the lemma sign **-golo** where a full treatment of the adjective is given. It would be totally inappropriate for the lexicographer to refer the user from **-golo** to one or more of the forms **mogolo**, **segolo**, **legolo**, etc. since no, or less information is given in the latter cases. However, in order to enable the user to retrieve as much information as possible unidirectional cross-referencing should be used sparingly. Where cross-referencing occurs, e.g. from the less frequently used lexical item to the more frequently used one, from

a dialectical form to the standard language variant or from a form representing old spelling to the latest norm, the lexicographers will do well to include these less frequently used forms as part of the treatment in the articles of the more frequently used forms. This will help to enhance the data presentation and to explicate relations holding in the lexicon.

Bi-directional cross-referencing should be seen as the default cross-referencing procedure, also in cases where a comprehensive lexicographic treatment is given to both lexical items and where the user will benefit from the information given in the 'other' entry. Compare the entry for **tone** versus **Bantu languages**:

DGT

**tone** (*segalô, toon*)

Tone can be defined as *pitch variations* which affect the meaning and function of words. **Tone** is one of the distinctive features of the Bantu language family (see **Bantu languages**), and in these languages differences in **tone** between words which have exactly the same shape, result in a difference in meaning. Two basic tones (also called *tonemes*) are usually distinguished, namely a **high tone** and a **low tone**, although more detailed distinctions are often drawn between, for example *rising* and *falling tones*, *mid*, *mid-high* and *mid-low* tones, etc. A **tone** (or *toneme*) is always associated with a particular **syllable**, i.e. there are as many **tones** in a word as there are syllables since **tones** realise on **vowels**. This is one of the reasons why vowels are often referred to as *syllable nuclei*. (See: **nucleus**.) ...

**Bantu languages** (*\*maleme a Babaso, Bantoetale*)

A linguistic term used internationally to refer to a **language family** of which the members exhibit certain common characteristics such as that they are mainly agglutinating in nature (see **agglutinating languages**), that they all have a system according to which nouns are grouped into classes as well as a system of grammatical agreement (see **agreement**) while tone plays a distinctive role. (See **tone**.) More than 400 languages belong to this family, and they are distributed over a very wide area in Africa which roughly lies south of an imaginary line which stretches from the Cameroon area on the West coast to more or less the Lake Victoria area and Kenya on the East coast. Although the term **Bantu** is used internationally to refer to this family of languages, it became stigmatised in South Africa due to political reasons. Consequently, writers have recently tried to avoid the use of the term **Bantu**, by *inter alla* replacing it with the term **African languages**. The latter term is, however, inappropriate, since not all languages which are spoken in Africa belong to the **Bantu language family**. (See **African languages**.) Other terms such as **Sintu**, **Kintu** and **Ntu** have also been suggested, but none of these have become generally accepted. At the 7th biennial international conference of the African Language Association of Southern Africa held at the University of the Witwatersrand from 6 to 9 July 1993, several internationally acclaimed scholars from the United States, Germany and France, as well as scholars from African universities such as the Universities of Lubumbashi and Dar-es-Salaam, freely used the term **Bantu** in their conference papers. In his paper, Prof Kamba-Muzenga of the University of Lubumbashi expressed the hope that his free and unbiased use of the term will contribute to it being rid of the negative connotations which have come to be associated with it in South Africa. These scholars have also pointed out that there is no internationally acceptable term which can be introduced instead of **Bantu** to refer to this family of languages. The view is generally held that, when used in the proper context, i.e. as a purely **linguistic** term which refers to **a family**

of *languages* and nothing else, there should be no reason at all for the term **Bantu** to be offensive.

In the article of **tone**, for example, explicit reference is made to **Bantu languages** and **nucleus**. At the reference address, **Bantu languages**, the user finds more useful information on tone in the African Languages. Likewise, the user who firstly consulted the entry **Bantu languages** will find, in addition to other useful information given there, “tone plays a distinctive role. See **tone**”. This is good lexicography since for User A who consulted the lemma **Bantu languages**, as well as User B who looked up **tone**, the cross-references were useful because they obtained more information at the respective reference addresses in respect of two important and closely related issues such as *African Languages* and *tone*. The same holds true for the explicit cross-reference made to *tone* in the article of *syllabic nasal*.

### 12.14 Implicit versus explicit cross-referencing

A cross-reference entry can be either implicit or explicit. The implicit entry does not contain a cross-reference marker and is therefore a more condensed entry. However, there is no fundamental difference between implicit and explicit cross-referencing. Compare once again the derivations of *reka* given here as an example of implicit cross-referencing where the user who looks up any derived noun in Sesotho sa Leboa is cross-referred to the verb without the explicit use of cross-reference markers.

NSDN

**thekollano**, (n-)/di- v. REKA

**thekollelano**, (n-)/di- v. REKA

**thekollelo**, (n-)/di- v. REKA

**thekollišano**, (n-)/di- v. REKA

**thekollišo**, (n-)/di- v. REKA

**thekollo**, (n-)/di- v. REKA

It is however important that the user should be able to distinguish clearly between cases of implicit cross-referencing and other conventions used in the dictionary. (S)he should not for example confuse italics used for the sake of emphasis with implicit cross-referencing as is the case in DGT. An important statement in the article of **tone** in DGT reads: “Tone is always associated with a particular *syllable*”. The user of DGT consulting **tone** could easily perceive the italicized word *syllable* as an implicit reference entry but find it to be non-functional since in looking up **syllable** (s)he is referred to *another* address namely **nucleus**. Also note that although **tone** is one of the key issues discussed in the article of **nucleus**, no *explicit* cross-reference is given to *tone*.

Key terms used in the treatment of the lemma **tone** which are italicized such as *pitch variations*, *tonemes*, and especially *syllable* are not treated in the dictionary. The user expects a clearer distinction between *implicit cross-reference* to a different cross-reference address, on the one hand, and mere instances of *emphasis* on the other.

This does not mean that the lexicographer should solely utilise explicit cross-references to distinguish between emphasis and cross-referencing since there is no fundamental difference in value between explicit and implicit cross-reference systems. The former is simply more obvious than the latter. Thus it is suggested

that the lexicographer should utilise both as long as implicit cross-references can be clearly distinguished from mere emphasis. The implicit cross-reference strategy must however be clearly apparent. This simply means that terms used within the articles of entries which are themselves lemmatised and treated elsewhere in the dictionary must stand out and be treated consistently. Yet again the type of dictionary, the target users and the reference skills of the target users will play a decisive role. Implicit cross-references demand more skills from the user than explicit cross-references.

### **12.15 In conclusion**

It can be concluded that application of the mediostructure is a useful and essential strategy available to the lexicographer to enhance the quality of dictionary articles by referring users to reference addresses where more information can be retrieved. (S)he should maximally utilize both implicit and explicit cross-reference strategies and be aware of all the pitfalls outlined above such as dead cross-references, misleading cross-references, or cross-references to the wrong cross-reference address, etc.

In DL the compilers regard the mediostructure as so important that they even use uppercase as implicit reference addresses for the defining vocabulary.

**Lexicography** The professional activity and academic field concerned with DICTIONARIES and other REFERENCE WORKS. It has two basic divisions: lexicographic practice, or DICTIONARY-MAKING, and lexicographic theory, or DICTIONARY RESEARCH. ...

It can rightfully be argued that the lexicographer should guard against excessive text condensation. However, (s)he should utilise opportunities to strengthen the coherence of the dictionary by optimal organization of the mediostructure as guiding structure to assist the user and to enhance access to data in order to ensure successful dictionary consultation procedures.

# Bibliography

---

## Dictionaries

- (ANNA) Martin, W. Gouws, R.H., Renders, L., Maks, I. (eds.) (in preparation) *Woordeboek Nederlands-Afrikaans/Afrikaans-Nederlands*.
- (BASIS) Gouws, R., I. Feinauer, F. Ponelis. 1994. *Basiswoordeboek van Afrikaans*. Pretoria: J.L. van Schaik.
- (CDE) Hanks, P. (ed.) 1979. *Collins Dictionary of the English Language*. London: Collins.
- (CED) Collins Cobuild, 1997. *New Student's Dictionary*. Birmingham: Harper Collins.
- Cluver, A.D. de V. 1993. *A Dictionary of Language Planning Terms*. Pretoria. University of South Africa.
- (COBUILD) Sinclair, J. (ed.) 1988. *Collins COBUILD English Language Dictionary*. London: HarperCollins.
- (COBUILD 2) Sinclair, J. (ed.) 1995. *Collins COBUILD English Dictionary*. London: HarperCollins.
- (COD) Allen, R.E. (ed.) 1990<sup>8</sup> *The Concise Oxford Dictionary*. Oxford: Clarendon Press.
- (CSD) Rycroft, D. K. 1982. *Concise SiSwati Dictionary*. Pretoria: J.L. van Schaik.
- (CZD) Dent, G. R. and Nyembezi, C. L. S. 1961 *Compact Zulu Dictionary*. Pietermaritzburg: Shuter & Shooter.
- (DGT) Louwrens, L.J. 1994. *Dictionary of Northern Sotho Grammatical Terms*. Pretoria: Via Afrika.
- (DL) Hartmann R.R.K. & G. James 1998. *Dictionary of Lexicography*. London: Routledge.
- (DS) Snyman, J.W. et al. 1990. *Dikišinare ya Setswana English Afrikaans Dictionary*. Pretoria: Via Afrika.
- (ECC) *Electronic Collins COBUILD*, Lingea Lexicon Version 3.1 1997. Birmingham: HarperCollins.
- (ETTE) *English - Tsonga Tsonga – English Pocket Dictionary*. Johannesburg: Sasavona. 1982.
- (EWA) Van Wyk, G. et al (eds.) 2003. *Etimologiese woordeboek van Afrikaans*. Stellenbosch: Bureau of the WAT.
- (EZD) Doke, C. M., Malcolm, D. McK., and Sikakana, J. M. A. 1958. *English and Zulu Dictionary*. Johannesburg: Witwatersrand University Press.
- (GW) Eksteen, L.C. et al. (eds.) 1997<sup>14</sup>. *Groot Woordeboek/Major Dictionary*. Cape Town: Pharos.
- Harper Collins. *Electronic Collins COBUILD*, Lingea Lexicon Ver 3.1 1997.
- (HAT) Odendal, F.F & R.H. Gouws 2000<sup>4</sup>. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Doornfontein: Perskor.
- Johnson, S. 1755. *A Dictionary of the English Language*. London: J. & P. Knapton, T. & T. Longman et al.
- (KNS) Ziervogel, Dirk & Mokgokong, Pothinus C.M. 1988<sup>4</sup>. *Klein Noord-Sotho woordeboek, N.-Sotho-Afrikaans-English, Afrikaans-N.-Sotho, English-N.-Sotho*. Pretoria: J.L. van Schaik.
- (KW) Kromhout, J. 1997<sup>13</sup>. *Klein woordeboek / Little Dictionary, Afrikaans-Engels, English-Afrikaans*. Cape Town: Pharos.

- (LCE) Botne R & Kulemeka A.T. 1995. *A Learner's Chichewa and English Dictionary*. Cologne: Rüdiger Köppe.
- (LDOCE) Procter, P. (ed.) 1978. *Longman Dictionary of Contemporary English*. Harlow: Longman.
- Martin, W. et al. (eds.) 1989. *Van Dale Groot woordenboek Nederlands-Engels/Engels-Nederlands*. Utrecht/Antwerpen: Van Dale Lexicografie.
- (MED) *Macmillan English Dictionary for Advanced Learners*. 2002. Macmillan Publishers Limited.
- (NEN) Kriel, Theunis J. 1976<sup>4</sup>. *The New English – Northern Sotho Dictionary, English – Northern Sotho, Northern Sotho – English*. Johannesburg: Educum Publishers.
- (NeW) Du Plessis, M. 1999. *New Words Nuwe Woorde*. Cape Town: Pharos.
- (NSD) Prinsloo, D. J. and Sathekge, B. P. 1996. *New Sepedi Dictionary, English – Sepedi (Northern Sotho), Sepedi (Northern Sotho) – English*. Pietermaritzburg: Shuter & Shooter.
- (NSDN) Ziervogel, D. and Mokgokong, P. C. M. 1975. *Comprehensive Northern Sotho Dictionary, Northern Sotho–Afrikaans/English*. Pretoria: J.L. van Schaik.
- (NSSD) Chapole, S. R. 1997. *New South Sotho Dictionary*. Pietermaritzburg: Shuter & Shooter.
- (NSTO) Department of Education and Training, 1988. *Northern Sotho Terminology and Orthography No. 4*. Pretoria: Government Printer.
- (NW) De Villiers, Meyer. et al. *Nasionale Woordeboek*. Cape Town: Nasou, 1987<sup>7</sup>.
- (NWSG) Gouws, R.H., M. Stark, L. Gouws 2004. *Nuwe woordeboek sonder grense*. Cape Town: Maskew Miller Longman.
- (OALD) Crowther, Jonathan. ed. 1995. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press.
- (OED) Murray, J et al. (ed.) 1933, 1989<sup>2</sup> Simpson, J. & E. Weiner et al. (eds.) *Oxford English Dictionary*. Oxford: Oxford University Press.
- (POD) *Pocket Oxford Dictionary* (on CD ROM). 1994. Oxford: Oxford University Press.
- (POP) Kriel, T.J. 1988. *Popular Northern Sotho Dictionary*. Pretoria: J.L. van Schaik.
- (PUKU 1) Kriel, T.J. 1983. *Pukuntšu Dictionary*. Pretoria: J.L. van Schaik.
- (PUKU 2) Kriel, Theunis J. & Van Wyk, Egidius B. 1989<sup>4</sup>. *Pukuntšu woordeboek, Noord-Sotho–Afrikaans, Afrikaans–Noord-Sotho*. Pretoria: J.L. van Schaik.
- (PUKUTLH) Mojela, M.V., M.P. Mogodi, M.R. Selokela and M.C. Mphahlele. 2004. *Pukuntšuthaloši ya Sesotho sa Leboa*. Pretoria: University of Pretoria.
- (RD) Grobbelaar, P. et al. (eds.) 1987. *Reader's Digest Afrikaans-Engelse Woordeboek/English-Afrikaans Dictionary*. Cape Town: The Reader's Digest Association, South Africa (Pty) Ltd.
- (SASD) Chambers-Macmillan, 1996. *South African Student's Dictionary*. Manzini: Macmillan Boleswa.
- (SED) Brown, J., 1985. *Setswana English Dictionary*. Pula Press.
- (SESD) Matumo, Z.I., 1993. *Setswana-English-Setswana Dictionary*. Gaborone: Macmillan.
- (SSED) Mabile, A. and Dieterlen, H. 1988. *Southern Sotho-English Dictionary*. Revised by R.A. Paroz. Morija: Morija Sesotho Book Depot.
- Summers, D. (ed.) 1982. *Longman Dictionary of English Language and Culture*. Harlow: Pearson Education.
- (SZD) Dent, G.R. and Nyembezi C.L.S. 1988. *Scholar's Zulu Dictionary*. Pietermaritzburg: Shuter & Shooter.
- (TAW) Du Plessis, M. 1993. *Tweetalige Aanleerderswoordeboek/Bilingual Learner's Dictionary*. Cape Town: Tafelberg.

- (TDV) Wentzel, P.J. and Muloiwa, T.W. 1982. *Improved Trilingual Dictionary. Venda – Afrikaans/English*. Pretoria: UNISA.
- (TFW) Joubert, P.A. 1997. *Tweetalige frasewoordeboek/Bilingual Phrase dictionary*. Cape Town: Pharos.
- (THAN) Kgasa, M.L.A. & Tsonope, J. 1995. *Thanodi ya Setswana*. Botswana: Longman.
- (THAND) Kgasa, M.L.A. (ed.) 1976. *Thanodi ya Setswana ya Dikole*. Cape Town: Longman Penguin Southern Africa.
- (TW) Bosman, Daniël B. et al (eds.) 1984<sup>8</sup>. *Tweetalige woordeboek/Bilingual Dictionary*. Cape Town: Tafelberg.
- (VAW) Labuschagne, F.J. en L.C. Eksteen (eds.). 1993<sup>8</sup>. *Verklarende Afrikaanse Woordeboek*. Pretoria: J.L.van Schaik.
- (VDGW) Martin, W. & G.A.J. Tops (eds.) 1989<sup>2</sup>. *Van Dale Groot Woordenboek Engels-Nederlands*. Utrecht: Van Dale.
- (WAT) Schoonees, P.C. et al (eds.) 1951 - *Woordeboek van die Afrikaanse Taal*. Stellenbosch: Bureau of the WAT.
- (WAZ) Dekker, A. M. and Ries, J. H. 1958. *Woordeboek. Afrikaans – Zoeloe, Zoeloe-Afrikaans. Afrikaanse Pers Boekhandel*.
- (WDF) Kempcke, G. et al. 2000. *Wörterbuch Deutsch als Fremdsprache*. Berlin, New York: Walter de Gruyter.
- (WLWF) Beißwenger, M. et al. (eds.) to appear. *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research*. Berlin: De Gruyter.
- (WNT) De Vries et al. (eds.) 1864-1998. *Woordenboek der Nederlandsche Taal*. 's-Gravenhage: SDU Publishers.
- (W9) Mish, F.C. (ed.) 1987. *Webster's Ninth New Collegiate Dictionary*. Springfield, Massachusetts: Merriam-Webster.
- (ZED) Doke, C. M. and Vilakazi B. J. 1948. *Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.

## Other literature

- Al-Kasimi, A.M. 1977. *Linguistics and bilingual dictionaries*. Leiden: E.J. Brill.
- Almind, R. and Bergenholtz, H. 2000. Die ästhetische Dimension der Lexikographie. In: Fix, U. & H. Wellmann (eds.) *Bild im Text – Text und Bild*. Heidelberg: C. Winter: 259-288.
- Almind, R. 2005. Dictionaries through databases. In: Gottlieb, J.E. Mogensen, A. Zettersten (eds.) 2005. *Symposium on Lexicography XI*. Tübingen: Max Niemeyer: 101-109.
- Atkins, B.T. Sue, Michael Rundell and Edmund Weiner. 1997. *Salex'97. A Training Course in the Compilation of Monolingual Dictionaries*. Unpublished course material of a tutorial held at the Dictionary Unit for South African English, Rhodes University, Grahamstown, 15-26 September 1997.
- Atkins, B.T.Sue, (Ed.). 1998. *Using Dictionaries. Studies of Dictionary Use by Language Learners and Translators*. Max Niemeyer Verlag. Tübingen.
- Atkins, B.T. Sue. 1998a. Some discussion points arising from Afrilex-Salex'98. (Unpublished course evaluation document, University of Pretoria.)
- Atkins, B.T. Sue and Varantola, Krista. 1998. Monitoring Dictionary Use. In Atkins B.T. Sue (ed.) 1998: 21-82.
- Barnhart, C. L. 1962. Problems in Editing Commercial Monolingual Dictionaries. In: Householder, F.W. & S. Saporta (eds.) 1962: 161–81.

- Bergenholtz, H. 1990. Lexikographische Instruktionen für ein zweisprachiges Wörterbuch. In: *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 43: 19-37.
- Bergenholtz, H. 1995. Layout. In: Bergenholtz, H. & S. Tarp (eds.) 1995: 224-230.
- Bergenholtz, H. and Tarp, S. (eds.) 1995. *Manual of Specialised Lexicography*. Amsterdam: John Benjamins.
- Bergenholtz, H. and Tarp, S. 2002. Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchsgegenstände verstehen. In: *Lexicographica* 18: 253-263.
- Bergenholtz, H., Tarp, S. and Wiegand, H.E. 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. In: Hoffmann, L. et al. (eds.) *Fachsprachen. Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research*. Berlin, De Gruyter: 1762-1832.
- Bolinger, D. 1985. Defining the indefinable. In: Ilson, R. (ed.) 1985: 69-73.
- Burkhanov, I. 1998. *Lexicography. A dictionary of basic terminology*. Rzeszów: Wydawnictwo.
- Busane, M. 1990. Lexicography in Central Africa: the User Perspective with Special Reference to Zaïre. In: Hartmann, R.R.K. (ed.) 1990: 19-35.
- Cowie, A. (ed.) 1987. *The Dictionary and the Language Learner*. Tübingen: Max Niemeyer.
- Dagut, M. 1981. Semantic Voids as a Problem in the Translation Process. In: *Poetics Today* 2/4: 61-71.
- De Schryver, Gilles-Maurice. 1999. *Bantu Lexicography and the Concept of Simultaneous Feedback, Some preliminary observations on the introduction of a new methodology for the compilation of dictionaries with special reference to a bilingual learner's dictionary Cilubà-Dutch*. Unpublished MA Dissertation, Ghent University.
- De Schryver, G-M. & D. Joffe. 2004. On How Electronic Dictionaries are Really Used. In: Williams, G. & S. Vessier (eds.). 2004. *Proceedings of the Eleventh EURALEX International Congress, EURALEX 2004, Lorient, France, July 6-10, 2004*. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud: 187-196.
- De Schryver, G-M. & D.J. Prinsloo. 2000. The compilation of electronic corpora, with special reference to the African languages. In: *Southern African Linguistics and Applied Language Studies* 18/1-4: 89-106.
- De Schryver, G-M. & D.J. Prinsloo. 2000a. Electronic corpora as a basis for the compilation of African-language dictionaries, Part 1: The macrostructure. In: *South African Journal of African Languages* 20/4: 291-309.
- De Schryver, G-M. & D.J. Prinsloo. 2000b. Electronic corpora as a basis for the compilation of African-language dictionaries, Part 2: The microstructure. In: *South African Journal of African Languages* 20/4: 310-330.
- De Schryver, G-M. & D.J. Prinsloo. 2000c. Dictionary-Making Process with 'Simultaneous Feedback' from the Target Users to the Compilers. In: Heid, U., S. Evert, E. Lehmann & C. Rohrer (eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart: 197-209.
- De Schryver, G-M. & D.J. Prinsloo. 2003. Compiling a lemma-sign list for a specific target user group: The Junior Dictionary as a case in point. In: *Dictionaries*: 24: 28-58.
- Fox, Gwyneth. 1987. The Case for Examples. In: Sinclair, John M. (ed.) 1987: 137-149.
- Francis, W. Nelson & Kucera, Henry. 1964. *Manual of Information to Accompany 'A Standard Sample of Present-Day Edited American English, for Use with Digital Computers'*. Providence: Brown University.

- Gallardo, A. 1980. Dictionaries and the Standardization Process. In: Zgusta, L. (ed.) 1980. *Theory and Method in Lexicography*. Columbia: Hornbeam Press: 59-69.
- Geeraerts, D. 1984. Ontwikkeling in de lexicografische theorievorming. In: *Jaarboek van de Stichting Instituut voor Nederlandse Lexicologie 1981-1982*: 56-65.
- Geeraerts, D. 1996. Entfernung, Fortschritt und Unvollständigkeit: das historische Verhältnis von Lexikologie und Lexikographie. In: Kondering K-P. & A. Lehr (eds.) *Linguistische Theorie und lexicographische Praxis*. Tübingen: Max Niemeyer: 7-15.
- Gouws, R.H. 1986. Afrikaans Lexicography. In: Hartmann, R.R.K. (ed.) 1986. *The History of Lexicography*. Amsterdam: John Benjamins: 75-83.
- Gouws, R.H. 1988. Die gebruik van etikette as leksikografiese hulpmiddel. In: *SA Tydskrif vir Taalkunde Geleentheidsuitgawe 6*: 1-56.
- Gouws, R.H. 1989. *Leksikografie*. Cape Town: Academica.
- Gouws, R.H. 1991. Toward a Lexicon-based Lexicography. In: *Dictionaries 13*: 75-90.
- Gouws, R.H. 1993. Afrikaans learner's dictionaries for a multilingual South Africa. In: *Lexikos 3*: 29-48.
- Gouws, R.H. 1994. Ostensiewe adressering in vertalende woordeboeke. In: *Lexikos 4*: 61-85.
- Gouws, R.H. 1996. Bilingual Dictionaries and Communicative Equivalence for a Multilingual Society. In: *Lexikos 6*: 14-31.
- Gouws, R.H. 1996a. Idioms and Collocations in Bilingual Dictionaries and their Afrikaans Translation Equivalents. In: *Lexicographica 12*: 54-88.
- Gouws, R.H. 1999. *Die maatskaplike gerigtheid van die metaleksikografie in 'n meertalige samelewing*. Inaugural address. University of Stellenbosch.
- Gouws, R.H. 1999a. *A theoretically motivated model for the lexicographic processes of the National Lexicography Units*. Unpublished report submitted to the Pan South African Language Board.
- Gouws, R.H. 2000. Strategies in Equivalent Discrimination. In: Zettersten, A. et al. (eds.) *Symposium on Lexicography IX*. Tübingen: Max Niemeyer: 99-111.
- Gouws, R.H. 2000a. Doeltaalgerigtheid teenoor lemmagerigtheid in vertalende woordeboeke. In: *Tydskrif vir Geesteswetenskappe 40* (1): 39-47.
- Gouws, R.H. 2000b. Toward the Formulation of a Metalexicographic Founded Model For National Lexicography Units in South Africa. In: Wiegand, H.E. (ed.) *Wörterbücher in der Diskussion 4*. Tübingen: Max Niemeyer: 109-133.
- Gouws, R.H. 2001. Teorie word praktyk - en 'n nuwe woordeboek is op pad! In: *Stilet. XIII*: 3: 44-59.
- Gouws, R. H. 2001a. Der Einfluß der neueren Wörterbuchforschung auf einen lexicographischen Gesamtprozeß und den lexicographischen Herstellungsprozeß. In: Lehr, A. et al. (eds.) *Sprache im Alltag*. Berlin: De Gruyter: 521-531.
- Gouws, R.H. 2001b. Lexicographic training: approaches and topics. In: Emejulu J du P (ed.) *Elements de Lexicographie Gabonaise*. Tome I. New York. Jack Hillman Publishers: 58-94.
- Gouws, R.H. 2002. The outer texts in the De Gruyter Wörterbuch Deutsch als Fremdsprache. In: Wiegand, H.E. (ed.) *Perspektiven der pädagogischen Lexikographie des Deutschen II*. Tübingen: Max Niemeyer: 609-619.
- Gouws, R.H. 2002a. Equivalent relations, context and cotext in bilingual dictionaries. In: *Hermes 28*: 195-209.
- Gouws, R.H. 2002b. Using a frame structure to accommodate cultural data. In Emejulu, J. du P. (ed.) *Elements de Lexicographie Gabonaise*. Tome II. New York: Jimacs-Hillman Publishers: 54-69.

- Gouws, R.H. 2003. Aspekte van mikrostrukturele verskeidenheid en inkonsekwentheid in woordeboeke. In: *Lexikos* 13: 92-110.
- Gouws, R.H. 2003a. Types of articles, their structure and different types of lemmata. In: Van Sterkenburg, P.G.J. (ed.) 2003. *A Practical Guide to Lexicography*. Amsterdam: John Benjamins: 34-43.
- Gouws, R.H. 2004. Outer texts in bilingual dictionaries. In: *Lexikos* 14: 67-88.
- Gouws, R.H. 2004a. Monolingual and Bilingual Learners' Dictionaries. In: *Lexikos* 14: 264-274.
- Gouws, R.H. 2004b. Milestones in Metalexigraphy. In: Van Sterkenburg, P.G.J. (ed.) 2004. *Linguistics Today – Facing a Greater Challenge*. Amsterdam; Johan Benjamins: 187- 205.
- Gouws, R.H. (to appear). Sublemmata or main lemmata – a critical look at the presentation of some macrostructural elements. In: *Proceedings: Copenhagen International Symposium on Lexicography X11*.
- Gouws, R.H. et al. (eds.) in preparation. *Dictionaries. An international encyclopedia of lexicography. Supplementary volume: recent developments with special focus on computational lexicography*. Berlin: W. de Gruyter.
- Gouws, R.H. & D.J. Prinsloo. 1997. Lemmatisation of Adjectives in Sepedi. In: *Lexikos* 7:45-57.
- Gouws, R.H. & D.J. Prinsloo. 1998. Cross-referencing as a Lexicographic Device. In: *Lexikos* 8:17-36.
- Gouws, R.H. and D.J. Prinsloo. 2005. Left-expanded Article Structures in Bantu with Special Reference to isiZulu and Sepedi. In: *International Journal of Lexicography* 18: 25 - 46.
- Gouws, R.H. & F.A. Ponielis. 1992. The development of Afrikaans and the lexicographic tradition. In: Zgusta, L. (ed.) *History, Languages and Lexicographers*. Tübingen: Max Niemeyer: 77-104.
- Gouws, R.H. & H.E. Wiegand. (to appear) Strukturtraditionen in der zweisprachigen Lexikographie mit Afrikaans. Ein Beitrag zur Geschichte der Lexikographie in Südafrika. In: Wiegand, H.E. *Studien zur zweisprachigen Lexikographie mit Deutsch*.
- Gove, P.B. 1966. Self-explanatory words. In: *American Speech* 41: 182-198.
- Haas, M.R. 1962. What belongs in the bilingual dictionary? In: Householder, F.W. & S. Saporta (eds.) 1962: 45-50.
- Hartmann, R.R.K. (ed.) 1983. *Lexicography: Principles and Practice*. London: Academic Press.
- Hartmann, R.R.K. (ed.) 1984. *LEX'eter '83 Proceedings*. Tübingen: Max Niemeyer.
- Hartmann, R.R.K. 1989. Sociology of the Dictionary User: Hypothesis and Empirical Studies. In: Hausmann, F.J. et al. (eds.) 1989-1991: 102-111.
- Hartmann, R.R.K. (ed.) 1990. *Lexicography in Africa*, Volume 15, Exeter Linguistic Studies. Exeter: University of Exeter Press.
- Hartmann, R.R.K. 1992. Lexicography, with particular reference to English learners' dictionaries. In: *Language Teaching* 25:151-159.
- Hartmann, R.R.K. 1994. Bilingualised versions of learners' dictionaries. In: *Fremdsprachen lehren und lernen* 23: 206-220.
- Hausmann, F.J. 1977. Einführung in die Benutzung des neufranzösischen Wörterbücher. Tübingen: (Romanistische Arbeitshefte 19).
- Hausmann, F.J. 1986. Grundprobleme des zweisprachigen Wörterbuchs. In: Hyldgaard-Jensen, K. & A. Zettersten (eds.) *Symposium on Lexicography III*. Tübingen: Max Niemeyer: 137-154.
- Hausmann, F.J. 1989. Die gesellschaftlichen Aufgaben der Lexikographie in Geschichte und Gegenwart. In: Hausmann, F.J. et al. (eds.) 1989-1991: 1-19.

- Hausmann, F.J. 1989a. Wörterbuchtypologie. In: Hausmann, F.J. et al. (eds.) 1989-1991: 968-981.
- Hausmann, F.J. 1992. Die zweisprachigen Wörterbücher. In: Hyldgaard-Jensen, K. & A. Zettersten (reds.) 1992: *Symposium on Lexicography V*. Tübingen: Max Niemeyer: 408-411.
- Hausmann, F.J. and Wiegand, H.E. 1989. Component Parts and Structures of General Monolingual Dictionaries: A Survey. In: Hausmann, F.J. et al. (eds.) 1989-1991: 328-360.
- Hausmann, F.J. et al. (eds.) 1989-1991. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography*. Berlin: De Gruyter.
- Householder, F.W. & S. Saporta (eds.), *Problems in Lexicography, Report of the Conference on Lexicography held at Indiana University, November 11-12, 1960*. Bloomington: Indiana University.
- Hulstijn Jan H. and Atkins B.T. Sue 1998. Empirical research on dictionary use in foreign-language learning: survey and discussion. In Atkins B.T. Sue (ed.) 1998: 7-20.
- Hupka, W. 1989. Das enzyklopädische Wörterbuch. In: Hausmann, F.J. et al. (eds.) 1989-1991: 988-999.
- Ilson, R. (ed.) 1985. *Dictionaries, Lexicography and Language Learning*. Oxford: Pergamon Press.
- Jackson, H. 1985. Grammar in the Dictionary. In: Ilson, R. (ed.): 53-59.
- Johnson, S. 1747. *The Plan of a Dictionary of the English Language*. Facsimile edition - 1970. Menston: The Scolar Press.
- Kammerer, M. and Wiegand, H.E. 1998. Über die textuelle Rahmenstruktur von Printwörterbüchern: Präzisierungen und weiterführende Überlegungen. In: *Lexicographica* 14: 224-237.
- Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. London: Longman.
- Kilgarriff, Adam. 1997. Putting Frequencies in the Dictionary. In: *International Journal of Lexicography* 10(2): 135-155.
- Kromann, H-P, T. Riiber, P. Rosbach 1984. 'Active' and 'Passive' bilingual Dictionaries. The Scerba Concept Reconsidered. In: Hartmann, R.R.K. (ed.) 1984: 207-215.
- Kromann, H-P, T. Riiber, P. Rosbach 1984a. Überlegungen zu Grundfragen der zweisprachigen Lexikographie. In: Wiegand, H.E. (ed.) *Studien zur neuhochdeutschen Lexikographie*. Hildesheim: Georg Olms Verlag: 159-238.
- Kromann, H-P. et al. 1991. Principles of Bilingual Lexicography. In: Hausmann, F.J. et al (eds.) 1989-1991:2711-2728.
- Kruyt, J.G. & Dutilh, M.W.F. 1997. A 38 Million Words Dutch Text Corpus and its Users. In: *Lexikos* 7: 229-244.
- Landau, S.I. 1989. *Dictionaries: The art and craft of lexicography*. Cambridge: Cambridge University Press.
- Landau, S. I. 2001 *Dictionaries: The Art and Craft of Lexicography*. Second Edition. Cambridge: Cambridge University Press.
- Laufer, B. 1992. Corpus-based versus Lexicographer Examples in Comprehension and Production of New Words. In: Tommola, H. et al (eds.) 1992. *Euralex '92 Proceedings*. Tampere: University of Tampere: 71-76.
- Lemmens, M. & H. Wekker. 1986. *Grammar in English Learners' Dictionaries*. Tübingen: Max Niemeyer.
- Lombard, F.J. 1990. *'n Metaleksikografiese fundering van Afrikaanse skoolwoordeboeke*. Unpublished doctoral dissertation. University of Stellenbosch.
- Lombard, F. 1991. Die aard en aanbieding van die leksikografiese definisie. In: *Lexikos* 1: 158-180.

- Louw, J.P. (ed.) 1985. *Lexicography and Translation*. Cape Town: Bible Society of South Africa.
- Louw, P.A. & R.H. Gouws. 1996. Lemmatiese en nielemmatiese adressering in Afrikaanse vertalende woordeboeke. In: *S.A. Tydskrif vir Taalkunde* 14(3): 92-100.
- Martin, W. & R.H. Gouws. 2002. A new dictionary model for closely related languages: The Dutch-Afrikaans dictionary project as a case-in-point. In: Heid, U., S. Evert, E. Lehmann & C. Rohrer (eds.). 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart: 783-792.
- MacLeod, Catherine, and Ralph Grishman. 2000. The Influence of Corpora on Lexicons: Corpora Use in the Creation of COMLEX Syntax and NOMLEX. In Heid, Ulrich *et al.*, eds. 2000. *Proceedings of the Ninth EURALEX International Congress, EURALEX 2000*. Stuttgart: Universität Stuttgart: 141-8.
- McArthur, T. 1986. *Worlds of reference*. Cambridge: Cambridge University Press.
- Media24. <http://www.media24.co.za>.
- Nielsen, S. 1995. Alphabetic macrostructure. In: Bergenholtz, H. and Tarp, S. (eds.) 1995: 190-195.
- Nielsen, S. 2003. Changes in Dictionary Subject Matter. In: Wiegand, H.E. (ed.) 2003. *Untersuchungen zur kommerziellen Lexikographie der deutschen Gegenwartssprache I*. Tübingen: Max Niemeyer Verlag:109-114.
- Nielsen, S. 2003a. Towards a general theory of bilingual legal lexicography. In: Kastberg, P. (ed.) 2003. *LSP Translation in the new Millennium*. Aarhus: Aarhus School of Business: 165-188.
- Nkadimeng, H.M., (Ed.) 1985. *Tšwelopele*. Department of Co-operation and Development. Johannesburg: Perskor 26(3): 5.
- Otto, A.N. 1989. *Kriteria vir 'n Afrikaanse aanleerderwoordeboek*. Unpublished doctoral dissertation. University of Stellenbosch.
- Prinsloo, D.J. 1991. Towards computer-assisted word frequency studies in Northern Sotho. In: *SA Journal of African Languages* 11(2): 54-60.
- Prinsloo, D.J. 1994. Lemmatization of verbs in Northern Sotho, In: *SA Journal of African Languages* 14(2): 93-102.
- Prinsloo, D.J. 2004. Revising Matumo's Setswana – English – Setswana Dictionary. In: *Lexikos* 14:158-172.
- Prinsloo, D.J. 2005. Electronic Dictionaries viewed from South Africa. *Hermes, Journal of Linguistics* 34: 11-35.
- Prinsloo, D.J. & G-M. de Schryver. 2000 The Concept of Simultaneous Feedback as Applied to the South African Context, with special reference to the Sepedi Dictionary Project (SeDiPro). 5th International Conference of the African Association for Lexicography, University of Stellenbosch, South Africa: 3-5.
- Prinsloo, D.J. & G-M. de Schryver. 2001. Monitoring the Stability of a Growing Organic Corpus, with special reference to Sepedi and Xitsonga. In: *Dictionaries* 22: 85-129.
- Prinsloo, D.J. and Gilles-Maurice de Schryver. 2002. Designing a Measurement Instrument for the Relative Length of Alphabetical Stretches in Dictionaries, with special reference to Afrikaans and English. In: Braasch, A. & C. Povlsen (eds). *Proceedings of the Tenth EURALEX International Congress, EURALEX*. Copenhagen: Center for Sprogteknologi, Københavns Universitet: 483-494.
- Prinsloo, D.J. and G-M. de Schryver. 2005. Managing eleven parallel corpora and the extraction of data in all official South African languages. In: Daelemans, W., T. du Plessis,

- C. Snyman & L. Teck (eds.). 2005. *Multilingualism and Electronic Language Management*. Pretoria: J.L. van Schaik: 100–122.
- Prinsloo, D.J. & R.H. Gouws. 1996. Formulating a new dictionary convention for the lemmatization of verbs in Northern Sotho. In: *South African Journal of African Languages* 16(3): 100-107.
- Prinsloo, D.J. & R.H. Gouws. 2000. The use of Examples in Polyfunctional Dictionaries. In: *Lexikos* 10. 138-156.
- Scerba, L.V. 1940. Towards a general theory of lexicography. In: *International Journal of Lexicography* 8 (4): 315-350.
- Sinclair, J. (ed.) 1987. *Looking up*. London: Collins.
- Sledd, J. & W.R. Ebbitt (eds.) 1962. *Dictionaries and That dictionary*. Chicago: Scott, Foresman and Co.
- Steyn, M. 2004. The Access Structure in Learner's Dictionaries. In: *Lexikos* 14: 275-298.
- Steyn, M. & L. Gouws 2005. Nuwe woordeboek sonder grense: a typological and communicative bridge. In: *Lexikos* 15:164-178.
- Summers, Della. 1993. Longman/Lancaster English Language Corpus – Criteria and Design. In: *International Journal of Lexicography* 6(3): 181-208.
- Summers, Della. s.d. [1996-1998]. Corpus Lexicography – The Importance of Representativeness in Relation to Frequency. In: *Longman Language Review* 3: 6-9.
- Tarp, S. 1994. Funktionen in Fachwörterbüchern. In: Schaefer, B. & H. Bergenholtz (eds.) *Das Fachwörterbuch: Fachwissen und seine Repräsentation in Wörterbüchern*. Tübingen: Narr: 229-246.
- Tarp, S. 2000. Theoretical challenges to LSP lexicography. In: *Lexikos* 10: 189-208.
- Tarp, S. 2002. Translation dictionaries and bilingual dictionaries – two different concepts. In: *Journal of Translation Studies* 7: 59-84.
- Tarp, S. 2002a. Functions in de Gruyter Wörterbuch Deutsch als Fremdsprache. In: Wiegand, H.E. (ed.) *Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des »de Gruyter Wörterbuch Deutsch als Fremdsprache«*. Tübingen: Max Niemeyer Verlag: 609-619.
- Tarp, S. 2004. Reflections on dictionaries designed to assist the users with text production in a foreign language. In: *Lexikos* 14: 299-325.
- Tarp, S. 2004a. Basic problems of learner's lexicography. In: *Lexikos* 14: 222-252.
- Tarp, S. & Gouws, R.H. 2004. Wie leer wat uit Afrikaanse (aan)leerderwoordeboeke? In: *Tydskrif vir Geesteswetenskappe* 44 (4):276-298.
- Trautmann, C. 1993. *Ekwivalensieverhoudinge in Afrikaanse vertalende woordeboeke*. Unpublished MA-thesis. University of Stellenbosch.
- Van der Colff, A. 1996. *Zur Konzeption eines einsprachigen deutschen Lernerwörterbuchs*. Unpublished doctoral dissertation. University of Stellenbosch.
- Van der Merwe-Fouché, M. 1999. *Gebruikersperspektief in die leksikografie*. Unpublished doctoral dissertation. University of Stellenbosch.
- Van Sterkenburg, P.G.J. (ed.) 2004. *Linguistics Today – Facing a Greater Challenge*. Amsterdam; Johan Benjamins.
- Van Wyk, E.B. 1995. Linguistic Assumptions and Lexicographical Traditions in the African Languages. In: *Lexikos* 5: 82-96.
- Wiegand, H.E. 1983. Überlegungen zu einer Theorie der lexikographischen Sprachbeschreibung. In: Hyldgaard-Jensen, K. & A. Zettersten (eds.) *Symposium zur Lexikographie/Symposium on Lexicography*. Hildesheim: Georg Olms: 35-72.

- Wiegand, H.E. 1983a. Ansätze zu einer allgemeinen Theorie der Lexikographie. In: Schildt, J. & D. Vieweger (eds.) *Die Lexikographie von heute und das Wörterbuch von morgen*. Berlin: Akademie: 92-127.
- Wiegand, H.E. 1983b. Was ist eigentlich ein Lemma? In: Wiegand, H.E. (ed.) *Studien zur neuhochdeutschen Lexikographie*. Hildesheim: Georg Olms Verlag: 401-474.
- Wiegand, H.E. 1984. On the structure and contents of a general theory of lexicography. In: Hartmann, R.R.K. (ed.) 1984: 13-30.
- Wiegand, H.E. 1989. Der gegenwärtige Status der Lexikographie. In: Hausmann, F.J. et al. (eds.) 1989-1991: 246-280.
- Wiegand, H.E. 1989a. Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: alphabetische Anordnungsformen und ihre Probleme. In: Hausmann, F.J. et al. (eds.) 1989-1991: 371-409.
- Wiegand, H.E. 1989b. Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. In: Hausmann, F.J. et al. (eds.) 1989-1991: 462-501.
- Wiegand, H.E. 1989c. Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. In: Hausmann, F.J. et al. (eds.) 1989-1991: 409- 462.
- Wiegand, H.E. 1991. Printed Dictionaries and Their Parts as Texts. An Overview of More Recent Research as an Introduction. In: *Lexicographica* 6: 1-126.
- Wiegand, H.E. 1996. Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher. In: Wiegand, H.E. (ed.) 1996. *Wörterbücher in der Diskussion II*. Tübingen: Max Niemeyer Verlag: 1-82.
- Wiegand, H.E. 1996a. Zur Einführung. In: Wiegand, H.E. (ed.) *Wörterbücher in der Diskussion II. Tübingen*: Max Niemeyer Verlag: VII-XV.
- Wiegand, H.E. 1996b. A theory of lexicographic texts. An overview. In: *SA Journal of Linguistics* 14/4: 134-149.
- Wiegand, H.E. 1996c. Über die Mediostrukturen bei gedruckten Wörterbüchern. In Zettersten, A. & V.H. Pedersen (eds.) 1996. *Symposium on Lexicography VIII*. Tübingen: Max Niemeyer: 11-43.
- Wiegand, H.E. 1997. Die gesellschaftliche Verantwortung der wissenschaftlichen Lexikographie. In: *Hermes* 18, 117-202.
- Wiegand, H.E. 1998. *Wörterbuchforschung*. Berlin: De Gruyter.
- Wiegand, H.E. 2001. Was eigentlich sind Wörterbuchfunktionen? Kritische Anmerkungen zur neueren und neuesten Wörterbuchforschung. In: *Lexicographica* 17: 217-248.
- Wiegand, H.E. 2002. Über textuelle Strukturen der Wörterbuchartikel und Artikelnischen im DE GRUYTER WÖRTERBUCH DEUTSCH ALS FREMDSPRACHE. In Wiegand, H.E. (ed.) 2002b: 413-442.
- Wiegand, H.E. 2002a. Zur Makrostruktur und zu den äußeren Zugriffsstrukturen im DGWDaF. In Wiegand, H.E. (ed.) 2002b: 413-442.
- Wiegand, H.E. (ed.) 2002b. *Perspektiven der pädagogischen Lexikographie des Deutschen II*. Tübingen: Max Niemeyer.
- Wiegand, H.E. 2003. Wörterbuch zur Lexikographie und Wörterbuchforschung (WLWF)/ Dictionary of Lexicography and Dictionary Research. Eine kurze Projektbeschreibung. In: Botha, W.F. (ed.) 2003. *'n Man wat beur*. Buro van die WAT: Stellenbosch: 368-384.
- Wiegand, H.E. 2003a. Überlegungen zur Typologie von Wörterbuchartikeln in Printwörterbüchern. Ein Beitrag zur Theorie der Wörterbuchform. *Lexicographica* 19: 169-313.
- Wolski, W. 1989. Formen der Textverdichtung im allgemeinen einsprachigen Wörterbuch. In: Hausmann, F.J. et al. (eds.) 1989-1991: 956-967.

- Wolski, W. 1989a. Das Lemma und die verschiedenen Lemmatypen. In: Hausmann, F.J. et al (eds.) 1989-1991: 360-371.
- Wolski, W. 1991. Formen der Textverdichtung im zweisprachigen Wörterbuch. In: Hausmann, F.J. et al. (eds.) 1989-1991: 2837-2854.
- Zgusta Zgusta, L. 1971. *Manual of Lexicography*. The Hague: Mouton.
- Zgusta, L. 1987. Translation Equivalence in a Bilingual Dictionary. In: *Dictionaries* 9:1-47.
- Zgusta, L. 1988. *Lexicography today*. Tübingen: Max Niemeyer.
- Zgusta, L. 1989. The role of dictionaries in the genesis and development of the standard. In: Hausmann, F.J. et al. (eds.): 70-79.
- Zgusta, L. 1989a The Influence of Scripts and Morphological Language Types on the Structure of Dictionaries. In: Hausmann, F.J. et al. (eds.): 296-305.

## A

abbreviation 117, 125, 172  
 absolute synonyms 155, 156  
 access alphabet 97, 98, 99, 100, 101, 102, 103  
 access structure 5, 6, 8, 19, 50, 51, 63, 64,  
 117, 157, 165, 166, 167, 168, 169, 170,  
 171, 172, 175, 177  
 addressing structure 8, 63, 64, 134, 135, 138,  
 162  
 article 7, 15, 16, 17, 41, 44, 48, 50, 51, 52, 54,  
 55, 57, 58, 59, 60, 62, 63, 64, 67, 72, 73,  
 76, 81, 84, 85, 88, 89, 90, 91, 92, 93, 94,  
 95, 96, 97, 98, 99, 101, 102, 103, 104,  
 105, 106, 107, 109, 111, 112, 115, 116,  
 117, 118, 119, 125, 126, 127, 128, 129,  
 132, 133, 134, 135, 136, 137, 138, 139,  
 140, 141, 142, 143, 145, 146, 148, 151,  
 152, 153, 154, 156, 157, 158, 160, 162,  
 163, 165, 167, 168, 169, 170, 171, 172,  
 173, 175, 178, 179, 180, 181, 182, 183,  
 184, 185, 187, 189, 191  
 article block 107, 109  
 article slot 16, 88, 89, 117, 141, 151, 172  
 article stretch 57, 62, 89, 90, 98, 99, 101, 103,  
 106, 111, 112, 135, 167, 169

## B

back matter 6, 16, 18, 39, 41, 56, 57, 58, 59,  
 60, 62, 63, 96, 97, 143, 146, 166, 177,  
 180, 187  
 bi-directional 9, 189  
 bias 3, 5, 49, 86, 136, 166  
 bifunctional 54, 161  
 bilingual 3, 5, 7, 13, 14, 16, 34, 42, 48, 52, 53,  
 54, 55, 56, 57, 61, 63, 71, 91, 96, 103,  
 106, 118, 119, 126, 127, 129, 131, 132,  
 134, 136, 138, 139, 151, 152, 153, 154,  
 157, 158, 159, 160, 161, 162, 195, 196,  
 197, 198, 199, 200, 201  
 bilingualised 55  
 bisopal 54, 55, 56

## C

carrier of text 18, 19

central list 4, 16, 18, 51, 56, 57, 58, 59, 60, 62,  
 63, 64, 65, 85, 97, 98, 117, 119, 166,  
 167, 179, 180, 187, 188  
 chronolectic 50, 130, 131, 181  
 chronolectic label 130  
 Circular definition 145  
 cluster 28, 89, 90, 95, 105, 107, 108, 109, 169,  
 184  
 collocates 35, 144  
 comment on form 119, 120, 121, 122, 123,  
 124, 125, 127, 132, 151, 173  
 comment on semantics 51, 52, 92, 93, 108,  
 111, 119, 125, 126, 127, 129, 131, 132,  
 133, 134, 141, 142, 145, 146, 151, 152,  
 153, 154, 157, 161, 173, 175  
 communication-orientated functions 8, 14,  
 132  
 communicative equivalence 55, 131, 136,  
 162  
 complex article 91, 92, 142, 182  
 components 4, 5, 10, 50, 57, 58, 59, 62, 63, 64,  
 74, 99, 106, 119, 123, 144, 165, 177  
 conceptualisation plan 10, 12, 14, 15, 17, 18,  
 19, 58, 96, 97, 100, 103, 118, 129, 165  
 concordance line 150  
 condensation 52, 65, 89, 90, 104, 105, 110,  
 121, 123, 124, 136, 139, 192  
 condensed 121, 122, 123, 124, 132, 135, 136,  
 180, 191  
 congruence 154, 155, 156, 160  
 conjunctiveness 68  
 conjunctivism 78, 79  
 container of knowledge 16, 19  
 context 4, 13, 26, 27, 33, 54, 116, 127, 128,  
 129, 136, 138, 139, 145, 148, 153, 156,  
 162, 163, 181, 190, 197  
 continuum 34, 42, 56  
 contrastive examples 118, 171  
 corpus i, 4, 17, 21, 22, 23, 24, 25, 26, 27, 28,  
 29, 30, 31, 32, 33, 34, 35, 36, 37, 49, 50,  
 51, 52, 53, 65, 71, 74, 79, 84, 111, 113,  
 129, 148, 149, 150  
 corpus line 34  
 cotext 28, 51, 52, 127, 128, 129, 138, 139, 140,  
 141, 153, 156, 157, 162, 175, 197  
 cross-reference 9, 41, 44, 93, 94, 134, 137,  
 145, 177, 178, 179, 180, 181, 184, 185,  
 186, 187, 188, 189, 191, 192

cross-reference address 9, 177, 178, 179, 180,  
184, 185, 186, 187, 188, 189, 191, 192  
cross-reference entry 93, 94, 137, 177, 178,  
180, 184, 185, 191  
cross-reference marker 177, 178, 180, 191  
cross-reference position 177, 178, 180, 181,  
184, 185  
cultural 4, 14, 19, 60, 63, 92, 159, 175, 197

## D

data category 16, 47, 64, 117, 119, 142, 165,  
170, 172, 173, 175  
data density 49, 50, 51, 52  
data distribution i, 5, 8, 14, 18, 48, 49, 50, 54,  
58, 59, 60, 61, 62, 91, 93, 97, 115, 124,  
133, 138, 141, 166, 172  
data distribution structure 5, 18, 48, 50, 58,  
60, 62, 91, 97, 115, 124, 133, 138, 141,  
166, 172  
data types 3, 45, 48, 49, 50, 52, 60, 62, 91,  
115, 116, 118, 119, 124, 125, 133, 142,  
151, 166, 170, 174  
dead references 186  
decode 51  
decontextualisation 9, 181  
default article 63, 91, 127, 133, 141, 172  
definiens 51, 52  
defining vocabulary 143, 192  
definition 3, 33, 48, 49, 50, 92, 116, 126, 127,  
141, 143, 144, 145, 146, 147, 148, 149,  
150, 151, 153, 185, 189  
degrees of comparison 123  
derivation 32, 60, 67, 75, 87, 110, 123  
derived form 186  
descriptive 2, 14, 48, 49, 50, 53, 55, 56, 60,  
61, 65, 96, 97, 102, 103, 106, 110, 146,  
150, 180  
diacritic 97, 98, 102  
dictionary article 7, 17, 54, 55, 60, 64, 85, 89,  
91, 95, 115, 116, 119, 129, 132, 136,  
138, 139, 153, 154, 160, 162, 165, 170,  
171, 172  
dictionary basis 16, 17, 21  
dictionary conceptualisation plan 10, 12, 14,  
15, 17, 18, 19, 58, 96, 97, 100, 103, 118,  
129, 165  
dictionary culture 12, 42, 45, 46, 50, 60, 123,  
175  
dictionary external cross-reference 9, 179,  
180  
dictionary plan 7, 9, 10, 12, 13, 15, 18, 49, 98,  
124, 125, 165, 172

dictionary skills 41, 42, 43  
Dictionary typology 7, 45, 46  
dictionary typology 10, 47, 48, 63, 105  
dictionary use 4, 6, 19, 39, 41, 42, 43, 46, 85,  
91, 117, 138, 142, 170, 175, 177, 199  
disjunctiveness 68  
disjunctivism 78, 79  
divergence 155, 156, 157, 158, 160, 161, 162,  
163  
document types 23  
double addressing 137

## E

electronic dictionaries 8, 80  
encoding 37, 39, 144  
encyclopedic 3, 14, 48, 50, 63, 93, 142, 148  
entrance lemma 105, 107, 109  
entry 4, 18, 48, 56, 62, 64, 68, 80, 88, 93, 94,  
103, 111, 115, 116, 117, 118, 120, 122,  
127, 132, 134, 135, 137, 138, 139, 140,  
141, 157, 159, 160, 165, 172, 177, 178,  
180, 184, 185, 186, 187, 189, 190, 191  
equivalent 8, 33, 52, 55, 56, 63, 80, 115, 116,  
118, 126, 127, 128, 129, 132, 136, 138,  
139, 140, 141, 152, 153, 154, 155, 156,  
157, 158, 159, 160, 161, 162, 163, 188  
etymological data 132  
etymological dictionaries 47  
etymology 3, 49, 132, 133  
explicit 9, 44, 50, 52, 129, 158, 178, 179, 185,  
189, 191, 192  
extended data distribution structure 58  
external cross-reference 9, 179, 180, 181  
extra-linguistic 48, 50, 147, 148, 159

## F

feedback 25, 43, 148  
frame structure 7, 4, 6, 16, 18, 50, 57, 58, 59,  
60, 61, 62, 64, 65, 96, 166, 167, 197  
frequency 13, 25, 26, 29, 30, 31, 32, 50, 52, 60,  
61, 74, 76, 94, 100, 126, 131, 143, 184,  
185, 187, 188, 200  
front matter 7, 16, 57, 62, 98, 101, 117, 120,  
166, 167, 171, 181, 186  
full equivalence 154, 155, 156

## G

gap 159, 161  
general dictionary 16, 47, 59, 61, 62, 86, 129,  
144, 167

general preparation phase 15, 16, 17  
 genuine purpose 5, 13, 14, 59, 60, 61, 134,  
 138, 175  
 genus and differentiae definitions 146  
 glosses 127, 128  
 guiding element 63, 85, 89, 94, 95, 96, 104,  
 107, 115, 119, 135, 151, 153, 165

## H

hapax legomena 21  
 headers 167, 168  
 historical i, 1, 4, 9, 50, 100, 126, 132  
 homonym 100, 101, 109, 187  
 horizontal ordering 90, 91, 95, 104, 105, 106,  
 107, 108, 168, 169  
 hybrid 55, 56, 78, 81, 182  
 hyphen 80, 97, 98, 99, 102  
 hyponymy 181  
 hyponyms 93

## I

idiom 36, 47, 88, 89, 117, 150, 156, 172  
 indicator 98, 117, 118, 120, 171, 172, 174,  
 175, 178  
 inner access structure 6, 64, 117, 165, 170,  
 172, 175  
 inserted inner text 133, 134, 137  
 inserted text 133, 137, 187  
 instruction book 15, 97  
 integrated 5, 50, 58, 59, 61, 138, 139, 140,  
 141, 162  
 integrated microstructure 138, 139, 140, 141,  
 162  
 integrated outer texts 7, 58, 59  
 internal cross-reference 9, 179, 180, 188  
 internet 7, 8, 18, 22  
 intuition 33, 34, 52, 152, 166

## K

keyword 27, 28, 117, 172  
 Knowledge-orientated functions 14  
 knowledge-orientated functions 132  
 knowledge elements 177

## L

label 129, 130  
 language for general purposes 6  
 layout 6, 19, 157, 168, 175  
 learners 32, 41, 43, 44, 76, 92, 122, 125, 198

left-expanded article structures 81  
 lemmalist 30, 31, 32  
 lemmatic addressing 64, 135, 136, 137, 138,  
 153, 162  
 lemmatisation 15, 37, 40, 65, 67, 68, 74, 75,  
 76, 78, 79, 80, 81, 82, 83, 84, 86, 103,  
 184, 186  
 lemma cluster 107, 108  
 lemma sign 29, 33, 49, 51, 52, 55, 59, 60, 63,  
 64, 85, 88, 89, 90, 91, 92, 93, 94, 95, 98,  
 99, 100, 101, 102, 104, 105, 106, 109,  
 111, 116, 117, 118, 119, 120, 124, 126,  
 127, 128, 129, 130, 134, 137, 138, 139,  
 140, 141, 142, 148, 151, 154, 156, 157,  
 158, 165, 167, 168, 169, 178, 179, 180,  
 181, 184, 185, 189  
 lexical divergence 156, 157, 160  
 lexical gaps 158, 159  
 lexical item 47, 49, 51, 52, 63, 86, 87, 88, 90,  
 91, 92, 93, 94, 97, 100, 101, 103, 106,  
 107, 108, 109, 111, 119, 123, 125, 126,  
 133, 139, 142, 151, 153, 156, 157, 158,  
 159, 162, 163, 170, 178, 180, 189  
 lexicographic functions 7, 8, 10, 13, 14, 109,  
 175  
 lexicographic labels 50, 127, 130  
 lexicographic process i, 9, 10, 11, 12, 14, 15,  
 16, 17, 18, 19, 20, 43, 45, 46, 53, 55, 56,  
 58, 63, 107, 138, 165, 168, 175  
 lexicographic road signs 167  
 lexicographic rulers 111  
 LGP 6  
 linguistic dictionaries 13, 48  
 log files 30  
 low density 51  
 LSP 6, 7, 200, 201

## M

main article 96  
 marked form 98, 101  
 material acquisition phase 15, 17  
 material collection i, 16, 17  
 material preparation phase 15, 17, 18  
 material processing phase 15, 18  
 mediostructure 63, 177, 181, 184, 192  
 metalanguage 15  
 metalexigraphy 3, 4, 5, 6, 12, 39, 57, 63  
 micro-architecture 5, 50, 158, 172, 173, 174,  
 175  
 microstructural programme 15, 16, 18, 118,  
 119  
 mini-grammar 4

mono-equivalence 160  
 monodirectional 54  
 monofunctional 54  
 monolingual dictionary 5, 45, 48, 49, 55, 56,  
 65, 106, 110, 116, 126, 127, 138, 139,  
 141, 151, 153, 157  
 monoscopal 54, 55, 56  
 monosemous 126, 155, 156  
 multicultural 53, 63  
 multilexical 86, 88

## N

nested 62, 65, 94, 95, 107, 109, 111, 135, 169  
 neutrality 147, 148  
 neutral form 131, 156  
 niched 62, 65, 94, 95, 107, 108, 109, 135, 169,  
 182, 184  
 non-contrastive examples 171  
 non-integrated microstructure 138, 139, 140,  
 141  
 non-integrated outer texts 5, 50, 58, 59  
 non-lemmatic addressing 135, 136, 138, 153,  
 162  
 non-typographical structural indicators 6,  
 15, 116, 170, 171  
 non-typographical structural markers 117  
 normative 50  
 numerical marker 100

## O

obligatory microstructure 91, 93, 94, 141,  
 142  
 obligatory microstructures 138, 141  
 one to more than one relation 156, 160, 161  
 one to one relation 153, 154, 155, 156  
 optical salience 168  
 ordering 8, 47, 58, 62, 65, 80, 88, 89, 90, 91,  
 95, 96, 97, 98, 99, 100, 101, 102, 103,  
 104, 105, 106, 107, 108, 109, 110, 111,  
 117, 125, 126, 135, 138, 140, 167, 168,  
 169, 170, 173, 175, 181  
 organisation plan 10, 12, 13, 14  
 ostensive definitions 146  
 outer text 57, 59, 60, 61, 62, 93, 166, 167

## P

paraphrase of meaning 8, 48, 49, 51, 52, 55,  
 88, 91, 92, 93, 107, 108, 115, 116, 126,  
 127, 128, 133, 138, 139, 140, 141, 143,  
 151, 152, 153, 157, 172

partial equivalence 154, 155, 156, 157  
 partial lemmata 90, 105, 169  
 part of speech 93, 100, 101, 115, 116, 117,  
 118, 124, 125, 127, 132, 141, 142, 171,  
 174, 178  
 pedagogical 49, 51, 53, 139  
 place keeping symbol 94, 124  
 Polyfunctional 53, 201  
 polyfunctional 7, 54, 55  
 polysemy 126, 128, 153, 157, 158  
 pragmatic 23, 47, 48, 125, 127, 128, 129, 137,  
 181  
 prescriptive 2, 30  
 primary sources 16  
 primitive microstructure 141  
 pronunciation 3, 47, 49, 91, 100, 116, 119,  
 120, 121, 129, 144, 151  
 publication 1, 2, 3, 9, 12, 18, 19, 22, 71, 115,  
 172, 173  
 publishing preparation phase 15, 18, 19  
 punctuation 127, 157

## Q

query 27, 28, 29, 32, 33, 35

## R

rapid access 117, 157, 158, 168, 170, 172  
 rapid retrieval of information 139  
 reference needs 11, 14, 43, 85  
 reference skills 3, 5, 11, 85, 90, 120, 138, 192  
 referential gap 159  
 reflexive component 9, 19  
 register 56, 62, 97, 130, 154, 155, 156  
 remote addressing 137, 140  
 restricted dictionary 47, 132  
 reversibility 56

## S

salient entry 94  
 scanning 22  
 search area 8, 63, 64, 91, 118, 125, 142, 165,  
 170, 172, 175, 8  
 search field 89, 91, 116, 117, 119, 125, 139,  
 142, 171  
 search fields 5, 15, 64, 91, 116, 117, 118, 133,  
 135, 165, 172, 175  
 search route 5, 64, 157, 165, 167, 169, 170,  
 172  
 search zones 58  
 secondary sources 16

selection strategy 71  
 semantic divergence 156, 157, 158, 160, 161, 162  
 semantic gaps 158  
 semantic opaqueness 107  
 semantic relations 41, 47, 50, 65, 108, 133, 142, 181  
 semi-colon 111, 127, 157, 158  
 senses 15, 28, 33, 34, 87, 115, 126, 127, 128, 129, 131, 139, 140, 141, 149, 150, 151, 152, 153, 155, 157, 158, 161, 162, 163, 180  
 simple data distribution structure 58  
 Simultaneous Feedback 43, 196, 200  
 single article 91, 93, 142, 151, 152, 153, 160  
 single articles 62, 63, 91, 93, 94, 141  
 sinuous lemma file 65, 89, 91, 95, 103, 104, 106, 107, 135  
 slang 50, 106, 130, 131  
 source language 16, 48, 52, 53, 55, 56, 126, 131, 136, 139, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 171, 187  
 space limitations 15  
 special field 6, 61, 129  
 Spoken data 21  
 spoken data 22  
 Standard 22, 34, 50, 196  
 standard 2, 13, 14, 23, 34, 37, 48, 49, 50, 51, 53, 54, 55, 56, 61, 69, 70, 71, 130, 131, 132, 139, 190, 203  
 statistical information 28  
 stress 49, 119, 120  
 structural indicator 98, 118, 171, 172, 174, 175  
 style guide 15  
 stylistic label 130  
 subarticle 94, 96, 135  
 subcomment 108, 126, 127, 128, 130, 138, 139, 140, 141, 157, 172  
 subject field labels 130  
 sublemma 64, 89, 96, 108, 109, 135, 136, 169  
 sublemmas 74  
 sublemmatic addressing 135, 136  
 sublexical 86, 87, 99  
 subtypological 49, 52, 53, 54, 55  
 superordinate 93, 145  
 superscript 100, 120  
 surrogate equivalent 159, 160  
 synchronic 50, 51, 100, 126, 132, 177  
 synonym 94, 126, 139, 145, 181, 185  
 synopsis articles 48, 63, 93, 94

## T

taboo 131  
 target user 3, 7, 10, 11, 13, 17, 30, 32, 39, 41, 43, 44, 45, 46, 52, 53, 57, 60, 61, 67, 74, 75, 76, 96, 107, 111, 121, 124, 140, 143, 144, 149, 152, 158, 161, 165, 170, 172, 177, 187, 188, 189, 196  
 terminology 46, 47, 61, 63, 180, 196  
 textual coherence 177  
 textual condensation 52, 65, 89, 90, 104, 105, 110, 121, 123, 124, 136, 139  
 textual elements 115  
 text block 92, 105, 108, 109, 111, 117, 134, 137, 139, 140, 141, 173, 175  
 text elements 157  
 text production 7, 14, 54, 107, 120, 124, 125, 127, 128, 129, 131, 132, 161, 162, 163, 175, 201  
 text reception 7, 14, 54, 86, 107, 120, 124, 127, 128, 132, 154, 161, 162, 163, 175  
 thematic ordering 47  
 theoretical lexicography i, 1, 2, 4, 5, 6, 7, 64, 85  
 thesaurus 47  
 thumb index 167  
 thumb tags 112  
 tilde 44, 52, 105, 123, 169  
 tokens 21, 28, 30, 31  
 tone 49, 120, 121, 190, 191  
 topic switching 136  
 top to bottom relation 175  
 translation equivalent paradigm 33, 115, 152, 153, 156, 157, 160, 161, 162  
 transparency 65, 106, 107  
 treatment unit 51, 60, 63, 64, 119, 135, 136, 142, 151, 162  
 treatment units 13, 18, 45, 63, 86, 87, 88, 89, 106, 117, 135, 136  
 typographical structural indicator 118, 171  
 typographical structural indicators 6, 15, 116, 117, 170, 171, 172  
 typology 7, 8, 10, 45, 46, 47, 63, 94, 96, 105, 48

## U

unexplained lemmata 106, 107  
 unmarked form 98, 99, 101, 102  
 user-friendly 39, 46, 60, 67, 76, 78, 85, 104, 115, 121, 173  
 user-perspective 3, 6, 7, 11, 39, 41, 78, 85, 158

**V**

Vertically ordered 90

vertically ordered 106, 107, 108, 109

**W**

word book 57

word list 26, 57, 177, 187

written communication 21, 29

written data 21

**Z**

zero equivalence 154, 159, 160

# Principles and Practice of South African Lexicography

*Principles and Practice of South African Lexicography* is directed at experts in the field of practical and theoretical lexicography in South Africa, applying the general theory of lexicography to the South African lexicographic environment.

The authors of this book are leaders in the field of South African lexicography and active participants in the international lexicographic arena, publishing regularly in national and international journals and giving papers at international conferences and workshops. Both are NRF-rated researchers, founder members and former presidents of AFRILEX, the African Association for Lexicography, and both have played a major role as PanSALB consultants in the training of the staff members and in guiding the South African National Lexicography Units.

**Rufus Gouws** is Professor and co-ordinator of the Programme for Lexicography in the Department of Afrikaans and Dutch at Stellenbosch University. He has published extensively in the field of theoretical lexicography. For his book *Leksikografie* he received the Linguistic Society of Southern Africa's Award for Linguistics. As practical lexicographer he co-edited a number of Afrikaans dictionaries, including the authoritative HAT (the *Verklarende Handwoordeboek van die Afrikaanse Taal*). He currently is an editor of the comprehensive *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research* and of the international encyclopaedia of lexicography: *Dictionaries. An international encyclopedia of lexicography*. At the University of Stellenbosch he is a recipient of the Rector's Award for Excellence in Research.

**Danie Prinsloo** is Professor in and Head of the Department of African Languages at the University of Pretoria and Chair of Research of the Faculty of Humanities. He is well known nationally and internationally for his ground-breaking research in the field of African language lexicography. Danie is the author and co-author of ninety articles on especially African Language lexicography and a number of Northern Sotho dictionaries. He started the corpus revolution in Africa a decade and a half ago and plays a major role in the development of corpus lexicography and in the design of lexicographic tools such as the Rulers and Block Systems for the balancing of alphabetical categories. He has been awarded thrice a medal for Exceptional Performer of the University of Pretoria.

