

# Chapter 1

## Introduction

### 1. Background to the study

It is a generally accepted fact that most dictionaries for African languages are not products of high lexicographic achievement. Gouws (1990:55) maintains that:

"the majority of dictionaries for African languages are the products of limited efforts not reflecting a high lexicographical achievement ... with a few exceptions these dictionaries offer only restricted translating equivalents and reflect a complete lack of lexicographic planning".

Snyman et al. (1990: preface) highlight another problematic aspect of dictionaries compiled without the use of a corpus namely that the dictionary team is aware that common and even essential words may easily be omitted during the compilation of a dictionary. The reason for this can be because the dictionary compiler had not encountered such words before, and merely hope that there are not too many examples of this kind.

For Setswana in particular, the dictionaries that are currently available are products of limited efforts with a minimum standard of lexicographical achievement. Typical examples are the inconsistencies found on the macrostructural level and inferior treatment of lemmas on the microstructural level. Many of these inconsistencies can be attributed to not making use of written as well as spoken electronic corpora that have become available to the lexicographer during the past two decades. If African

lexicography is to take its rightful place in the new millennium, the active compilation, querying and application of corpora should become an absolute priority, cf. De Schryver and Prinsloo (2000a and 2000b) and Prinsloo and De Schryver (1999 and 2001b).

Modern-day lexicography emphasises the importance of being guided by the user-perspective when compiling dictionaries. However, consideration of the user-perspective does not seem to play a role in most dictionaries of African languages.

A user-oriented dictionary should lead to enhanced information retrieval procedures. (See Prinsloo and Gouws 2000:138). 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. It is believed that 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. According to Prinsloo (1994:93), previously the user-perspective was not seriously considered in the compilation of African language dictionaries in general and Setswana dictionaries in particular. The role of the lexicographer was, to a large extent, seen to prescribe to the users what was presumed to be the 'correct' meaning and use of words. The challenge to the modern Setswana lexicographer is to compile practical, descriptive and user-friendly dictionaries. Good dictionaries can be used as linguistic instruments by their respective user groups. The best dictionaries are those that are the most user-friendly (see Gouws and Prinsloo 2005).

The mission of the lexicographer should be to enhance the quality of dictionaries along the way to the perfect dictionary.

According to Prinsloo and Gouws, (1996:103), the lexicographer is the mediator between linguistics and the dictionary user. Given the complicated morphosyntactic system of African languages, the task of the lexicographer as mediator is even more challenging.

## **1.1 Aims of the study**

As so far no attempt was made to design a comprehensive, modern day approach for Setswana lexicography, this study aims at providing the necessary knowledge, in terms of implementing internationally accepted standards to the specification of African languages. Firstly, it is hence aimed at delivering a broad theoretical description of issues of today's dictionary compilation. Secondly, current African language dictionaries will have to be compared to this standard in order to demonstrate in detail where the African languages lexicography approach still has to be developed. Our general aim is to deliver a guideline for lexicographers on how to compile corpus-based Setswana dictionaries in combining both issues.

The aims of the study can be divided into three categories:

- a broad theoretical survey of core aspects of dictionary compilation.
- a critical evaluation of African language dictionaries with specific focus on Setswana dictionaries.
- an attempt to design a comprehensive lexicographic approach for the compilation of corpus-based Setswana dictionaries.

These aims will now be analysed in more detail:

## **1.2 Broad theoretical survey of core aspects of dictionary compilation**

The aim is to describe the principles and practice that are characteristic of the compilation of modern dictionaries and then to evaluate dictionaries of the major languages of the world such as English, French and German, according to these principles. Issues like the collection of textual data, corpus compilation, access structure, and macrostructural and microstructural aspects will all be examined. Specific emphasis will be placed on the macrostructural aspects such as the

compilation of the lemma list in the central text and the design of front matter and back matter texts, since these are regarded as the most common shortcomings in African language dictionaries (De Schryver and Prinsloo, 2000b and 2000c). On the level of the microstructure, the focus will be, inter alia, on clear and comprehensive sense distinction, and on the selection of appropriate examples. Lastly, we will examine definitions (paraphrases of meaning) of a high lexicographic standard.

### **1.3 Critical evaluation of African language dictionaries with specific focus on Setswana dictionaries**

The principles and practices that are characteristic of the compilation of modern dictionaries and knowledge gained from the evaluation of the dictionaries of major languages of the world, as mentioned above will be applied in a critical evaluation of African language dictionaries with specific focus on Setswana dictionaries. This evaluation will be focused on the macrostructural and microstructural aspects as given in the background study of this chapter with special attention to the lexicographic aspects unique to African languages such as tonal indication and lemmatization strategies. Critical analysis of all currently available Setswana dictionaries will be done with special reference to the dictionaries of Brown (1987),(SESD) Snyman, et al. (1990),(SEAD) Matumo (1993),(MSED) Kgasa, (1976) (THAND)and Kgasa and Tsonope (1995).(THAN)

### **1.4 Designing a comprehensive lexicographic approach for the compilation of Setswana dictionaries**

The research outcomes in 1 and 2 will finally be utilized with focus on the compilation of macrostructures and microstructures of high lexicographic standards for both monolingual and bilingual dictionaries of Setswana.

Conclusions reached from the study of core literature. (Zgusta 1971, Landau, 2001, Gouws and Prinsloo, 2005) and the preliminary analysis of Setswana dictionaries, indicate that the major shortcomings in Setswana dictionaries are:

- lack of proper lexicographic planning
- absence of written and spoken corpora for Setswana and corpus querying for macrostructural and microstructural application
- overemphasis on the comment on form, e.g. grammatical information
- insufficient semantic information
- imbalances in alphabetic categories, i.e. over versus under treatment
- deviation from a normal alphabetic ordering
- inconsistencies in lemmatization, e.g. stem and word lemmatization within the same dictionary
- lack of a selection strategy for lemmas
- absence of user feedback.

The envisaged research aims are therefore extended to the compilation of corpus-based Setswana dictionaries that can be broken down into the following phases:

- build an oral corpus for Setswana.
- enlarge the current Pretoria Setswana text corpus.
- design lexicographic tools such as a ruler and block system for the oral as well as for the extended written corpus.

## **1.5 Research methodology**

The study will:

- present a comprehensive theoretical conspectus of the creation of electronic corpora followed by a practical exploration for the African languages.

- focus on various applications of lexicographic principles and practices in the broad field of lexicography.
- use informants for the compilation of oral corpora, such as mother-tongue speakers of Setswana, as sources of spoken data, which will be processed to form an oral corpus.

## 1.6 Scope and limitation of the study

The study consists of six chapters.

**Chapter 2** will focus on the origin and historical development of Setswana as a language. It will further highlight the missionaries's contribution and the role played by various stakeholders such as the Setswana Language Board towards the development of Setswana. The chapter also indicates how other languages or dialects have affected or influenced the orthography of Setswana and addresses the challenges the Language Board faced in the standardization of Setswana. Finally, the chapter illustrates how an oral Setswana corpus was compiled and how language changes led to new terminology being developed and eventually enriching the lexicon of the Setswana corpus.

**Chapter 3** first presents a comprehensive theoretical conspectus of the creation of electronic corpora followed by a practical exploration of the African languages. It also focuses on various applications of lexicographic principles and practices in the broad field of lexicography and outlines the different steps needed to compile a structured corpus. This corpus should in due course, take an organic form containing texts from the genres and domains used including those that are spoken and written in private as well as in public. Furthermore the chapter highlights how corpora are mainly queried to obtain or generate word lists reflecting overall and comparative counts or concordance lines giving keywords in context. Prinsloo (2004:42) regards the keyness function in WordSmith Tools, for example, as ideal for selecting the so-called keywords by comparing, for instance, a dedicated Setswana corpus with a general

corpus. Illustrative contrasts between the South African oral and text corpora and the Botswana text corpus will be demonstrated using the frequency word lists. Finally, studies will be conducted to monitor the stability of the growing organic Setswana corpora.

**Chapter 4** will highlight gaps and inconsistencies in Setswana dictionaries on the macrostructural level. Specific emphasis will be placed on macrostructural aspects such as lemma lists in the central texts and frequency indication strategies in the dictionary, given that these aspects form the most common shortcomings in African language dictionaries, particularly in Setswana dictionaries. Frequency lists, culled from corpora, will be used to evaluate the lemma lists on the macrostructural level as well as for frequency indication.

**Chapter 5** will focus on the importance of how the utilisation of a corpus can enhance the quality of the microstructure of the Setswana dictionaries' articles. Four major issues around the importance of corpora will be covered. Firstly, corpora as an aid to sense distinctions. Secondly, corpora as an aid to retrieve typical collocations. Thirdly, corpora as an aid to pinpoint frequent clusters. Finally, as an aid to select typical and natural examples. The chapter will also highlight problems related to the treatment of verbs, polysemous multiple meaning, synonyms and the treatment of the Setswana months in the currently available Setswana dictionaries. Thus, each section will conclude with suggestions for the improvement of the respective Setswana dictionaries by means of a corpus-based microstructure.