A PRACTICAL APPROACH TO THE STANDARDISATION AND ELABORATION OF ZULU AS A TECHNICAL LANGUAGE

by

LINDA VAN HUYSSTEEN

submitted in accordance with the requirements

for the degree of

DOCTOR OF LITERATURE AND PHILOSOPHY

in the subject

LINGUISTICS

at the

UNIVERSITY OF SOUTH AFRICA

PROMOTER: PROF L. A. BARNES

JOINT PROMOTER: PROF. C. T. MSIMANG

November 2003

TABLE OF CONTENTS

CHAPTER 2
LANGUAGE STANDARDISATION AND ELABORATION OF THE AFRICAN LANGUAGES

CHAPTER 3
ORTHOGRAPHICAL TERMINOLOGICAL STANDARDISATION PROBLEMS REGARDING ZULU

CHAPTER 4

THE METHODS OF WORD-FORMATION THAT FACILITATE LANGUAGE AND TECHNICAL ELABORATION IN ZULU

CHAPTER 5
THE VALUE OF WRITTEN ZULU CORPORA IN SEMI-
AUTOMATIC TERM EXTRACTION FOR STANDARDISATION
PURPOSES

CHAPTER 6
THE VALUE OF ORAL CORPUS ANNOTATION FOR IMPROVING THE ACCEPTABILITY OF TECHNICAL TERMINOLOGY IN ZULU

CHAPTER 7
CONCLUSION

TABLES

BIBLIOGRAPHY   240

ADDENDA

DECLARATION

I declare that **A PRACTICAL APPROACH TO THE STANDARDISATION AND ELABORATION OF ZULU AS A TECHNICAL LANGUAGE** is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

....

L. VAN HUYSSTEEN

ACKNOWLEDGMENTS

I should like to extend my sincerest gratitude to the following people who contributed in whichever way to the completion of my thesis:

S   My promoter, Prof. L. A. Barnes for his encouragement and insightful guidance throughout the course of my study, despite his own tight schedule. He allowed me to explore several options and work independently in the Zulu subject domain. Yet, he also inspired me to formulate explicitly and seek linguistically sound solutions in theory and practice. Even during difficult times he remained calm, collected and patient.

S   My co-promoter, Prof. C. T. Msimang (Thabizolo), who assisted me with the correct interpretation and linguistic application of Zulu texts and terminology and who also pointed out specific cultural considerations inherent in the elaboration process. His firsthand experience and knowledge of the standardisation process in Zulu proved invaluable.

S   Prof. A. D. de V. Cluver, the 'father' of this study.

S   The medical personnel who facilitated the conduct of interviews and distribution of questionnaires during my hospital visits:
Matrons Landers, Mdima and Sosibu and Dr Vos at King George V Jubilee Hospital in Durban,
Matron Mageba at McCord Zulu Hospital in Durban,
Matrons Radebe and Sosibu at Prince Mshiyeni Hospital in Umlazi,
Matrons Bachus and Potgieter at Newcastle Provincial Hospital in Newcastle,
Dr Valentine, Matron Eckersley and Sister Ngqolungu at Vryheid Hospital in Vryheid, and
Matron Hattingh at Amajuba Memorial Hospital in Volksrust.

iii

- Zulu informants among hospital staff for their contribution by way of intuitive oral responses.
- Mrs R. Koekemoer of the Department of Arts and Culture whom I consulted and who made available to me official medical terminology lists.

- Dr M. Alberts of PanSALB who kept me informed about the language planning scenario in South Africa.

- Prof. D.J. Prinsloo and Mr G. de Schryver who introduced me to *WordSmith* Tools.

- My father-in-law, Prof. J. A. van Huyssteen who willingly and expertly proofread and edited the manuscript.

- My colleague and friend, Dr N. Mollema who helped me with the final electronic layout and format of the thesis.

- My colleague and friend, Prof. S. E. Bosch who always encouraged me.

- My mother, Mrs M. E. Mulder, who helped me on a daily basis with household chores and who never failed to inspire me to keep faith even in times of distress.

I should also like to thank the Research Fund of the University of South Africa for providing a financial grant in order to enable me to conduct research in Kwa-Zulu Natal and Mpumalanga.

Last, but by no means least, I wish to thank my husband Evert for his unfailing moral support despite his own busy study and work programme. I dedicate this thesis to him and our two children, Johan and Ingrid, who often had to suffer lack of attention and household chaos as a result of my study. However, their presence and support never failed to uphold me.

SUMMARY

The lack of terminology in Zulu can be overcome if it is developed to meet international scientific and technical demands. This lack of terminology can be traced back to the absence of proper language policy implementation with regard to the African languages. Even though Zulu possesses the basic elements that are necessary for its development, such as orthographical standards, dictionaries, grammars and published literature, a number of problems exist within the technical elaboration and standardisation processes:

- Inconsistencies in the application of standard rules, in relation to both orthography and terminology.
- The lack of standardisation of the (technical) word-formation patterns in Zulu. (Generally the role of culture in elaboration has largely been overlooked).
- The avoidance of exploiting written technical text corpora as a resource for terminology. (Text encoding by means of corpus query tools in term extraction has just begun in Zulu and needs to be properly exemplified).
- The avoidance of introducing oral technical corpora as a resource for improving the acceptability of technical terminology by, for instance, designing a type of reusable corpus annotation.

This study contributes towards solving these problems by offering a practical approach within the context of the real written, standard and oral Zulu language, mainly within the medical terminological domain. This approach offers a reusable methodological foundation with proper language exemplification that can guide terminologists in terminological research, or to some extent even train them, to achieve effective technical elaboration and eventual standardisation.

This thesis aims at attaining consistent standardisation on the orthographical level in order to ease the elaboration task of the terminologist. It also aims at standardising the methods of word- (term-)

formation linking them to cultural factors, such as taboo. However, this thesis also emphasises the significance of using written and oral technical corpora as terminology resource. This, for instance, is made possible through the application of corpus linguistics, in semi-automatic term extraction from a written technical corpus to aid lemmatisation (listing entries) and in corpus annotation to improve the acceptability of terminology, based on the comparison of standard terms with oral terms.

ABBREVIATIONS

BUV   Black Urban Vernacular

CZD   Central Zululand Dialect

DAC   The Department of Arts and Culture

DACST   The Department of Arts, Culture, Science and Technology (has now divided into two Departments, the Department of Arts and Culture (DAC) and the Department of Science and Technology (DST)

DACST 1997(a) *Draft List*: *Basic Health Terms* compiled by the *National Terminology Services* (*NTS*), specifically by the *Divis*

ion of Biological and Agricultural Sciences of t

he *Department of Arts', Culture', Science and Tech*

n
o
l
o
g
y

DACST 1997(b) *Draft List*: *Sex Education* compiled by the *National Terminology Services* (*NTS*), specifically by the *Division of Biological and Agricultural Sciences* of the *Department of Arts, Culture, Science and Technology*

DBE   Department of Bantu Education (in the previous dispensation)

DET   The Department of Education and Training (in the previous dispensation)

DET 1993   Department of Education and Training: *IsiZulu terminology and orthography No*. 4 (latest)

DOE   The Department of Education (since the 1994-democracy)

ISO   International Organisation for Standardisation

KWIC   Keyword in context. This is another name for the *concordance* tool in WST

LANGTAG   Final Report of the Language Plan Task Group - Towards a National Language Plan for South Africa - 1996

LIEIP   Language-in-Education Implementation Plan - 1997, made available by the Department of Education to see to the implementation of the LIEP below

LIEP   Language in Education Policy - 1997, made available by the Department of Education

LNC   Language of Narrower Communication

LWC   Language of Wider Communication

MUWA   Multiterm Website Access

NEPI   National Education Policy Investigation - Language Policies for Medium of Instruction: An Information Document for discussion by Parents, Teachers and Principles - 1992

NLB(s)	National Language Body/Bodies
NLP	National Language Project - Negotiations and Language Policy Options in South Africa - 1993
NLPF	National Language Policy Framework - finalised in 2002 by the Department of Arts and Culture
NLS	National Language Service - established in 1998, earlier known as **NTS** (see below)
NLU(s)	National Lexicography Unit(s)
NTS	National Terminology Services now known as **NLS** (see above)
PanSALB	Pan South African Language Board **-** established in 1995
PLC(s)	Provincial Language Council(s)
STANON	Standard and Non-Standard African Language Varieties in the Urban areas of South Africa - Main Report for the STANON Research Programme 1996
TCS	Terminology Coordination Section
ZLC/B	Zulu / IsiZulu Language Committee / Board.
ZNLB	IsiZulu National Language Body
WST	*WordSmith Tools* - Computerised query tools for text analysis

KEY TERMS

corpus planning, standardisation, Zulu language elaboration, technical language, orthographical standard, word-formation, culture related aspects, corpus linguistics, frequency count, actual frequency, concordance, written corpus, oral corpus, text encoding, oral corpus annotation.

CHAPTER 1

INTRODUCTION

1.1   General introduction to the study

In South Africa the complexity of corpus planning practice, particularly in the African languages including Zulu, has been underestimated to a great extent. To date it has been easier for researchers to come up with theoretical rather than practical solutions and workable exemplified methods. In the underdeveloped African languages, suffering from the lack of terminology, solutions are not easy to find. Studies in the African languages generally focus on the shortcomings of the previous Apartheid-driven Language Boards, the recognition of non-standard varieties, phonological adaptation and natural term development within the speech community. Although a contribution, these studies did not really aim at finding methods that could apply to the real language situation in order to solve the problem of underdeveloped terminology.

The focus of this study falls on a practical approach to the standardisation and elaboration of Zulu as a technical language, specifically in the medical field. Practical solutions to standardisation and elaboration problems are offered through proper methodology and exemplification, the point of departure being the real written and spoken Zulu language in the professional domain. Orthographical problems, for instance, can only be identified when the application of the standard orthography is investigated in written standard and technical texts. Yet, some of the solutions may also be exemplified in these very same texts. In addition, this thesis aims at providing a practical methodological approach as basis on which future research on technical elaboration and standardisation in Zulu can be built. This approach basically treats the standardisation and elaboration of Zulu within the real technical language

domain considering the orthography, the methods of word-formation and the utilisation of relevant written sources to extract terminology by means of corpus query tools, and the utilisation of relevant oral sources for term comparison in order to arrive at corpus annotation (analysis). The latter two methods specifically relate to the application of (computational) corpus linguistics to the Zulu language.

The purpose of this chapter is to introduce problems and issues which are related to the research conducted for this thesis, i.e. the background to the research problem, statement of the research problem, aims of the research, nature of research, research methodology, an exposition of chapters and the scope of this study.

1.2   Background to the research problem

Although Language Boards, presently called *National Language Bodies* (NLBs)*,* have been set up to guide the standardisation process in coordination with other terminology structures, the standardisation and elaboration processes have not been very successful in the African languages. The reason for this is that in the African languages terminologists who were assigned to develop terminology for use in schools did not make use of the real language as point of departure. There was also not much consultation with interested parties. It is a fact that most African languages, including Zulu, have not progressed beyond the point of very basic terminology even though they possess the tools that are necessary for their development such as grammars, dictionaries and other publications. This lack of terminology has serious implications for the Zulu language which is not being optimally used in education and the technological field. Consequently, it is regarded as having a low status.

The lack of terminological development in the African languages can be attributed mainly to the lack of language and educational policy implementation, but also to the lack of coordination in the national language standardisation and elaboration processes.

1.2.1 Language and educational policy implementation

At this stage it is appropriate to refer to the South African Language Policy which is not a separate official document but forms part of the South African Constitution contained in *Act 108 of 1996*. From the language-related sections in the Constitution, such as 'official languages', 'equality in language', 'language in education', 'language and culture', it is clear that the language rights of the individual, the elevation of the status and development of the previously disadvantaged (African) languages, and cultural equality within the framework of multilingualism, are entrenched.

General language policy, which mainly has to do with the choice of (a) national/official language(s) is closely related to the educational language policy of a state. The educational language policy of South Africa is referred to in the Constitution (*Act 108 of 1996, Section 29*) but is fully discussed in a separate official document by the *Department of Education* (DOE) and titled *Language-in-Education Policy* (LIEP), dated 1997.

Yet, however noble the principles or aims of the language (educational) policy are, they would remain mere theory and not norm if not practically applied in society. Cluver (1994) warns that this multilingual language policy is an idealistic model which is difficult to implement. According to him the positive challenge of multilingualism lies in the equal development of all the South African languages. The African languages were historically disadvantaged by the previous South
African Nationalist government and did not enjoy full official status until 1994. Ironically, however, the recognition of 11 (9 African) official languages today is a direct result of the implementation of Apartheid policies to facilitate the idea of separate homelands, in each of which a different indigenous African language became the official language alongside English (Kaschula & Anthonissen 1995).

In South Africa the African languages are officially still used in many ' Black' schools as language of

instruction for at least the first four years. However, in practice the African languages are presently being used to an ever lesser extent in such schools, since parents strive for their children to be educated in English. Yet, the generation of knowledge, be it scientific and technological, comes through education in understandable conceptualised forms through the medium of the home languages (*Language-in-Education Implementation Plan* (LIEIP) in DOE1997).

Not only African politicians have a low regard for their national languages but also African parents and pupils. This is largely due to previous colonial educational policies. Recently many African parents have become increasingly reluctant to teach their children any African language. This trend is also becoming prominent in the new South Africa where African parents prefer to communicate with their children in English. De Klerk (1995a) points out that many African language-speaking and Afrikaans-speaking pupils struggle with English before even having mastered their own language. Furthermore, English is seen as the language of empowerment, especially in the job market.

The only way in which the African languages can develop, is at the grass roots level through education. Therefore the use of primary languages as *Language/s of Learning and Teaching* (LOLT) at all levels of schooling is to be encouraged. The LIEIP (DOE 1997) states that a way in which to redress the underdevelopment of the African languages, is to stop favouring the previous colonial languages. This means that speakers of these languages should start valuing their own languages as languages of empowerment.

The LIEIP (DOE 1997) further states that language development of the national languages can only be achieved through the integration of own language learning into all learning processes, irrespective of the subject taught. To sum up, for Zulu to elaborate to its capacity, a proper modern discourse dealing with technical and scientific topics needs to develop in the language. This modernisation should be initiated at the primary educational level before it can be attained at the secondary, or even the tertiary level.

1.2.2 National language standardisation and elaboration processes

The lack of terminological development in the African languages can be attributed not only to the lack of language and educational policy implementation but also to the lack of coordination in language standardisation and elaboration processes. The latter processes were investigated and found to be lacking coordination between the different language interest structures, such as the *National Language Bodies* (NLBs), the *Provincial Language Councils* (PLCs) and the *National Language Service* (NLS), a situation that should be rectified by the national language authority, the *Pan South African Language Board* (PanSALB).   If one were to standardise terms in the African languages, or in Zulu for that matter, one would not know how to go about it and what channels to follow. Neither would one know which linguistic elaboration methods to utilise.

In South Africa this uncertain situation can partly be attributed to the fact that many changes took place since 1994 in the naming and functions of official language structures. For instance, the name of the *National Terminology Services* (NTS) changed to the the *National Language Service* (NLS). Where there was previously only one language authority (Board) for each African language, e.g. the *(Isi)Zulu Language Board*, there are now basically two, the NLBs which function on the national level , e.g. the *IsiZulu National Language Body* (ZNLB), and the PLCs which function on the provincial level.

The lack of terminology in the African languages, also Zulu, can be overcome only if standardisation and elaboration activities are coordinated through effective language planning via the mentioned official language agents and structures. Consultation and coordination in the term-creating activities among the different African languages should also be promoted.

Nevertheless, language policy cannot be fully implemented by the coordination of language interested structures or even the government if the speakers of the African languages themselves do not do their

part. Modernisation, by implication standardisation and elaboration in the technical domain, will become reality only if the African languages are maintained and their status thus improved by their speakers - before it is too late!

1.3   Statement of the research problem

Although Language Boards, presently called *National Language Bodies* (NLBs), have been set up to develop terminology in coordination with terminology structures, such as the NLS, the elaboration of terms has not been sufficient in the African languages to meet the global needs of technological development.

The African languages possess the basic tools that are necessary for their development such as orthographical standards, terminology lists, dictionaries, grammars and publications. Yet, there are some serious deficiencies in these tools that get in the way of effective technical elaboration and standardisation. It is thus sad but true that most African languages, including Zulu, have as resource only the most basic terminology which is hardly sufficient to cope with national educational demands, let alone international technological and scientific demands. In order to overcome such deficiencies, they need to be approached from a practical perspective.

Four main factors (problem areas) underlying the poor implementation of language policy and inefficient coordination of corpus planning initiatives, have been identified:

i)   There are serious inconsistencies in the formulation and interpretation of orthographical rules in the Zulu language in general and in terminology. Problems in application can be attributed to the lack of proper exemplification in the formulation of these rules. Particular problem areas such as word division, capitalisation and changing trends in the language and terminology which are not reflected in the orthography, need to be ironed out. Furthermore, proper dissemination

of information did not take place. It is a fact that the orthographical rulings of the previous ZLC/B (which are still valid today) did not reach all the relevant educational or public sectors. In KwaZulu-Natal where research was conducted medical workers were mostly unaware of the existence of the previous ZLC/B, let alone its rulings.

ii) For a long time African language development focussed on the establishment of an orthography, Christian terminology and basic school terminology of a Western register (LANGTAG in DACST 1996). In the Sixties African terminologists developed elementary terminologies in the African languages to enhance the use of these languages as medium of instruction in the initial school years in accordance with Apartheid policies. However, these terminologists had to rely on their own intuition since no documentation of the methods of word-formation existed then.

Even today indigenous word-formation methods unique to the African languages, such as derivational patterns, rules for compounding or adoption of loans, are not clearly reflected in grammars or in more advanced analytical works. In an advanced linguistic analysis of Zulu by Poulos and Msimang (1998), for instance, word-formation is not treated very extensively. Thus LANGTAG (DACST 1996) suggests that a planned research programme needs to be introduced that will expose the underlying patterns of word-formation in the African languages to prevent these languages from losing their derivational transparency.

Alberts (1997) basically agrees that word-formation principles for all African languages need to be established since a knowledge of these principles is essential for the training of terminologists. However, what the terminologist also has to bear in mind, is the way in which such elaboration methods are linked to Zulu culture; an aspect which has also been largely overlooked.

iii)   In the STANON-Report, Calteaux (1996) puts standardisation in the African languages thus far in a realistic sociolinguistic perspective, by asking where the sources of such standards are and whether terminological work on 'standards' done by the now disbanded Language Boards have been looked at critically enough. When these official standard terminological sources (mainly terminology lists) prove inadequate, research data in the form of other written texts should be incorporated. For instance, the latest existing texts available in a specific technical field such as medical pamphlets should be gathered. As far as could be established, this has not been done to any great extent. Thus far the value of relevant written sources in terminological elaboration and standardisation has been neglected, as far as term extraction, for instance, is concerned. The extraction of terms aided by the application of computational linguistics has just begun but needs to be adjusted and exploited further to suit lemmatisation (the listing of basic forms) in the African languages, especially the complex morphology of Zulu. Extracting technical terms from written texts can determine which 'new' terms have been coined by the Zulu professional community.

iv)   In the STANON-Report, Calteaux (1996) asks whether the evidence (written official terms) of the national standard is oral-aural based.   Not only research data in the form of written technical text sources should be incorporated, but also oral sources. Thus far the value of oral sources in terminological elaboration and standardisation has been neglected. There is a lack of verification to determine to what extent existing standard technical terms have been accepted and used by the Zulu-speaking community in actual oral use as evidenced by everyday communication in the work domain.

1.4   The aims of the research

The general aim of this study is to provide a practical methodological approach with proper exemplification to the standardisation and elaboration of Zulu as a technical language.   Indirectly, its

aim is also to determine how recognised, used and understood the existing standard terminology is within the professional (medical) community it serves. The ultimate aim is to contribute towards a foundation which will make the elaboration and standardisation process in South Africa, with particular emphasis on Zulu, a more acceptable and coordinated process.

This study has two main aims, one theoretical and one practical. The theoretical aim is to critically examine standardisation rules and elaboration practices in Zulu in order to provide a reference framework in terminological work. The practical aim is to constitute a methodological kernel illustrating the utilisation of the real sources of the language in term creation.

The theoretical and practical objectives can specifically be summed up in two aims:

1.1   The first aim is towards proper standardisation in relation to the applicability of the Zulu orthography and elaboration methods, i.e. word-formation.

1.2   The second aim of a methodological nature, is to illustrate how the written and oral sources of the Zulu language in a specific technical (medical) field can be utilised for terminological work. This second aim is closely linked to the application of corpus linguistics, especially the development of corpus linguistic methodology in Zulu.

In order to achieve these two aims for this thesis, they can be divided into five specific research goals to effect the following:

- Identify the inconsistencies in the application of the Zulu orthography, also in relation to terminology and propose solutions through proper formulation and exemplification of rules.
- Document, with the intention to standardise, the existing word-formation methods that facilitate Zulu language and technical elaboration.
- Illustrate how the written sources of the Zulu language in a specific technical (medical) field can

be utilised as resource for elaboration. Terms are, for instance, extracted semi-automatically by means of corpus query tools.

- Illustrate how the oral sources of the Zulu language in a specific technical (medical) field can be utilised to improve the acceptability of (standard) terms.

- The fifth implicit goal is to contribute towards the training of terminologists in standardisation procedures, for instance, in critical linguistic analysis, and in the gaining of insight into elaboration methods as well as in the application of corpus linguistics.

1.5 The nature of research in this study

From the onset it must be made clear that this thesis does not contain a specific hypothesis, but rather a practical methodology obtained from exploratory research. As far as could be established no similar practical methodological research has been undertaken with regard to Zulu thus far. This study focusses on a practical approach, providing methodology with proper exemplification towards more acceptable standardisation and elaboration of Zulu as a technical language.

All research methodology basically falls into either qualitative or quantitative research. Leedy (1993:139,143) makes it easy to distinguish between the two: "If the data is verbal it is *qualitative,* if it is numerical it is *quantitative.*" Qualitative studies tend to be investigative and descriptive in nature, while quantitative studies usually construct hypotheses which have to be tested against hard facts.

According to the views of Bogdan and Biklen (1992:29-31, 51-52) the research conducted for this thesis is essentially of a qualitative nature. In qualitative research the natural setting is the direct source of data and the researcher's insight plays a major role in the undertaken research.

This research is descriptive, often practical, and it is about the process and not about outcomes. Furthermore, the data is analysed inductively and a report often follows after the data has been

collected.   In this thesis, data is not used to prove or disprove a specific hypothesis. For this study it is also true that: "The study itself structures the research, not preconceived ideas or any precise research design" (Bogdan & Biklen 1992:58).

However, even when conducting qualitative research, the value of quantitative research in the form of numbers may not be disregarded (Bogdan & Biklen 1992:147). It has become quite common to blend these two research methods to what has become known as *triagulation* (Leedy 1993:147). Even in this study both these methods were fruitfully combined, although the quantitative aspect was dealt with to a lesser extent.

To explain the specific nature of this research fully, reference should be made to a *practical approach*. The practical approach does not claim to be absolute or prescriptive but it will at least serve as an example for those concerned with elaboration and standardisation and pave the way forward. Sager (1990) actually also implies that guidelines for language development should be laid down since, unlike international standardisation bodies, national standardisation bodies rarely lay down standards and guidelines for the naming, compilation, selection and publication of terminology. The aim of the practical approach is to guide language planners and/or terminologists, or to some extent even train them, to follow certain guidelines or methods towards effective standardisation and elaboration. Above all, this approach explores practical ways of implementing technical elaboration by providing a reference framework on the orthography and word-formation while also enhancing natural term development.

In this study the focus is on medical terminology in Zulu. The study attempts to apply a practical approach to the issue of term creation.

1.6   Research methodology

The following research methodology, based on the theory of qualitative and quantitative methods, was designed for this thesis:

- Firstly, the formulation and exemplification of earlier and present standard orthographical rules were critically examined in official documents of the previous ZLC/B.   Also standard terminology lists of the previous ZLC/B and NTS and present technical (medical) texts were scrutinised in order to determine how these rules were interpreted and applied. In this manner inconsistencies in the formulation, application and exemplification could be identified and solutions proposed.

- Secondly, standard terminology lists of the previous ZLC/B and NTS, present technical (medical) texts and Zulu grammars were scrutinised in order to gain insight into the existing general language and technical elaboration processes in Zulu. An overview of the methods of word-formation was documented with the aim of standardising them as a resource for elaboration. Cultural factors that play a role in elaboration were also identified.

- Thirdly, corpus linguistic theory such as corpus design, text collection and encoding were applied to written Zulu texts.   Available, relevant written technical (medical) sources were utilised as terminology resource. For instance, a method for basic term extraction in Zulu by means of computerised tools to aid lemmatisation was followed and fully exemplified. Term extraction for both simple and complex (mutli-) terms, considering both linguistic and technical factors, were exemplified.

- Fourthly, corpus linguistic theories were also applied to oral technical (medical) sources so that they could be utilised as terminology resource. Oral (medical) terms were obtained by means

of Zulu health-term questionnaires distributed to medical health workers in hospitals which mainly serve the Zulu speech community such as in KwaZulu-Natal. See Addendum 1 for detailed research procedures and the questionnaire. An intuitive natural response on terminology was obtained for research purposes, such as term comparison, by interviewing mother-tongue speakers (medical staff), in their practical work domain. This comparison then led to corpus annotation (analysis) aimed at improving the acceptability of terms, i.e. giving an indication of what terms are actually used and accepted in a specific technical (medical) field. The basis of this annotation was the comparison of standard written and oral medical terms. See Addendum 2 for oral corpus annotation exemplification and statistics.

From the description of the methodology above it should be clear that both qualitative and quantitative research methods were used. For the first two research themes the qualitative method was mostly used since insights were mainly deduced from standard official documents and terminology lists, grammars and publications in order to come up with the identification of standardisation problems and suggested solutions. For the last two research themes the quantitative method was used to a large extent since methodology was exemplified along numerical lines such as frequency counts of written and oral terms. For the third research theme computerised corpus query tools were used to illustrate a method for term extraction. For the fourth research theme corpus annotation was based on the frequency count of term comparison. Yet, the naming of corpus annotation tendencies, such as 'indigenous coinage', based on the researcher's observation, was arbitrary and can thus be considered as belonging to qualitative research. Another reason why the fourth research theme can also be considered as qualitative, is the fact that the natural setting (hospitals and clinics) was the direct source of data and that the researcher's insight played a major part in the research (the utilisation of the questionnaire for interviews to obtain feedback on medical terms).

1.7   Exposition of chapters

Following the introductory **Chapter 1**, **Chapter 2** offers an overview of the literature concerning standardisation and elaboration with the focus on national language planning, specifically the corpus planning perspective. Standardisation is discussed with reference to different aspects such as definition, models, norm, purpose, stages, agents, limitations and problems. These aspects are applied to the context of the African languages, in particular Zulu.

Chapter 3 focuses on the inconsistency in the formulation, application and exemplification of orthographical (terminological) rules in Zulu.   It also suggests ways of improving the formulation of orthographical rules. It captures adjustment in the orthography to simplify its application to the latest trends in terminology development in Zulu.

Chapter 4 attempts to remedy the lack of documentation of the methods of word-formation by providing an overview of the existing word-formation methods in Zulu, including those that apply to technical language. This section also considers cultural factors which cannot be separated from these elaboration methods.

Chapter 5 provides methodology to demonstrate how technical written text sources can be used as a resource for terminology by applying (computational) corpus linguistics. It shows how texts in a specific technical field can be used to compile corpora which can be utilised for the purpose of semi-automatic term extraction to aid lemmatisation. This chapter exemplifies how computerised corpus query tools, called *WordSmith Tools* (WST), in particular *wordlist* and *concordance*, can be used for the extraction of both simple and complex (multi-) terms, including nominal, verbal and qualificative/adjectival terms, considering both linguistic analytical and technical factors.

Chapter 6 explores methodology to demonstrate how oral sources can be used as a resource for terminology by applying the theory of corpus linguistics. It explains how existing standard terms can be compared to oral terms (obtained through questionnaire-related interviews - see Addenda 1 and 2) to provide a type of corpus annotation (analysis). It motivates how this annotation can be (re)used for the purpose of improving the acceptability of technical terminology. For instance, the corpus annotation tendency, named 'accurate designation' can be used to improve the accurate designation of concepts.

Chapter 7 draws conclusions from the researched practical approach to the standardisation and elaboration of Zulu as a technical language. This chapter also outlines some contributions and limitations of this thesis and some recommendations for future applications and research are made.

1.8 General scope of this study

This study does not attempt to be prescriptive in its practical approach to the standardisation and elaboration of Zulu as a technical language. Rather, it provides some guidelines towards an understanding of national standardisation and elaboration of the African languages, in particular Zulu.

In view of the lack of previous guidelines, a methodological practical foundation is provided, with sufficient exemplification in Zulu in order to encourage further application and research. The methodology focuses on the application of corpus linguistics to a specific set of data. This data is mainly selected written and oral medical corpora which cannot be considered representative of the entire Zulu medical corpus. Yet, the methodology and insights gained in this study can be applied to other technical fields where elaboration is needed.

CHAPTER 2

LANGUAGE STANDARDISATION AND ELABORATION OF THE AFRICAN LANGUAGES

2.1   Introduction

This chapter concerns language planning activities in mainly developing countries where, as logically stated by Fishman (1974:85), most language planning occurs. Standardisation within a language planning perspective is discussed with reference to the common linguistic division of status and corpus planning. This chapter focusses briefly on status planning as far as national and educational language policies are concerned. Standardisation and elaboration are discussed from a corpus planning perspective. However, no type of corpus planning (language standardisation) can be successful if status planning (policy) is not thoroughly implemented.

The input of official bodies and appointed committees on the implementation and improvement of language policy, as set out in the LANGTAG Report and in the *Language-in-Education Implementation Plan* (LIEIP), for instance, is also discussed.   Against the background of such input and the present general perceptions of language issues, the status of the African languages is evaluated in terms of the constitution.   Shortcomings are pointed out as far as the actual implementation of language policy in the public and educational sectors is concerned and possible solutions are proposed to enhance the process of functional development.

An attempt is also made to define the two interrelated concepts of standardisation and elaboration. However, it should be noted that elaboration is seen as part of the standardisation process and not as a separate concept. Since standardisation is perceived as quite a complex linguistic issue, it is discussed with reference to different aspects such as definition, models, norm, purpose, stages,

agents, limitations and problems. All these aspects are approached from an (inter)national perspective, paying special attention to the African languages, with specific language exemplification in Zulu.

2.2 Language planning

It is difficult to formulate a precise theory for language planning since it is such a complex activity influenced by many changing social factors such as economics and politics, for instance (Cooper 1989:182). There are many different approaches towards language planning, and many definitions have been attempted. A rather satisfactory general definition of language planning is the one formulated by Toffelson (1991:16):

> The commonly-accepted definition of language planning is that it refers to all conscious efforts to affect the structure or function of language varieties. These efforts may involve creation of orthographies, standardisation and modernisation programmes, or allocation of functions to particular languages within multilingual societies. The commonly accepted definition of language policy is that it is language planning by governments.

In the context of this thesis this definition is preferred, since it specifically refers to a multilingual context which is relevant to South Africa. Language planning can be divided into two main sections, *status planning* and *corpus planning*. Status planning deals mainly with language policy and its implementation as well as the selection of languages used for official purposes and education. Corpus planning, however, deals with codification, standardisation and language elaboration with regard to technical development for the educational and public sectors.

Heugh, Siegrühn and Plüddemann (1995:vii) feel that language planning has thus far been planned hierarchically from the top down, i.e. by those in power, whereas it should happen from 'below', so that it is less exclusive, in order to fulfil the needs of society. Bamgbose (1991:145) also supports this in saying that decision-making should happen at all levels of society, including non-governmental agencies. This type of language planning, in which the specific needs of individual communities are addressed, by means of consultation and participation in policy formulation, is called community-based language planning (LANGTAG in DACST 1996:215). Bamgbose (1991:109) reasons that language planning requires the identification of major language problems in a nation along with outcomes to solve these problems. Neustuphy (1974:39) uses the practical terms *policy approach* and *cultivation approach* instead of status and corpus planning, seeing these as basic types of treatment of language problems.

In the following section some attention is given to status planning since the interrelationship between status and corpus planning (specifically standardisation in this context) cannot be overlooked. In this regard Ohly (1987:55) states that if a language has low status its terminography will be underdeveloped. Furthermore, it must be noted that even aspects of status planning cannot be discussed in isolation, since national language policy and educational language policy are interrelated.

2.2.1 Language policy as part of status planning

Before discussing policy planning, it is perhaps apt to cite a definition of the term language policy seen as an integral part of language planning, by Heugh, et al. (1995: vii):

> A country's **LANGUAGE POLICY** is a set of principles conceptualised within an overarching framework of values, usually embodied in the constitution. If it is to be effective, the language policy has to be congruent with a country's national development plan. **LANGUAGE PLANNING** as a term refers to the process of

implementing a particular language policy.

Bamgbose (1991:111) provides a clearer, more multilingual-type of definition of language policy, which also warrants citing:

> ... a language policy may be defined as a programme of action on the role or status
> of a language in a given community. In a multilingual situation, a language policy
> decision necessarily involves the role or status of one language in relation to other
> languages.

There are different language policies, each of which is applicable to a specific language scenario prevailing in a specific country, associated with particular scholars or committees. The formulation of different types of language policies can aid governments in adopting the best possible national language policy. Scholars who contributed greatly towards the field of language policy in Africa, are, for instance, Nhlapo (1945), Alexander (1991), Marivate (1992), Bamgbose (1991), Cluver (1993,1994), Crawhall (1993) who coordinated the *National Language Project* (NLP) and Kaschula & Anthonissen (1995). Although language policy is approached differently by these scholars, several parallels can be drawn between them, especially concerning implementation in the South African multilingual context. General language policy, which mainly has to do with the choice of (a) national/official language(s) is closely related to the educational language policy of a state. It is thus appropriate to refer to both the South African language policy and the educational language policy below.

2.2.1.1 The South African language policy

The South African Language Policy is not a separate official document but forms part of the South

African constitution contained in *Act 108 of 1996* and basically makes provision for the following language-related issues:

- Official languages (Section 6)

- Equality in language (Section 9)

- Language in education (Section 29)

- Language and culture (Section 30)

- Cultural, religious and linguistic communities (Section 31)

- Language with regard to arrested, detained and accused persons (Section 35).

From the language-related sections it is clear that the language rights such as the elevation of the status and development of the previously disadvantaged languages, language equality and cultural and community rights within the framework of multilingualism, are entrenched in the constitution.

2.2.1.2 The educational language policy

The educational language policy, which every country should have, is about "... setting out the relationship between the teaching of the various languages and the levels at which they are taught" (Bamgbose 1991:106). According to Bamgbose (1991:62) language mainly has three objectives in education, i.e. literacy, medium of instruction and subject. Msimang (1992:41), in this regard, reasons that the school is still the best place to develop and maintain a language. The latter statement is in agreement with the LANGTAG Report (DACST 1996:70) stating that "...the education system is the main mechanism used to spread the developed form of the language."

The educational language policy of South Africa is being referred to in the Constitution - *Act 108 of 1996, Section 29* - but is fully discussed in a separate official document by the *Department of Education* (DOE) and titled *Language-in-Education Policy* (LIEP), dated 14 July 1997.

Reference should in this regard be made to the *Language-in-Education Implementation Plan* (LIEIP), an important document for the interpretation and implementation of the educational language policy for South Africa, issued by the DOE in 1997 in terms of the *South African Schools Act, Act 84 of 1996*.

However, the educational language policy for South Africa cannot be seen in perspective without referring to yet other important official documents such as the LANGTAG Report (1996) and the *National Education Policy Investigation* (NEPI). The LANGTAG Report (DACST 1996:26,124,125), sets the goals for a language-in-education policy, as basically being access to meaningful education, the promotion of nation-building through multilingualism, and modernisation in the African languages. The NEPI issued by the previously (somewhat stigmatised) *Department of Education and Training* (DET) in 1992 reflects the history of the languages used as medium of instruction in education in South Africa. This document also impresses upon parents and teachers the need to make informed decisions about which language should be used as medium of instruction.

The various language-in-education policies will not be discussed in detail. The problems in status planning implementation and solutions to such problems, however, form the focal point of the following section.

2.2.2 Problems with status planning implementation in South Africa and possible solutions

The identified problems with status planning implementation are twofold. They mainly deal with the implementation of the language policy in general, and specifically with the implementation of the educational language policy.

2.2.2.1 Problems and solutions with regard to the general language policy

According to 1990 statistics provided by the Human Sciences Research Council, 75% of all South Africans have an indigenous African language as mother-tongue and the other 25% of South Africans speak 'other' languages such as Afrikaans, English and Portuguese, for instance (Kaschula & Anthonissen 1995:97). Yet, if we look at languages actually used in official, educational and societal settings, English is by far the most utilised prestigious language in South Africa - a language not representing the actual majority of indigenous African language speakers! However, this situation will prevail if African language speakers themselves do not take action in order to value and empower their languages.

Bamgbose (1991:111), being an African, and after having evaluated the situation on the continent, recognises typical causes for language policy problems in Africa:

> Language policies in African countries are characterised by one or more of the following problems: avoidance, vagueness, arbitrariness, fluctuation, and declaration without implementation.

Most scholars of language planning would agree that Bamgbose's statement could be applicable to South Africa. Even though the language policy looks very impressive on paper it is often not applied successfully. The LANGTAG Report (in DACST 1996:36,37; 45,47 and 194,195) identified a number of problems regarding the implementation of the South African language policy, particularly with regard to language equity. The main problem areas, particularly in relation to the African languages, are identified followed by solutions/recommendations to these problems. It should be noted however, that the 'problems' and 'solutions' cannot be regarded as absolute. They are merely an arbitrary attempt to systematise the problematic South African language policy scenario:

i)   There is a gap between the language policy adopted by government and its implementation. "There should be monitoring to ensure that constitutional promises become final proposals" (LANGTAG in DACST 1996:150) by applying language equality and multilingualism.

ii)   Language services lack an adequate infrastructure and language workers enjoy a very low status.
      The proper infrastructure for language development and language awareness should be created.   Language workers (including teachers) should enjoy a higher status with corresponding higher compensation.

iii) There is a lack of trained language workers such as translators, teachers and interpreters in the African languages.
      Language teaching and teachers' training methods should be improved and be more Afrocentric. This also applies to literacy programmes. Such efforts should be coordinated between all the African languages and adjusted to be more needs-based.

iv)   Standard-setting procedures do not function or coordinate properly.
      These should be transformed, developed and modernised in order to provide a meaningful language service.

v)   Language equity is lacking in the private and public service and at governmental and at provincial level because of the widespread use of English as trade and administrative language.
      The dignity of all languages in South Africa must be respected in all sectors of society. This specifically applies to the marginalised African languages such as Tsonga and Venda.

2.2.2.2 Problems and solutions with regard to the educational language policy

Bokamba (1991:81) slams most African educational policies, regarding them as tools for colonising the Africans. Bamgbose (1991:69,70) goes on by referring to the *inheritance situation* since colonialism continues to shape education throughout Africa, exemplified by the choice of a colonial language of wider communication (LWC) as language of education. Alexander (1995:39) quite clearly understands what is meant by a colonial educational policy in the South African context:

> In South Africa, this legacy of the colonised mind manifested in the underestimation
>
> of the indigenous African languages and the overvaluing of English and, sometimes,
>
> Afrikaans is all too well-known to educators.

The result of colonialism in Africa is that the colonial languages, mostly French and English (both LWC languages), still greatly influence education policies in the independent countries and are regarded as the languages of academic function and scholarly discourse (Gonzalez 1993:12). Bokamba (1981,1991) and Gonzalez (1993) regard this continued use of the colonial language as medium of instruction (especially at the primary level) as counter-developmental. Furthermore, there are many poorly qualified teachers some of whom have a poor command of English or any other colonial language.

Not only African politicians have a low regard for their national languages but also parents and pupils. Many African parents, some belonging to the same mother-tongue group, never teach their children any African language. Breton (1991:169) calls this tendency a "...voluntary shift of mother-tongue."   This is also true in the new South Africa where many African parents prefer to communicate with their children in English and insist on English education.

It is worthwhile noticing that those parents who reject mother-tongue instruction, sometimes argue that the use of African languages as medium of instruction would disqualify African children from entering the former Model C schools which teach mainly in English (and Afrikaans), eventually also putting them at a disadvantage in the labour market. They also fear that insisting on mother-tongue instruction could entrench the previous Apartheid educational language policy which divided schools along ethnic lines.

Bokamba (1981:19) and Gonzalez (1993:16) basically agree that a bilingual educational policy is the best solution for developing countries. In order to solve developmental problems, an applicable national language or lingua franca must be used at the primary and secondary school levels and a LWC (mostly French or English) at university level and even at this high level the national language can be used for selected subjects (Bokamba 1981:19). Gonzalez (1993:16) proposes the same policy as Bokamba but opts for the national language as medium for the first four years only and thereafter the international language.

Cluver (1993:41) positively suggests that the South African government make funds available to develop all South African languages up to the Standard 10 level (now Grade 12) in order to overcome the unequal status, especially of the African languages in education. Prah (1998:35) sees the solution to the African educational crisis as mother-tongue tuition, stating that the language in which it is possible to conceptualise "...with the greatest ease and mental agility..." is the mother-tongue and "If science and technology are to reach the overwhelming rural masses of Africa, this can only be done in the languages of the masses."

Another solution is that Africans adopt a positive attitude towards their languages. Such is the view of Nkondo (1987:72) who believes that every African language must be regarded and studied in its own right, with the African language itself as medium of instruction.

Notwithstanding the discussed problems, the following are some helpful definite strategies proposed by the DOE (1997a:13-17) in their LIEIP Document in order to promote language development in the national languages:

*   the promotion of language awareness through multilingualism and language equity in cooperation with PanSALB,
*   the responsibility of parents to link economic empowerment with the best interests of their children's learning,
*   the cooperation of different government departments in South Africa and of neighbouring countries,
*   the provision of meaningful language interaction environments such as for sensible reading, writing, speaking and listening,
*   the integration of own language learning into all learning processes, irrespective of the subject taught.

Equipped with the proper background of language (status) planning, the topics of standardisation and elaboration can now be discussed. These topics are important, not only for language development, but also for proper language policy implementation. Although standardisation deals with both status and corpus planning, the emphasis in this thesis falls on the corpus planning.

2.3   Standardisation

Status planning is concerned mainly with the selection of language(s) used for official and educational purposes while corpus planning is more concerned with development of the language, the expansion of the lexicon, the creation of terms, codification and standardisation (Calteaux 1996:161). Standardisation is seen by Cooper (1989) and many other scholars such as Calteaux (1996) as part

of corpus planning. Hudson (1980:32) adds a language engineering perspective since he sees standardisation as a process of intervention by society in the normal development of language. For standardisation to be effective, a sizable part of society must take part in it. Standardisation is also seen as a process in this general definition by Matthews (1997:352):

> Standardisation is the process, often in part at least deliberate, by which standard forms of a language are established. Forms and varieties which are not standard are simply 'non-standard'.

Since standardisation is such a complex concept, it is, in an attempt to systemise it, to be discussed with reference to different aspects such as 'definition', 'models', 'norm', 'purpose', 'stage', 'agents', 'limitations' and 'problems'. This approach is fit since "Language standardisation is a process and is therefore essentially dynamic" (Ansre 1971:681).

2.3.1   Defining language standardisation

The *Longman Dictionary of Applied Linguistics* (Richards, Platt & Platt 1992:351) defines a standard language in simple terms, including its application in the language environment:

> It is also called the **standard dialect**, **standard language** or **standard**. A standard language is the variety of a language which has the highest status in a community and which is usually based on the speech and writing of educated native speakers of the language. A standard variety is generally:
> (a)   used in the news media and in literature
> (b)   described in dictionaries and grammars
> (c)   taught in schools and taught to non-native speakers when they learn the language as a foreign language.

This definition by Richards et al. (1992) is more descriptive than the general one provided by Matthews (1997) in 2.3 earlier.   Garvin's (1993) definition, however, is more concise than the previous two (see 2.3.2.3). According to Cooper (1989:135,137) the language of the elite is usually ideally perceived as correct and standard. This language then becomes the language of controlling institutions and schools.

Written varieties of language are generally rated more highly than spoken forms to such an extent that the written form is taught at school but never used in ordinary conversation (Cooper 1989:137). Furthermore, variety in language poses a problem for publishers who benefit from a large population sharing the same norm (Cooper 1989:137). Ideally the spoken language should be on a par with the written standard; this has not happened for most languages and the written form has always been the most valued (Njogu 1992:69). For standardisation to be beneficial for the general population, standardised forms need to be made available to as broad a section of the population as possible (Njogu 1992:69).

LANGTAG (DACST 1996:69) has a positive attitude towards the official standard African languages and promotes spreading the standard variety and modernising it, in order to enhance language development. Still, what is needed is that the standard African languages should start functioning outside their traditional domain of the immediate speech community and should also be used by government, industry and legal practice, for instance (LANGTAG in DACST 1996:69).

However, to define standardisation is far from an achievable task and in order to complement the definitions above more should be said about the properties of a standard language in relation to a 'non-standard language' (see 2.3.8.2). What emerged from the definitions and discussion above is

that a standard language is essentially a written selected variety which enjoys a particular status in the speech community it serves.

2.3.1.1 The properties of standard languages

The properties of standard languages have also been basically discussed under models of standardisation (see 2.3.3).   The important properties of standard languages, some of which also feature in the given definitions above, are put forward by Van Wyk (1992:26-32) :

- They are superordinate varieties co-existing with other non-standard varieties.
- They are used for higher forms of language.
- Standardisation of a given variety can be described as the result of historical incident. The given variety serves as standard for the spoken and written form or both. They contain linguistic norms recognised and striven for by a speech community.
- They represent formal style in language.
- They are difficult to define (as may have become clear from the preceding discussion - *own emphasis*).
- They may or may not be official (In South Africa all standard languages are official - *own emphasis*).
- They may or may not be based on vernacular varieties.
- They may or may not have subvarieties.
- They may or may not be used, or not used at all by a speech community.
- They are not static.

However, after considering the properties of a standard language, standardisation can be further defined by adding the aspects discussed below, such as models, norm, stages, agents, etc.

2.3.2 Models of standardisation

Although various models of standardisation have been designed, only three of the most significant are highlighted for this study, one by Haugen (1966) and the others by Crystal (1993) and Garvin (1993). Haugen's model mainly comprises four stages for the description of the standardisation process. Crystal's model, like Haugen's also comprises four stages although he names them somewhat differently and refers to his model as 'language planning in practice'. Garvin's model has a slightly different focus from that of the other two since it is based on the answering of three questions from a language planning perspective. Although these models of standardisation are approached from different viewpoints, there is some extent of overlap between them.

2.3.2.1 Haugen's model

Haugen's (1966) model is probably the best known and most widely used model. As a result many scholars such as Hudson (1980) base their models of standardisation on that of Haugen (1966:341-352) which comprises the following four stages:

- *Selection of a norm*

This stage deals with the selection of one out of a number of dialects/languages to become the standard. It could mean adjusting an existing language variety or creating a new variety that will become the new standard.

- *Codification of form*

Nationhood ideally demands that there should be a linguistic code through which communication can be facilitated (Haugen 1966:345). The norm is established by devising an appropriate script for the selected language variety. Codification entails establishing the orthography and describing the grammar and lexicon of the chosen language. This stage is also commonly known as graphisation by scholars like Ferguson (Ansre1971:680). Haugen (1966:348) concludes by defining codification

as "...*minimal variation in form...*" . Minimal variation can only be attained in a stable language with the least variation in norm.

iii)   *Elaboration of function*

This stage entails the expansion of the lexicon and the creation and development of scientific and technical terms to meet the demands of complex communication. It involves modernisation of the language to reach the highest level of expression in all possible domains such as politics, law, medicine, technology, education, culture, etc. Thus, according to Hudson (1980:33) elaboration is linked to "...all the functions associated with central government and with writing, for example in parliament and law courts, in bureaucratic, education and scientific documents of all kinds and ... various forms of literature." Garvin (1993) refers to this stage as the intellectualisation phase (see 2.3.2.3 ii,iii).

Haugen (1966:348) observes that elaboration is "...*maximal variation in function*...", especially attained in a fully developed language. This third elaboration stage can be compared with Garvin's second question of how a standard language serves its users. See 2.3.2.3 (ii). Terminology needs to be created and developed to meet the needs of a rapidly changing global world. Furthermore, terminology has to be documented in glossaries, dictionaries and other normative documents or manuals. This can only happen in the written form as explained by Haugen (1966:348):

> Writing, which provides for the virtually unlimited storage and distribution of vocabulary, is the technological device enabling a modern standard language to meet the needs of every specialty devised by its users.

iv)   *Acceptance by the community*

The last stage deals with the acceptance by the speech community of the previous three stages of development of the language. For a language to be accepted it must have a body of users so that it can spread as standard in the social system of a nation.

Haugen (1966:350) summarises that the first two features of the model refer to the form and the last two to the function of standard language. Yet, it is made clear by Haugen (1966:349) that the above stages depend on one another. Codification and elaboration, for instance, will not progress unless the community can agree on the selection of the norm.

2.3.2.2 Crystal's model

There is a strong indication that Crystal's (1993:364) model is also based on Haugen's (1966) model although the four stages are named somewhat differently:

i)   *Selecting the norm*

In this regard Crystal clearly states the factors that need to be taken into account when selecting a variety. In one instance it is necessary to choose a particular variety of a language; in another instance a new variety can be created, considering social factors such as literary use and social class.

ii)   *Codification*

Codification deals with devising an alphabet and writing system for the oral form of the language. Provision will also have to be made for standard pronunciation, grammar and vocabulary.

iii)   *Modernisation*

The lexicon is to be modernised so that foreign material in, for instance, the consumer market and scientific field can be translated.

iv)   *Implementation*

The chosen standard language will have to be officially implemented in the governmental and educational sectors and be regarded as the language of educational progress and social status.

2.3.2.3 Garvin's model

Garvin (1993:37-54) bases his model on the asking and answering of the following three questions:

i)   *What is a standard language?*

In order to answer this question Garvin (1993:41) gives the following concise but accurate definition:

> One can define standard language as a codified variety of language that serves the
> multiple and complex communicative needs of a speech community that has either
> achieved modernization or has the desire of achieving it.

Garvin discusses two important properties of standardisation, namely flexible codification and intellectualisation. Codification is not only the written code but does imply that rules of correctness are codified in documents which are accessible to the speech community such as grammars and dictionaries. Yet, this codification also has to be flexible in order to accommodate changes in the language. However, the property of intellectualisation (emphasised in his definition) is more far-reaching since it deals with the capacity of language to develop more accurate means of expression in, for instance, technology, higher education and politics.

ii)   *How does a standard language serve its users?*

A standard language serves the speech community through officialisation (Garvin 1993:39,40). This means that the language is officially recognised by law or by a constitution in order to serve as an identity symbol in the public and governmental sector, in transactions and on postage stamps, for instance.

Garvin's second question (1993:43) also concerns the domains of language use. The domains of modern life that are served by the standard language are technology, science, government and politics, higher education, culture, the law and religion. These domains can be associated with the elaboration step of intellectualisation as becomes clear in the following statement which forms an answer to the second question:

> ...a standard language serves above all for the cultural and intellectual
> communication of a speech community and allows it to use its own language to deal

with these important domains (Garvin 1993:44).

iii)   *What are the conditions required for the development of a standard language?*

In a way this third question is closely related to the previous question, since the functions of the standard language discussed below also relate to the purpose of standardisation (serving the speech community). Garvin (1993:45) emphasises the fact that there are no civilised and primitive languages. Any language can be developed such that it becomes standardised. Haugen (1966:344) also maintains that inherently handicapped languages do not exist. Garvin (1993:45-48) sets out three conditions that relate to the degree of standardisation, namely the properties, functions and evoked attitudes of a standard language.   The properties have already been mentioned, namely flexible codification and intellectualisation. He postulates five functions of a standard language, these being the unifying, separatist, prestige and participatory functions, all symbolic, and the objective frame-of-reference function.

After standardisation has occurred, the unifying function unifies, in spite of differences, several dialects in a single speech community. The separatist function singles out a speech community with a separate identity within neighbouring speech communities. The prestige function adds prestige to the possession and mastering of a standard language. The participatory function promotes the use of the own language to participate in modern communication within different domains such as the scientific domain, for instance. The frame-of-reference function serves as reference, mainly in matters of language correctness.

Garvin links these five functions to different language attitudes in the following manner:

•   The unifying and separatist functions are linked to the attitude of loyalty. The unifying
       function concerns the loyalty of a speech community to the standard language. The

separatist function concerns the loyalty of a speech community to its own standard language, and not to any other variety.

- The prestige function links up with the pride of a speech community to possess a 'real' language and not just a dialectal variety.

- The participatory function links up with a desire to take part in modern life through intellectualisation in the field of technology and science, for instance.

- The frame-of-reference function links up with an awareness of norm. Members of a speech community are aware of a model that they can follow for matters in language and literary creation, for instance.

Some of the stages of these three models of Haugen (1966), Crystal (1993) and Garvin (1993) are discussed in more detail with exemplification in the African languages and in Zulu in particular in the sections below.

2.3.3   Standardisation as a norm

Norm in a language can also be seen as a type of habitual or obligatory language use but does not exist if a language is underdeveloped (Drozd & Roudny 1980:34). One can thus assume that "Standard languages refer to the written formal form of the language. They are taught in the schools and used in publications and the radio" Calteaux (1996:36). Sager (1990:118) basically agrees with the normative viewpoint of these three scholars since he sees a standard language as the written form of educated speech and as part of a standardised system. Haugen (1966:346) agrees that the written format is a 'crucial requirement' for a standard language. This means that, for instance, rules are laid down for spelling, grammar and pronunciation. This standardisation process is carried out by governmental and educational institutions. Also see 2.3.6 for agents of standardisation. 'Norm' usually has the connotation of prescription as Calteaux (1996:38) rightfully observes. This means that standard languages mostly have notions of 'correct' and 'incorrect'. It is for this reason that

Garvin (1993) associates norm with frame-of-reference. See also 2.3.2.3 (iii).

Rubin (1977:158-168) breaks 'norm' down into the following six logical interrelated processes (of which the first three should always co-occur):

a) isolating the norm,

b) evaluating the norm by a significant group who perceives it as correct,

c) prescribing the norm for specific contexts,

  and for standardisation to be effective the norm must be

d) accepted,

e) used, and

f) remain in use until replaced by another norm.

These interrelated processes serve to describe the complexities of a language standardisation process. However, scholars warn that one should be careful in establishing the norm of a language: In this regard Drozd and Roudny (1980:34,35) suggest that when the norm is shaped by artificial intervention the stability of the language should not be disturbed. Since language planning cannot fulfil all needs, there must be a norm to measure deviation. However, such a norm cannot be static since language is dynamic and changing. The real challenge is to set the norm in a creative and realistic manner (Calteaux 1996:172).

2.3.4   The purpose of standardisation

According to Njogu (1992:70) the purpose of standardisation is "... to provide an authoritative reference point in the form of grammars and dictionaries and it can also contribute to nationhood." Nationhood it seems, is not an aim in itself but may become a reality if language standardisation is successful.

Sager (1990:122) summarises the purpose of standardisation very effectively. He also specifically refers to the standardisation of terms which is quite relevant for this study:

> Standards are economical because they establish prior agreement of reference among the participants and therefore assist in the achievement of effective communication among specialists by speeding up the process of communication. Standards are precise because they eliminate misunderstanding by establishing a clear one-to-one equivalence between terms and the region of the conceptual system referred to.

In the African context, a developing standard language will enjoy stability and mainly fulfil the following functions: unifying, separating but regarding other languages, enjoying prestige, and being normative, in the sense of having a frame of reference (Chimhundu 1992:86,87). Also see 2.3.2.3 (iii) for Garvin's model concerning the functions (also in combination with the attitude of the speakers) of the standard language in the community.

2.3.5 The stages of standardisation

A language can only be regarded as standardised once it has undergone certain stages of development. Yet, these stages are named arbitrarily by different scholars (discussed under 2.3.2). The result is that very often different terms designate more or less the same concept, e.g. 'elaboration' and 'modernisation'. Although 'graphisation' and 'codification' are very closely related concepts, the former is a prerequisite for the latter as becomes clear in the ensuing discussion in the following sections. In the following sections the different stages are discussed and exemplified in terms of the African languages, particularly Zulu.

2.3.5.1 Selection

The term selection refers to the process whereby a language variety must be chosen or varieties must be identified out of existing varieties to become the standard language. LANGTAG (DACST 1996:220) sees this stage as a process "...of making some aspects of language conform to a standard variety." Calteaux (1996:37) observes that standard is defined in agreement with a language community, i.e. that to a Zulu, standard means 'standard Zulu', etc. In South Africa the standard forms of the African languages are based on regional dialects spoken in specific rural areas (Calteaux 1996:38).

Very often missionaries applied the principle of dialect harmonisation as, for instance, in the case of Tsonga, Xhosa and Sotho (LANGTAG in DACST 1996:73). However, as Chimhundu (1992) argues, such dialect unification is not always generally accepted. The name 'Shona', one of the official languages of Zimbabwe, was given by the South African linguist Doke who was assigned to develop a standard language which was based on a unification of the dialects of Mashonaland, including Karanga, while other dialects were not considered with the result that this standard variety and its orthography were not readily accepted by local officials.

As early as 1847 the Wesleyan missionary Appleyard, in a contribution in a South African missionary magazine, classified Bantu into two dialects, namely 'Kafir' being Xhosa and 'Kafir proper' being Zulu (Doke & Cole 1961:38). This is the first written evidence of dialect selection as far as Zulu is concerned. Although the linguistic-historical origins of the Nguni are somewhat difficult to trace, owing to fragmentations and contradictions of oral and diachronic accounts (Msimang 1989:35), different sources such as Kubeka (1979:83,84, 91,93), Msimang (1989: 39,49) and Poulos & Msimang (1998) seem to agree that the Zulu dialect that basically forms standard Zulu, is the *Ntungwa* variety.   LANGTAG (DACST 1996:73) also confirms this stating that the

American Board of Commissioners for Foreign Missions selected the Ntungwa variety of Dingane as standard. Kubeka (1979:92) calls this *Ntungwa* variety the *Central Zululand Dialect* (CZD) since it "... more or less approximates to the original *IsiNtungwa*, but naturally has incorporated new sound features." Generally speaking, Zulu today is regarded as the name of the language spoken by Blacks from Zululand and Natal (Kubeka 1979:1). However, it could be said that all the dialectal varieties of Zulu spoken in KwaZulu-Natal today are regarded as Zulu.

Zulu is a fairly established and uniform language in comparison to other African languages in South Africa, perhaps because Döhne's (1857:xv) observation that Zulu is as old as Shaka's reign, is true. Shaka, through his conquests of tribes in Central Zululand, despite the hordes of other tribes fleeing to the south, also contributed towards this uniformity. This is confirmed by Kubeka (1979:228) who states that although the conquered tribes did not initially willingly accept the language (Zulu) of the conquerors, Zulu nationhood gradually contributed to the acceptance of the Zulu language. Furthermore, Shaka decreed that the language used in his presence, by other tribes, such as the *Lala*, for instance, only be his (the Ntungwa) variety (Msimang 1989:49).

2.3.5.2 Graphisation

After the selection of a speech variety or dialect, graphisation follows. This entails the process of reducing a language to writing (Cooper 1989:125). As far as graphisation is concerned it was only recently discovered that the first evidence of written Zulu (1844) was the work of a Frenchman, Adulphe Delegorgue (Davey & Koopman 2000:134,135). He compiled a vocabulary of the language which he called 'la Langue Zoulouse' before1847 by which time very little written evidence of Zulu had existed. However, the most important written contributions came from the pen of the

early missionaries of the American Board in Natal, Bryant and Grout (Doke & Cole 1961:40,41). In 1848 Bryant's pioneering grammatical contribution titled *The Zulu Language* was published.. This was followed by a journal article titled *The Zulu and other dialects of Southern Africa* by Grout (1849) and a grammar titled *A grammar of the Zulu language* (1859) by the same author. In 1857 Döhne's Zulu dictionary appeared, being the first scientific dictionary on a South African Bantu language.

In the Nguni languages graphisation was characterised by orthographical change due to language development. In some of the earliest Zulu orthographies phonetic symbols are used alongside roman symbols , e.g. **6** in **u6a6a** instead of **b** in **ubaba** (my/our father) for the bilabial implosive consonant. Presently all languages in South Africa use the roman script. However, as the African languages developed, changes occurred in the orthography. In the case of Zulu the most common changes are:

dhl is replaced by **dl**, e.g. **-dhlala** > **-dlala** (play),
h is replaced by by **hh** (for the voiced glottal fricative), e.g. **ihashi** > **ihhashi** (horse).

In aspirated plosives **h** must always follow the plosives **p**, **k** and **t** which was not the case in the earliest orthographies, e.g. **-peka** > -**pheka** (cook); -**kula** > **-khula** (grow); -**tela** > -**thela** (pour).

Disjunctive writing (separating morphemes in a word) is replaced by conjunctive writing (combining morphemes to form a word), e.g. **kwa Dukuza** > **kwaDukuza** (Zulu name for Stanger).

2.3.5.3 Codification

Codification is the next stage which is closely related to the former (graphisation) and it typically refers to the written form of a standard language evidenced by the existence of published dictionaries, grammars, spellers and style manuals (Cooper 1989:145). However, to scholars such as Haugen (1966) and Garvin (1993) graphisation and codification are seen as a single process. Multiple evidence of Zulu's established codified status can be found in its many publications. Many dictionaries have appeared, mostly Zulu-English ones and a few Zulu-Afrikaans ones. The dictionary that marked a new era in the Zulu lexicography was undoubtedly the major contribution by Doke & Vilakazi (1949) containing only Zulu entries with English explanations. What marks Zulu's development even further, is the appearance of a series of monolingual explanatory Zulu dictionaries by Nkabinde (1982,1985) and Nyembezi 1992); and even the appearance of a monolingual explanatory Zulu dictionary of synonyms, by Shabangu (1987). Besides dictionaries, many Zulu grammars, both in Zulu (mostly for schools) and in English, appeared. Many studies based on Zulu linguistics and literature were/are being conducted. Literary publications including the genres novel, short story, essay, drama, poetry, children's story and traditional literature (praise poetry, songs, folktales, etc.), are unparalleled in number by any other African language in South Africa.

2.3.5.4 Elaboration

The next logical stage of standardisation is elaboration. See 2.3.2.1 (iii) for Haugen's definition of elaboration and 2.3.2.2 (iii) for Crystal's view on modernisation. Elaboration is a very important part of the standardisation process, also for this thesis, since it deals with the evidence of language development. It should be stated that the standardisation and elaboration processes in Zulu are treated as coordinated processes. In fact, elaboration is a phase in the standardisation process and therefore part of it.

Elaboration implies that expansion on technical vocabulary must take place and that formal writing

conventions must be established so that the new standard can be used in law, health care, government, etc. Although Zulu is a well-established language, there is still a shortage of technological terminology in the educational and public sector. Basic terminological work was produced for educational purposes by the previous ZLC/B. The previous NTS and the present NLS contributed, for instance, in the form of medical and specific AIDS-related terminology in all official languages.

The reason for the lack of terminology is that Zulu is used as medium of instruction for scientific technical subjects only for primary school tuition. See also see 2.2.2.2 for educational language problems. As far as the subject Zulu for mother-tongue speakers is concerned, it is taught through the medium of Zulu up to matric level. Mother-tongue speakers are at present taught through the medium of Zulu only at a minority of South African universities. Most universities, however, employ both Zulu (mainly for literature) and English (mainly for linguistics) in Zulu offered as subject for mother-tongue speakers. When Zulu is offered as subject for non-mother-tongue speakers the medium of instruction could be Afrikaans and/or English depending on the language(s) of tuition chosen by the specific university.

When the objectives of term development in the African languages, and specifically Zulu, are considered, it becomes evident that no clear-cut policy, if any, has ever been formulated by South African policy makers. Following Abdulaziz' reasoning (1989:36), it would then serve little purpose to develop more complex scientific terminology for the tuition of science through the medium of Zulu at university level because science is not even taught via Zulu at secondary school level.

Modernisation is another term which is employed to refer to language elaboration by scholars such as Crystal (1993). Also see 2.3.2.2 (iii). Modernisation "...refers to the process whereby a language becomes an appropriate medium of communication for modern topics and forms of

discourse" (Cooper 1989:149). This stage is basically referred to by Garvin (1993) as intellectualisation (see 2.3.2.3). If this notion of communicative modernisation by Cooper is applied to the African languages, including Zulu, one can come to the conclusion that modernisation (by implication also standardisation) has not been fully achieved. The 'modern discourse' referred to by Cooper does not exist in the true sense of the word, especially not as far as technology and science are concerned.

2.3.5.5 Acceptance

A logical outcome of elaboration is acceptance which means that the standardised variety must be accepted as the national language by inhabitants living in a specific area and be seen as a unifying drive. See also 2.3.2.1 (iv) for Haugen's model and 2.3.2.2 (iv) for Crystal's model. However, instead of acceptance Crystal prefers to call this stage 'implementation'. Galinski (1982:193) is of the opinion that team-work in terminology standardisation will contribute to acceptability:

> The more people are involved in the preparation of standardized terminology the more an ideal of terminological standardization is reached: a terminology based on general acceptance.

Acquisition planning, called such by (Cooper 1989:159), is another important additional stage of standardisation that needs to be added to the acceptance stage. He distinguishes three goals, a) acquisition of a foreign language b) reacquisition of a language which was once a vernacular such as Irish and c) language maintenance through acquisition by the future generation. In the South African context of the African languages, Zulu in particular, the latter goal is quite important. For the African languages to gain value and status, language maintenance through acquisition by the future generation is of the utmost importance. A point that has to be remembered by language planners

is clearly put forward by Cooper (1989:159): "Prevention of decline requires maintenance of acquisition."

2.3.6   Agents of standardisation

As in any other country, term development and standardisation is facilitated by agents or institutions. These agents work on national as well as international level to make standardisation possible. In most countries official bodies serve as instruments of standardisation. For national standardisation technical committees of the national standard organisations take the responsibility.   There are countries where such standardisation bodies or activities hardly exist, of which Namibia is an example, especially with regard to the indigenous African languages.

2.3.6.1   International standardisation

The *International Organisation for Standardisation* (ISO) is an important body which has published guidelines for technical language elaboration. It established some standardisation principles between 1963-1970. Another one of the internationally established well-known standardisation bodies is that of the Federal Republic of Germany, known as the DIN (Sager 1990:116,117). This body takes great care in the preparation of concept systems, international harmonisation and presentation of publications.

Although the ISO has international status, a major shortcoming is that it is perceived as an Indo-European organisation favouring the industrial countries (Sager1990:117).   Some of its recommendations are being redrafted and have not kept up with modern developments in technology. Furthermore, international cooperation is hampered by the use of very few languages such as French, English and Russian in documentation. In addition, the extent to which a country

lends priority to language development, can create further problems. Obviously no action can be taken if funds are not made available for terminology development.

Another international body for the standardisation of terminology is the *International Information Centre for Terminology* (INFOTERM) which was established in 1971 and sponsored by Unesco. This body uses the framework of the UNISIST technical programme in liaison with the ISO and is affiliated to the *Austrian Standards Institute* (ON). The *International Network for Terminology* (TermNet) is a terminology business network and an international cooperation forum for companies and institutions with the aim of developing a world market for terminological products, tools and services. It is accessible on the world- wide web and is valuable for universal information and knowledge management in terminology.

Nevertheless, the ISO principles as mentioned by Sager (1990:117-120) specifically deal with terminology development in languages and therefore merit mentioning in this study:

Principle 1

Standardisation of objects should come before the standardisation of terms. Terminographers should then reduce the number of designations.

Principle 2

Standardisation of terminology is a socio-economic activity involving all interested parties, who should eventually reach consensus.

Principle 3

The application of standardisation is more important than the mere publication of terminology.

Principle 4

Standardisation implies the choice of the appropriate term which is followed by fixation of this term in the form of a definition.

Principle 5

'Standard' must be re-examined and revised on a regular basis.

Principle 6

Verification is needed to establish whether the term conforms to specifications. A certain method has to be followed by term creators in order to apply verification to a set of terms.

Principle 7

It is important to legalise standards within the context of a sensitive understanding of economical convenience and language use for social, regional and subject groupings.

2.3.6.2 National standardisation

In most countries official bodies serve as agents of standardisation. In Tanzania, for instance, the standardisation body for Kiswahili is known as BAKITA. In South Africa standardisation and terminology development in the African languages have been hampered by many educational, political and historical drawbacks, as outlined by Mtintsilana and Morris (1988:109). In the former Republic of South Africa the emphasis in term development was mostly on Afrikaans, to bring it on par with English, and on English. This is supported by Msimang (1992:139) who points out that the impression existed that South Africa is a bilingual state when in fact it is a highly multilingual state with nine standard African languages.

In South Africa, term standardisation in the African languages was originally the function of the former *Department of Bantu Education* (DBE) which later co-ordinated the different *Language Committees/Boards*, e.g. the *Zulu* or *Xhosa Language Committee/Board*. However, the structures and focus of the Language Boards in the African languages have changed over the years.

The Nationalist Government of this country played a big role in the standardisation of the African languages as clearly outlined in the LANGTAG Report (DACST 1996:78-82). In 1948 the Nationalist government started its official standardisation campaign in the African languages through *Language Committees* which were established by the *Department of Native Affairs* (*Section*: *Bantu Education*). Although the state contributed towards codification and standardisation, another task was also the promotion of ethnicity to enhance Apartheid policies.

In order to cope with the introduction of mother-tongue education in 1953, the then DBE had the task of developing and establishing orthographies and terminologies in the African languages. This education policy created an uproar in education because the African languages (including Xhosa), according to Jafta (1987:130) had not yet reached the stage of development to produce technical terms.

The establishment of the homeland states in South Africa (such as Venda) led to the adoption of African languages as official languages in these states. In turn, this led to the need for term creation in the legal and administrative fields where communication was conducted in the vernacular. The development of the Black media and Black consumer market led to an even greater demand for terminology in the African languages. The result was that terminologists created terms out of necessity. It is thus hardly surprising that some African scholars such as Jafta (1987:127,131) slam the terminology development and standardisation effort of the Language Boards, calling it an artificial process of term manufacturing aimed at serving Apartheid policies.

By the 1960's different *Language Boards/Committees* were appointed for each African language and by the 1970's most of these committees became autonomous within the various homeland structures. The main task of these Boards was to develop these languages: to codify, standardise and maintain the orthography, to elaborate on terminology but also to select suitable books for use in education. They also had to liaise with newspapers, broadcasters, the Bible Societies, cultural

organisations and be involved with place names. Clearly their task was too wide to focus on actual language development. The outcome was that they published a short trilingual (African language directed, Afrikaans and English equivalents) glossary of terminology for use in primary schools and by the national broadcasters.

Furthermore, the terminology lists compiled by these Language Boards contain little more than fairly elementary technical vocabulary (Cluver 1993:30). However, despite all the criticism, the Language Boards contributed to a great extent to the standardisation of orthography and terminology in the African languages. Compare in this instance the Zulu terminology lists prepared by the previous *Language Boards* under the auspices of Education departments in the older dispensation:

Department of Native Affairs (1957) *Zulu-Xhosa terminology and spelling No. 1*,

Department of Bantu Education (1962) *Zulu terminology and orthography No. 2*,

Department of Bantu Education (1976) *Zulu terminology and orthography No. 3*,

Department of Education and Training (1993) *IsiZulu terminology and orthography No. 4.*

African language scholars have thus far been highly critical of the previous Language Boards also because these Boards were associated with Apartheid policies, without noticing their worthwhile contributions. In future it is expected that scholars will be just as critical towards the new NLBs. Already these NLBs have very little to offer as far as publications and serious promotion of the African languages are concerned since the new political dispensation came about in 1994.

At present the former African language Boards are known as *National Language Bodies* (NLBs), thus the body for Zulu is known as the *IsiZulu National Language Body* (ZNLB) and the body for *Xhosa* is known as the *IsiXhosa National Language Body*, etc. In order to facilitate the process of political transformation in South Africa, all nine official African languages are referred to by their indigenous names: *Sepedi* (Northern Sotho), *Sesotho* (Southern Sotho), *Setswana* (Tswana),

SiSwati (Swazi), *Tshivenda* (Venda), *Xitsonga* (Tsonga), *IsiNdebele* (Ndebele), *IsiXhosa* (Xhosa) and *IsiZulu* (Zulu). These NLBs including Afrikaans and English, are under the jurisdiction of the national language authority PanSALB, established in 1995.

Africans mostly do not perceive multilingualism as a major stumbling block but rather as a realistic challenge. The constitution recognises eleven official languages but also promotes the development of the Khoi and San and Nama languages and the Sign Languages, used in South Africa (NLPF 2002:8).

In South Africa terminological standardisation is mainly taken care of by the *National Language Service* (NLS), previously known as the *National Terminology Services* (NTS). The NLS, for instance, prepares term lists in all the official languages such as a term list of medical terms and contributes in this manner towards African language development. Overseeing all language activities and terminological standardisation in this country, is PanSALB under the jurisdiction of the *Department of Arts and Culture* (DAC).

In the meantime, after the new dispensation in 1994 and the establishment of LANGTAG in 1996, significant language development structures were established. The *Provincial Language Councils* (PLCs) were established to monitor the usage and development of the languages being used in a specific province, e.g. in KwaZulu-Natal the languages concerned are Zulu, English and Afrikaans; in the Western Cape they are Xhosa, English and Afrikaans. The NLBs of each language (including the Sign Languages) have the function of standardising the orthography and development of a specific language. In the year 2000 the *National Lexicography Units* (NLUs) for all 11 official languages were founded. Their task is to compile general dictionaries for all these languages and to document and preserve material in these languages. At present the *Multiterm Website Access* (MUWA) is being developed as an interface being used to connect a multi-term database such as the *National*

Termbank with the internet, in order to make data available online.

According to the LANGTAG Report (DACST 1996:36) it had found on a fact-finding investigation that standard-setting bodies and their procedures need to be transformed. PanSALB should seek agreement between the different language interest structures, be they official or voluntary as far as corpus planning, for instance dictionary projects, is concerned (LANGTAG in DACST 1996:57).

However, PanSALB has since its establishment in 1995 not fully succeeded in achieving coordination between the different language interest structures, such as the NLBs, the PLCs and the NLUs , specifically as far as the specific function of each structure is concerned. From national conference attendances and contact sessions concerning the language scenario in South Africa it was alarming to come to the conclusion that there is much overlapping and little coordination between the mentioned structures. Hopefully the recently established *Term Coordination Section* (TCS) will coordinate the functions and management of the mentioned standard-setting structures, especially in the African languages, where there is a lack of terminology, in order to ultimately develop and modernise these languages.

If one should try to systematise the elaboration and standardisation processes in the African languages in South Africa, it would be quite difficult to give an explicit overview.   This uncertain situation can partly be attributed to the fact that many changes have taken place since 1994 in the naming and functions of official language structures. For instance, the name of the *National Terminology Services* (NTS) changed to the *National Language Service* (NLS), the *Department of Arts, Culture, Science and Technology* (DACST) has now split into two Departments, the *Department of Arts and Culture* (DAC) and the *.Department of Science and Technology* (DST). The NLBs function on the national level for each African language, e.g. the *IsiZulu National*

Language Body (ZNLB) whereas the PLCs function on the provincial level for each African language, while the previous Language Boards functioned on both these levels. Fortunately, however, Alberts (2003), a manager of *Lexicography and Terminology Development* at PanSALB succeeds in a recent publication to capture a much needed overview of language development in South Africa, sketching the collaboration between PanSALB (of the DAC) and terminology structures.

This vagueness in corpus planning activities also has to do with a misconception in this country that government must develop languages and make funds available for such purposes, whereas what should actually happen is that communities should initiate the process and then ask for official assistance (LANGTAG in DACST 1996:70-71).

2.3.7 The limitations of standardisation

Working intensively with terminology one cannot but come to the same conclusion as Sager (1990:123-128) that standardisation has its limitations with regard to what extent and under what circumstances it can be achieved. Also Rubin (1997:165) agrees that there are no universal criteria for evaluating the success of standardisation - "*How many need to use it for it to be called standardization?*" However, Sager (1990:123) reasons that standardisation can be successful on at least the levels of spelling, pronunciation, morphology and syntax.

However idealistic standardisation may be, many scholars agree that variation in language is a fact. Standardisation is wrong, according to Chimhundu (1992: 87) when it is considered a process where language is fitted into a mould with the purpose of uniformity in speech and writing. Cooper (1989:133) emphasises that variation in language will always be there as is evident in the variety of dictionaries in the same language. Finally he rightfully reasons that it is impossible to freeze the forms

of a living language, which transforms itself continuously even as it is itself transformed.

Standardisation is a continuing process and never finished (Njogu 1992:76). Therefore standardised items should be codified and regularly disseminated as widely as possible (Njogu 1992:76).

2.3.8   Problems in the standardisation of the African languages

Problematic issues in the standardisation (including elaboration) of the African languages, specifically Zulu, are discussed here to point out their relevance for this study. See also 1.3 for the research problem. However, another problem which has no immediate relevance for this study but significance for the national standardisation process as a whole, is the issue of non-standard languages. This issue is generally perceived as a threat to proper standardisation. However, it cannot be ignored since it can influence the standardisation process considerably. If addressed properly it can contribute towards a more acceptable national standardisation process in future.

2.3.8.1 Problematic issues in the standardisation of the African languages, specifically Zulu

The STANON-Report (Calteaux 1996:42) puts standardisation in the African languages thus far in a realistic sociolinguistic perspective, rightfully pointing out some thought-provoking questions:
* whose standard?
* where are the sources of this standard?
* is the evidence of this standard oral-aural based?
* is it not the case that each individual is a sort of 'standard' or standard bearer given the oral-aural background of African communities?
* has the work on 'standards' done by the now disbanded Language Boards been looked at critically enough?

As far as could be established, problems in the process of (technical) standardisation have been identified to a great extent. However, in the African languages problems have mostly been listed without offering a workable practical methodology not only to solve the problems, but also to enhance development through proper standardisation procedures and verification.

As is already known from the introductory discussion, language planning can be divided into corpus planning and status (policy) planning. This study indirectly deals with testing to what extent corpus planning, in particular elaboration and standardisation, have been successful. Unfortunately the development of terminology in the African languages, including Zulu, is mostly evidenced by basic terminology lists compiled by the African Language Boards. Furthermore, slow elaboration processes add to the impression that African languages cannot function at the same level as world languages (LANGTAG in DACST 1996:69).

Although the African languages possess the basic tools that are necessary for their development such as orthographical standards (set by the previous *Language Committees/Boards* and the present *Language Bodies/ Councils*), terminology lists, dictionaries, grammars and published literature, these tools have some serious deficiencies:

i)

There are serious inconsistencies in the application and interpretation of orthographical rules in general and in terminology.

ii)

Indigenous derivational patterns, rules for compounding or adoption of loans, also in relation to culture, are not clearly reflected in Zulu grammars. Basically the standardisation of word-formation patterns is largely lacking in the African languages, including Zulu.

iii)

There is a lack of utilising relevant written sources in a specific technical field for terminological work.

These written sources should be utilised to the fullest instead of solely relying on the intuition of terminologists and translators who were/are assigned the task of terminology development.

iv)

There is a lack of utilising relevant oral sources in a specific technical field in the practical domain for terminological work, by for instance interviewing professional workers. There is a lack of verification to determine to what extent existing standard technical terms, in comparison with oral terms, are accepted and used by the Zulu-speaking community in everyday communication.

In order to overcome the mentioned deficiencies that get in the way of effective technical standardisation and elaboration, a proposed method should be followed, based on research and the latest trends in corpus planning and language development.

The proposed method followed in this study is called a practical approach to the standardisation and elaboration of Zulu as a technical language. See 1.5 for the nature of the research. This approach does not claim to be absolute but could serve as a guideline for language planners and/or terminologists towards realistic technical elaboration and standardisation in Zulu.

2.3.8.2 The properties of non-standard languages

Unlike those of the standard languages, the properties of non-standard languages are not easily identified since these varieties are mostly not available in the written form. A non-standard language is a language which does not conform to the regulated norm and is thus not always socially acceptable (Calteaux 1996:38). Such languages thus do not have the so-called rules of standard languages which are applied in formal language contexts where they would be perceived as

appropriate.

Non-standard varieties are alive and well according to Khumalo (1995) and Zungu (1995), especially in urban areas of South Africa, such as Johannesburg and Durban. A misconception may now arise that standard languages are not spoken at all in urban townships. Usually one African language is the dominant over the others, which is the case with Zulu in Tembisa, where it forms a 'basis-language' or lingua franca (Calteaux 1996:50). *Tsotsitaal* and *Iscamtho* are considered 'bad' languages since they are stigmatised by speakers of standard Zulu and educators (Calteaux 1996:61).

According to Calteaux (1996:76) the mere description of African language varieties in the urban areas of South Africa (STANON-Project) is already an indication of standard languages becoming subject to change in order to modernise - so that speakers of these standard languages can participate in a changing world.

Since multilingualism is a very pertinent phenomenon in the urban areas of South Africa, Calteaux (1994) appropriately coined the term *Black Urban Vernacular* (BUV) for the speech varieties commonly used in black urban areas in order to facilitate communication between speakers of different African languages. Examples of such BUV-vernaculars are *Tembisa Mixed Language* and *Pretoria-Sotho*. Malimabe (1990) gives other examples of language influence in the Pretoria area where Tswana is influenced by Northern-Sotho but also how Tswana influences Zulu and Southern Ndebele.

"Language contact in the urban areas has a severe impact on the language use of pupils in schools" (Calteaux 1996:147). A pupil is sent to the school nearest to his/her home by the parents, and in that school the home language of the particular pupil is not offered (Calteaux 1996:150). Very often

Tsonga and Venda parents find it difficult to find an appropriate school where these languages are offered and as a result their children are compelled to attend a school where a Nguni or Sotho language is offered. The situation of finding an appropriate school is even worse when the child is used to using a BUV at home.

In urban schools the use of standard African languages is virtually reserved only for the mother-tongue classes (Calteaux 1996:148). Standard language use is difficult to maintain since most urban Black schools are mixed in that not everyone in the same vernacular class speaks the same home language (Calteaux 1996:148); the result being code switching by the teacher.

Attitudes towards the teaching of non-standard languages seem to be generally regarded as negative since some uneasy unanswered questions remain, such as: If these varieties can be taught, how will it be done and at which level (Thipa 1989:163)? Khumalo (1995:109) however, thinks that non-standard varieties should also gain recognition, but should then be efficient and adequate as far as terminology is concerned. The concerns are that these languages are socially stigmatised and are not formally documented in grammars and dictionaries, are unstable and thus they cannot be used productively in educational spheres.

The standard languages have enjoyed the most prestige in the townships and are regarded by the older generation as carriers of their culture while the younger generation opts for the use of English (Calteaux 1996:193). However, the non-standard varieties have expanded their range of acceptability and cannot be further ignored by language planners and policy makers. It is a fact, for instance, that varieties such as *Iscamtho* and *Tsotsitaal* are commonly used in homes and schools in the Soweto township. Furthermore, the media exert considerable influence on language. Very often the media use slang which is very quickly picked up by the youth (Mathumba 1993:211) who cannot always distinguish between correct and incorrect language use (Malimabe 1990:75).

2.4   Conclusion

Standardisation, within the broader framework of language planning, specifically corpus planning, was put into (inter)national sociolinguistic perspective. Status planning, being related to corpus planning, was discussed with regard to language policy and educational language policy with relevance to their actual implementation in the South African constitution. Eventually problems relating to status planning were summarised and possible solutions offered.

The basic problem with (educational) language policy lies in its implementation. Unilingualism (English) instead of multilingualism is favoured in the governmental, local and private sector owing to the low regard the African language speakers and officials have for the status of their languages, especially in the job market. One of the main reasons why English is preferred is because it was fostered by previous colonial educational policies in Africa. However, if the African languages are to develop, multilingualism and language equity should be encouraged through language awareness campaigns in all educational institutions. African languages should increasingly be promoted as languages of both learning and teaching.

Within the (inter)national perspective, standardisation was discussed with reference to the aspects of definition, models, norm, stages, purpose, agents, limitations and problems. The standardisation models of Haugen (1966), Garvin (1993) and Crystal (1993) formed the focus of this discussion, especially as far as the identified stages in the standardisation process are concerned. It was found that Zulu was successful as far as achieving the first stage, selection, since it is commonly known by scholars and speakers alike that the *Ntungwa* variety or the (CZD) was selected to become the acceptable standard Zulu language.

However, Zulu was also fairly successful in achieving the second and third stages. Graphisation and codification went fairly smoothly allowing for the development of some changes in the orthography such as the indication of aspiration after plosives. Yet, a prerequisite for codification is that it must be flexible (Garvin 1993). It is in this regard that the Zulu orthography should cater for changes and technical development in the language. However, flexibility also concerns non-standard varieties. In this regard Ansre (1971:697) warns that if the written standard does not become the spoken standard fast enough, native speakers of different varieties, especially when they are in the majority, will reject what they consider to be an imposed standard and will seek to reestablish a more acceptable variety.

However, the next stage, elaboration, was not fully achieved even though Zulu has a proud literary tradition with an abundance of publications in and about the language. There still is a lack of terminology in Zulu since full intellectualisation (communication in all possible technical and scientific fields) has not materialised as yet. Consequently the next stage, acceptance, has also not been fully achieved. First and foremost is the lack of proper motivation by speakers and officials to implement the language policy by, for instance, maintaining the language.

From a situational analysis of national language standardisation procedure it was found that it is quite a vague process which generally lacks coordination between the different language interest bodies such as the NLS and the NLBs. It also became clear that although the necessary linguistic tools are available in the African languages for technical elaboration and standardisation, some serious deficiencies ought to be addressed, these being the interpretation of orthographical rules, the lack of standardisation and insight in the word-formation mechanisms in relation to African culture and the lack of utilisation of relevant written and oral sources for terminological work and verification of terms. In this study an attempt is made to address the mentioned deficiencies by offering a practical approach to the standardisation and elaboration of Zulu as a technical language, the point of departure being the real language situation.

Sager (1990:123) warns about idealistic perceptions of standardisation; it should rather be seen as a communicative informative device, exploiting language but simultaneously respecting its limitations. Cooper (1989:134) in this regard also warns that standardisation for all people at all times in all contexts is an illusion. Standardisation, however, is not an aim in itself and can only be successful if it is a functional process of language development.

It is true that perceptions on standardisation have been quite negative in recent language studies. In this regard Garvin (1993:38) states that in America standardisation has been given a secondary status since much of the research concentrates on non-standard varieties. This is also partially true for South Africa with reference to the STANON-Project (1995) based on standard and non-standard varieties in South Africa. However, this thesis, like Garvin (1993:41), argues for the standardisation of the African languages since the officialisation of a language implies its use in the public and governmental sector. Furthermore, one should develop one's own standard languages and thus use them in complex discourse rather than using an ex-colonial language. It should become clear then that the approach towards standardisation makes it negative or positive. If standardisation is approached from a practical perspective, i.e. if it enhances usefulness and communication in a language, despite its limitations, it can be regarded as purposeful.

The development of the African languages is characterised by limited know-how in the theory of term development and a lack of documented terms. However, this lack of terminology in the African languages, including Zulu, can be overcome only if action is taken towards effective management of standardisation and elaboration structures and if coordination in the term-creating activities is promoted among the African languages.

CHAPTER 3

ORTHOGRAPHICAL TERMINOLOGICAL STANDARDISATION PROBLEMS REGARDING ZULU

3.1   Introduction: An overview of standard orthographical and terminological development in the African languages, in particular Zulu

To initiate the discussion on orthographical standardisation issues, it may be appropriate to cite a definition of what orthography actually is: "Spelling, including letters (upper and lower case) and diacritics" (LANGTAG in DACST 1996:220). In this chapter, with Zulu as language of exemplification, spelling and letters in the upper and lower case are dealt with. Spelling deals with a number of linguistic issues such as old versus new orthography, writing conventions, notation and phonological trends in the language. Diacritics are not dealt with here since they are not part of the Zulu orthography as they are of other African languages such as Northern Sotho and Venda, for instance.

The concept orthography in the African languages is quite different from orthography in other languages of South Africa since it deals not only with spelling but also with terminology which is listed in the form of an addendum. It is for this reason that official orthographical publications, for instance, carry names of the following nature: *IsiZulu terminology and orthography No. 4* (DET 1993).   However, these two remain different concepts but are treated together in the same publication for the sake of convenience only.

The main agents of orthographical standardisation in the African languages played a major role in establishing a written standard for these languages, although they were politically stigmatised and often regarded as having done very little by some scholars, when they were still known as the *Language Committees/Boards*. At present the main agents of standardisation in the eleven official languages are the *National Language Bodies* (NLBs). The body responsible for Zulu standardisation, for instance, is known as the *IsiZulu National Language Body* (ZNLB).

However, the NLBs have the responsibility, under the auspices of the *Pan South African Language Board* (PanSALB), to standardise the orthography of the African languages and develop these languages. It thus seems that the NLBs have inherited the traditional functions of the previous *Language Committees/Boards*. See also 2.3.6.2 for national standardisation.

The STANON Report (Calteaux 1996:42) puts standardisation in the African languages thus far in a realistic sociolinguistic perspective, rightfully pointing out some thought-provoking questions, of which one specifically reads: "Has the work on 'standards' done by the now disbanded Language Boards been looked at critically enough?" Needless to say, this question is discussed at length in this chapter, specifically as far as orthographical consistency is concerned, also with reference to technical terminology. The official orthographies referred to were prepared by the previous *Language Committees/Boards* under the auspices of Education Departments in the older dispensation, all of them containing Zulu terminology lists:

Department of Native Affairs (1957) *Zulu-Xhosa terminology and spelling No. 1*,
Department of Bantu Education (1962) *Zulu terminology and orthography No. 2*,
Department of Bantu Education (1976) *Zulu terminology and orthography No. 3* and
Department of Education and Training (1993) *IsiZulu terminology and orthography No. 4*.

It must be noted that the latest Orthography No. 4 is the one, for the sake of immediate reference and applicability, mostly referred to. It is also significant that the name *IsiZulu* is used for the first time in this type of official publication in order to indigenise the language.

To get an overview of orthographical standardisation in the African languages, the existing developmental tools such as orthographies including official terminology lists, grammars, dictionaries, published literature and technical (medical) leaflets distributed for primary health care, were investigated. Although other African languages are referred to, the language of exemplification is Zulu and the field mainly medical terminology. However, some terms of other fields are also included to offer a general perspective.

The Zulu health terms referred to or discussed in this chapter are those found or listed in the following official sources, i.e.

i)   Department of Education and Training: *IsiZulu terminology and orthography No. 4*, henceforth abbreviated as DET 1993;

ii)   *Draft List*: *Basic Health Terms* compiled by the *National Terminology Services* (*NTS*), specifically by the *Division of Biological and Agricultural Sciences* of the *Department of Arts, Culture, Science and Technology*, henceforth abbreviated as DACST 1997a);

iii)   *Draft List*: *Sex Education* compiled by the *National Terminology Services* (*NTS*), specifically by the *Division of Biological and Agricultural Sciences* of the *Department of Arts, Culture, Science and Technology,* henceforth abbreviated as DACST 1997b);

iv)   Leaflets on general health care issued by the *Department of Health* and

v)   A few commercial leaflets distributed in the primary health care sector (consult the list following the bibliography).

3.1.1   Orthographical terminological standardisation problems in Zulu

After investigating the standard orthographical and terminological Zulu sources mentioned in 3.1 above, problems in the process of orthographical standardisation, including technical terminology, were identified. These problems mainly relate to inconsistencies in the application and interpretation of orthographical rules in general and in terminology. Particular problem areas concerning the orthography reflected in the general writing style of Zulu writers/terminologists are:

1)   old versus new Zulu orthography;

2)   writing disjunctively or conjunctively;

3)   the lack of accuracy in morphological notation;

4)   capitalisation and

5)   changing linguistic trends in the language which are not reflected in the orthography.

It is self-evident that in any type of publication, proper editing is of utmost importance. It is for this reason that editing is not discussed at length here but briefly referred to. If editing is not properly conducted, it may cause problems for the (un)informed user of terminology lists. In a single official publication such as DACST (1997a), for instance, quite a number of spelling errors, some of them glaring, were identified, e.g.

umtwana ozalwe eseshonile instead of **um_n_twana ozalwe eseshonile** (stillborn baby);

ukungasebenzi kahl_a_ kwenso instead of **ukungasebenzi kahl_e_ kwenso** (kidney failure);

ukukhipha iq_h_anda instead of **ukukhipha iqanda** (ovulate);

ukuvikela komphakathi ekuncoleni instead of **ukuvikela komphakathi ekungcoleni** (sanitation) and

uklamba instead of **uk_u_lamba** (starvation), etc.

If there is a spelling mistake as in the examples above, it means that a term cannot readily be used or referred to when needed. In some of the leaflets distributed by the local councils for primary health care, it is quite shocking to find a much worse situation of language neglect; a situation which is obviously not monitored by local authorities.

3.2 The identification of orthographical terminological standardisation problems in Zulu and possible practical solutions

The whole process of identification of orthographical terminological standardisation problems in Zulu, is discussed alongside the posing of solutions. The five identified problems, being old versus new roman orthography, writing disjunctively or conjunctively, the lack of accuracy in morphological notation, capitalisation and changing linguistic trends in the language which are not reflected in the orthography, need to be addressed since they can prevent terminologists from effectively fulfilling their task as language practitioners. In this chapter the aforementioned orthographical problems/inconsistencies in Zulu are identified by questioning the logic behind some rules formulated by the ZLC/B. Problems have to be addressed by offering a workable practical methodology that would enhance development in Zulu. This methodology is called a practical orthographical standardisation approach in this chapter. See also 1.5 for the nature of research. In its application recommendations to bring about adjustments and change in the

orthography are put forward to language interest structures for discussion in order to remedy the situation. These recommendations may involve, for instance, the acknowledgment of certain phonological and morphological changes and trends in the Zulu language. After proper research into the language situation and after agreement has been reached, recommendations must be put in writing and be properly exemplified before submission to the national language authority PanSALB for final approval before implementation.

A practical orthographical standardisation approach, in which problems are identified and solutions posed, does not claim to be absolute or prescriptive but it will at least serve as an example for those concerned with orthographical standardisation and pave the way forward. Its aim is to guide language planners or terminologists, or to some extent, even train them, to be linguistically accurate and consistent, when dealing with orthographical standardisation issues.

However, even after orthographical problems have been identified and addressed through a practical approach, it will not mean much if these adjustments/changes are not made known by the ZNLB to all the concerned language interest structures in the educational and public sectors. The practical standardisation approach aims to solve the five identified orthographical problems in a logical linguistic manner.

3.2.1   Old versus new Zulu orthography

The expression 'old versus new' is used arbitrarily to indicate that the Zulu orthography has changed over the years; 'old' meaning orthography that used to be applied and 'new' meaning the current standard orthography.

In order to establish what is old and what is new orthography, the development of the Zulu orthography is traced , i.e. by investigating examples of Zulu words/phrases representative of old orthography and those representative of new orthography as they appear in the earliest and more recent grammars, dictionaries, terminologies and official orthographies of the ZLC/B (mentioned earlier in 3.1). These examples are then

compared and orthographical variations/changes are presented in the form of a survey. Since many spelling problems in Zulu can be attributed to the fact that 'old' instead of 'new' orthographical rules are still being applied, it is necessary that inconsistencies regarding 'old versus new orthography' are pointed out. The ultimate aim of this presentation is to have the old orthography replaced with the latest orthography.

In the Nguni languages graphisation was characterised by orthographical change due to language development. In some of the earliest Zulu grammars and dictionaries phonetic symbols were used alongside roman symbols , e.g. [ɓ] in **uɓaɓa** as in Doke & Vilakazi (1949) instead of **b** in **ubaba** (my/our father) for the bilabial implosive consonant. At present all languages in South Africa use the roman script and therefore this aspect need not be discussed any further. However, as the African languages developed, changes occurred in the orthography. It is interesting to note that the old orthography varies considerably from grammarian to grammarian. In the case of Zulu the most common changes are: **dhl** is replaced by **dl**, **h** is replaced by **hh** (for the voiced sound) and the lack of aspiration is replaced by the othographical inclusion of aspiration (**h**). Examples of old orthography are commonly found in older grammars and dictionaries. However, even if these old spelling rules are not applied any more, it is advisable to be aware of them since traces of the old spelling forms still occur in surnames and place names.

3.2.1.1 dhl is replaced by dl, e.g. *-dhlala* > *-dlala* (play)

Early grammarians such as Colenso (1882) and Samuelson (1925) use **dhl** instead of **dl**, in for instance, **ba<u>dhl</u>lulile** (Colenso1882:7) instead of the current **ba<u>dl</u>lulile** (they have passed). It may be worth mentioning here that in the earliest grammars no distinction was made between **<u>hl</u>** and **<u>dl</u>** and therefore Döhne (1857:29) gives the spelling -**hlala** for both 'stay' and 'play' instead of   -**hlala** and -**dlala** respectively. The use of **dhl** in the surname **Dhlomo** (instead of **Dlomo**) also evidences traces of the old spelling forms.

3.2.1.2 <u>h</u> is replaced by <u>hh</u> (for the voiced glottal fricative), e.g. *-hahama* >
-hhahhama (growl like a dog)

For many years the voiced glottal fricative, for which the phonetic symbol is [ɦ], was represented in the Zulu orthography by either **h** or **hh**. The orthographical rules governing this voiced glottal fricative were inconsistent and varied over the years, causing confusion, even today, as is evident in the examples that follow.

Samuelson (1925:14) is one of the early grammarians who distinguished between the voiceless fricative calling it a "soft H" and the voiced fricative, calling it "the baritone sound of the throat." Doke (1945:16) recommends that there should be orthographical distinction between the voiced fricative **hh** and the voiceless fricative **h**. However, he was compelled to use a single **h** for both sounds as prescribed by the University Committees affiliated to the University of Natal.

The first three Zulu orthographies *Zulu-Xhosa terminology and spelling No. 1* (1957), *Zulu terminology and orthography No. 2* (1962) and *Zulu terminology and orthography No. 3* (1976) ruled that the single **h** be used for both the voiceless and voiced fricative. Finally however, this rule was changed and, in the latest *IsiZulu terminology and orthography No. 4* (1993), the voiced glottal fricative is written as **hh**, e.g. **ihhashi** (horse), similarly to Doke & Vilakazi (1972). The voiceless fricative is to remain a single **h** as in -**hamba** (go/leave).

3.2.1.3 The lack of aspiration is replaced by the orthographical inclusion of aspiration, e. g. -*peka* > *pheka* (cook)

The lack of aspiration is still one of the most common errors evidencing old orthography today. Aspiration (**h**) was not always orthograpically indicated after the plosives **p**, **k** and **t** in the earliest publications e.g. -**peka** > -**pheka** (cook); -**kula** > -**khula** (grow) and -**tela** > -**thela** (pour). However, even if these old orthographical rules are not applied any more, it is advisable to be aware of them since traces of the old spelling forms still occur in surnames like **Kumalo** instead of **Khumalo**, for instance. The lack of aspiration is a serious error because the distinction between these plosives with or without aspiration bring

about differences in meaning, e.g. **-thetha** (scold) and **-teta** (carry on back). This is supported by Colenso (1905:ix) who exemplifies that roots which appear to be identical in spelling may differ in meaning e.g. **-tenga** (sell) and **-tenga** (waver). The former word should, according to the latest orthography be spelt with aspiration, -**thenga** (sell). The inclusion of aspiration in speech is not a problem for Zulu speakers as they add it intuitively but it could be for non-mother-tongue speakers. Needless to say then, the pronunciation of mother-tongue speakers should be taken into account for the inclusion of aspiration in a word or newly coined term.

It must be noted that **bh** in a word such as -**bheka** (look/watch) is a delayed breathy voiced speech sound and not an aspirated plosive. However, it is conveniently dealt with here because breathiness is not manifested in the written form and the same word without the **h** can bring about change in meaning, e.g. -**beka** (to put down). The **b** in the latter word is a voiced bilabial implosive and should not be confused with the breathy voiced bilabial plosive **bh.** Scholars of Zulu must be aware of the fact that Doke & Vilakazi (1949, 1972) in their monumental Zulu-English dictionary, write **bh** as **b**, e.g. -**beka** (look /watch) , in order to distinguish it from the voiced bilabial implosive for which the phonetic symbol [**ɓ**] is used**,** e.g. **-ɓeka** (put down).

In the earliest Zulu grammars aspiration is seldom indicated orthographically. There are those grammarians who do not indicate aspiration, those who are inconsistent and those who do indicate it. Roberts is one of the grammarians who does not indicate aspiration in his version of the *Zulu-English Dictionary* (1915). In Döhne's *Zulu-Kafir Dictionary* (1857) and Grout's *Grammar of the Zulu Language* (1893) no aspiration is indicated either.

Wanger (1917:3) is inconsistent as he does not always indicate aspiration in the orthography but states that it does exist in the pronunciation of words. At times Wanger (1917:2) indicates aspiration by means of a small superscript [h], e.g. **ukup[h]ila** (live). Bryant (1905:93) is also inconsistent as he does not always indicate aspiration in his examples of place names, e.g. after **t** in **umHlatuze** and **uTukela**. Yet elsewhere Bryant (1905:754, 760) writes place names with the aspiration included, e.g. **uThukela** and **umHlathuze**.

After a period during which aspiration was not indicated or used inconsistently it was eventually recognised to the extent that it was indicated consistently orthographically. Grammarians such as Samuelson (1925), Doke (1945, 1984), Malcolm (1966), Van Eeden (1956), Ziervogel, Louw & Taljaard (1981), and TaIjaard & Bosch (1988) indicate aspiration, e.g. after **t** in **uThukela** (Tugela River).

The lack and inconsistency in the orthographical indication of aspiration found in Zulu grammars is also evident in well-known official place names which are seen on a daily basis on name boards on our public roads, e.g.

Kyalami	instead of	**Khayalami**
Mapumulo	instead of	**Maphumulo**
Tokoza	instead of	**Thokoza**
Tembisa	instead of	**Thembisa**.

However, in some place names aspiration is indicated, e.g. **eZakheni**, **Zwelethu** and **Phola**.

3.2.2   Writing disjunctively or conjunctively

Another issue in the Zulu orthography is the question of writing disjunctively or conjunctively.   In the disjunctive method of writing linguistic units are generally written separately from one another, e.g. **si ya khulum a** (we are talking), while the conjunctive method requires these units to be joined. This results in morphologically complex words (Wilkes 1985:148), e.g. **siyakhuluma**.   The disjunctive method is generally used in the Sotho languages.   However, the disjunctive manner of writing also occurs in Zulu, although Zulu is known as a language to which the conjunctive writing method applies.

Although the conjunctive writing method is currently used for the Nguni languages, the disjunctive method was applied by many early Zulu grammarians. The question of writing disjunctively or conjunctively also deals to some extent with old versus new orthography but in the strictest linguistic sense actually with word division. However, strict linguistic principles concerning word identification, such as phonological and syntactical considerations, as applied by Van Wyk (1958) are not discussed here since the written word

(on face value) is dealt with in this chapter. Disjunctivism and conjunctivism do therefore, for the purposes of this chapter, refer to the writing system (orthography) and not to word division.

In older Zulu grammars and dictionaries either the disjunctive or conjunctive method of writing is generally applied. However, inconsistency in the application of the disjunctive or conjunctive method is found in some grammars. Furthermore, writers following the same approach do not always indicate word boundaries in the same manner (Wilkes 1985:149). In this regard, compare the following examples:

> **ngi ya kuya** (Döhne 1857:ixxvii)
>
> (I'm going to go) and
>
> **Lizaku puma** (Roberts 1899:67)
>
> (It - the sun - will come through).

Samuelson (1925:45) also displays inconsistency in the orthography of place names. He writes the locative prefix **kwa**- disjunctively from the noun, e.g. **Kwa Dukuza** (Stanger) and **Kwa Zulu** (Zululand). On the other hand, he also writes place names conjunctively (as they should be written) with the prefix forming part of the name, e.g. **eMngeni** and **Umzimkulu**.

The conjunctive method of writing is followed by Nguni grammarians such as Colenso (1905), Bryant (1905), Wanger (1917), Samuelson (1925) and Doke (1945). Wanger (1917:1) considers the conjunctive method of writing for Zulu as the least complicated. Bryant (1905:91) goes further by criticising those writers who want to apply European orthographical rules to Zulu. Bryant (1905:92) explains why the conjunctive approach is preferable: "...in the word *wahamba,* for instance, the particle *wa-* on its own would be meaningless and unintelligible by the Native mind." Samuelson (1925:17), in agreement with the latter two grammarians, regards this method as correct since the 'Zulu word' constitutes in the Zulu mind "... a complete thought, under one controlling accent and enunciation conveying one undivided meaning."

Doke (1945:33) follows the conjunctive approach based on the same reasons mentioned by Bryant (1905:23), namely that the non-isolatable parts be treated as formatives and not as parts of speech,

defining the Zulu word as follows:

> The complete word, ... contains one and only one main stress; but when analysed, it is
> found that the resulting formatives do not possess any main stress, cannot stand alone, and
> therefore are not complete words (Doke 1945:33).

Even today the notion of the Zulu word is quite complex and is normally a phrase/sentence in other languages, e.g. **Ngisazokubona** (I shall still see you). Doke's finding (as set out) above:

> ... had a profound effect on the writing system of especially the Nguni languages, where its
> acceptance led to the final adoption of conjunctivism (at the expense of disjunctivism) as
> the sole method of word division (Wilkes 1985:150).

Doke's example of conjunctivism is also followed by later grammarians such as Van Eeden (1956), Ziervogel et al. (1981), TaIjaard & Bosch (1988) and Poulos & Msimang (1998).

Yet, even today conjunctivism is not followed throughout and deviations do occur, even in an official publication such as the Draft list: *Basic Health Terms* (DACST:1997a), e.g. **iredi bladi seli** (red blood cell) and **iyelo fiva** (yellow fever). It is quite obvious that the disjunctive manner of writing English was carried over to the almost similar Zulu words. Since there is no pronunciation problem these words could easily be written conjunctively as **iredibladiseli** and **iyelofiva**. Terminologists and lexicographers should constantly apply the conjunctive principle of writing in line with the orthographical rules of Zulu so as not to confuse the users of term lists and dictionaries.

The topics that merit discussion concerning the disjunctive or conjunctive manner of writing are the demonstrative pronoun and the use of the apostrophe and the hyphen.

3.2.2.1 The demonstrative pronoun

One of the most argued orthographic issues in Zulu deals with the manner of writing the demonstrative pronoun - either disjunctively or conjunctively. This problem came about because the orthographic rules formulated by the ZLC/B concerning the demonstrative pronoun varied from conjunctive to disjunctive, creating confusion to such an extent that uncertainty still persists today. Compare, for instance, how confusing and inconsistent the following ruling in *Zulu-Xhosa terminology and spelling No. 1*. (1957:4) is:

> (1) Demonstrative pronouns may be written either conjunctively or disjunctively. In school books, especially for the primary school, they will be printed disjunctively, e.g. *lezi zinkomo*, *lowaya muzi*.

The problem with the above rule is that it leaves room for inconsistency in its application (an option for primary school pupils to write the demonstrative disjunctively). Orthographical rules need to be applied consistently by all members of a speech community, regardless of social factors such as school grading in order to maintain a standard.

Yet, the latter rule was later changed in *Zulu terminology and orthography No. 3*. (1976:15):

> (a) A demonstrative pronoun preceding a noun should be written conjunctively with the noun, e.g. -
> *Lezizinkomo*, *lomfana*,
> but if the noun precedes the demonstrative pronoun, it should be written disjunctively, e.g. -
> *izinkomo lezi*, *umfana lo*.

It is because of the first part of the rule above that the conjunctive manner of writing is found in earlier grammars and literary works and even in the Zulu Bible, e.g.

lesisitsha (this plate),   **labobantu** (those people) (Doke 1984:92).

Lelisu ngilenze ukuze **lencwadi** ifundeke kalula futhi ibemnandi nasendlebeni (Dhlomo

1961:i - *own emphasis*)

(I applied this view to make this book easy to read and pleasant on the ear).

UIsaka watshala ku**lelozwe**, wathola ngawona **lowomnyaka** okuphindwé kayikhulu,

ngokuba uJehova wambusisa (UGenesise 26 verse 12 - *own emphasis*)

(Isaac planted crops in that land and the same year reaped a hundredfold, because the

Lord blessed him).

The second part of the latter rule in Orthography No. 3 never really caused any problem and the

disjunctive manner of writing the demonstrative following the noun is commonly found even in earlier literary

works:

UMkabayi **lona** kwakuyinkosikazi emangalisayo ngezindlela eziningi (Dhlomo 1961:1 -

own emphasis).

(This Mkabayi was a fascinating lady in many ways).

Eventually the rule of writing the demonstrative, preceding or following the noun, was simplified in *IsiZulu*

terminology and orthography No. 4 (DET1993:xii):

Demonstratives

All demonstratives are written as separate words, e. g.

lo mfana, lelo tshe, labaya bantu, etc.

Poulos and Msimang (1998:14) supply a logical reason for the latter ruling stating that the demonstrative

should be written separately from the noun since it is regarded as a separate word category, commonly

known as a demonstrative <u>pronoun</u> (*own emphasis*). It is because of this simplified rule above that the

manner of writing the demonstrative disjunctively from the noun, is adhered to (although not consistently)

in recent grammars, dictionaries, publications and literary works, e.g.

> **lo mfula** (this river) and **le nto** (this thing) (Poulos & Msimang 1998:14);
>
> Asiside **lesi sikhathi**, Bheki-Bheki (Buthelezi 1993:38 - *own emphasis*)
>
> (It is not long this time, Bheki-Bheki);
>
> Ngize ngisuke ku**leli litshe** ngizwa nginokukhathala kodwa kukhona injabulo engingeyichaze emoyeni wami (Ntuli 1994:56 - *own emphasis*)
>
> (I then left this stone and felt exhausted but my soul was filled with joy which I could not explain).

Seemingly, the writing of the demonstrative is not to be discussed as an orthographical problem since it is not a term in itself, without the noun which it qualifies and therefore does not concern the terminologist. However, it does concern him/her since terms are not merely listed in draft lists, glossaries or dictionaries but are eventually used in written communication such as leaflets, for instance. In primary health care, leaflets are used to convey information to the public and as such the demonstrative should be written correctly in its syntactic context. In the mentioned leaflets however, writing errors pertaining to the demonstrative do occur, to mention but a few:

> Uma uzimiselele ukukhokhela **lesevisi**, ungaxhumana neklininki ye-Marie Stopes, ku**lenamba** elandelayo: (011) 337-8020 ukuthola izeluleko (*Ukukhipha isisu* developed by the *Reproductive Health Materials Package by PHRU, SPH AND AMREP - own emphasis*)
>
> (If you so wish to pay for this service you can contact the Marie Stopes clinic at this following number to find advice: (011) 3378020).
>
> **Leligciwane** lengculaza landa kuphela ngokuya ocansini (*Uthando…ukuvikela umndeni wakho kungculaza* developed by the *Department of National Health and Population Development - own emphasis*)
>
> (This virus can only be spread by sexual intercourse).

In most of the very recent leaflets such as the two quoted from above, the demonstrative is commonly written conjunctively with the noun it precedes in sentences, e.g. **lesevisi**, **kulenamba** and **leligciwane** almost as if the latest orthographical rule, to write it disjunctively, e.g. as **le sevisi**, **kule namba** and **leli gciwane** respectively, never came into effect. It must be stressed that according to the latest orthographic rule (DET 1993) the demonstrative must be written disjunctively.

3.2.2.2 The apostrophe

The use of the apostrophe can be regarded as part of the conjunctive manner of writing since it indicates elision of vowels and not separation of units within a lexical item. The apostrophe is not a prominent orthographic feature in Zulu and is only used in exceptional cases. According to the latest Zulu orthography, No. 4 *(*DET 1993:viii) the apostrophe can only be used to indicate elision, mainly as it occurs in dialogue and poetry, e.g. **Angesabi 'nja** (I do not fear any dog). In the latter sentence the **i** in the word **inja** (dog) is, for instance, elided. Yet, in ordinary writing this rule is not strictly adhered to and generally elision is seldom indicated, but not perceived as incorrect, e.g.**Uthanda kudla kuni**(What food do you like?); here the elision of **u** in **ukudla** is not indicated and **Asiboni muntu** (We don't see anyone); here the elision of **u** in **umuntu** is not indicated.

In a few isolated examples found in official sources such as DACST (1997a) the apostrophe is not used in accordance with Zulu orthographic rules. In this source the use of the apostrophe is not altogether clear, e.g. **iyul'sa** (ulcer), **ifaren'si** (pharynx) and **iyul'na** (ulna). Vowels need not be added (which are supposed to be left out where the apostrophe appears) to form the vowel-consonant-vowel pattern since this pattern is not strictly adhered to any more (also see 3.2.5 for changing linguistic trends). These words could easily be written without the apostrophe as **iyulsa**, **ifarensi** and **iyulna** since there is no apparent pronunciation problem. Sound combinations like **ls**, **ns** and **ln** in these respective words **iyulsa**, **ifarensi** and **iyulna** have become acceptable in the Zulu speech community. It can thus be concluded that since the rule to use the apostrophe to indicate elision is not adopted consistently, its use should be made optional by the ZNLB.

It is interesting to note that Colenso (1882:23, 24) uses the apostrophe after the prefix **kwa**- and **ka**- in a manner similar to the indication of the possessive in English, e.g. **kwa'Dukuza**. In Zulu the prefixes **kwa**- and **ka**- are possessive concords morphologically, which can also denote locality (place) semantically. The name **kwa'Dukuza** literally means 'at the homestead/place of Dukuza'. This homestead was one of Shaka's where Stanger is now situated. Wanger (1917:213) uses the apostrophe in the same manner, also to indicate possession, e.g. **uthando luka'Nkulunkulu** (the love of God). However, the apostrophe in Zulu as in **kwa' Dukuza** can by no means be used to indicate any type of possession as it does in English. Thus the spelling of **kwaDukuza** is correct and in accordance with the present Zulu orthography.

Furthermore terminologists and lexicographers should be consistent in the application of the orthographical rules of Zulu and use the apostrophe sparingly so as not to confuse the users of term lists and dictionaries.

3.2.2.3 The hyphen

The use of the hyphen as punctuation mark can be regarded as part of the conjunctive writing system, since it prevents lexical items from being entirely separated. The hyphen is not used that often except in earlier writings of official place names employing it after the prefix **kwa**-, e.g. **Kwa-Magwaza** and **Kwa-Mbonambi** (*National Place Names Committee Gazetteer*: 1951). This practice is not followed any more except in the new spelling of the province of **KwaZulu-Natal**. Perhaps the spelling of the latter name should be clarified and revisited in the next Zulu orthography publication.

According to the latest orthography (DET 1993:viii), Rule 8 dealing with the hyphen states that it will mainly be used in the following instances:

- when a numeral is preceded by an (inflected) prefix to join concords to numerals, e.g. **zingu-10** (they are ten);
- to separate two vowels combining with a glottal stop between them, e.g. **ama-apula** (apples):
- in enclitics, e.g. **woza-ke** (come then) and

•   for practical reasons in lengthy compound words, e.g. **ikhaboni-dayoksayidi** (carbon dioxide - *own example*).

The phrase 'for practical reasons' is very vague and one cannot but question the logic behind the use of the hyphen, even more so because no examples are included in the rule to explain it. Furthermore, on what grounds is a compound considered lengthy enough to merit the use of a hyphen? Even in some of the terms quoted from official sources one cannot always understand why a hyphen is used, e.g. **umthola-mpilo** (clinic), **ukuhlola-ngculazi** (HIV antibody test) or not used, e.g. **isidakamizwa** (drug), **isibulala mabhakithiriya** (antibiotic) (DACST:1997b). More questions could be asked such as: Why is a word like **umthola-mpilo** (clinic) hyphenated and **isidakamizwa** (drug) not? Should the latter word not be hyphenated as follows: **isidaka-mizwa** (drug) since it has exactly the same length as **umthola-mpilo** (clinic)? Why can **umthola-mpilo** (clinic) not be written without the hyphen as **umtholampilo** (clinic) as it is used in many medical leaflets? Furthermore, why is **isibulala mabhakithiriya** (antibiotic) not hyphenated in the same manner as **ukuhlola-ngculazi** (HIV antibody test) but written disjunctively as two separate words? In the same manner one can argue the use of the hyphen in other technical terms from the latest Zulu orthography (DET 1993), e.g. **ibizo-ningi** (collective noun), **ungwaqa-mfuthwa** (fricative consonant) and **inzalo-mpinda** (compound interest).

One can keep on arguing in this fashion only to conclude that rules concerning the hyphenation of words or terms should be reformulated by the present ZNLB since it clearly creates confusion amongst terminologists and users of terminology lists alike. Perhaps the rule to include the hyphen (in lengthy compounds) should be made <u>optional</u> or at least an example should be included to explain the rule properly. Note that an own example, **ikhaboni-dayoksayidi** (carbon dioxide) is included to explain the rule above.

In all the consulted Zulu grammars the manner of writing compounds is not specifically treated as an issue as the rule of conjunctivism applies throughout; neither is the insertion of a hyphen mentioned as an alternative manner of dividing lengthy compounds. On the contrary, the examples found in Poulos &

Msimang (1998:88) are quite lengthy compounds which are simply written conjunctively, e.g. **ukhandalimtshelokwakhe** (wayward person) and **umlomungathethimanga** (king, i.e. he whose word is final).

It has been observed that the hyphen is increasingly being used in foreign or scientific loan terms that have been incorporated into the Zulu language. In many such loan terms from DACST (1997a) the rule to use the hyphen for the separation of vowels is no problem since the hyphen is correctly inserted according to the latest orthographic rules of Zulu, e.g. **i-adenoyidi** (adenoids), **i-insulin** (insulin), **i-okisijini** (oxygen), etc.

However, where the separation of vowels is not an issue, a hyphen is used to separate the Zulu noun class prefix from the foreign Zulu loan terms, e.g. in **i-bhakhithiriya khonjakhithiva** (bacterial conjunctivitis); **i-homofiliya** (haemophilia), **i-x-reyi** (x-rays) quoted from DACST (1997a); **i-gypsum** (gypsum), **i-vise versa** (vice versa), **i-domino** (domino), **i-foramen magnum** (foramen magnum) quoted from the *Zulu orthography* No 4 (DET1993); **i-sterilisation** (sterilisation), **i-vasectomy** (vasectomy) quoted from a leaflet *Ukukhipha isisu* developed by the *Reproductive Health Materials Package by PHRU, SPH AND AMREP* and **i-semen** (semen) quoted from a leaflet *Ukuvalwa kwenzalo kowesilisa* developed by the *Reproductive Health Materials Package by PHRU, SPH AND AMREP*.

This latter rule can initiate the formulation of a rule as regards hyphenation in technical terminology since this aspect is not treated in the latest Zulu orthography although quite a number of examples occur in this very same publication. It can be observed that in the example of **i-x-reyi** (x-rays), two hyphens are used since the hyphen already appears in the English loan word. Yet, the hyphenation of technical loans will undoubtedly appear more and more in the current globalisation context and will therefore have to be treated orthographically by the present ZNLB and the Boards for the other African languages for that matter, even if only mentioned briefly.

Another aspect in the use of the hyphen in Zulu has to do with its use in foreign or scientific acronyms that

have been incorporated into the Zulu language. It must be noted that acronyms used here are mostly international abbreviations in capital letters. See also 4.4.7 on word- formation patterns, specifically abbreviations. In examples found in DACST (1997a) the hyphen is inserted between the initial vowel prefix and the acronym, e.g. **i-HIV** (HIV- human immunodeficiency virus), **i-AZT** (zidovudine - drug used for the treatment of AIDS), **i-GP** (GP - general practitioner). This insertion is also found in an example quoted from DACST (1997b), e.g. **i-ELISA** (enzyme-linked immuno sorbent assay - blood test for AIDS). The Zulu acronym **i-AIDS** (AIDS) is commonly used, simply because it is shorter and catches on, alongside the Zulu coinage **ingculazi** (AIDS) in most of the medical pamphlets distributed by the *Department of Health.* The same applies to the acronym **i-IUD** (IUD -intrauterine device) which is widely used in leaflets distributed for primary health care.

However, up to now the ZNLB has not yet formulated an orthographical rule concerning this modern trend of hyphenation in the formation of acronyms. Needless to say, it should be addressed as soon as possible, since acronyms occur more and more worldwide. Such a rule can be brief with proper exemplification and could, for instance, read as follows: For acronyms the hyphen is inserted between the initial vowel prefix and the acronym, e.g. **i-HIV** (HIV).

3.2.3   The lack of accuracy in morphological notation

On the one hand, the lack of accuracy in morphological notation is an issue that specifically concerns terminologists, since the notation of terms is actually the concrete evidence of a long process of consultation and research. On the other hand, notation also concerns the users of such terminology lists who do not necessarily have sufficient linguistic knowledge of the language concerned.

Notation in the context of this chapter specifically deals with the way in which entries or lemmas (basic forms) are listed in terminology lists. Most inconsistencies in notation may be attributed to terminologists having insufficient linguistic insight into the morphological structure of the Zulu language. Perhaps it can be said that some language practitioners are not properly trained or qualified to fulfil their task. Another

alarming factor is that some compilers of these terminology lists assume that the users are supposed to be informed Zulu linguists, which is not necessarily the case. Several official terminology (draft) lists were investigated and the lack of accuracy and consistency in morphological notation can be exemplified as follows in three different publications:

i) *IsiZulu orthography No 4* (DET 1993)

In i) nouns, being considered full words which can be used as such, are correctly listed, mostly in the singular, e.g. **ikhensa/isimila/umhlaza** (cancer), **uqhuqho** (malaria) and **ithumba** (abscess). The traditional manner of notation followed for nouns in Zulu dictionaries such as Doke & Vilakazi (1972), namely to prefix a hyphen to noun stems (leaving the prefix since it can vary for singular and plural), e.g. -**thumba** (abscess) however, is not practical for a terminology list since users of such lists usually need immediate access to full terms. A stem mostly consists of a morpheme or morphemes which is/are not considered a complete word. In accurate morphological notation a hyphen is written in front of such a stem, be it verbal, qualificative or nominal. However, this accurate practice is not generally followed in i) where the hyphen precedes adjective and relative stems (used as qualificatives), e.g. -**hlanzekile** (sanitary) but not verbal stems, e.g. **gaya** (digest) and **goma** (immunise). The reasoning behind this could be that the hyphen is used to distinguish between stem types, i.e. *verbs* such as **gaya** (without the hyphen) and *qualificatives* such as **-hlanzekile** (with a hyphen). Since Zulu dictionaries such as Doke & Vilakazi (1972) also notate verb stems such as **jova** (inject) and qualificative stems such as –**ncane** (small) in the same manner, this practice followed in (i) is acceptable.

ii) The *Draft list*: *Basic Health Terms* (DACST 1997a) and

iii) *Draft list*: *Sex Education* (DACST 1997b)

In ii) and iii) the hyphen is also used as in publication i) above, not preceding verb stems but preceding qualificative stems as explained above.

However, in these two publications the inconsistent use of the prefix in morphological notation occurs. This

inconsistency concerns the notation of the verb, specifically verb stems, infinitive verbs and deverbative nouns. The use of the prefix **uku-** in front of only some verb stems such as **ukuguga** (age) and **ukuncelisa** (**ibele**) (breast-feed), and not in front of others such as the verbs **bandisha** (bandage) and **nquma** (amputate) quoted from DACST (1997a) creates confusion. The same applies to examples quoted from DACST (1997b) where **ncelisa** in **ncelisa** (**ibele**) (breast-feed) is listed as a stem but other verbs are prefixed by the class 15 class prefix **uku-**, e.g. **ukuvikela** (protect) and **ukubuyisa** (vomit). The discrepancy in the form of notation lies in the fact that the prefix **uku-** can precede any Zulu verb stem, forming an infinitive implying the meaning of 'to', e.g. **ukuguga** (<u>to</u> age), etc. What complicates the matter even further is that in other contexts the adding of the prefix **uku-** can bring about change in the word category and eventually the meaning, e.g.

> **-ncelisa** (breast-feed) can form an infinitive verb **ukuncelisa** (to breast-feed) or a deverbative noun **ukuncelisa** (breast-feeding);
> **-khulelwa** (fall pregnant - containing a passive verbal extension **-w**-) can form an infinitive verb, **ukukhulelwa** (to fall pregnant) or a deverbative noun **ukukhulelwa** (pregnancy).

For the sake of logical morphological consistency, to prevent confusion and to guide the uninformed user, it would be advisable to additionally notate such similar words belonging to different word categories with <u>v</u> for <u>verb</u> and <u>n</u> for <u>noun</u>, e.g.

> **ukuncelisa** (to breast-feed) <u>v</u> and **ukuncelisa** (breast-feeding) <u>n</u>.

Unfortunately, however, this linguistic practice of marking word categories is not consistently followed throughout these publications.

From the foregoing discussion it has become clear that the simplest manner of morphologically notating the verb is that it should not be preceded by the **uku-** (class 15) prefix, e. g. **ncelisa** (breast- feed), so as not to confuse the user as to whether **ukuncelisa**, for instance, is an infinitive verb (to breast-feed) or a noun (breast-feeding). The compilers of terminology lists under the auspices of the ZNLB should thus be advised to follow proven traditional methods of listing instead of varying the norm.

3.2.4   Capitalisation

When dealing with capitalisation, the traditional issues concerning place names, deity, titles, etc. are discussed. Thus far, other aspects concerning capitalisation, especially as far as the naming of technical terms is concerned, have not been discussed by the ZNLB. In order to address this problem, examples of the capitalisation of technical terms in the latest orthography (DET:1993:x), two official term lists and leaflets distributed for primary health care were consulted.   Capitalisation as far as (medical) terminology is concerned in Zulu can roughly be based on capitalisation in English, except that the linguistic structure of Zulu must be adhered to:

i) The first letter (after the initial vowel prefix) in the name of a specific commercial product or name of an international technical (medical) term or illness is capitalised. This is quite an easy rule since the first letter of such a name would be capitalised in English too, e.g.

> **iP̲urity** (Purity - health cereal) quoted from a 'Purity' commercial leaflet; **ibhengela leMedical Alert** (Medical Alert bracelet), **ukubakhona kwefekitha yeRhesus** (Rhesus positive) quoted from DACST (1997a) and **umdlavuza wesikhumba kaKaposi** (Kaposi sarcoma - type of skin cancer which often affects AIDS patients) quoted from DACST (1997b).

Since the ZNLB has not yet formulated any rule for the capitalisation of the name of an international technical term, the rule formulated above in i) could initiate the formulation of such a rule.

ii) In the case of acronyms, which are usually written in capital letters, the initial vowel prefix preceding the acronym is in the lower case, e.g. **i T.B.** (TB. - tuberculosis), **i-GP** (GP - general practitioner) and **i-HIV** (HIV - human immunodeficiency virus) quoted from DACST (1997a).
Since the ZNLB has not yet formulated any rules for the capitalisation of acronyms in international technical

terminology, the rule suggested in ii) above could initiate the formulation of such a rule.

iii) In loans that concern place names the first letter after the initial vowel prefix is usually capitalised. The initial vowel may also be followed by a hyphen before the capital, e.g. **i-Japan** (Japan), **iCape Province** (Cape Province) quoted from DET (1993). In the latter two examples above it is quite obvious that the hyphen is inconsistently used. If a hyphen is used after the initial vowel prefix in **i-Japan** (Japan) because it is a loan, then certainly the hyphen should also have been used after **i-** in **iCape Province** (Cape Province) which is also a loan. Perhaps then a rule should be formulated to make the hyphen optional. Also see the discussion on the hyphen earlier in 3.2.2.3.

3.2.5 Changing linguistic trends in the language which are not reflected in the orthography

Hlongwane (1995) and Mthembu (1996) warn that the present orthography of the African languages does not reflect current trends in the phonology (and consequently the vocabulary) of these languages. Sounds which were previously uncommon may now have become quite common. In order to substantiate the above-mentioned change in the African languages, it is relevant to mention a few phonological and morphological changes, since these are the very ones that will have to be reflected in the orthography and that will eventually also have an influence on the lexicon of a language.

3.2.5.1 Phonological trends

In Zulu most syllables are open, i.e. they end on vowels, e.g. **i/ ki / la / si** for **ikilasi** (classroom). However, the modern version of **ikilasi** (classroom), namely **iklasi**, rather conforms to the CCV-pattern in the second syllable, e.g. **i** / **kla** /**si**. Other examples of similar modern adoptives quoted from Hlongwane (1995:61) are **istradi** (street) instead of **isitaladi**, and **idrobha** (town, Afrikaans 'dorp') instead of **idolobha**. It is interesting to note that the **r** is increasingly being used in modern loans where it was previously regarded as a non-Zulu sound and mostly substituted by **l**, e.g. **imajarini** (margarine) and **igaraji** (garage) as opposed to the earlier **imajalini** and **igalaji** respectively.

Koopman (1992:110) cites examples of Zulu adoptives used by urban modern educated Zulus, where even the double **rr** is used , e.g. **i-rrisidi** (receipt) and **i-lorri** (lorry). After having consulted modern data, Koopman (1994:163) further confirms that the **r** "...has now become a 'regular' phoneme in the language, albeit restricted to lexical items assimilated from English and Afrikaans". Nkabinde (1968:20) admits that although the phonetic structure of Zulu is somewhat disturbed by the adoption of speech sounds such as **dr**, **gr** and **st**, these sounds bear the evidence of continual linguistic change. In certain linguistic terms Kumalo (1987a, b) uses the sound combination **ksi,** which is not that common in standard Zulu, e.g. **isimenthiksi** (semantics) and **isinthe ksi** (syntax), also employing this CCV-pattern. Kumalo seemingly disregards the phonological rules of Zulu by ignoring the rule that an aspirated plosive such as **th** cannot be preceded by the nasal **n** to form the sound combination **nth**. Since the nasal causes aspirated plosives to assimilate to ejective plosives, the grammatical terms **isimenthiksi** (semantics) and **isintheksi** (syntax) should thus read **isimentiksi** (semantics) and **isinteksi** (syntax) respectively, reflecting the sound combination **nt** (without the **h**). However, if the **n** is regarded as being part of a syllable the former two terms (with aspiration - **h**) can be regarded as correct.

Koopman (1992:111) also exemplifies other types of somewhat 'foreign' consonant clusters in Zulu, e.g. **thr** in **i-bhe thri** (battery) and **khr** in **i-khrimu** (cream). Phonologically, consonant clusters, such as the ones underlined below, which were previously unacceptable, have become quite common, e.g. **inkontraka** (contract), **ikne btange** (pliers from Afrikaans 'knyptang') and **isipringi** (spring). It is already obvious that standard Zulu syllabic structures are not adhered to in these loans which are used by those who are constantly exposed to Western culture (Zungu1995:156). The syllables are becoming closed, since they do not necessarily end in vowels, e.g. **ikneb-tange**. Thipa (1992:81) observes the same trend in the occurrence of foreign consonant clusters such as **pr** and **fr** in Xhosa. He (op. cit.) ascribes this occurrence to Xhosa's exposure to "...Western cultural influences and experiences", a viewpoint which can very well also be made applicable to Zulu.

Nevertheless, it is important that such phonological changes or trends as are discussed above should be evaluated for the sake of determining how commonly they occur in the spoken form. If uncommon sound

combinations like **khr** and **ksi** are continually being used, for instance, they should be proposed as possible changes/inclusions in the orthography by all language interest structures, the ZNLB being the main initiative authority.   In this manner, after consensus has been reached, such changes may eventually be included in the orthography of the language so as to represent the living language. Hlongwane (1995:60) supports this type of action that can lead to change, reasoning that what is non-standard today may become standard tomorrow.

3.2.5.2 Morphological trends

Morphological changes in the Zulu language also occur of which a few are mentioned here. Hlongwane (1995:61-62) notices that some adopted verbs employ the sound combination
-isha, previously quite uncommon in Zulu, for no apparent reason, e.g. **-tadisha** (study) and
-filisha (fill out a form).   Koopman (1994 :244) however, sees this **-isha** (or **-ish-**) as a verbal suffix specifically used for adoptive verbs.

Another example of morphological change deals with an irregular derivational process whereby verbs are derived from adopted nouns, e.g. **-farisa** (act hypocritically) derived from **umfarisi** (Pharisee/hypocrite) (Koopman1994:238).

It is important that such morphological changes or trends as are discussed above should be evaluated for the sake of determining how commonly they are used by the speech community. If uncommon verbal suffixes like **-ish-** are continually being used, for instance, they should be proposed as possible changes in the orthography. Eventually, after consensus has been reached and approval obtained from the ZNLB, such changes may become part of the orthography of the language.

Although syntax and semantics do not deal with orthography (spelling) in a strict linguistic sense, they indirectly have a role to play in the course of time and therefore need to be mentioned here. Syntactic changes are for instance evident in code switching in a sentence like **Ngiyahamba** *today* (I'm leaving

today) and in the use of English conjunctions such as 'but', 'then' and 'because' as mere fillers in Zulu sentences (Khumalo 1995:63:100). Semantic change, for instance, is evident in words which convey semantic shift, e.g. **isitezi** (stair) has now acquired the meaning of 'double storey house' (Koopman 1994:136). See also Chapter 4 for word-formation patterns.

3.3 A practical approach to solving orthographical terminological standardisation problems in Zulu

After orthographical standardisation in Zulu was investigated by consulting the existing developmental tools such as orthographies compiled by the previous *Language Committee/ Board,* official terminology lists, grammars, dictionaries, published literature and technical (medical) leaflets distributed for primary health care, it became apparent that there are still some deficiencies as far as orthographical development is concerned.

These orthographical deficiencies mainly relate to the following problem areas:

1)   old versus new Zulu orthography;

2)   writing disjunctively or conjunctively;

3)   lack of accuracy in morphological notation;

4)   capitalisation and

5)   changing linguistic trends in the language which are not reflected in the orthography.

In this chapter the above-mentioned orthographical issues and inconsistencies in Zulu have generally been identified, by questioning the logic behind some rules formulated by the ZLC/B. Problems have not only to be identified but also to be addressed by a workable methodology, called a practical standardisation approach, that would enhance development in the Zulu language (see 1.5 for the nature of the research). In its application, adjustments and change in the orthography of Zulu have to be discussed by the ZNLB. It is imperative that orthographical inconsistencies be addressed since they can prevent terminologists

from effectively fulfilling their task as language practitioners. After proper research into the language situation and after wide consultation, recommendations must preferably be put in writing and submitted to the national language authority PanSALB for final approval before implementation.

Recommendations may involve, for instance, the motivation of improved formulation of existing orthographical rules, perhaps also more explicit exemplification in juxtaposition of these rules and the acknowledgment of certain phonological and morphological trends in the language.

This practical standardisation approach, in which problems are identified and addressed, does not claim to be absolute. However, its aim is to guide or train language practitioners, to some extent at least, to be linguistically accurate and consistent in formulation and application. In this manner an example is set to pave the way forward towards continual effective standardisation.

In order to solve the five identified problems (1-5 mentioned in 3.3 above) concerning Zulu (technical) orthographical standardisation, the following recommendations, based on a practical standardisation approach, are posed in the following sections.

3.3.1   Old versus new Zulu orthography

Old orthography used in earlier Zulu publications should consistently be replaced by new orthography since there are still traces of old orthography evident in, for instance, Zulu surnames. The following are the major changes in the Zulu orthography:

dhl in **Dhlomo** should be replaced by **dl** in **Dlomo** for instance;

h in -**hahama** should be replaced by   **hh** in **-hhahhama** (growl), for instance, in the case of the voiced glottal fricative sound;

aspiration (**h**) should be added to the plosives **k**, **p** and **t** where applicable after a period during which aspiration was omitted or used inconsistently in earlier Zulu publications. Its indication is important since

the lack of aspiration can bring about change in meaning, e.g. **-teta** (carry on back) and **-thetha** (reprimand). Traces of old orthography that are still found in well-known place names should be changed to new orthography, e.g. **Tokoza** should become **Thokoza** (including aspiration).

3.3.2   Writing disjunctively or conjunctively

Zulu gradually developed from a language with a disjunctive writing system to a language with a conjunctive writing system - although both systems were arbitrarily used by grammarians.   Generally speaking, in the disjunctive system morphemes are written separately to form a word/phrase, e.g. **si ya gul a** (we are ill), whereas in the conjunctive method morphemes are combined to form a word/phrase, e.g. **siyagula** (we are ill). However, it is now a fact that Zulu employs the conjunctive writing system.

For a long period of time the previous ZLC/B constantly varied the rules for the manner of writing the demonstrative from conjunctive to disjunctive and back again so that it remains problematic, even today.
   It is common to find the demonstrative written conjunctively as part of the following noun, e.g. **lesisikhwama** (this bag - one word) instead of the orthographically correct **lesi sikhwama** (two words). However, the latest Zulu orthography simplifies the manner of writing the demonstratives to a disjunctive one, whether it precedes or follows the noun it qualifies, e.g. **lo mfana** (this boy) or **umfana lona** (this boy).

The use of the apostrophe in words is part of the conjunctive manner of writing since it is used to indicate elision and not separation, e.g. **angesabi 'nja** - **angesabi inja** (I do not fear any dog) where the **i** in **inja** is elided. Yet, in ordinary writing it is quite common and perceived as correct not to use the apostrophe to indicate elision, e.g. **angesabi nja** (I do not fear any dog).   Furthermore, the use of the apostrophe to indicate elision of vowels is not altogether clear in examples such as **iyul'sa** (ulcer) and **iyul'na** (ulna) in DACST (1997a).   These words should actually be written without the apostrophe as **iyulsa** and **iyulna** since the rule to make use of the apostrophe to indicate elision is not strictly adhered to. Thus the use of the apostrophe should be made optional.

The use of the hyphen also deals with the conjunctive manner of writing since linguistic units are not entirely separated (i.e. not written disjunctively). In earlier writings the hyphen occurs in the writing of official place names employing the prefix **kwa**-, e.g. **Kwa-Mbonambi**. This type of practice is now only found in the new spelling of the province of **KwaZulu-Natal**.

However, according to the latest Zulu orthography, the hyphen will mainly be used to join concords to numerals, to separate two vowels in juxtaposition and in enclitics. Yet, it is also used for practical reasons in lengthy compound words, e.g **ikhaboni-dayoksayidi** (carbon dioxide - *own example*). The logic behind the phrase 'for practical reasons' can be questioned since it is a very vague statement, even more so because examples are not included in the rule to explain on what grounds a compound can be considered lengthy enough to merit the use of a hyphen. Poulos and Msimang (1998:88) do not seem to have a problem with quite lengthy compounds which are simply written conjunctively, e.g. **ukhandalimtshelokwakhe** (wayward person). Clearly the rule concerning the hyphenation of lengthy words should be revisited (considering conjunctivism as a given and including proper exemplification) by the ZNLB since it creates confusion.

In the latest Zulu orthography hyphenation in foreign Zulu loan terms is not treated although quite a number of examples occur in this very same publication. This aspect of hyphenation will undoubtedly appear more and more in the current globalisation context and can therefore not be ignored by the ZNLB. After the investigation of technical loans a possible rule can be formulated as follows: The hyphen is inserted to separate the vowel class prefix from the foreign loan word (loan), e.g. **i-homofiliya** (haemophilia).

Another orthographical aspect concerning the hyphen in Zulu has to do with its use in technical /scientific acronyms that have been incorporated into the Zulu language. Acronyms are becoming more popular worldwide since they are international, shorter and catch on. For a start, the following rule with proper exemplification could, for instance, be formulated for hyphenation in acronyms by the ZNLB: The hyphen is inserted between the initial vowel prefix and the capitalised acronym, e.g. **i-HIV** (HIV - human

immunodeficiency virus).

The use of the apostrophe and the hyphen in words can be regarded as part of the conjunctive writing system, since they prevent lexical items or sounds from being entirely separated. However, in order to maintain consistent conjunctivism, the use of the hyphen and apostrophe should be minimised.

3.3.3   The lack of accuracy in morphological notation

Notation specifically deals with the way in which lemmas are listed in terminology lists.   Most inconsistencies in notation may be attributed to language practitioners having insufficient linguistic insight into the language.   Specific official terminology lists were investigated and the main notational morphological inaccuracy that could be found is the inconsistent use of the prefix **uku-** (class15) in the notation of verbs.

The class 15 prefix **uku-** preceding only a few verb stems such as **ukuguga** (age) and **ukuncelisa** (**ibele**) (breast-feed), and not most other verb stems such as **nquma** (amputate) in word lists such as DACST (1997a) creates confusion. The reason for this is that the prefix **uku-** can precede any Zulu verb stem, forming an infinitive implying the meaning of 'to' or bringing about change in the word category and eventually in the meaning, e.g.

> **ncelisa** (breast-feed) can form an infinitive verb **ukuncelisa** (to breast-feed) or a deverbative noun **ukuncelisa** (breast-feeding).

For the sake of logical morphological consistency, the simplest manner of morphologically notating the verb is that it should not be preceded by the **uku-** prefix, e.g. **ncelisa** (breast-feed). In this manner the user will not be confused as to whether **ukuncelisa**, for instance, is a verb (to breast-feed) or a noun (breast-feeding).

3.3.4   Capitalisation

After the investigation of examples of capitalisation in relevant official and commercial publications, it was found that capitalisation as far as technical (medical) terminology in Zulu is concerned, can roughly be based on capitalisation rules in English, except that the linguistic structure of Zulu must be adhered to. The following rules could be formulated so as to deal with the latest trends in capitalisation:

i) The first letter after the initial vowel in the name of a specific commercial product or name of an international technical (medical) term or illness is capitalised. This is quite an easy rule since the first letter of such a name would be capitalised in English too, e.g. **iPurity** (Purity - health cereal).

ii) In technical acronyms which are usually capitalised terms, the initial vowel preceding it is in the lower case, e.g. **i-GP** (GP - general practitioner) and **i-HIV** (HIV - human immunodeficiency virus).

3.3.5   Changing linguistic trends in the language which are not reflected in the orthography

The present ZNLB should see to it that orthographical rules be adjusted to accommodate significant developments and change in the language, be they, for instance, phonological or morphological. As LANGTAG in DACST (1996:69) puts it:

> In each standard variety it is necessary, from time to time, to adjust the spelling system and
> this might also be necessary in some of South Africa's languages.

In Zulu, the traditional open syllabic system of consonant vowel (CV) in a syllable , e. g. **ikilasi** (classroom) does not necessarily hold any more. Many modern adoptives conform to the CCV-pattern in a syllable , e.g. **istradi** (street) instead of **isitaladi**.   It is noted that the **r** is increasingly being used in modern loans where it was previously regarded as a non-Zulu sound and mostly replaced by **l**, e.g. **imajarini** (margarine) as opposed to the earlier **imajalini** (margarine). Koopman (1992:110) even cites examples of Zulu adoptives where the double **rr** is used , e.g.
i-rrisidi (receipt). He also exemplifies other 'foreign' consonant clusters in Zulu, e.g **thr** in

i-bhethri (battery) (Koopman 1992:111). It is a fact that closed syllables (not ending in vowels) are now readily accepted, e.g. **ikneb-tange** (pliers).

Besides phonological changes, morphological changes are also quite common. One such example deals with an irregular derivational process where verbs are derived from adopted nouns, e.g.
-farisa (act hypocritically) derived from **umfarisi** (Pharisee, hypocrite) (Koopman1994:238).   Another relatively new morphological tendency is evident in the use of the verbal extension **-ish-** in adopted verbs, e.g. **-filisha** (to fill out a form) (Koopman 1994:244).

Nevertheless, it is important that recent phonological and morphological trends should be evaluated to determine how commonly they are used in the spoken form. If uncommon sounds like **khr**, **r,** and suffixes like **-ish-**, etc. are continually being used, for instance, they should be incorporated in the orthography of the language so as to represent the living language.

3.4   The way forward

After orthographical problems have been identified and addressed through a practical standardisation approach, it won't mean much if these adjustments /changes are not made known to all the concerned language interest structures. What Mathumba (1993:125) says about the rulings of the *Tsonga Language Board*, namely that they did not always reach educators or the general public, is also true for the other African languages, including Zulu. New rulings should be effectively disseminated to speakers of the language to get feedback on which final decisions can be based.

However, the lack of terminology in the African languages can be overcome if action is taken towards standardisation in the first instance, i.e. on the orthographic level. This level is the easiest to attain as put forward by Sager (1990:123) who reasons that standardisation can be successful on at least the levels of spelling, pronunciation, morphology and syntax.

Scholars such as Thipa (1989:179-180) and Mathumba (1993:208-210) make some valuable recommendations regarding the *Language Boards/Committees* (now known as *Language Bodies*), one of them being that the composition of the *Board* (now NLBs) should be changed to include more members who are knowledgeable and qualified in linguistics and language planning. This can be supported since such composition will definitely lead to a decrease of inaccuracies in the linguistic formulation of orthographical rules. Such an improved situation can only enhance terminological development in the African languages.

CHAPTER 4

THE METHODS OF WORD-FORMATION THAT FACILITATE LANGUAGE AND TECHNICAL ELABORATION IN ZULU

4.1   Introduction: Corpus planning as part of language planning

The sociolinguistic field of language planning, for the purpose of this study, can be divided into two aspects, namely <u>status</u> <u>planning</u> and <u>corpus</u> <u>planning</u>. Status planning (see 2.2) deals mainly with decisions taken by governments regarding language policy and its implementation and includes language development. Corpus planning, on the other hand, deals with language standardisation and elaboration or modernisation. Language elaboration refers to the creation of new terms in order to meet the scientific, educational and technical demands of a language. This can be achieved by following certain methods of word-formation existing in a specific language. Obviously, since this chapter specifically deals with word-formation methods that facilitate language elaboration, the focus is on corpus planning.

The two main divisions of language planning, referred to as status and corpus planning by Fishman (1977:37), stand in direct relation to each other according to Ohly (1987:55) in that a language with low status has underdeveloped terminology. Since this is particularly true for the African languages in South Africa including Zulu, the problem of underdevelopment needs to be addressed by, for instance, following an approach of 'language cultivation' which implies the development of language at all levels. Many scholars such as Neustuphy (1974) and Gonzalez (1993) use the term 'language cultivation' instead of 'language elaboration' to refer to corpus planning. According to Gonzalez (1993:17,19) the cultivation of a language refers to the cultivation of both its literature and its science. Cluver (1989:13) is clearly of the same opinion in this regard, stating:

> Thus the language of a modern, industrialized speech community will not only reflect its literary and cultural achievements but also its scientific and technical achievements.

The type of cultivation referred to above will eventually lead to intellectualisation, implying the use of a (national) language "... as medium of scholarly discourse" (Gonzalez 1993:18). Gonzalez (1993 :19) sees this intellectualisation of the national language as "... symbolic of the last conquest of development." Needless to say, this type of intellectualisation is still to be achieved in the African languages.

4.1.1 Corpus planning in South Africa

This chapter briefly deals with corpus planning in South Africa, specifically with reference to Zulu and the roles of national authorities such as the former *Language Boards/Committees* in facilitating term development in the African languages. In addition to this, the methods of word-formation in Zulu that facilitate (technical) elaboration are discussed and exemplified with ample reference to medical terminology.

The emphasis in this chapter is on the state of corpus planning in the African languages of South Africa, specifically Zulu. However, for the sake of exemplification and universality, corpus planning in other countries is also referred to. It is a fact that the only fully developed languages in South Africa are English and, to a lesser extent, Afrikaans. This is also mostly true for the rest of Africa where the only fully developed languages are the European languages and the only developing languages that are used as medium for secondary and tertiary education are Swahili and Yoruba (Cluver 1993:40). Since the African languages have not been developed to their full potential, they are deficient in terms. This is confirmed by Cluver (1993:30) stating that the terminology lists compiled by the previous *Language Committees/Boards* contain little more than fairly elementary technical vocabulary. However, some progress has been made in the medical field, even if available only in draft form, with the compilation of the *Draft list*: *Basic Health Terms* (1997a) and *Draft List*: *Sex Education* (1997b) compiled by the *National Terminology Services* - NTS, now known as the *National Language Service* - NLS, of the Department of Arts, Culture, Science and Technology - DACST, now divided into two departments known as the Department of Arts and Culture - DAC and the Department of Science and Technology - DST. Terminological practice obviously resorts under the former, the Department of Arts and Culture.

As in any other country, term development is facilitated by an institution (Ohly 1987:59). See also agents of standardisation in 2.3.6. In South Africa term creation in the African languages was originally the function of the former *Department of Bantu Education* (DBE) which later co-ordinated the different *Language Committees/Boards*, e.g. the *Zulu* or *Xhosa Language Committee/Board*. Terminology development in the African languages was hampered by many educational, political and historical drawbacks, as outlined by Mtintsilana and Morris (1988:109). In order to cope with the introduction of mother-tongue education in 1953 and, later, the establishment of the homeland states in South Africa (such as Venda) term creation was needed in the legal, administrative and educational sectors. The result was that terminologists, although not always acquainted with the processes of term development, created terms out of necessity. It is thus hardly surprising that some African scholars slam the term creation effort of the Language Boards, calling it an artificial political process of term manufacturing (Jafta 1987:27, 131). For a long time in African language development, corpus planning was approached from a Eurocentric perspective (LANGTAG in DACST 1996:71). Initially such development focussed on the establishment of an orthography, Christian terminology and basic school terminology of a Western register.

The present language authority in South Africa is known as the *Pan South African Language Board* (PanSALB), established in 1995 under the new dispensation. The activities of the former *Language Boards/Committees* were terminated and new bodies formed by the new government. These are known as *National Language Bodies* (NLBs), thus the body for Zulu is known as the *IsiZulu National Language Body* (ZNLB) and the body for Swazi is known as the *SiSwati National Language Body*, etc. These African language Bodies including those for Afrikaans and English, were under the PanSALB Act authorised to set up *Provincial Language Councils* (PLCs) to promote the development of specific languages in each province. Furthermore, the NLBs are to promote the standardisation of the orthography and terminology in each relevant language. The NLBs and PLCs should work in close association with the *National Language Service* (NLS) and the *Terminology Coordination Section* (TCS) of the DAC, especially in the development and coordination of terminology.

However, in South Africa little is known about the language elaboration and standardisation processes in general and structures which administer these. It is alarming to note that there seems to be an overlapping of tasks in some of these above-mentioned language interest structures, since the task of each is not clearly spelt out. Fortunately, Alberts (2003) succeeds in a recent publication in capturing a much needed overview of language development in South Africa, sketching the collaboration between PanSALB and terminology structures.

Another issue that warrants mentioning is that without proper language policy implementation successful development of terminology cannot be achieved. See 2.2.2 for problems with status planning implementation. In Africa in particular proper language policy implementation has remained a problem. South Africa is no exception. It has an ideal language policy imbedded in the constitution; however, its real problem lies with proper implementation particularly as far as the African languages are concerned.

4.1.2 Language elaboration facilitated by methods of word-formation in relation to culture

In this chapter different types of word-formation methods, including borrowing in Zulu are discussed and exemplified in order to give an overview of all possible types. This is done since there is an absence of work done in the field of word-formation patterns in the African languages (LANGTAG in DACST 1996 and Alberts 1997) - see also 1.3 for the statement of the research problem. This task is a practical approach to standardise, at least to some extent, the word-formation methods of Zulu. This can fulfil a need in the training of terminologists and lexicographers who first of all have to have the necessary linguistic insight in the formation strategies of words or terms. In another respect such discussion on word-formation methods can lead to the possible publication of a style or reference manual that could be used during such training.

These methods of word-formation draw either on the internal resources of the language or on the external resources which are borrowings from other languages. The advantages and disadvantages of the application of these discussed methods of word-formation are then examined.

When terminologists want to coin terms, they basically use the following <u>word-formation patterns of lexical expansion</u>:

i)   derivation - word-formation through affixation, i.e. creating the term from indigenous roots

ii)   semantic shift - giving a new meaning to an existing word

iii)   compounding - conjoining two or more words by means of combinations such as

        NP + NP, VP + NP, etc.

iv)   loan translation/calquing, i.e. translating the new term into the target language

v)   deideophonisation - coining new terms by means of ideophones (onomatopoeic words)

vi)   borrowing - using lexical items from a donor language

vii)   abbreviation - blending and clipping of phrases or words or using acronyms.

Besides the effective use of such word-formation patterns, the study of Zulu language elaboration would not be complete without our mentioning the way in which such elaboration is linked to extra-linguistic factors such as culture. World view and taboo for instance, are two culture-related sociolinguistic aspects which should be taken into account in any type of terminological or lexicographical development (Van Huyssteen 2002). Taboo is an umbrella term to refer to terms that are unsuitable for use in a specific social context. According to Zulu culture it is taboo to refer to terms with a sexual connotation in a direct manner. A culturally bound word, considered as part of a world view, such as a 'reeds mat/sleeping mat' **ucansi**, is therefore used to indirectly and evasively refer to most words with a sexual connotation.

Before proceeding further with a discussion of word-formation methods, it is important to note the differences between ordinary and technical language.

4.2   Technical language

The emphasis in this chapter is on technical elaboration and therefore there are certain aspects such as 'concept', 'term', 'terminology' and 'terminography' that have to be understood in order to grasp the concept of technical language.

4.2.1   Towards a definition of *concept*, *term*, *terminology* and *terminography*

Before one can understand what a term is, it is necessary to understand what a <u>concept</u> is. Felber (1982:14) reasons that there are two aspects of communication, the concept (meaning) and its form (linguistic symbol).   Felber (1982:14) thus views 'concept' as a "...meaning strictly distinct from neighbouring meanings..."; a concept is what remains in the memory about an object or situation in aid to its   identification.   Concepts exist independently of terms but need terms for their coding in a comprehensible form (Felber 1982:14).

The next step is to proceed to a definition of <u>term</u>. Drozd and Roudny (1980:33) probably give the shortest and most sensible definition of a term:

> The smallest unit of functional language is the term. The term is a naming unit for the
> technical or scientific concept. The terminological system is a naming system of concepts.
> Terms function in a technical or scientific language as the opposite of non-terms, ...

As will be noticed in the course of the discussion, a term is either a single word or a phrase (word group). The use of a term depends to a great extent on how clearly its meaning has been understood (Batibo 1992:92). It may not be used if the intended meaning is feared to be distorted. This problem can be countered if term lists or glossaries include clear definitions of such terms, e.g.

> <u>Term</u>:   **isithuthwane** (epilepsy)
>
> <u>Concept</u>:   definition - *A disorder of the brain which causes sudden attacks of*
> *uncontrolled, violent movements of the body and loss of*
> *consciousness* (DACST 1997a:8).

Cluver (1989:337) also sees the supplying of definitions as important and reasons that terms cannot be

standardised without definition. Alberts (1997:186) is in agreement with Batibo (1992) and Cluver (1989) but makes it specifically applicable to legal terminology in Northern Sotho. Other concepts related to term that merit discussion are terminology and terminography. Broadly speaking, the vocabulary of technical language, pertaining to a certain subject field, is known as terminology. Felber (1982:12) sees terminology as the basis for the ordering of knowledge, the transfer of knowledge, the formulation of subject information, the condensing of subject information and the storing of such information. Perhaps it would be appropriate to quote a more accurate definition of terminology by Sager (1990:2):

> Terminology is the study of and the field of activity concerned with the collection, description, processing and presentation of terms, i.e. lexical items belonging to specialised areas of usage of one or more languages.

From Cluver's (1989:8) definition of terminography quoted below, it is clear that there are some parallels with terminology but that terminography is rather a subdivision of lexicography:

> Terminography is generally seen as the scientific processing of technical languages and particularly the standardisation and lexicographical representation of technical terms.

However, for the purpose of this study, the interrelated concepts of terminology, lexicography and terminography, have been put into concise perspective and are thus not pursued any further. What is important, however, in terminological development is the aim of terminography as put forward by (Cluver 1989:8): "The aim of terminography is to make sure that each concept is clearly identified, defined and named by a proper technical term". Sager (1990:21,22) adds a practical perspective to show how Cluver's aim can be achieved:

> The terminologist describes the concepts of any one discipline in three ways: by definition, by their relationship to other concepts - as expressed by the conceptual structure and realised in linguistic forms - and by the linguistic forms themselves, the terms, phrases and

expressions chosen for their realisation in any one language.

This latter statement by Sager (1990) is quite relevant here since it actually summarises the workings of terminology as discussed above within a framework of closely related linguistic aspects without which language elaboration is not possible.

4.2.2   The properties of technical language

Now that it has been established that technical language is popularly known as 'terminology' and the technical word as 'term', the following general properties of technical language insofar as it differs from ordinary language, can be seen in perspective:

*   Contrary to ordinary language, technical language shows a one to one correlation between the concept and the term. Preferably only one term should be used to denote a single concept and synonyms should be reduced (Cluver 1989:5). This one to one correlation, however, cannot be followed in Zulu since many synonymous terms are used interchangeably alongside one another such as **imbo/imfiva** (fever), **ithelevishini/i-TV /umabonakude** (television), **umfundisi/uthisha/uthishela** (teacher).

*   Furthermore, the language of science is based on universal concepts and logic (Abdulaziz 1989) and the linguification of such concepts (Gonzalez 1993:19). In this regard Alberts (1999b:3) goes further stating:

> **Terms** are **exact** and should have **no emotional connotations** attached to them. When emotional connotations are attached to **terms** these terms become **words** and therefore part of the **general vocabulary** (the terrain of **lexicography**).

This statement however, cannot generally apply to the African languages since in medical terms with a sexual connotation, for instance, a type of avoidance language (*taboo*) which is somehow linked to emotion, is used. Yet these taboo forms are still terms and not words - see 4.5 for culture-related aspects in elaboration.

* Technical language contains a high percentage of international terms and loan words (Cluver 1989:5).

* Technical language expresses itself in an economical way (Cluver 1989:6).

* Technical language "... has a faster growing vocabulary" (Cluver 1989:6). It must be noted here, however, that the terminology of the African languages is not growing as fast as the vocabulary of the standard/ordinary language, as is the case with the more developed (European) languages.

* Technical language contains more nouns than the standard language (Cluver 1989:6).
It must be remembered, however, that the terminology of the African languages is not larger than the vocabulary of the standard/ordinary language, as is the case with the more developed (European) languages, and as such would not necessarily contain more nouns. Cluver's naming of 'standard language' in the latter two properties of technical language does not hold since technical terms can also be standardised - and not only words in the ordinary language.

It becomes clear from the above discussion that the properties of technical languages that form part of well developed European languages cannot necessarily be made applicable to the African languages, including Zulu, because of differences at the structural, sociolinguistic and developmental levels. The African languages have their own linguistic nature and cannot be compared to European languages on a one to one basis. This is evident in the following discussion on word-formation in which some of the unique properties of Zulu language elaboration are clarified through exemplification.

4.3   Motivation towards the standardisation of methods of word-formation in language and technical elaboration in Zulu

At this point it has become necessary to motivate a study in the methods of word-formation in language and technical elaboration in Zulu for this thesis, i.e. to formulate a statement of the problem (see also 1.3).

The African languages possess the basic tools that are necessary for their development (LANGTAG in DACST 1996:81). These tools include an orthographical standard, dictionaries, grammars and published literature. However, as outlined by Van Huyssteen (1999), these tools have some serious deficiencies. In the dictionaries an Eurocentric missionary approach is still evident; as far as vocabulary is concerned, it is to a large extent Western-orientated without reflecting word-formation patterns unique to the African languages such as derivational patterns, rules for compounding or indigenisation of loans. In grammar books and even in more advanced linguistic-analytical works such as Poulos & Msimang (1998) and Doke (1984) word-formation is not treated very systematically. Basically, following Doke's approach, Poulos and Msimang (1998) treat word-formation in Zulu in an informative manner and with morphological accuracy, but with the emphasis on 'nominal deriving mechanisms' - which unfortunately does not include all existing word-formation methods. Even though compounding, adaptations (loans) and suffixal derivation (by means of feminine, diminutive and augmentative suffixes) are also treated under the same heading of derivation in this scholarly work, some other word-formation methods such as semantic shift and loan-translation are not discussed.   In all other Zulu grammars such as Taljaard & Bosch (1988) the situation is much worse since word-formation methods are virtually not treated and hardly any reference is made to them.

Basically work on word-formation patterns, with the exception of Madiba's (2000) thesis on modernisation in Venda, is largely lacking in the African languages; thus LANGTAG (DACST 1996:82) suggest:

> To prevent African languages from losing their derivational transparency, a planned
> research programme needs to be introduced that will expose the underlying patterns of

African word-formation and borrowing. This research programme should also include the analysis of African textual strategies.

These two projects should then be converted into language development manuals. Language developers need to know what word-formation patterns are available to them. They must also have an idea of what the inherent stylistic mechanisms are that can be used to develop new registers such as the one used in court or in journalism.

...Once these basic tools are available, the training of language developers can commence.

As far as corpus planning in Zimbabwe is concerned, Chiwome (1992:89) is basically in agreement with LANGTAG in DACST (1996) above, stating that principles of word-/term- formation and standardisation need to be addressed in Shona so that those concerned with coinage may at least be acquainted with them. These principles in Shona, and in other African languages for that matter, need to be consistent and precise while also adhering to linguistic rules of the target language, even if some issues need to be addressed by academics (Chiwome 1992:91).

Alberts (1997:190) adds a practical perspective by mentioning the need of the terminologist to have such principles of word-formation at his/her disposal:

...the terminologist has to apply specific terminological principles when denoting concepts. He/she also has to apply certain linguistic principles. No terminologist can coin a term if he/she does not know the basic word-formation principles of a language. Unfortunately, the basic word-formation principles for all African languages have not yet been established. There is a dire need for such principles. The National Terminological Services would like to form a working relationship with any person in any language group who has the linguistic background and knowledge to assist the office with this research.

It should be mentioned here that in order to acquire such mentioned linguistic background and knowledge concerning the basic word-formation principles of a language, an aspirant terminologist/lexicographer has to undergo formal linguistic training, aimed at a career as a professional language practitioner. Unfortunately, this type of strict linguistic training has been done away with at most South African universities, especially in the African languages, since it appears to have been perceived as Eurocentrically based, too formal and too analytical. In fact, this issue of formal training needs to be redressed by tertiary institutions (in accordance with LANTAG above) so as to avoid a shortage of properly trained terminologists in the near future. It is obvious that a shortage of trained manpower can hamper effective language elaboration and development.

In the African languages, translators are often erroneously transferring the register of the source language to the target language since manuals or guidelines are largely absent (LANGTAG in DACST 1996:81). Needless to say, language interest structures or individuals need to initiate an effort to start compiling such essential documents containing guidelines for future use.

This initiative can be found in this chapter in which different methods of word-formation in Zulu, including borrowing, are discussed and exemplified in order to give an overview of all possible existing methods. This practical approach can be considered as an effort to generalise, even standardise, at least to some extent, the word-formation patterns of Zulu. In this manner terminologists and lexicographers can gain the necessary linguistic insight in word-/term- formation strategies be they morphological, phonological or syntactical, to eventually coin original terms in accordance with the natural linguistic elaboration mechanisms inherent in the culture of the language.

4.4   Methods of word-formation that facilitate language and technical elaboration

In this chapter, the methods or trends of word-/term-formation generally found in Zulu, are discussed. However, some technical methods of term-formation in Zulu, especially pertaining to medical terminology, are emphasised. The methods of word-formation form an important part of language elaboration because

they are the very linguistic tools that make technical modernisation and expansion of the lexicon possible. In the creation of terms in the African languages not only the general theory of word-formation should be considered, but also the rules of coinage already existing in these languages (Ohly 1987:61). In order to elaborate on the terminology of a language one has to draw from two possible sources, namely <u>internal</u> <u>resources</u> and <u>foreign</u> <u>resources</u> (loans from other languages) as stated by Mtintsilana and Morris (1988:110). Although formulating his argument somewhat differently, Cooper (1989:151) in principle agrees with the latter two scholars stating two alternatives for language elaboration, i.e.

> i) create the term from indigenous sources by
>> a) giving a new meaning to an existing word or
>> b) creating a term from an indigenous root or
>> c) translating the new term
> ii) create the term by borrowing from another language.

These two alternatives are actually conflicting in nature, since the first one has an indigenous goal while the latter has an international communicative goal (Jernudd 1977 in Cooper 1989:151).

However, these two alternatives do not form the point of departure for this chapter since such division is far from clear-cut as a degree of overlapping occurs. The method of semantic shift, or instance, occurs in both the indigenous and foreign resources of the language. Other scholars, for instance Akida (1974) in Ohly (1977:124), categorise word-formation methods differently, using morphological divisions of basic word forms on the one hand and compound word forms on the other hand. Also included in this chapter, is the examination of the advantages and /or disadvantages of the application of each method of term creation. This is done against a sociolinguistic background, i.e. by determining how these methods are perceived by scholars, some of whom are mother-tongue speakers. It must be remembered that the methods of word- formation as well as the sources of new lexical items differ from language to language due to differences in structure. However, there are some parallels to be drawn between the word-formation patterns of different languages.

It must be mentioned here that not all examples of Zulu (medical) terms are standardised and that some terms originate from general usage unless mentioned otherwise. Most of the terminology used is listed in the official term lists, *Zulu Terminology and Orthography No. 3* (1976) and *IsiZulu Terminology and Orthography No. 4* (1993) of the ZLC/B, the *Draft List: Basic Health Terms* (1997) and the *Draft List: Sex Education* (1997). Both draft lists were compiled by the previous *National Terminology Services* now known as the *National Language Service*.

The methods of word-formation that draw on either the internal and/or foreign resources of the language are derivation, semantic shift, compounding, loan-translation, deideophonisation, borrowing and abbreviation. Reliance on the existing resources is generally preferable to borrowing in the coinage of terms and is a method which has been applied successfully to retain the character of the Somali language (Andrzejewski 1979:105), and for that matter, the character of other languages too.

4.4.1 Derivation

A method of word-formation that mostly draws on the internal resources of the language is derivation. In employing the method of derivation new terms are coined from roots by adding affixes. Akida (1974) in Ohly (1977:124) calls a similar process in Swahili "prefixal and suffixal verbal derivation." Some scholars such as Mtintsilana and Morris (1988) do not consider derivation as a separate method of word-formation but rather see it as an inherent part of the morphological structure of the African languages. Strictly speaking this is morphologically correct, as affixation also occurs in almost all the other methods of word-formation (see later). Poulos and Msimang (1998) have the same view on derivation as the previous two scholars and therefore treat the whole process of word-formation in Zulu quite extensively under the heading 'Nominal deriving mechanisms'. Derivation is also referred to as 'native language derivation' by Ferguson (1977), 'the native trend' by Ohly (1981) and 'outright coinage' by Matšela & Mochaba (1986) because of the use of internal morphemic language resources in this method. Chiwome (1992) and Temu (1984) define it somewhat differently, calling it a method by means of which new terms are created from existing words. At this stage it is important to distinguish between the concepts *word*, *root* and *stem* in

Zulu by means of exemplification. In the word **siyathenga** (we are buying) there is a basic form **-theng-** called the root which carries meaning (buy) and a stem **-thenga** (root plus ending). Derivation is a method of word-formation which is used abundantly in the African languages, employing prefixes and suffixes as in the following Zulu examples:

isiqaphelisi, (lighthouse) derived as follows:

isi-	class 7 noun prefix
-qaphel-	verbal root (watch/look out for)
-is-	applied verbal extension
-i	impersonal nominal suffix

ubudakwa, (alcoholism) derived as follows:

ubu-	class 14 noun prefix
-dak-	verbal root (be intoxicated)
-w-	passive verbal extension
-a	verbal suffix

ukubeletha, (childbirth) derived as follows:

uku-	class 15 noun prefix
-beleth-	verbal root (carry on back)
-a	verbal suffix

isiguli, (patient) derived as follows:

isi-	class 7 noun prefix
-gul-	verbal root (be ill)
-i	personal nominal suffix

isikhuthaza, (stimulant) derived as follows:

isi-   class 7 noun prefix

-khuthal-   verbal root (be diligent)

-is-   causative extension (**l** becomes **z** when **-is-** is added to it)

-a   impersonal nominal suffix

The example **-farisa** (act hypocritically), derived from the nominal loan **umfarisi** (Pharisee) recorded by Koopman (1994:238) is rather exceptional since a verb is derived from a noun, just the opposite of the derivational process that occurs in the examples listed above. The reason for this is that the original noun was taken from English and no such root form exists in Zulu. In Zulu and other African languages, however, words can also be derived from other parts of speech (besides verbs) such as adjectives and relatives (qualificatives), e.g.

ubunzima, (difficulty) derived as follows:

ubu-   class 14 noun prefix

-nzima relative stem (difficult).

However, the opposite, which postulates that adjective and relative stems have their origin in nouns, may also be true, e.g. that **-nzima** (relative stem - difficult) has its origin in the noun **ubunzima** (difficulty). It is for this reason that adjective and relative stems are also called adnoun stems by Taljaard and Bosch (1988).

It must be noted, that in the process of derivation, word category change is taking place. It is for this reason that Cluver (1989) also views this method as 'conversion'. The change in word category in each of these examples can, for instance, be compared:

-daka (be intoxicated/drunk) a verb   becomes   **isidakwa** (alcoholic) a noun

-nzima (difficult) a relative   becomes   **ubunzima** (difficulty) a noun

umfarisi (Pharisee) a noun   becomes **-farisa** (act hypocritically) a verb.

Derivation can be regarded as an advantageous method of term creation because it mainly draws from the internal resources of language (Adam & Geshekter 1980 and Andrzejewski 1979). By employing these internal resources ordinary people can also have access to technology, regulations and training, making the need for foreign experts unnecessary. Therefore the national language is continually being used and maintained. Another advantage of this method, mentioned by Matšela and Mochaba (1986:139,140) with reference to Sesotho and applicable to all the African languages of South Africa, is that it greatly simplifies the task of the term creator. Many a Zulu term may, for instance, be derived from a single root or stem, e.g.

from **-funda**   (learn) the following words are derived:

umfundisi   (teacher/preacher)

isifundo   (lesson)

imfundo   (study)

umfundi   (scholar).

4.4.2   Semantic shift

Semantic shift is a means of term creation whereby the existing meaning of a word usually acquires an expanded or modified meaning in order to name a new, generally related concept. In this manner ordinary words acquire specialised meanings (Andrzejewski 1979:5). Semantic shift, as referred to by Andrzejewski (1979), Mochaba (1987) and others is also referred to as 'semantic expansion' by Batibo (1992), Akida (1974) in Ohly (1977) and Chiwome (1992), 'semantic transfer' by Ohly (1987) and even 'discovery' by Baker (1987). In the older languages, such as Arabic and Somali, archaisms are revived as part of semantic shift (Andrzejewski 1979; Baker 1987 and Ekwelie 1971).

Many Zulu examples of terms coined by means of semantic shift can be found, e.g. the original meanings of the words **umnyango** (door) and **izilwane** (animals) have now been extended to mean 'department' and

'fauna' respectively. Also in medical terms, which is of particular significance in this chapter, many examples of semantic shift occur. Such an interesting example of semantic shift is **izinkobe** (pills). The singular form of the latter word **ukhobe** literally means 'mealie grain'. Clearly, the kernels of the mealie remind one of pills. The medical term **ezinganeni** (children's ward), commonly used in hospitals in KwaZulu-Natal, is also an example of the application of semantic shift because it literally means *at the children*.

Another inventive medical term that exemplifies semantic shift is **isigoga** (quadriplegic). The noun **isigoga** is derived from the verb **-goga** (obstruct/prevent/disable). The connotation of the latter word with the former is apparently that a quadriplegic moves around with difficulty because such a person is physically disabled. The term **ukudaka imizwa** (to *intoxicate the senses*) is another example of a Zulu medical term where the literal meaning has been shifted to coin a scientific term for 'anaesthesia'.

A word such as **amafutha** (animal fat) has now extended its meaning to include other related meanings, such as 'margarine', 'grease', 'ointment' and 'motor oil'. One of the most interesting examples of extended meaning encountered in the same word, can be found in the medical term **emafutheni** (ultrasound clinic), for instance. This term was recorded by Zungu (1995) as part of *isiHosi* (an informal Zulu hospital language used in Durban). It originated in the following way: When a pregnant patient goes for an ultrasound or sonar test, a sticky gel is rubbed onto her stomach prior to the test. This type of gel reminds one of **amafutha** (animal fat). **Emafutheni** (at the place of fat) is then regarded as the venue - ultrasound clinic - where the sonar test is conducted. Yet another excellent *isiHosi* example of semantic shift is **esithombeni** (*the place where pictures are taken*) for the 'X-ray department'. This locative is derived from the basic noun **isithombe** (picture/photograph).

Mochaba (1987:140) mentions interesting cases of semantic shift which originate from the younger Sesotho generation. **Lebatooa,** of which the traditional meaning is 'constituency', has now acquired the meaning of 'girl friend', the connotation being 'something permanent', like having a steady relationship with a girl. Mkhulisi (1996) notices the same trend in the younger Zulu generation. New terms have their origin in the rural areas but are used by the youth in the urban environment. She remarks that these terms catch on

because the young people use them to adapt to the new situation so as not to be regarded as being old-fashioned. Mkhulisi (1996) cites the following 'trendy' (*own designation*) examples in this regard. See Table 1 below:

TABLE 1

'Trendy' semantic shift

term used by youth	*original meaning*	*new meaning*	*normal equivalent term*
insimbi	iron	gun	**isibhamu**
ingane	child	young girl	**intombi**
isiguga	**-guga** (be old/ disabled)	old lady	**isalukazi**
ukucanda	**-canda** (chop up wood)	to eat	**ukudla**
isichamtho	medicine used for an enema	talk	**ukuqamunda**
ukuphotha	twist	deceive	**ukukhohlisa**
edladleni	in a temporary hut	at home	**ekhaya**

The development of terms such as these in the first column in Table 1 above have to be regarded as an important sociolinguistic development in language elaboration.   Mochaba (1987:140) views them as advantageous since they are a true reflection of natural term development; they come from the people, catch

on easily and are then absorbed into the language. An advantage of employing the method of semantic shift is that the terms are transparent to the users because this method mostly draws on the internal resources of the language. The examples above of terms used by the youth (see the first column), for instance, are pure Zulu words.

Ohly (1987:65) points out that some of the disadvantages of the application of semantic shift could be that terms may become over polysemous and ambiguous. He uses the Swahili example **mkahawa,** which originally meant 'coffee-house' or 'café', and whose meaning has now been extended to include 'snack bar', ' tearoom' or 'restaurant'. The user of this particular Swahili term may easily confuse some of these extended meanings with one another.

Although semantic shift commonly occurs as purist coinage, it also occurs in borrowing. Examples of semantic shifts which are borrowings are discussed in two rather interesting articles by Louwrens (1993) and Hlongwane (1995). Hlongwane (1995:65) explains that there is no one-to-one correspondence between the adopted word and the donor word when semantic shift occurs, e.g. **umesisi** no longer means 'Missis' but now means *white married woman* instead. Each of the following examples of loans also illustrates semantic shift of a similar kind:

igeli no longer means 'girl' but now means *female servant*;
upondo no longer means 'pound' but now means *two rand*;
usheleni no longer means 'shilling' but now means *ten cents* and
ibhokisi no longer means 'box' but *stand in a court of law*, *coffin* or *dustbin* (Louwrens 1993:3).

Yet another such interesting example was found in collected fieldwork data. The commonly used Zulu loan word **ibhodlela** no longer means 'bottle' but now means 'incubator'. The origin of this semantic shift is twofold. On the one hand **ibhodlela** is a glass container in which items are kept, in this case, a premature baby; on the other hand babies which are kept in such incubators are taken away from their mothers and

as a result cannot be breast-fed but are bottle-fed instead. A rather unusual example of semantic shift can be found in the medical term **izikelemu** (intestinal worms). This loan word has its origin in the Afrikaans word 'skelms' (rogue/shifty person), the implication being that people are mostly unaware of them.

The following examples show how commercial trade mark names have, through semantic shift, become inclusive or general:

ushekazi ('Checkers' plastic bag) > any plastic bag

ihuva ('Hoover' vacuum cleaner) > any vacuum cleaner.

4.4.3   Compounding

Compounding is another productive method of word-formation in which one word or term is formed from two or more words or terms. Such combinations conform to the existing patterns of derivation (Andrzejewski 1979:105). This view is clearly shared by Poulos and Msimang (1998) who treat compounding quite extensively as a nominal deriving system. This view is correct when approached from a morphological perspective since compounding is also a process of affixation (adding prefixes and/or suffixes) to word combinations. However, not only derivation occurs in compounding but also an element of semantic shift (see examples below). What is interesting is that most Zulu compounds are excellent examples of purist coinage from the internal resources of the language since Zulu words rather than foreign words are used, e.g.

i-fuza (resemble) + **umsindo** (noise)   >   **ifuzamsindo** (onomatopoeia)

isi-qeda (finish) + **iphunga** (smell)   >   **isiqedaphunga** (deodorant)

u-ma-bona (see) + **kude** (far)   >   **umabonakude** (television)

isi-vikela (defend) + **umzimba** (body)   >   **isivikela-mzimba** (antibody)

isi-nika (give) +**amandla** (strengh)   >   **isinikamandla** (energy food)

izin-siza (help) + **ukuzwa** (to hear)   >   **izinsizakuzwa** (hearing aid)

in-diza (fly) + **umshini** (machine)   >   **indizamshini** (aeroplane)

isi-daka (be drunk/intoxicated) + **imizwa** (senses)   > **isidakamizwa** (drug)

uku-gayeka (grindable) + **kokudla** (of food) > **ukugayekokudla** (digestion)

u-ma-khala (cry) + **ekhukhwini** (in the pocket) > **umakhalekhukhwini** (cellular telephone).

According to Ungerer (1983) the components of Zulu compounds lose their word autonomy within the compound and act as morphemes. This observation is correct since the meaning of the compound is not necessarily a combination of the meanings of its lexical components, as in **indlulamithi** (giraffe) derived from **in-edlula** (pass) + **imithi** (trees) and in the medical term **um-thola** (find) + **impilo** (health) > **umtholampilo** (clinic). It should be noted that in this chapter, a hyphenated word such as **isivikela-mzimba** (antibody) is treated as a compound word and not as a loan-translation. Had it been two separate words however, it would have been regarded as a loan-translation. However, it must be made clear yet again that the division between compounding, loan-translation, abbreviation (blending - see 4.4.7.1), derivation and even semantic shift, is far from clear-cut as an extent of overlapping occurs.   A clear-cut division of word-formation methods in Zulu is not actually possible but their categorisation is attempted in this chapter for the sake of systematisation.

Compounding, referred to as such by Mtintsilana and Morris (1988) and Chiwome (1992), is also referred to as 'composition' by others such as Temu (1984) who rightly points out that the components of such compositions may belong to different word categories. Akida (1974) in Ohly (1977:124) therefore divides compounds into uncumulative and cumulative compounds. The uncumulative compound is formed by noun head + qualifier (NP + NP), e.g. **idlebelendlovu** (plant - Trimeria alnifolio/elephant ear) < **idlebe** (ear) **-la-** (of) + **indlovu** (elephant) while the cumulative compound is of a verbo-nominal structure (VP + NP), e. g. **uvuthondaba** (climax - literary term) < **u-vutha** (ripe) + **indaba** (story).   It is obvious from the examples of compounds listed thus far, that most Zulu compounds morphologically display a verbo-nominal structure (VP + NP).

Andrzejewski (1979:105) is of the opinion that phrases that are used to coin new terms are also part of compounding. Mtintsilana and Morris (1988:111) basically agree with him, stating that paraphrasing/loan-translation is closely related to compounding. In this chapter, however, these two concepts are discussed

separately.

Hlongwane (1995), who calls Zulu compounds 'adaptions' for no apparent reason, states that the type of compounds that are coined from the internal resources of the language (such as the representative examples discussed) are advantageous in that they are more understandable to all speakers, mainly because Zulu words rather than foreign words are used.

4.4.4 Loan-translation

Another productive method of word-formation used in the African languages is loan-translation, also known as 'paraphrasing' or 'calquing'. Loan-translation occurs when new terms are created by a translation of the meaning of a foreign term into the target language (Zulu), e.g.

'object', a grammatical term, is translated as **umenziwa** (literally, *the one acted upon*)

'borehole' is translated as **umgodi wokudonsa amanzi** (literally, *hole for the drawing of water*)

'blood test' is translated as **ukuhlolwa kwegazi** (literally, *the testing of blood*)

'embryo' is translated as **isibindi sembewu** (literally, *the liver/core of the seed*)

'antibiotic' is translated as **isibulala magciwane** (literally, *the killer of germs*)

'varicose vein' is translated as **umthambo ovuvukele** (literally, *the* swo*llen vein*).

In Zulu many loan-translations are possessive constructions, e.g. **ukuhlolwa kwegazi** (the testing of blood - blood test) and **isibindi sembewu** (the liver/core of the seed - embryo). Had the term **isibulala magciwane** (killer of germs - antibiotic) been written as one word, it could have been considered a compound. This implies that (wrong) orthography can determine categorisation. However, since the division between compounding and loan-translation is far from clear-cut, translated terms written as two or more words are for the purposes of this chapter, regarded as loan-translations. See also 4.4.3 'compounding' in this regard.

'Loan-translation' is the term used by scholars such as Chiwome (1992), and 'translation' is the term used

by Matšela and Mochaba (1986), while 'calquing' is used by both Abdulaziz (1989) and Temu (1984). The term 'paraphrase' is used by Mtintsilana and Morris (1988:110), seemingly when a single term from the source language is translated into the target language by means of a phrase (two or more words), e.g. the English term 'manual' translated into Zulu is **incwadi yokuchazisa** (*a book of explanation*).

According to Ekwelie (1971) loan-translation is not simply the translation of a term into the target language, but it is also the summing up or definition of such a term. The term 'casualty', for instance, is translated into Zulu as **umuntu olimele noma oshonile engozini** (*a person who got hurt or who died in an accident*) and the term 'dehydration', for instance, is translated as **ukuphela kwamanzi emzimbeni** (*the lack of water in the body*). In addition, Matšela and Mochaba (1986 :138 ,139) state that extra-linguistic factors such as culture and environment must also be considered to make a translated term relevant and transparent to the speakers of the target language. See also 4.5 for culture-related aspects in Zulu language elaboration.

A main concern of using the method of loan-translation is expressed by Tumbo (1982) and Fourie (1993) who warn that very often terms instead of concepts are translated. In this regard Tumbo (1982) advises that the concept should first be identified, then defined, and finally converted into a term. The following examples can be regarded as satisfactory 'conceptual' translations:

isihlambululi sisu **-** Zulu (literally, *the rinser of the stomach*) for 'antacid' - this is not a marked possessive construction but implied**,**

K. F. (**Khokha ifika**) - Zulu (literally, *pay when it arrives*) for 'COD' (cash on delivery)

isivimbo - Zulu (literally, *the stopper*) for 'bath plug',

umenzi - Zulu (literally, *the doer*) for 'subject' - a grammatical term,

igwinya - Zulu (literally, *something to swallow*) for 'vetkoek', from Afrikaans - a type of deep-fried bread dough which may be difficult to swallow.

The latter example also shows an element of semantic shift and derivation which proves yet again that the division between methods of word-formation tend to overlap and are thus far from clear-cut. Another

possible problem with loan-translation is equivalence. Newmark (1981) in Fourie (1993:82) rejects the concept of total equivalence between two languages. What, however, could be striven for, is "... an equivalence in content of message" (Snell-Horby (1988) in Fourie (1993:82).

In loan-translations that are borrowings it is usually the first word that is of Zulu origin while the following words are borrowings, e.g.

isifo sikashukela (diabetes, literally translated as *illness of sugar*)

umbala i-ultra violet (ultra-violet ray, literally translated as *ray of ultra violet*)

isibulali mabhakithiriya (antibiotic, literally translated as *killer of bacteria*).

4.4.5   Deideophonisation

The method of deideophonisation is a unique method of word-formation found in the African languages. Mtintsilana and Morris (1988) are the first South African scholars who identified this as a method of elaboration. Deideophonisation deals with the coinage of terms from sounds that can be associated with the object or action that has to be named. This process of deideophonisation shows parallels with onomatopoeia in English. In Zulu and Xhosa the process involves the prefixing of a class prefix to a sound, e.g.

isi-bhamu (*the sound heard when a gun is fired*) > **isibhamu** (gun)

isi-thuthuthu (*the sound of a running engine*) > **isithuthuthu** (motorcycle) - in Zulu and Xhosa

u-gandaganda (*the sound of a running tractor engine*) > **ugandaganda** (tractor)

i-hayihayi (*the sound of coarseness/roughness of breathing*) > **ihayihayi** (hyperventilation /high blood pressure).

The process in the latter three examples cannot only be regarded as deideophonisation but strictly speaking also as a process of compounding since the stems **thu**, **ganda** and **hayi** are reduplicated respectively. However, not all ideophones are onomatopoeic. The **ngqi** in the example **umhlathi-ngqi** (tetanus) indicates the state of stiffness of the jaw muscles and not a sound. It is obvious that deideophonisation as

a method of word-formation can only be advantageous in that it draws on a unique word category in the African languages, namely the ideophone (a word that resembles a sound). Drawing on the internal resources of the language to this extent can only enhance proper understanding of terminology by mother-tongue speakers.

4.4.6 Borrowing

Unlike the methods of word-formation which make use of the internal resources of the language, borrowing is a method of term creation that makes use of foreign language resources. It is usually the less developed language that borrows from the more developed one (Hlongwane 1995 and Matšela & Mochaba 1986), e.g. **ivayirasi** in Zulu is borrowed from 'virus' in English. No language is self-sufficient (Nkondo1987:70) because no perfectly homogeneous language group exists (Jafta 1987:127). As soon as languages come into contact, they start borrowing from one another, i.e. they start sharing terms and concepts (Matšela & Mochaba 1986:145). Furthermore, borrowing in a language is the result of the changing culture of a society as Kaschula and Anthonissen (1995:17) put it:

> Of course if culture is reflected in language, we may expect to find that social change produces linguistic change. So in a community which is becoming increasingly multicultural we can expect to find traces of the new forms of interaction in the languages of different cultural groups.

Borrowing becomes quite common in a multilingual society such as South Africa because of the needs of cross-cultural communication (Fourie 1993:84). Furthermore, borrowing is a very productive word-formation method whereby a foreign term is incorporated into the target language, by an indigenisation process, inevitably involving the phonology of the target language. African languages borrow from Afrikaans, English and other African languages, e.g. **imfaduko** (drying cloth, from Afrikaans 'vadoek'), **ikhensa** (cancer), **igozi** (gauze bandage) and **ukuqeqesha** (Zulu term borrowed from Xhosa, meaning ' training' taken from Xala in Mtintsilana & Morris 1988). The process of borrowing, however, differs

from (African) language to (African) language. Swahili draws from traditional languages such as Arabic, ethnolects (contact languages) and internationalisms (Greek and Latin) as in the case of **mofimu** (morpheme, an internationalism) (Ohly 1981:102). However, Zulu borrows from English for this purpose which in turn borrowed from the classical languages, i.e. Greek and Latin for international scientific terminology. In the following international loan words the Greek and Latin prefixes and suffixes are underlined (in the English) to show their classical origin as exemplified by Cluver (1989 205-305):

isayensi yeBiology (<u>Bi</u>ology)

i-oksijini (oxyg<u>en</u>)

iphenisilini (penicill<u>in</u>)

idemathithisi (dermat<u>itis</u>)

i-antiserumu (<u>anti</u>-serum)

iganjirini (gangr<u>ene</u>)

imethabholisimu (<u>meta</u>bolism)

idawuni sinidromu (Down <u>syn</u>drome). .

Adapting classical terms to the target language is an international trend in coinage and as Cluver (1989:290) states: "...it becomes clear that a fairly substantial sector of the scientific world uses this classical terminology".

From the examples **imfaduko** (drying cloth, Afr. 'vadoek') and **ikhensa** (cancer) it becomes evident that new terms have to adopt the morphological, phonological and orthographical rules of the target language. To suit the morphology of Zulu, for instance, the respective class prefixes **im-** and **i(li)-** have to be prefixed to the respective noun stems **-faduko** and **-khensa** and to adopt the phonology and orthography of Zulu, for instance, the Afrikaans **v** and English **c** have to be replaced by the Zulu **f** and **kh** respectively. Adoptions also have to be made to loan words which have their origin in Greek and Latin in order to suit the Zulu morphological and phonological structure, e.g. the adding of a prefix **i-** to the stem **-estrogen** in **i-estrogen** (oestrogen).

From the above-mentioned examples it thus becomes clear why scholars such as Hlongwane (1995) and Chiwome (1992) prefer the term 'adoption' to 'borrowing' with the latter being used by Temu (1984) and Andrzejewski (1979). The term 'adoptives' is also used by Koopman (1994) while 'loan-borrowing' is used by Fellmann (1979). Professional terminology is usually adopted from English for the English-speaking African countries (Mochaba 1987), as in the Zulu examples **ivalvu** (valve), **ijondisi** (jaundice) and **imfiva** (fever).

Terminologists should not only take care of the morphology and phonology of new terms but should also be aware of sociolinguistic development and change in a language. Hlongwane (1995) makes a major contribution in this regard. He remarks that the use of modern adoptives in Zulu is a marker of the social class of its users. These adoptives are used mainly by the educated and are therefore closer to the donor language (Afrikaans or mostly English) which is generally associated with advancement.

It is interesting to note that in modern loans established rules of adoption in Zulu are not adhered to but have changed. See also 3.2.5 for linguistic trends which are not reflected in the orthography. In contrast to traditional adoptives which adopt the open syllabic system of consonant vowel (CV), e. g. **ikilasi** (classroom), the modern adoptives conform to the consonant consonant vowel-pattern (CCV) which is **iklasi**. Other examples with a similar syllabic system to the latter example are **istradi** (street, from Afrikaans 'straat') instead of **isitaladi**, and **idrobha** (town, from Afrikaans 'dorp') instead of **idolobha** (Hlongwane 1995:61). It is interesting to note that the **r** is increasingly being used in modern loans, for instance in **itrankwilaza** (tranquilliser), where it was previously regarded as a non-Zulu sound and replaced by **l**. Having consulted modern data, Koopman (1994:163) further confirms that the **r** "... has now become a regular phoneme in the language, albeit restricted to lexical items assimilated from English and Afrikaans." Nkabinde (1968:20) admits that although the phonetic structure of Zulu is somewhat disturbed by the adoption of speech sounds such as **dr**, **gr** and **st,** these sounds bear the evidence of continual linguistic change. Hlongwane (1995:61,62) notices the trend that some adopted verbs employ the morpheme **-ish-**, quite uncommon in Zulu, for no apparent reason, e.g. in **-tadisha** (study). Koopman (1994:244), however, sees **-ish-** as a suffix specifically used for adoptive

verbs. This -**ish-** seems to show some morphological resemblance to the causative verbal extension **-is-** but the matter needs to be investigated further.   Koopman (1992:111) also exemplifies other types of somewhat 'foreign' consonant clusters in Zulu, e.g **thr** in **i-bhe**thri (battery) and **khr** in **i-khr**imu (cream). Thipa (1992:81) observes the same trend of the occurrence of foreign consonant clusters such as **pr**, **fr** and the like in Xhosa. He (op. cit.) ascribes this occurrence to Xhosa's exposure to "... Western cultural influences and experiences."   His viewpoint may very well also be applied to Zulu.

Apart from phonological changes, morphological changes also occur in loan words, particularly influencing the noun class system in Zulu (Hlongwane 1995:62). Modern loans such as **i-ogani** (organ) tend to belong to class 9 with the class prefix **i-** instead of class 5 with class prefix **i(li)-**. Although the class 9 noun prefix should be **in-** or **im-,** the noun **i-ogani** (organ) is recognised as a class 9 noun by its subject concord **i-** in syntactic context and not by its prefix. The reason why the simple class prefix **i-** is used is that it is closer to English or Afrikaans which has no class prefixes. Another interesting phenomenon is the plural formation of [+ human] loans. The plurals of class 5 nouns **imeneja** (manager) and **imeya** (mayor) are not [1]**amameneja** and [1]**amameya** (class 6) as would be expected, but **omeneja** and **omeya** respectively, perhaps to conform to the [+ human] plural form **o-** of class 2(a).

It is important that the phonological and morphological changes or trends discussed above be evaluated so as to determine how commonly they occur in the spoken form. If uncommon sounds like [**thr**] are continually being used, for instance, they should be taken up in the orthography of the language so as to reflect the living language.

Some scholars like Cluver (1989) and Sager (1990) see 'new word coinage' as another separate method of word-formation, which occurs when new terms have to be coined for new inventions in technology. However, in this chapter new word coinage is not discussed separately since it overlaps to an extent with borrowing, e.g. **i-imeyili** (e-mail), **ikhompiyutha** (computer),
i-website (website) and **i-internet** (internet) and to some extent with the other methods already discussed

---

[1]   ungrammatical

such as compounding, e.g. **umakhalekhukhwini** (cellular telephone). It is interesting to note that many new terms such as **i-website** (website) and **i-internet** (internet) are almost entirely taken over from English. Zungu (1995:150) calls such loans 'domesticated adoptives'. She remarks that they usually only attach one formative at the beginning of the word. In the latter two terms, for instance, the prefix **i-** is the only Zulu morphological constituent. This prefix is usually a marker of class 9 nouns, the noun class many such new coinages belong to.

The use of borrowing as a method of word-formation can be advantageous, pursued within an indigenous framework. However, unrealistic unproductive conservative purism can impede the standardisation process in cases where adopted terms are already in use, especially in the advanced level of education where students are more inclined to use such loan terms (Chiwome 1992:90). Such terms are, for instance, the scientific terms which have their origin in Greek and Latin discussed earlier. Even in ordinary language use certain purist terms are unacceptable, since they are hardly ever used in communication. Such unused purist standardised examples in Zulu are **isiqandisisi** (*something that makes cold* - fridge) and **isiqhebezo** (*something that switches on* - switch) taken up in *Zulu Terminology and Orthography, No. 3* (1976). As far as could be established, Zulu speakers frown upon the use of these words and wonder where they come from since they are more accustomed to the generally used loan words **ifriji** (fridge) and **iswishi** (switch) respectively.

4.4.7   Abbreviation

Although some methods of word-formation, such as abbreviation, are used to a limited extent in the African languages, they nevertheless deserve mentioning, especially in the context of globalisation where they are becoming more and more popular. Abbreviation is termed 'compression' by Sager (1990:263) who captures it as a method of word-formation where complex terms are compressed/reduced to a more compact standardised form, e.g.

amazinga C (degrees C - degrees <u>C</u>elsius)

ifekthamvama elikhulu (F. V. K.) - greatest common measure (G. C. M.)

i-radar (radar - radio detecting and ranging).

The full forms of such abbreviations are hardly ever used and these abbreviations function as normal word forms in the language (Cluver 1989:265), e.g. the abbreviated form PC is almost always used instead of its full form 'personal computer'. The same applies to 'cell' that is preferred to the full form 'cellular phone'. Even in the African languages these abbreviations are adapted to suit the language concerned and used in the same manner as in English. The Zulu abbreviations in the form of **isel** (cell) and **i-PC** (with the addition of the class prefix **i-** to adapt to the Zulu morphology), for instance, are commonly heard in everyday communication. Cluver (1989:263-266) treats abbreviation as a form of 'word manufacturing' which also includes acronyms, abbreviations and shortenings. He does not categorise blending and clipping as abbreviations and discusses each as a separate process in the creation of new words. In this chapter, however, abbreviation is categorised somewhat differently with the sub-categories of blending, clipping and acronyms for the sake of logical systematisation (since they are actually all abbreviations).

4.4.7.1 Blending

A method of word-formation, rightfully considered an abbreviation by Temu (1984) but called 'blending' by Cluver (1989), is evident when the parts of two different words are combined, e.g.

umntwana wami (*child of mine*)   > **umntanami** (my child)

uthisha omkhulu (*big teacher*)   > **uthishomkhulu** (headmaster)

isinqanda ukuvunda (*attacker of the rot*) **> isinqandakuvunda** (antiseptic - noun)

izakha umzimba (*builders of the body*)   > **izakhamzimba** (building food/protein)

isidaka imizwa (*intoxicator of the senses*) > **isidakamizwa** (narcotic/drug - noun).

It is evident yet again that a great deal of overlapping occurs between blending, compounding and loan-translation as methods of word-formation in Zulu (see also 4.4.3 and 4.4.4 for compounding and loan-translation). Blending as such, it seems, need not be considered as a separate method of word-formation in Zulu but rather as a phonological process of elision that facilitates compounding. The apparent reason

why blending occurs in the examples above is that Zulu is an agglutinating language that employs a conjunctive writing system.

4.4.7.2 Clipping

Clipping is another method of word-formation which has to do with abbreviation in the sense that a term is reduced to one of its parts, e.g. **ikhondense** 'condensed milk' (Koopman 1994:258) where the latter part (milk) is elided and not translated. Other Zulu examples of clipping in which the same type of elision occurs are: **ikhwashi** (kwashiorkor), **iselula/isel** (cellular phone), **iskizo** (schizophrenic patient) and **indizamshini** > **indiza** (aeroplane). Strictly speaking, the latter example can also be seen as an internal process of derivation (see 4.4.1) because **indiza** is derived from the verb stem **-ndiza** (fly).

4.4.7.3 Acronyms

Abbreviations in the form of acronyms occur universally and are, to some extent, enhanced by the mass media. According to Cluver (1989:264) "Acronyms are formed by using one or more initial letters of an expression and forming a new word with them...", e.g. **i-TB** (tuberculosis) and **K. O.** (**ukheshi oda**) - (literally, *cash on order*) for CWO (cash with order). Cluver (1989:265) further states that the full words, such as 'television' for instance, are rarely used and abbreviations such as TV, function as ordinary words in the language.

Some of the following examples of acronyms in medical terminology occur not only internationally but also in Zulu, with the addition of the necessary prefixes to conform to the Zulu language structure:

i-HIV (human immunodeficiency virus)

i-MC (mentally handicapped patient/mental case) - (Zungu 1995)

i-X-reyi (X-rays)

i-AZT (zidovudine - AZT)

i-IUD (intrauterine device - IUD)

i-GP (general practitioner - GP)

i-ELISA (enzyme-linked immuno sorbent assay blood-test for AIDS - ELISA).

4.5   Culture-related aspects in Zulu language elaboration

The discussed methods of word-formation that facilitate language elaboration in Zulu cannot be perceived fully without mentioning how they relate to extra-linguistic factors such as culture. In fact, the acceptance of a term may even depend on such extra-linguistic factors (Batibo 1992:99). It must be said however, that for the purposes of this chapter culture refers to its widest general application. It is thus clear that,

> Language can be studied not only with reference to its formal properties...but also with regard to its relationship to the lives and thoughts and culture of the people who speak it (Gregerson 1977:56).

Equipped with the necessary background on the facilitation of language elaboration in the African languages, Zulu in particular, two culture-related aspects, being world view and taboo can now be discussed and exemplified in the light of the needs of language development in South Africa.

4.5.1   World view

According to Whorf, an American linguist and anthropologist, "... every language represents and creates a distinct reality "(Seymour-Smith 1986 in the *Macmillan Dictionary of Anthropology*).
No wonder then that Kaschula and Anthonissen (1995:21) state that "The relationship between language and culture is well reflected in the vocabulary of a language."
This actually means that a person's mother-tongue offers him/her a framework for his/her perception or world view. It could also, to the extreme extent, mean that: "There is no thought without language" (Hudson 1980:104). One must, however, admit that ideas do shape language. However, as Hudson (1980:83,104)

observes "Most of language is contained in culture" and "We dissect nature by our communicative and cognitive needs rather than by our language." Although this hypothesis of world view is yet to be proven, many examples of its application in term development can clearly be evidenced in this chapter.

World view, or the Zulu world view for that matter, is continually determined by cognitive and communicative needs which have to be fulfilled by developing new terminology in the language.

To fill the gap of the lack of terminology, existing linguistic items can be adjusted to fit individual needs by metaphorical extension (Hudson 1980:05).   However, for the purposes of this study the term 'semantic shift' suffices, since not all such extensions are metaphoric in nature. Needless to say, terms created by semantic shift correlate with a certain cultural world view. Semantic shift is a means of word-formation whereby the existing meaning of a word in the general lexicon acquires a modified meaning in order to name a new, mostly related concept. Many Zulu examples of terms coined by means of semantic shift can be found in which the original meaning of a concrete word such as **umnyango** (door) has now been extended to mean something abstract, namely 'department' as in **Umnyango Wezempilo** (Department of Health). Another interesting example of semantic shift is found in **ukubeletha** (labour - noun). **Ukubeletha** literally means *to carry on the back*. The correlation with world view here is that in African culture labour is associated with the carrying of an infant on the back.

However, not only words or terms created by semantic shift correlate with a certain cultural world view, but also words or terms created by derivation, compounding and loan-translation. The term **umthungo** (operation) is an example of derivation since this noun is derived from the verb stem **-thunga** (sew). On the other hand, **umthungo** (operation) can also be regarded as an example of semantic shift since the meaning of 'sew' suggests some connection with 'finishing off an operation by stitching'.

Compounding is another productive method of term creation in which word combinations conform to the existing patterns of derivation, but also to the application of culturally bound semantic shift, as in the following terms:

um-qeda (finish) + **izwe** (land /nation)   > **umqedazwe** (epidemic)

u-vutha (ripen) + **indaba** (story)   > **uvuthondaba** (climax - a literary term)

u-ma-bona (see) + **kude** (far)   > **umabonakude** (television).

Yet again, as in the other methods of word-formation, world view (perception) is clearly reflected in loan-translation, at times by reference to cultural objects or concepts, as in the following examples:

'respiratory system' is translated as   **umgudu wokuphefumula** (*dagga pipe of breathing*)

'digestive tract' is translated as   **umgudu wokudla** (*dagga pipe of eating*)

'embryo' is translated as   **isibindi sembewu** (*the liver/core of the seed*)

'snack' is translated as   **isibambamoya** (*catcher of wind*).

From the examples above it is clear that world view is reflected in this type of coinage so that concepts (linked to culture) and not merely terms, are translated. To suit the Zulu idiom, it is interesting to note, many loan translations are possessive constructions, e.g. **umhlaza wesikhumba** (*growth <u>of</u> the skin* - skin cancer) or sometimes definitions, e.g. **umntwana ozalwe eseshonile** (*a child who is born dead* - stillborn baby).

For a full discussion of the relevant word-formation methods discussed here, see 4.4.1 'derivation', 4.4.2 'semantic shift', 4.4.3 'compounding' and 4.4.4 'loan-translation'.

4.5.2   Taboo

The second culture-related aspect of Zulu language development to be discussed is taboo. It may be relevant to quote a definition of the concept taboo as found in Crystal (1993:8):

> The word *taboo* has been borrowed from Tongan, where it means 'holy' or 'untouchable'.
> Taboos exit in all known cultures, referring to certain acts, objects, or relationships which
> society wishes to avoid - and thus to the language used to talk about them. Verbal taboos

are generally related to sex, the supernatural, excretion, and death, but quite often they extend to other aspects of domestic and social life.

The reason why the concept taboo has to be properly understood by terminologists is that they have to devise terms for important current health issues such as sex education, specifically with relation to AIDS. However, before one can proceed with citing taboo terms in Zulu, it is perhaps important to explain some of the following ethno-linguistic concepts, to which taboo is closely related such as euphemism and *inhlonipho*:

Euphemism: Decorous speech; a way of describing an offensive thing by an inoffensive expression (Doke & Vilakazi 1972 :xxii).

Inhlonipho: Respect, honour, reverence; a showing of respect (Doke & Vilakazi 1972).

A good example to illustrate taboo with regard to these ethno-linguistic concepts above in relation to culture, is the use of the Zulu term **isifo socansi** (sexually transmitted disease), a loan- translation with the literal meaning *the illness of the reeds mat /sleeping mat*. When Zulu culture is taken into consideration, it is taboo to refer to terms with a sexual connotation in a direct manner.   A cultural object such as **ucansi** (the reeds mat/sleeping mat), a culturally bound word, is therefore used to indirectly and evasively refer to 'sex'. Another good example to illustrate taboo is the use of the Zulu medical term **ijazi lomkhwenyana** (condom, literally translated as *the coat of the young man*). Clearly, the initial loan word **ijazi** (coat) is an example of semantic shift with the connotation of 'covering' linked to the use of a condom. Furthermore, **ijazi** (coat) avoids direct reference to a sexual object such as a condom and can thus be seen as a form of avoidance or euphemism, but at the same time as the showing of respect. Other examples of taboo are the Zulu terms **ukuba sesikhathini** for 'menstruation' (*to be in a certain time*), **ukuphela kwenzalo** for 'menopause' (*to quit giving birth*), **ukukhipha isisu** for 'abortion' (*to remove the stomach*), **ukukhipha iqanda** for 'ovulate' (*release an egg*). An extreme example of euphemism is found in the Ndebele example **ukuba namehlo amanengi/ukutjiswa ziingazi** (*to have many eyes/to be heated by blood*) which is an

equivalent for 'casual sex' (*Draft list: Sex Education* compiled by the NTS (DACST 1997b)).

Another interesting example of how a term with a sexual connotation was coined, applying linguistic taboo techniques, is the Zulu term for AIDS / HIV namely **ingculaza**/**ingculazi,** a term which created a lot of misunderstanding generally speaking, and in the press. The term for AIDS **ingculazi** is a widely accepted term in medical circles and used by Zulu medical staff in hospitals in Natal where research was conducted - it is also an official term contained in draft lists for medical terminology compiled by the *National Terminology Services* now known as the *National Language Service*. According to many informants this term was coined by a gifted Radio Zulu (*UKhozi FM*) announcer by the name of Thokozani Nene.

It was soon discovered that there was a mystery surrounding this term **ingculazi**. The first reason for this mystery is that since informants wanted to show respect (*inhlonipho*), they saw it as a cultural taboo term and thus decided not to talk about it - seeing it as a forbidden topic, and therefore avoided clarifying its meaning. The second reason has to do with the intuitive cultural linguistic means imbedded in the language to convey the concept of taboo. Since Zulu speakers, many of them academics, would or could not clarify the meaning of **ingculazi**, Doke & Vilakazi's *Zulu -English Dictionary* (1972) and other dictionaries were resorted to, alas without its origin being found. Fortunately, after some time, an entry with the stem **-ngculazi** was found in Nyembezi's (1992) monolingual Zulu dictionary, reading:

> **-ngculazi** (in) bz isifo esibulalayo okulukhuni ukuselapha. Siqeda amandla omzimba okuzivikela sibhebhetheka ikakhulu ngokuhlangana ocansini.
>
> (-**ngculazi** (in) n a fatal illness that is difficult to cure. It attacks the immunity of the body and is especially transmitted through sexual intercourse).

Eventually, the mystery of the actual origin of this word was solved by the appearance of an article in the *Daily News* of 23 August, 1999 p9, in which the term creator, Nene himself, was interviewed. **Ingculazi** he explains, has its origin in the compounding of two words, namely **ugcunsula** (sexually transmitted

disease) and **umgcalaza** (a poison used to kill an *induna* (headman) of a previous Zulu King, Cetshwayo)
. It seems that a type of phonological blending occurred between the first part of the former and the second
part of the latter word in order to form the compound, i.e. **ug<u>cu</u>nsula** + **umgca<u>laza</u>** **> ingculazi**. This
form of taboo evidenced in the formation of **ingculazi** can actually be seen as a type of linguistic avoidance
that is manifested in the morphology of the language.

Not only terms that originate from the internal sources of the language such as the examples discussed
above serve as taboo terms but also loan words. Such an example is the loan term **iluphu,** for instance, that
is commonly used for 'intrauterine device' (a loop placed in the womb of a woman to prevent pregnancy).
  However, what is even more interesting is that loan words are often used to replace taboo terms in the
target African language (Louwrens 1993). In accordance with Louwrens' observations, the following
examples where loan terms replace the Zulu taboo terms could be traced in Zulu: **ukuba phregi** (be
pregnant) replaces **ukuba nesisu** (literally, *to be with a stomach*), **amakaka** (from Afrikaans 'kak')
replaces **amasimba** (stool/excrement/faeces)**,** **ipipi** (from Afrikaans 'piepie') replaces
ubudoda/**umphambili** (penis), **idrophu** (from the English acronym DROP as origin) replaces
ugcosula/igonoriya (gonorrhoea). This is a way in which the actual language - Zulu - is avoided and the
loan language preferred - the utmost form of taboo!

It should be absolutely clear by now that Matšela and Mochaba (1986:138-139) have it right when they
state that extra-linguistic factors such as culture and environment must also be considered to make a
translated term relevant and transparent to the speakers of the target language - in short - culture-related
aspects, such as world view and taboo, must be taken into account in any type of language elaboration
process.

4.6   Conclusion

Corpus planning deals with language elaboration (more specifically word-/term-formation). This elaboration
may require the use of technical language. Contrary to ordinary language, technical language is more formal

and more inclined to show a correlation between the concept and the term.

It has been found that the methods of word-formation that facilitate language elaboration are derivation, semantic shift, compounding, loan-translation, deideophonisation, borrowing and abbreviation. To some extent, affixation (the adding of prefixes and/or suffixes) is evidenced in all these methods but particularly strongly in derivation. The verb root **-gul-** (be ill), for instance, derives the noun **isiguli** (patient) by adding the class prefix **isi-** and the nominal suffix **-i**.

Like derivation, semantic shift is also a method that produces terms that utilise the internal resources available in the language. Semantic shift is a method whereby the existing meaning of a word is extended in order to name a new related concept, e.g. **isithombe** (photograph) now also names a new concept **isithombe** (x-ray) because of parallels in meaning between the former and latter terms.   It has also been found that the method of compounding, the formation of one word from two or more words, produces original internal purist coinage, e.g. **u- vutha** (ripen) + **indaba** (story) > **uvuthondaba** (climax - literary term).

Deideophonisation is a method of word-formation which also exemplifies original purist coinage in that a term is formed on the basis of its onomatopoeic resemblance to a specific sound.   An example is **isithuthuthu** (motorcycle) which is formed on the basis of its resemblance to the sound of a running engine **thu-thu-thu**.

The methods of word-formation that draw on foreign resources in terms of loans from other languages are loan-translation and borrowing. Loan-translations occur when new terms are created by translating the meaning of a foreign term into the target language, e.g. 'embryo' is translated as **isibindi sembewu** (literally, *the core/liver of the seed*).   Such Zulu loan-translations are mostly possessive constructions and semantically speaking conceptual translations.

The method of word-formation that draws exclusively on foreign resources in terms of loans from other

languages, is borrowing. All borrowed terms must be phonologically and morphologically adapted to suit the structure of the target language. One such an example is **ibhulima** (bulimia), morphologically adapted by means of the class prefix **i-** and phonologically adapted by means of the introduction of a Zulu sound such as **bh**. In borrowings many new sounds and sound combinations occur which were previously uncommon in Zulu, such as the **r** (previously usually replaced by **l**) in **ikhalo̱ri** (calorie) and the **thr** in **ibhe̱thri** (battery). These should then, according to Koopman (1994) and Hlongwane (1995), be incorporated into the language.

Abbreviation is another method of word-formation that occurs in the blending and clipping of words and in acronyms. Blending occurs when the parts of two words merge and can therefore also be regarded as part of compounding, e.g. as in the case of **isidaka imizwa** (*intoxicator of the senses*) > **isidakamizwa** (narcotic/drug - n). Clipping occurs when a term is reduced to one of its parts, e.g. **iskizo** for 'schizophrenic patient'. Acronyms are initial words of an expression which are abbreviated in a sequence of letters in order to form a new term, e.g. **i-HIV** (human immunodeficiency virus - HIV). Acronyms occur universally and their use is enhanced by the mass media to a great extent.

After having discussed the seven methods of word-formation in Zulu it became evident that no clear-cut division exists between these methods and that overlapping occurs to a great extent.
Compounding, for instance, is a method of word-formation which conforms to the existing patterns of derivation (Andrzejewski 1979). It is for this reason that Poulos and Msimang (1998) treat compounding as a nominal deriving system. However, not only derivation occurs in compounding but also an element of semantic shift and abbreviation (blending) of words, e.g. **u-ma-bona** (see) + **kude** (far) > **umabonakude** (television).

Also the divisions between compounding and loan-translation are far from clear-cut. Had the loan-translation **isibulala magciwane** (*killer of germs* - antibiotic), for instance, been written as one word or hyphenated, it could have been considered a compound. The conclusion that can thus be drawn is that a clear-cut division of word-formation methods is not always possible in Zulu, but it has been pursued in this

chapter for the sake of systematisation.

According to many scholars reliance on the existing resources of the language, as in the methods of derivation, semantic shift and compounding, is preferable to borrowing in the coinage of terms. Most purists are not in favour of indiscriminate borrowing as they feel that it may eventually suppress the source language and pollute it (Ohly 1987). In Zulu, for instance, the following specific kinship terms have been suppressed, namely **udadewethu** (my/our sister), **udadewenu** (your sister) and **udadewabo** (his/her/their sister) by a single Afrikaans/English borrowing **usisi** (sussie/sister).

However, Mochaba (1987) points out that the advantage of using borrowing in the coinage of international terms is that similarity in the form of technical terms would promote acceptability. Also Fourie (1993) mentions some other advantages, i.e. that terms become available through spontaneous borrowing and that borrowing is faster and cheaper than purist coinage. According to Andrzejewski (1979) one cannot always prevent borrowing from occurring, especially when a language has to cope with the influx of Western concepts. The only way in which borrowing will not pose a threat to purist coinage is for the indigenous languages to be used and maintained continuously.

The African languages are not simply recipients of foreign terms as may be believed. On the contrary, they supply the Western world with some unique terms, some of which are untranslatable into the European languages (Ohly1987), e.g. **safari** (safari), **indaba** (story, case), **uphuthu** (thick porridge, pap), **imamba** (snake name) and **inyala** (antelope name).

In this chapter the manner in which methods of word-formation in Zulu can facilitate language elaboration, and also the way in which these methods are linked to extra-linguistic aspects such as culture, were investigated. World view and taboo are some of the culture-related aspects which were discussed and exemplified against the needs of terminological elaboration in Zulu.

The hypothesis of world view is based on the theory that a person's mother-tongue offers him/her a

framework for his/her perception of the world. Terms can, for instance, be created by semantic shift which correlate with a certain cultural world view, e.g. **ukubeletha** (labour-noun). **Ukubeletha** literally means *to carry on the back*. The connotation with world view here is that in African culture labour is associated with the action of carrying a child on the back after the birth.

The second culture-related aspect of Zulu language development discussed, was taboo. Taboo is a type of avoidance language that refers to terms that are not allowed to be used in a specific social environment because they would be considered offensive, vulgar or disrespectful. Since terminologists have to devise terms for important current health issues such as sex education and AIDS, the concept taboo has to be properly understood and applied by them.

A good example to illustrate taboo, for instance, in relation to culture, is the use of the Zulu term **ukukhipha isisu** for 'abortion' with the literal meaning *to remove the stomach*. According to Zulu culture it is taboo to refer to terms with a sexual connotation in a direct fashion. **Isisu** (stomach) is therefore used to indirectly and euphemistically refer to the uterus and -**khipha** (remove) to the abortion of the fetus.

The process of Zulu language elaboration, discussed with reference to methods of word-formation in relation to culture, will not be complete if a proper modern discourse dealing with technical and scientific topics does not develop in the language (Cooper 1989). The only way in which such a discourse can be achieved is through the continual development of the language, especially in the advanced technological and scientific environment. What is further needed is a properly verified standardisation process that will ensure the use of acceptable terms in the speech community. But what is needed above all else, is that the African languages be maintained and used by their speakers who should also contribute to the improvement of the status of their languages with the firm belief that their languages have sufficient elaboration capacity to participate in modern discourse.

CHAPTER 5

THE VALUE OF WRITTEN ZULU CORPORA IN SEMI-AUTOMATIC TERM EXTRACTION FOR STANDARDISATION PURPOSES

5.1   Introduction

The value of using written technical sources for term extraction for elaboration and standardisation purposes has previously not been fully exploited. The reason for this is that the point of departure was terminology created by appointed terminologists of the previous ZLB/Cs and the NTS. Ideally the written sources utilised for extraction purposes should stem from practical language use in a specific technical domain, such as medical pamphlets used at clinics and hospitals, for instance.

Before one can proceed with a discussion on research done on a written Zulu corpus (text) and its role in the extraction of technical terms for standardisation purposes, it is important that one should first come to terms with some basic concepts concerning corpus linguistics. Since the computer plays a significant role in current linguistic research (Oostdijk 1991:7), it is relevant that its role in corpus linguistics be discussed. Corpus linguistics deals with the study of language use by means of text corpora. A *corpus* is a body of language data, say texts in some language, which can be used as a basis for linguistic research and which can be stored electronically on a computer. This corpus can consist of samples of spoken (oral) or written text. In this chapter the written (published) Zulu technical (medical) texts constitute the corpus and therefore form the point of departure.

Although corpus linguistics has a wide field of practical application, such as translation, language teaching, the development of language processing machines, lexicography, etc. the emphasis in this chapter is on terminology. It must be noted that when reference is made to lexicography in this chapter, its application

to terminology is also implied.

Lexicographers and terminologists in the African languages are faced with a challenge to compile dictionaries according to the latest trends in modern lexicography (De Schryver & Prinsloo 2000b:291), since procedures have mostly been developed for the more developed European languages. The way to meet this challenge is by also using electronic corpora in order to obtain objective information about these languages.

The emphasis in this chapter is on the development of a methodology for the extraction of technical terms in Zulu in order to facilitate lemmatisation, and the criteria for the inclusion of such terms in a terminology list with the purpose of standardising them. To extract terms from a written corpus is to recognise and isolate terms in a running text. Yet, extraction is not entirely separated from the process of lemmatisation. To lemmatise is to be able to analyse the language on different levels, first and foremost, the morphological level. The terminologist lemmatises if s/he, equipped with sufficient formal grammatical insight, isolates the most basic morphological units in the Zulu word or term. For quite some time lemmatisation was performed by hand and was thus called manual lemmatisation. Today, however, the computer, through the utilisation of corpus query tools, has become a valuable aid in automatic term extraction, in so far as isolating words according to frequency, alphabetical order and context, is concerned. In the African languages, it would be safer to talk about semi-automatic term extraction, since the existing corpus query tools have not specifically been developed for the African languages, but rather for the European languages. In Zulu, being a morphologically complex conjunctive language, the role of manual term extraction cannot be ruled out.

For this chapter the latest existing written corpora available in a specific technical field were used as a basis for research. Medical pamphlets were gathered because they represent an existing medical written Zulu discourse which reveals the use of terms in context. These medical pamphlets are generally available at clinics and hospitals and are mainly distributed by the Department of Health. For the purposes of this chapter, however, only pamphlets covering health care issues such as cholera, tuberculosis and HIV/AIDS were selected.

Oostdijk (1991:3) describes in simple, yet accurate terms how linguistic research making use of corpora, is conducted:

> Once analyzed, the corpus constitutes a database that may be consulted in order to obtain information about linguistic structures, their frequency and distribution, as well as to gain insights into the co-occurrence restrictions that hold.

The practical application of this quotation in the context of this chapter is to develop a methodology for the extraction of terms from written technical texts (medical pamphlets). Eventually research analysis (text encoding) will determine which terms are worth lemmatising (listing in their most basic form) and eventually worth standardising.

5.2   The theory of corpus linguistics

Terminological research done in the field of written medical corpora in the scope of this chapter can be understood only within the framework of *corpus linguistics*, of which *computational linguistics* (CL) forms an essential part today. Computational linguistics deals mainly with the role the computer plays in the creation of corpora for linguistic research purposes (Oostdijk 1991:2,7). However, the main objective of corpus linguistics is very simply summarised by Oostdijk (1991:4), who claims it to be the study of language use and variation.

5.2.1   What is a corpus?

The word *corpus* is from Latin origin and literally means 'body'. The *Collins English Dictionary* (Hanks, McLeod & Urdang 1986) supply two significant definitions of *corpus*. One is a fairly general one, namely "the main body, section or substance of something" while the other one is more applicable to a written

corpus, namely "a collection or body of writings." However, in the context of corpus linguistics, the concept *corpus* is clearly defined by Kennedy (1998:1) 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 or description.

This corpus of language data, which can be used as a basis for linguistic research, can be stored electronically on a computer. Such corpus can consist of samples of spoken (oral) or written text. The computer corpora are fast becoming a universal resource for language research. The result is that over the past few years, corpora have increased dramatically in size and variety due to ease of access through available software (Leech 1997:2). However, corpora cannot be measured according to size only. Rather than size, of more importance is the diversity of texts used in the corpus (Leech 1997:2). In any case, scholars design corpora to suit their specific needs.

However, the method followed for text encoding (term extraction and analysis) in a small corpus can very well be made applicable to a more substantial corpus where research is conducted on a much bigger scale, on a national level, for instance. A big corpus has the capacity of ultimately becoming a growing, *organic* one in which, according to Atkins et al. (1992) in De Schryver & Prinsloo (2000a:92), the living language grows. This means that after a specific representative corpus has been built and analysed, some material will be added or deleted to enhance the corpus. Thereafter the whole process will start all over again. However, before an organic corpus can be attained, corpus compilers will have attempted to put some structure in the range of assembled texts. The compiler selects a number of texts from an available electronic range which may be grouped into sub-corpora, depending on the type of research conducted. This preliminary corpus is then called a *structured corpus*, the very type that exists in the African languages thus far compiled by De Schryver and Prinsloo (2000a:92).

However, although there are many advantages corpus linguistics can offer language research, a compiled

corpus cannot be regarded as the absolute representative of language, a fact motivated as follows by De Schryver and Prinsloo (2000a:92):

> ...linguists disagree whether a corpus should try to be balanced or representative. It seems as if a corpus will never be balanced because there are too many parameters, and it seems as if a corpus will never be truly representative of all language usage, either, as it is impossible to define the population.

Leech (1987:3) confirms that a corpus based approach towards language makes no claim to interpret language in an entirely human manner; but admits that the computer can analyse a relatively large amount of language data to compensate for this shortcoming. However, corpora can be put to good use in revising and improving existing dictionaries (and terminologies, which is the emphasis of this study) (De Schryver & Prinsloo 2000b:305). One can, for instance, in this manner ensure that a dictionary (and terminology list), reflects the language used by native speakers and not that of the lexicographer (and terminologist) (De Schryver & Prinsloo 2000c:311). Furthermore, it was proven by De Schryver and Prinsloo (2000c:327) that corpus based queries can be more accurate than the intuition of a mother-tongue speaker with lexicographical training on capturing the possible contexts of language use of a specific entry (lemma). Nevertheless, it has become a known fact that corpus lexicography (and terminology) has now become more relevant than manual lexicography (and terminology) (De Schryver & Prinsloo 2000c:311). Furthermore, querying corpora has become an essential part of modern dictionary (and terminology) compilation and also a feasible reality for the African languages (De Schryver & Prinsloo 2000c:327-328).

A logical step forward would now be to put the theory of corpus linguistics as briefly discussed above, to practical use by compiling a structured corpus.

5.3   The theory and practice of the compilation of a structured corpus

Once the need has been established to compile a structured corpus, there are three basic steps to be followed, namely corpus design, text collection and text encoding (De Schryver & Prinsloo 2000a:92 -96). In this discussion of structured corpus compilation these three steps are specifically applied to the field of research done on term extraction with a view towards eventual lemmatisation. Acceptable

terms have to be listed and eventually standardised for new fields in technology. In the following discussion a degree of overlapping occurs between corpus design, text collection and text encoding since they are interrelated processes, although not necessarily consecutive.

5.3.1   Corpus design

Different types of corpora can be distinguished which will eventually determine the design of a corpus. There is, for instance, a written corpus, a spoken corpus, a corpus of children's language and a subject specific corpus (e.g. electronics), etc. The dynamic type of corpus is an open-ended language bank in which the corpus constantly changes to include new material and replaces old material (Sinclair 1991 in De Schryver & Prinsloo 2000a:93). This type of corpus deals with a vast number of running words and is likely to exist in the undertaking of major national terminology and dictionary projects.   This dynamic corpus obviously does not feature in this chapter which caters for a small manageable corpus design. However, in the South African context a dynamic corpus may refer to the dictionary projects the *National Lexicography Units* (NLUs) are presently undertaking or to a major African language lexicographical research project at the University of Pretoria initiated by De Schryver and Prinsloo. The African language corpora built at this institution are thus known as *University of Pretoria Corpora*, e.g. the *University of Pretoria Sepedi Corpus* (PSC), the *University of Pretoria Zulu Corpus* (PZC), etc.

In this chapter the design of the corpus is directed towards a small written corpus which comprises eleven medical pamphlets as texts, chosen because they represent a medical Zulu discourse which reveals the use of terms in context. Eventually three sub-corpora, based on cholera, tuberculosis and HIV/AIDS, were

built by grouping together pamphlets dealing with the same medical sub-topic. To cater for variety, the advice of Leech (1997:7) was followed in the corpus design, namely to include a wide range of representative text types. As a result, as far as possible, pamphlets with different origins (produced by different institutions and translators) were chosen. Although the medical corpus in this chapter is rather small, the aim is to make this small corpus exemplary of the type of research that can be done in the field of written medical terminology. Furthermore, it is more convenient to practically exemplify by means of a small corpus within the scope of a single chapter.

For practical reasons it is at this stage important to render the structural design of this written medical corpus comprising eleven texts, subdivided into three sub-corpora as follows:

Main written medical corpus (with mentioned pamphlets as text sources) - total of 2187 words

Sub corpus 1 - Cholera   485 words

1   *ICholera isifo esingabulala abantu* (Cholera, the sickness that can kill people) - Pretoria: Department of National Health and Population Development;

2   *Cholera isifo sohudo - yazi amaqiniso ukuze uphile* (Cholera - know the facts and stay alive) - Gauteng Provincial Government: Department of Health;

3   *Isifo sohudo ikholera - ukuhlanzeka kuyasiza ekuvikeleni ikholera* (Cholera, hygiene helps you to prevent it) - Durban Local Councils: Health/Ezempilo KwaZulu Natal.

Sub corpus 2 - Tuberculosis   703 words

4   *Konke okufanele ukwazi ngesifo sofuba noma i-TB* (Everything you should know about tuberculosis (TB)) - Pretoria: Department of Health;

5   *Isifo sofuba siyalapheka* (Tuberculosis is curable) - Pretoria: City Council of Pretoria - TB Services;

6   *Hlolelwa i-TB mahhala* (Your free TB-test) - Gauteng Provincial Government: Department of Health;

7   *Isifo sofuba i-TB kanye ne-HIV / AIDS* (Tuberculosis and HIV / AIDS) - Pretoria: Department

        of Health, AIDS Helpline.

Sub corpus 3 - HIV/AIDS   1340 words

8   *Ingculazi emphakathini* (p 24-31) (Aids in the community) - Johannesburg: National HIV/ AIDS

        Programme Department of Health, sponsored by Old Mutual, BP, AIDS Helpline, European

        Union, the Open Society Foundation for South Africa and UNAIDS;

9   *Uthando...ukuvikela umndeni wakho kungculaza* (Love is ... to protect your family against

        AIDS) - Pretoria: Department of National Health and Population Development;

10   *Abesifazane, abantwana neHIV (Women, children & AIDS)* - Pietermaritzburg AIDS Action

        Group:Department of National Health and Population Development;

11   *Amaphuzu abalulekile nge-HIV /AIDS* (Key points about HIV/AIDS) - Pretoria: Department of

        Health, AIDS Helpline.

5.3.2   Text collection

The basis of the corpus in this chapter was medical pamphlets written in Zulu. A number of medical pamphlets meant for dissemination in the primary health care sector were selected from an available range of medical texts. The pamphlets, mainly distributed by the Department of Health, were collected at public clinics and state owned hospitals in Gauteng (Pretoria and Johannesburg) and KwaZulu-Natal (Durban). These health pamphlets cover some major issues of health care in this country, namely tuberculosis, cholera and HIV/AIDS.

In this chapter the purpose of compiling this medical corpus is to extract (or lemmatise) terms and then compare different synonymous terms with one another in order to establish which ones are actually used in the general written Zulu discourse to such an extent that they merit standardisation.

However, there are three ways of entering relevant collected written material into computer files, i.e. by

means of:

i)   'Electronic transfer' where texts are downloaded from the internet and documents retrieved which are already available on computer;

ii)   '(Re)keyboarding' where handwritten or printed texts are typed into computer files;

iii)   'Scanning' where *Optical Character Recognition*-tools (OCR) are used to scan printed matter into computer files using specific computer software.

In the context of this chapter, rekeyboarding and scanning options were used to enter medical texts into computer files. Basically processes ii) and iii) were reversed. Initially the scanning of text did not seem to pose a problem although the usual slight changes in format were expected. However, in many instances, for quite a number of texts, the painstaking rekeyboarding and/or correcting of whole sections was the only option to present the text in readable chronological format. However, only after relevant adjustments and editorial corrections were effected, could the text be utilised for encoding purposes.

5.3.3   Text encoding

When dealing with a corpus, one should also understand a basic concept that goes hand in hand with it, namely *coding* or *text encoding*, also known as *corpus annotation*. In this chapter, the term text encoding is used for the sake of consistency.   It must be clear that text encoding is not found in the original (also called *raw*) text, in this case the 'given' (Leech 1997:3) written corpus, but is information derived from the text. Encoding is thus the process of adding 'enriching information' to the text (Leech, Myers & Thomas1995:2).

Text encoding follows after the processes of text design and computerised text collection and takes place when the raw text is encoded by corpus processing annotations. Since the focus of this chapter is on corpora building in the African languages, complex corpus processing annotations which mostly apply to well studied languages such as word tokenisation, part-of-speech tagging, syntactic parsing and markup,

are not discussed in detail. Word tokenisation, for instance, cannot be made applicable to a conjunctive language such as Zulu where word boundaries are not marked by spaces and punctuation marks. Tagging is valuable since it can be used to draw up concordance lines. Lemmatisation is tagging combined with detailed morphological information which enables one to lemmatise a corpus. Nevertheless, as experienced corpora builders in the African languages by now, De Schryver and Prinsloo (2000a:95,96) rightfully feel that it is not worth pursuing the level of detailed text encoding in the corpora of European languages in the African languages:

> We are convinced that the level of detail in encoding an (entire) electronic corpus has to be related to the potential use of what will be made of that corpus. Considering the huge cost and huge manual effort, pre-analysing and marking up African language corpora right from the start (through word tokenisation, part-of-speech tagging, lemmatisation, syntactic parsing and markup) does not seem justifiable.

However, corpora can be put to better use with corpus query tools which have the capacity to deal with a large number of computer text files, handle files stored as plain texts and marked up format, calculate statistics, supply alphabetical and frequency word lists and supply concordance lines (De Schryver & Prinsloo 2000a:96). The computer program, *WordSmith Tools*, used for text encoding in this chapter, has this required capacity.

The conduct of a frequency count of a word or term can be regarded as an important and valuable tool for terminological and lexicographical purposes. By means of a frequency count, for instance, one can determine the following:

* The popularity of a term or its functional use in a specific subject field by frequency of occurrence. Such popularity is determined by the rank of items in frequency lists, overall counts (total number of occurrences) in the entire corpus and distribution of the term across the different sub-corpora.

* How many synonymous terms exist for the same concept.

* What terms to list as entries (lemmas) in a draft terminology list, glossary, thesaurus or dictionary, as far as relatively new medical, specifically AIDS-related, terminology is concerned, for instance. An item with a substantial overall count/frequency and with reasonable spreading across sub-corpora, is likely to be included in lemmatised format.

A frequency count conduct can overcome many shortcomings in traditional lexicographical (and terminological) practices as argued by De Schryver and Prinsloo (2000a:97):

> The value of frequency count is twofold. Firstly, to ensure that frequently used words are not accidentally omitted as in a typical shortcoming of the traditional way according to which lexicographers added words to the dictionaries 'as they crossed the compiler's way'. Secondly, to ensure that precious dictionary space is not utilized by words 'unlikely to be looked up' by the target user. 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.

The absence of commonly used words in a dictionary or terminology list, a common occurrence in most African language dictionaries/terminologies, could have been avoided if the compilers had had access to frequency counts even if only small corpora were utilised (De Schryver & Prinsloo 2000b:294). Lemmata which show low and uneven distribution do not qualify to be included in a dictionary/terminology list amongst the frequently used words (De Schryver & Prinsloo 2000b:302). Lemmata link up with another important action corpus query tools must deal with, namely lemmatisation.

5.3.3.1 Lemmatisation

The process of lemmatisation is defined by Kennedy (1998:297) as follows:

> ...lemmatization is a process of classifying together all the identical or related forms of a word under a common headword, just as in dictionary-making many of the various morphological inflections or derivations of a word are listed under a single entry.

Lemmatisation is a morphological analysis of a word in order to reduce it to its most basic form, also called a canonical form by some scholars. In English this would mean to reduce the verb to its infinitive, discarding inflections such as *-s*, *-ed*, *-ing,* e.g. in *produces*, *produced* and *producing* respectively, to form *produce*. Nevertheless, tools with automatic taggers exist to lemmatise corpora in languages with a simple inflectional morphology like English. However, for the African languages, including Zulu, no such electronic tools specifically developed for the purpose of lemmatising, exist. It is promising though, that lemmatisation tools may be developed in the near future, considering the research being conducted by Bosch and Pretorius (2002) on developing a computational morphological analyser for Zulu.

In the study of African languages there is thus no option left but to still lemmatise manually, even if computational lemmatisition is used. In the case of Northern Sotho for the *University of Pretoria Sepedi Corpus* (PSC) this was done by lemmatising the first top thousand items of the word frequency list, which was in itself no easy task (De Schryver & Prinsloo 2000b:300,301). For a language such as Zulu, this process of manual lemmatisation is even more complex than for Northern Sotho since Zulu is inflectionally and morphologically quite a complex language with a conjunctive writing system. The sentence **Basazoyisebenzisa** (They shall still use it), for instance, is one word in Zulu consisting of seven different morphemes **ba-sa-zo-yi-sebenz-is-a** whereas the English equivalent consists of five words. The fact that a language has a disjunctive or conjunctive manner of writing (see 3.2.2), has far-reaching implications on the size of a corpus. It follows logically that a raw (unlemmatised) corpus of a disjunctive language such as Northern Sotho will have about twice the size of a similar type of raw corpus of a conjunctive language

such as Zulu. Yet, English is still a more isolating language with clearer word boundaries than the African languages.

Morphological analysis, also known as 'hyphenation', is particularly necessary when lemmatising a Zulu corpus. To illustrate this point the following discourse examples quoted from the index of a pamphlet *Ingculazi emphakathini*:1 can be considered:

> **Abantwana abanengculazi** (Children who have AIDS)...
>
> **Izimpawu zengculazi zasemzimbeni** (Physical signs of AIDS)...

Initially the underlined words would appear to be different words. Lemmatisation, which involves a morphological analysis of each underlined word, illustrating the complexity of inflections appearing in each, proves that this is not the case:

aba-	relative concord class 2
-na-	associative formative (*note that the -a- of -na- coalesces with -i- to form -ne-*)
-i(n)-	class 9 noun prefix
-ngculazi	noun stem

za-	possessive concord class 10
-i(n)-	class 9 noun prefix (*note that the -a- of -za- coalesces with -i- to form -ze-*)
-ngculazi	noun stem

From the analysed examples above it becomes clear that the basic form appearing in both words above is -**ngculazi**. This basic form is the one which will appear as a lemma in a dictionary. In a term list, however, the full noun **ingculazi** (AIDS), complete with the prefix **-i(n)-** and the stem **-ngculazi** is listed.

Lemmatisation is also necessary and even more complex when dealing with other parts of speech such as verbs. Consider the following discourse quoted from the same pamphlet *Ingculazi emphakathini*:31:

> **Abanye abantu bathi ikhondomu <u>alikuvikeli</u> kwingculazi nakwezinye izifo ezithathelwana ngocansi. Akulona iqiniso.**
> **Ikhondomu <u>ikuvikela</u> kakhulu kabi uma uyisebenzisa ngendlela efanele.**
> (Some people say that a condom does not protect you against AIDS and other sexually transmitted diseases. This is not true.
> A condom gives you a great deal of protection if it is used correctly).

At first glance the underlined words would appear to be different verbs. Lemmatisation, which involves a morphological analysis of each underlined verb, illustrating the complexity of inflections appearing in each, proves that this is not the case:

a-   negative morpheme

-li-   subject concord class 5 present tense

-ku-   object concord 2nd. person singular

-vikel-   verb root

-i   negative verb ending

i-   positive subject concord class 9 present tense

-ku-   object concord 2nd. person singular

-vikel-   verb root   verb stem

-a   positive verb ending

In both the underlined verbs in the text above the basic form appearing in both sentences is

-vikel- (protect), the verb root. However, traditionally it is the form of the verb stem consisting of the verb

root and the positive verb ending, e.g. **-vikela** which appears as a lemma in a dictionary or terminology list. It is interesting to note that the same noun in Zulu can belong to two different noun classes, implying that they also have varying subject concords. In the first morphological analysis above the implied noun **ikhondomu** is considered a class 5 noun and thus has **li-** as its subject concord. Yet, in the latter analysis the same noun **ikhondomu** is considered a class 9 noun and thus has **i-** as its subject concord. Terminologist should be aware of this tendency of class variation in loan words which has through consistent oral use become acceptable in the Zulu language.

The point that has to be made is that lemmatisation is important for languages such as Zulu before a terminology list can be compiled. The lemma must first be recognised by the corpus compiler and this requires some linguistic insight acquired through analytical training in the relevant language. This process can simply be called manual lemmatisation. Needless to say, simply being a mother-tongue speaker does not necessarily equip one to lemmatise properly in one's language. What is needed, is formal linguistic training in the morphological and phonological structure of the relevant African language. Unfortunately, however, this type of enabling training has mostly been done away with at tertiary institutions in South Africa which offer the African languages as subjects.

As already mentioned, another important aspect corpus query tools must deal with, is to supply *concordance lines*. It was found that frequency count (*wordlist* tool) on its own cannot always be sufficient and should, especially in the African languages, be combined with *concordance* in order to grasp a term or phrase in its proper context.

In an application such as the *WordSmith Tools* the c*oncordance* or *concord* tool is helpful since it displays the use of a chosen keyword-in-context after a word search has been conducted. *Concordance* is executed when a word is extracted from the text and listed in alphabetical frequency order together with words preceding or following it so that the lexicographer (and terminologist) will have some native speaker intuition concerning the context of its use (De Schryver & Prinsloo 2000a:96). The descriptive acronym KWIC, which stands for 'keyword-in-context' is thus perhaps a better term to refer to *concordance*. The

application of the *concordance* function assists the corpus builder to see several contexts in which a word or term can be used at one glance. In this manner the builder can make decisions concerning sense (trying to cover all the sense distinctions of a lemma sign), frequent clusters and translation equivalents. The result is that representative examples (lemmas) from the living language can eventually be selected for inclusion in a terminology list or dictionary.

5.4   A proposed method for semi-automatic term extraction from a written Zulu corpus

Term extraction has to do with the recognition of terms in a specific text, here a medical corpus. Medical pamphlets were gathered to form a medical corpus and then divided according to several sub-topics concerning health care to form sub-corpora. First of all, these corpora were designed and compiled in order to be as representative as possible and to suit the compiler's needs. The application of computerised tools to a corpus is to establish what terms are actually used in the practical medical field of health care (as evidenced by the written word) and which ones should be extracted (lemmatised) and then taken up in a terminology list. In the African languages, Zulu in particular, the application of computerised lemmatisation is usually preceded by a process of manual lemmatisation, which involves morphological analysis, and therefore called semi-automatic term extraction. Refer to the earlier discussion about lemmatisation in 5.3.3.1. Eventually, after consultation with language and subject experts and dissemination to all language interested structures and institutions, it will be decided what terms merit standardisation.

The texts in the pamphlets were scanned, then edited, after which a computer program, called *WordSmith Tools* (Scott 1999) - henceforth abbreviated WST - was used in order to reveal certain tendencies in the usage of medical terms. The WST-package consists of six tools which are each devised for specific text analysis tasks, being the *Wordlist (WL)*, *Concord (C)*, *Keywords (K), Splitter*, *Text Converter* and *Dual Text Aligner*. However, for the purposes of the chapter only two of these tools, the *Wordlist* and *Concord* were used to exemplify the most basic type of text encoding in order to extract terminology from a Zulu technical (medical) corpus.

Only after a corpus file, converted to *text* (.txt) format has been selected, can WST be used. The execution of the *wordlist* tool in WST simultaneously generates three coordinated lists, namely a *frequency word list* (abbreviated F), an *alphabetical word list* (abbreviated A) and a *statistical list* (abbreviated S) dealing with bytes, tokens, average word length, sentence length, paragraph length, number of letters in a word, etc. Although the latter statistical list supplies interesting information about the corpus, it is not really a helpful tool in term extraction and is therefore not discussed any further. It should be mentioned briefly here that the settings of WST should be adjusted to suit the spelling of hyphenated words and to recognise lengthy conjunctive Zulu words.

In addition to the *wordlist* tool another handy tool of WST, called *concordance*, abbreviated *concord* or C, was used. After a corpus file has been selected this tool can be applied. This tool is helpful since it displays the use of a chosen keyword-in-context after a word search has been conducted. The frequencies (occurrences) of a particular term in all its possible immediate environment sentences can then be seen at first glance. This is indeed a helpful tool - but only if the terminologist has insight in the morphological structure of the language so that s/he can experiment with possible structures. WST can definitely help a great deal to extract and lemmatise medical terms (these two processes go hand in hand) and can even be regarded as semi-automatic term extraction tools, as was proven in Northern Sotho texts on a special field by Taljard and De Schryver (2002:57).

However, they cannot be considered the ultimate tools. The claim for semi-automatic term extraction would be much more difficult to prove in a morphologically complex conjunctive language such as Zulu. It would perhaps be relevant to conclude that the human factor cannot be discarded in terminological practice. This is confirmed by Taljard and De Schryver (2002:58):

> It will also be pointed out that human beings will always remain the final judges in any terminological activity, whether that endeavour be manual or computational.

A suggested consecutive practical method for the semi-automatic extraction of terms from a technical

(medical) written corpus for eventual selectional lemmatisation could schematically be summarised as follows:

➜	➜	➜	➜	
scan	edit text	apply WST *wordlist*	apply WST *concord*	extract terms/
pamphlets		frequency & alphabetical		lemmatise
		list		
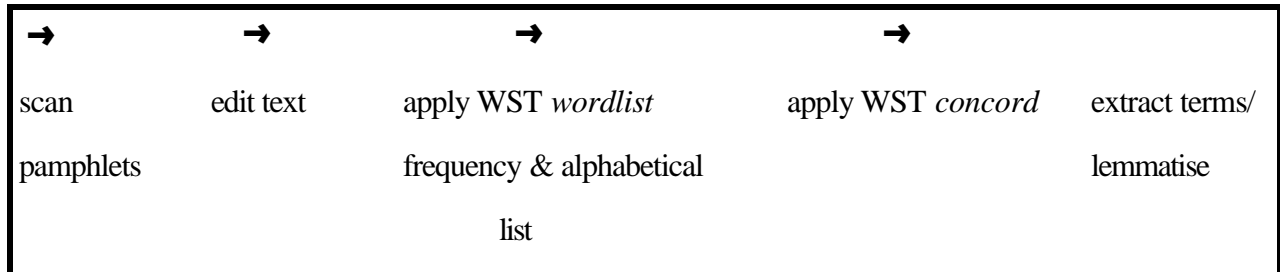
5.4.1 The use of the *WordSmith Tools*, in particular, *Wordlist* in term extraction

The application of the *WordSmith Tools, wordlist* in term extraction can best be explained in Tables 2 and 3 below. The use of the wordlist tool is discussed with relevance to two coordinated lists, namely a *frequency wordlist* and an *alphabetical wordlist*.

5.4.1.1 The frequency *wordlist*

Table 2 below is a typical random sample of single words extracted from the medical corpus according to the highest frequency of occurrence. It should be noted that the *wordlist* does not include the whole list of 2187 items since Table 2 merely fulfils the purpose of exemplification[2]. The words of the corpus are

---

[2] Note that the tables in this chapter are not necessarily complete and are merely random samples. Although all results are the outcome of the application of query tools on the corpus not all results are reflected in table format. Only selected representative tables are included for the sake of exemplification. The format in

arranged according to the highest frequency of occurrence. It must be remembered that not all words contained in the corpus are terms, e.g. in Table 2 **noma** (or) has the highest frequency, ranked number 1 (N1) with 107 frequencies/occurrences in the corpus but is merely a conjunction and NOT a term. The frequency *wordlist* (called F in short) would help to extract terms, i.e. establish the popularity of a specific term at first glance. The most obvious extracted terms, with high frequency, are **i-HIV** ranked Number 8 (N 8) with 35 frequencies, **i-TB** ranked N17 with 23 frequencies, **igciwane** ranked N 28 with 16 frequencies, **ingculaza** ranked N 29 with 16 frequencies and **izimpawu** ranked N 33 with 15 frequencies (see Table 2 below). The frequency *wordlist* would also help to extract synonymous or variant terms in order to compare their frequencies, e.g. **ingculaza** (AIDS) ranked N 29 with 16 frequencies and **ingculazi** (AIDS) ranked N 32 with 15 frequencies. Refer also to 5.5.2.2 to see how this problem can be addressed.

The ultimate goal in such term extraction is to establish which listed terms are worthy of standardisation. It must be noted that in a terminology list, the actual lemma of a Zulu noun is not listed but the noun itself. The lemma of the Zulu term for 'illness' **isifo** is listed without its prefix as **-fo** in a dictionary but as a full noun, complete with prefix, in a terminology list as **isifo**. This has been the practice in the compilation of terminology lists in the African languages thus far. It is sensible to list the full form, especially for foreign and uninformed users who need readily available terms at first glance.

TABLE 2

The Frequency Wordlist

WordSmith Tools -- 2002/09/11 10:29:46			
N	Word	Freq.	% Lemmas
1	NOMA	107	2.29
2	UMA	94	2.01
3	UKUTHI	67	1.43

which the tables are presented are not exactly the same format as in WST, although the contents is the same, due to the incompatibility of computer programs in the saving mode.

4	UMUNTU	54	1.16
5	WAKHO	43	0.92
6	ABANTU	39	0.83
7	UMNTWANA	38	0.81
8	I-HIV	35	0.75
9	SIFO	29	0.62
10	FUTHI	28	0.60
11	KUFANELE	28	0.60
12	OCANSINI	27	0.58
13	IZINGANE	26	0.56
14	ISIFO	25	0.54
15	UKUDLA	25	0.54
16	GCIWANE	24	0.51
17	I-TB	23	0.49
18	KANJANI	22	0.47
19	AIDS	21	0.45
20	KANYE	21	0.45
21	AMANZI	20	0.43
22	LOKHU	20	0.43
23	UKUBA	20	0.43
24	NGAPHAMBI	19	0.41
25	LELI	18	0.39
26	SOFUBA	18	0.39
27	LESI	17	0.36
28	IGCIWANE	16	0.34
29	INGCULAZA 16	0.34	
30	KOKUBA	16	0.34
31	NJALO	16	0.34
32	INGCULAZI	15	0.32
33	IZIMPAWU	15	0.32

5.4.1.2 The alphabetical *wordlist*

Table 3 below is a typical random sample of alphabetically arranged single words extracted from the medical corpus; also indicating frequency of occurrence. The alphabetically arranged words would help the terminologist to trace a specific term at first glance as well as its frequency of occurrence, e.g. **abaguli**

is the 6 th alphabetical word in the corpus with 2 frequencies. This alphabetical list can be used simultaneously with its coordinated frequency list, for instance, Table 2 above. The term **isifo**, for instance, is ranked N 14 with <u>25 frequencies</u> in Table 2 above and if this term is looked up in the alphabetical list, Table 3 below, it is ranked as N 724 which means that it is the 724 th alphabetical word in the corpus with <u>25 frequencies</u>.

The application of the *wordlist* to a corpus results in isolating each word that is used in the corpus (as seen in Tables 2 and 3). The task of the terminologist, however, is to extract terms from this list and eventually lemmatise those new or unknown technical items with a relatively high frequency of use in the written corpus, relevant for a specific medical field, in this case cholera, tuberculosis and HIV/AIDS. The ultimate goal is to establish which terms need to be standardised. Unlike a dictionary which would list the lemmas or lexicon of a language, per se, the technical terminology list would only include terms which are mostly not found in general dictionaries.

TABLE 3

The Alphabetical Wordlist

WordSmith Tools -- 2002/09/11 10:41:17

N	Word	Freq.	% Lemmas
1	ABACABANGA	2	0.04
2	ABADALA	4	0.09
3	ABAFUNA	1	0.02
4	ABAFUNDISA	1	0.02
5	ABAGULA	1	0.02
6	ABAGULI	2	0.04
7	ABAGULISE	1	0.02
8	ABAGULISWA	2	0.04
9	ABAHLENGIKAZI	1	0.02
10	ABAHLUSHWA	1	0.02
11	ABAKABUKEKI	1	0.02
12	ABAKHOLWA	1	0.02
13	ABAKHULELWE	1	0.02
14	ABALASHELWA	1	0.02
15	ABALULEKILE	1	0.02
16	ABANAKEKELA	1	0.02
17	ABANALELI	5	0.11
18	ABANALESI	1	0.02
19	ABANDAKANYA	1	0.02
20	ABANE-HIV	7	0.15
21	ABANE-TB	5	0.11
22	ABANEGCIWANE	1	0.02
23	ABANENGCULAZI	2	0.04
24	ABANESIFO	1	0.02
25	ABANEZINGANE	1	0.02
	....		
722	ISIDINGO	2	0.04
723	ISIDODA	2	0.04
724	ISIFO	25	0.54
725	ISIFUBA	1	0.02
726	ISIGAMU	1	0.02

5.4.2 **The use of the *WordSmith Tools*, in particular, *Concordance* in term extraction**

In addition to the *wordlist* tool another handy tool of WST, called *concordance* can be used. This tool is helpful since it displays the use of a chosen key-word-in-context after a corpus text file has been selected and a word search has been conducted. *Concordance* is executed when a word is extracted from the text and listed in alphabetical order together with words preceding or following it. The use of a particular term, here **izimpawu** (symptoms) in all its possible contexts (occurrences in sentences) reveals 15 occurrences - see Table 4 below but also refer back to Table 2 (see 5.4.1.1). By applying the *concordance* tool, i.e. by entering the search word **izimpawu**, the contexts in which this term is used in the corpus can be seen at first glance.[3]

The *concordance* tool allows for words to be broken off at the beginning or end of a line , e.g. **Ukuk** (N4) instead of the full word **Ukukhwehlela** and **la** (N6) instead of **zokugula**, etc. (see Table 4). Since the complete word, **izimpawu** is searched, it must be entered as such under the option 'new word search'. After a search has been conducted and data and samples have become available, one can make type of corpus annotation (observation) such as the following:

In Table 4 the noun **izimpawu** is mostly part of a possessive construction indicating 'symptoms/ signs of ...', e.g. **izimpawu zokugula** - signs of becoming sick (N3), **IZIMPAWU ZE-TB -** symptoms of TB (N4), **IZIMPAWU ZENGCULAZI** - symptoms of AIDS (N5), etc. Such combinational patterns can for instance help the terminologist in making predictions about the general use of a term in context, for instance, that **izimpawu** is generally part of a possessive construction.

---

[3] All the occurrences are put underneath each other and shaded with colour so that they can stand out. Since the same WST format could not be attained here, the search word is underlined to stand out.

TABLE 4

Basic Concordance

WordSmith Tools -- 2002/09/13 12:15:57

IZIMPAWU: 15 entries (sort:5L,5L)

N	Concordance	Set Tag	Word No.	File	%
1	atholakala endlini yomuntu. IZIMPAWU Uhudo. Ukuhla.		746	est~1.txt	16
2	ezifeni eziyingozi. Azikho izimpawu ezisobala ezikhomba	2,852	est~1.txt	60	
3	eminingi ngaphambi kokuvela izimpawu zokugula, kodwa	3,766	est~1.txt	81	
4	nomoya owaneleyo. YIZIPHI IZIMPAWU ZE-TB? Ukuk		1,072	est~1.txt	23
5	LAZI EMPHAKATHINI IZIMPAWU ZENGCULAZI BA		2,070	est~1.txt	44
6	la, kodwa abantwana bona izimpawu zivela bengakawuqedi		3,771	est~1.txt	81
7	khwehlela, ukujuluka nokonda Izimpawu zakamuva zengcul		2,219	est~1.txt	47
8	soshukela Ingculazi Ukuguga IZIMPAWU ZESIFO SOF		1,543	est~1.txt	33
9	hamba uyohlolwa mahhala manje. IZIMPAWU Uma unalezi		1,685	est~1.txt	36
10	ingaze ikubulale YIZIPHI IZIMPAWU ZE-TB? I-TB yisifo		923	est~1.txt	20
11	ekhulelwe zijwayele ukuveza izimpawu zalesi sifo zingaka	4,178	est~1.txt	89	
12	wo manzi angcolile asemfuleni. IZIMPAWU ZESIFO SECH		497	est~1.txt	11
13	lu kusho ukuthi usunengculazi. izimpawu zalesi sifo	2,171	est~1.txt	46	
14	phe emlonyeni noma ngaphansi izimpawu ze-TB - ukukhwehl		2,213	est~1.txt	47
15	ampilo noma esibhedlela usabona izimpawu zokuqala zanom		2,496	est~1.txt	53

5.5 Linguistic analytical and technical aspects in term extraction from a written Zulu corpus

When extracting terms from a written technical corpus, there are some important practical aspects to remember. These aspects can mainly be divided into linguistic analytical factors and technical factors. At this stage it is assumed that the terminologist has become skilled enough to experiment with computerised WST especially as far as the *wordlist* and *concordance* tools are concerned. See 5.4.1 and 5.4.2.

5.5.1 Linguistic analytical aspects

When dealing with linguistic analytical aspects in Zulu term extraction, a sensible starting point would be the word categories, these mainly being the nominal, verbal and adjectival categories.

5.5.1.1 The nominal category

When dealing with term extraction in the nominal category, uninflected nominal terms, inflected nominal terms, and nominal multi-terms should be considered.

i) Uninflected nominal terms

The easiest terms to extract and lemmatise would be single nouns in their uninflected form. One such an example is **izimpawu** (symptoms), a term to be considered for inclusion in a medical term list since it appears 15 times and has quite a high frequency - the 33rd ranked word in the corpus of 2187. See 5.4.1.1 Table 2.   In the alphabetical list (a continuation of Table 3, but not reflected here) it is the 773rd alphabetical word in the corpus with 15 frequencies. It must be noted that a term such as **izimpawu** is always used in the plural form and will therefore be listed as such since its singular **uphawu** is hardly ever used in this particular medical context.

Other terms which are also easily lemmatised would be the nouns **ingculaza** (AIDS) and **igciwane** (virus), also uninflected. After the *wordlist* tool has been applied both **ingculaza** and **igciwane** appear in the frequency wordlist (see 5.4.1.1 Table 2) with 16 frequencies, which is also quite a high frequency in the corpus of 2187 words.

ii) Inflected nominal terms

To extract or lemmatise inflected nominal terms is much more complex than to do the same for uninflected ones, especially in an agglutinating language such as Zulu. The frequency of **izimpawu** (15) derived from the statistics in a *wordlist* (frequency and alphabetical *wordlist*) would be much higher if all inflected forms

of this same noun were also included. In order to trace all the realisations of a term (uninflected or inflected) in context, here **izimpawu**, the application of the *concordance* tool would be of great help. By manipulating the *concordance* tool to suit the structure of the Zulu language, a word search can be conducted. The search word would have to be the basic morphological form such as the noun stem **-mpawu** which features in almost all its realisations. If a bound morpheme like the latter stem, in other words, not the full word is searched, it must be entered as ***mpawu** - the * indicating where the word is broken. Rather, the * provides for doing so called *wild cart* searches when one is searching for part of a word within a whole word. After *concordance* has been executed, both the uninflected and inflected forms of **izimpawu** appear. Such inflected forms appear in Table 5 as **zimpawu** (N10, N15, N21); **Kunezimpawu** (N13); **benezimpawu** (N16); **onezimpawu** (N17) and **enezimpawu** (N19). However, all these inflected forms are listed as words separately from **izimpawu**, each with its own frequency, derived from the statistics in the *wordlist*, e.g. **zimpawu** (3); **Kunezimpawu** (1); **benezimpawu** (1); **onezimpawu** (1) and **enezimpawu** (1). By utilising *concordance*, it has thus been proven that the frequency of the term **izimpawu**, has now increased to 23 which is much higher than the initial frequency of 15 (refer back to 5.4.1.1 Table 2 and 5.4.2 Table 4). If all the realisations (indicated as *entries*) of **izimpawu** in Table 5 are taken into account, it is thus in this context appropriate to call this increased frequency of 23 the *actual frequency* (own designation) of the term **izimpawu**. It is interesting to note that by searching for ***mpaw*** the locative forms such as **ezimpawini** (of **izimpawu**), if any, could also have been captured.

The same *concordance* method followed to trace the actual frequency of the term **izimpawu** can in principle be made applicable to a term such as **ingculaza** to prove that it has a higher frequency than 16 (as reflected in 5.4.1.1 in Table 2). The inflected forms of the latter term are listed as separate words, each with its own frequency derived from the statistics in the *wordlist*, e.g. **kungculaza** (2), **lengculaza** (4), **nengculaza** (1), **sengculaza** (9), etc. Yet, all these inflected nouns have a similar noun stem, namely **-ngculaza**. In order to trace all occurrences of this term in context (inflected or uninflected), the application of the *concordance* tool by entering a word search is of great help. The word searched for will thus be the radical morphological form, i.e. the noun stem **-ngculaza**. If a bound morpheme like the latter stem, in other words not the full word is searched, it must be entered as ***ngculaza**, the * indicating where the

word is broken.   It has thus been proven that the frequency of the term **ingculaza**, now has increased to 34 which is much higher than 16 if all its realisations (indicated as *entries*) are taken into account.

In the context of Zulu, being the morphologically complex language that it is, *concordance* has particular significance in the lemmatisation process since the frequency of words/terms or even parts of words, (stems, roots, etc.), i.e. lemmas can be searched in the discourse of the text by conducting a word search. In this manner the actual frequency of a word/term can be established at first glance when the *wordlist* tool proves insufficient.

TABLE 5

Concordance: Actual frequency of a nominal term

WordSmith Tools -- 2002/09/11 12:27:50

xMPAWU: 23 entries (sort:5L,5L)

N	Concordance	Set Tag	Word No.	File	%
1	tholakala endleni yomuntu. IZIMPAWU Uhudo. Ukuhlanza.		746	est~1.txt	16
2	ezifeni eziyingozi. Azikho izimpawu ezisobala ezikhomba	2,852		est~1.txt	60
3	eminingi ngaphambi kokuvela izimpawu zokugula, kodwa	3,766		est~1.txt	81
4	nomoya owaneleyo.YIZIPHI IZIMPAWU ZE-TB? Ukukhw		1,072	est~1.txt	23
5	GCULAZI EMPHAKATHINI IZIMPAWU ZENGCULAZI		2,070	est~1.txt	44
6	ukujuluka nokonda Izimpawu zakamuva zengculazi Ungaba		2,219	est~1.txt	47
7	gula, kodwa abantwana bona izimpawu zivela bengakawuqe		3,771	est~1.txt	81
8	soshukela Ingculazi Ukuguga IZIMPAWU ZESIFO SOFU		1,543	est~1.txt	33
9	- hamba uyohlolwa mahhala manje. IZIMPAWU Uma unalezi		1,685	est~1.txt	36
10	PAWU Uma unalezi zimpawu ezilandelayo, kufanele uye ekli		1,688	est~1.txt	36
11	ingaze ikubulale YIZIPHI IZIMPAWU ZE-TB? I-TB yisifo e		923	est~1.txt	20
12	unina ekhulelwe zijwayele ukuveza izimpawu zalesi sifo zing		4,178	est~1.txt	89
13	uma sengiqala ukugula? Kunezimpawu eziningi zengculazi,	2,079		est~1.txt	44
14	manzi angcolile asemfuleni. ZIMPAWU ZESIFO SECHOLE		497	est~1.txt	11
15	Ngisho noma olulodwa lwalezi zimpawu ezingenhla lungaba		1,122	est~1.txt	24
16	bathole usizo, uma betholakala benezimpawu zalesi sifo. Inin		1,187	est~1.txt	26
17	amasotsha omzimba. Umuntu onezimpawu zeTB kufanele aye	1,578		est~1.txt	34
18	kusho ukuthi usunengculazi. Izimpawu zokuqala zengculazi U		2,171	est~1.txt	46
19	noma kubanda. Uma umuntu enezimpawu ezimbili noma ezin		1,096	est~1.txt	24
20	umhlophe emlonyeni noma ngaphansi izimpawu ze-TB - ukuk		2,213	est~1.txt	47
21	ngamandla Uma ubona noma uzwa lezi mpawu, sheshisa uye		516	est~1.txt	11
22	Umuntu one-TB angaba nezinye zalezi zimpawu Ukukhwehle		1,924	est~1.txt	41
23	olampilo noma esibhedlela usabona izimpawu zokuqala zano		2,496	est~1.txt	53

iii) Nominal multi-terms

It must be remembered that a corpus of lemmatised medical terms does not only contain single terms such as **isifo** (illness) but also contains multi-terms such as **isifo sofuba** (tuberculosis, literally *the illness of the*

chest) which consist of two or more words. These multi-terms are listed as separate words with different frequencies in the same *wordlist*, i.e. as **isifo** (25) and **sofuba** (18). It thus proves that although WST is user friendly, the *wordlist* tool has limitations as far as extracting multi-terms and thus indicating their actual frequencies in the African languages, here Zulu, is concerned. Since the *wordlist* won't isolate multi-terms, the tool of *concordance* can perform this task with some manipulation. A *concordance* word search is conducted by entering the multi-term **isifo sofuba** to reveal a frequency of 10 in the medical corpus. See Table 6 below.

The frequency of another indigenous multi-term term **isifo sohudo** (cholera, literally *the illness of diarrhoea*) cannot be established by means of the *wordlist* tool since the terms are each listed separately with different frequencies as **isifo** (25) and **sohudo** (4). Yet again, the *concordance* tool solves this problem, similarly to that of the previous term **isifo sofuba**. A *concordance* word search is conducted by entering the multi-term **isifo sohudo** to reveal a frequency of 4 in the corpus.

TABLE 6

Concordance: nominal multi-terms

WordSmith Tools -- 2002/09/11 11:37:24

ISIFO SOFUBA: 10 entries (sort:5L,5L)

N	Concordance	Set Tag	Word No.	File	%
1	lesi sifo siyathelelana. IMBANGELA ISIFO SOFUBA isifo		1,464	est~1.txt	32
2	I-TB KANYE NE-HIV / AIDS Isifo sofuba i-TB siyisifo		1,840	est~1.txt	39
3	eziyisithupa noma ngaphezulu ukuthi isifo sofuba selapheke	1,238	est~1.txt	27	
4	aye emtholampilo ayoxilongwa olashwe, isifo sofuba siyala	1,585	est~1.txt	34	
5	abantu, njengabomndeni wakho ISIFO SOFUBA I-TB KA		1,833	est~1.txt	39
6	uma litholakala ukuthi liphethwe isifo sofuba. Empeleni ak	1,200	est~1.txt	26	
7	IMBANGELA ISIFO SOFUBA Isifo sofuba sibangelwa ng		1,466	est~1.txt	32
8	WELASHWA KWESIFO SOFUBA (TB) ISIFO SOFUBA		1,218	est~1.txt	26
9	usizo. SINGABONWA KANJANI ISIFO SOFUBA (TB)?		1,115	est~1.txt	24
10	lo wesifunda sakho kumbe nabakwa: ISIFO SOFUBA SIY		1,445	est~1.txt	32

5.5.1.2 The verbal category

It is commonly known by terminologists and lexicographers alike that in the African languages verbs are even more complex to lemmatise for the purpose of compiling a term list than nouns since verbs are almost always inflected in texts, except for the imperative singular. The verb root can add prefixes in the form of concords for different persons and different noun classes and can be used in different tense or modal forms, employing different types of verbal extensions and verbal endings. When dealing with term extraction in the verbal category basic verbal terms, complex verbal terms and verbal multi-terms are considered.

i)   Basic verbal terms

The complexity of extracting verbal forms already becomes evident when dealing with basic verbs such as -**vikela** (to protect, here in relation to serious illnesses). The most basic imperative form found in the corpus, i. e. **vikela** has a frequency of 1 in the corpus as established by means of the *wordlist*. It is thus obvious that the *wordlist* will not be of much use in the extraction of verbal terms. When *concordance* is resorted to it is clear that **-vikela** features in many other realisations. The word search is conducted by using the most basic radical form that carries meaning, namely the verbal root **-vikel-,** entered as ***vikel***, the * indicating where the root starts and where it ends. A total of 33 entries are found which indicate the actual frequency of this term. See Table 7 below. Each different realisation in the *concordance* table has its own frequency which can be found in the *wordlist*, e.g. **ikuvikela** (1), **ukuzivikela** (1) **wokuvikela** (1), **lokuvikela** (1), **ukuvikela** (6), **ungazivikela** (1), **ungayivikela** (1) **singazivikela** (1), etc.

The manner of listing verbs in terminology lists thus far has been to list the stem, mostly without the hyphen. The reason for listing the hyphen here is for morphological accuracy since the verb stem is a bound morpheme that cannot have meaning without other combining morphemes:

  -vikela (protect/prevent here in relation to serious illnesses such as TB and AIDS).

Concordance is not only very handy in establishing the actual frequency of a term but also in indicating how a certain word is used in context - in relation to other words in its immediate environment. If the 33 occurrences of -**vikel**- are considered e.g. **uku*z*ivikela** (N1), **unga*z*ivikela** (N6), **yoku*z*ivikela** (N7), **koku*z*ivikela** (N10), **singa*z*ivikela** (N17), etc. (see Table 7) it becomes obvious that the form -**z̲ivikel-** which has 16 within 33 frequencies of **-vikel-**, is quite popular. The reflexive prefix **-zi-** used in the latter contexts means 'to protect oneself', a very important notion in combating serious illnesses such as TB, HIV/AIDS and cholera.

TABLE 7

Concordance: basic verbs

WordSmith Tools -- 2002/09/11 11:48:58

xVIKELx: 33 entries (sort: 5L, 5L)

N	Concordance	SetTag	Word No.	File	%	
1	zimba wakhe awusakwazi <u>ukuzivikela</u> ezifeni eziyingozi.	2,848		est~1.txt	60	
2	silapheki lesi sifo. Awukho umuthi <u>wokuvikela</u> lesi sifo. Udo		3,171	est~1.txt	68	
3	Kodwa-ke uma izinga <u>lokuvikela</u> umzimba lehlile, leli gciwa		1,023	est~1.txt	22	
4	Ukuncelisa ibele kusiza ukuvikela <u>izivikelamzimba</u> zomntwan		4,022	est~1.txt	86	
5	ngocansi. Akulona iqiniso. Ikhondomu <u>ikuvikela</u> kakhulu kab		2,784	est~1.txt	59	
6	ingathelelana ngokuya ocansini <u>ungazivikela</u> nangegazi. Ingat		2,759	est~1.txt	58	
7	ANSI OLUPHEPHILE Indlela engcono <u>yokuzivikela</u> kusifo		3,029	est~1.txt	65	
8	HUDU IKHOLERA Ukuhlanzeka Kuyasiza <u>Ekuvikeleni</u> IKH		719	est~1.txt	15	
9	izingane zakho ukuthi cha ocansini <u>olungavikelekile</u> ukugwe		4,234	est~1.txt	90	
10	AZI OKULANDELAYO Ngaphandle <u>kokuzivikela</u> kanye no		3,124	est~1.txt	67	
11	nezimpahla zasesikoleni. <u>UNGAYIVIKELA</u> KANJANI INGC		2,996	est~1.txt	64	
12	into ezimbalwo abantu abangazenza ukuze <u>bavikeleke</u> kulesi		151	est~1.txt	3	
13	Ukuncelisa ibele kusiza <u>ukuvikela</u> izivikelamzimba zomntwa		4,024	est~1.txt	86	
14	i khuponi ubese uliposela kwa <u>Ukuvikela</u> Ingculaza Igama	3,326		est~1.txt	71	
15	ohlalisana naye, unganceda futhi <u>nangokovikela</u> izingane zak		3,132	est~1.txt	67	
16	angasebenzisa <u>okokuvikela</u> (icondom) uma ningathemban	3,020		est~1.txt	64	
17	ekonakele sesingaqala isifo. <u>SINGAZIVIKELA</u> KANJANI K		141	est~1.txt	3	
....	....					....
33	ahlanzekile noma abilisiwe. <u>Vikela</u> ukudla ezimpukaneni uzix		780	est~1.txt	16	

---

ii)   Complex verbal terms

The tool of *concordance* yet again becomes quite handy since the application of the *wordlist*, as has already been proven, will not help to extract a term or give an indication of its actual frequency. A verb root like -**thel**- (pour) has extended its meaning to include the notion of 'spreading a disease' by adding the applied verbal extension **-el-**. If a word search through *concordance* is conducted it is then advisable not to search for the root -**thel**- but rather for the root plus the extension, i.e. -**thelel**- (spread disease), thus entering ***thelel***.   In Table 8 below it becomes clear that -**thelel**- has a frequency of 10 in the corpus but in the *wordlist* each realisation is listed separately with its own frequency, e.g. **ingathelelana** (1) **kungakuthelela** (1) **siyathelelana** (1), etc. It is also interesting to note that another verbal extension, namely the reciprocal **-an-** with the notion 'to spread to one another' is added to the applied extension **-el-**, e.g. **siyathelelana**. However, the suffixed passive extension **-w-**, makes the verbal construction even more complex, adding the notion of 'being spread amongst one another', e.g. **ingathelelwana**.

What makes verbal term extraction and eventual lemmatisation even more complex, is when verbs bear evidence of underlying phonological change as in the formation of passives. The structure of a vowel verb such as **-elapha** (cure/treat) can be made more complex by adding the passive verbal extension **-w-** as in the realisation **ingelashwa** (it can be cured) which bears the evidence of an underlying palatalisation process, e.g.

-**elaph-** (verb root) + **-w-** (passive extension) + **-a** (verb ending) = -**elashwa** (**ph** > **sh** -palatalisation).
It has already been proven that the *wordlist* tool will not be sufficient for extracting verbal terms to form a lemmatised list. It is thus obvious that the tool of *concordance* will have to be reverted to. Through *concordance* a word search of -**elapha** has produced a frequency of 9 occurrences in the corpus (not

reflected here). However, if the word search of its passive, **-elashwa** is conducted, it produces a frequency of 13 in the corpus, thus adding to the frequency of **-elapha** to form an actual frequency of 22.

TABLE 8

Concordance: complex verbs with extensions

WordSmith Tools -- 2002/09/11 12:01:07

xTHELELx: 10 entries (sort:5L,5L)

N	Concordance	Set Tag	Word No.	File	%	
1	phuza imithi akasenako ukuthi <u>angathelela</u> abanye lesi sifo.		1,252	est~1.txt	27	
2	ocansini ungazivikela nangegazi. <u>Ingathelelwana</u> futhi isuka		2,761	est~1.txt	58	
3	Akulona iqiniso. Ingculazi <u>ingathelelana</u> ngokuya ocansini		2,756	est~1.txt	58	
4	nye bacabanga ukuthi ingculazi <u>ingathelelwana</u> ngamanzi na		2,734	est~1.txt	58	
5	Abantu abaningi abalashelwa i-TB <u>abatheleli</u> abanye abantu			1,855	est~1.txt	40
6	ukuthinta umuntu o-HIV positive <u>kungakuthelela</u> ngengculazi			2,751	est~1.txt	58
7	zifo. I-HIV ibhebhetheka <u>ithelelana</u> ngokusuka komunye umu			4,436	est~1.txt	94
8	ongakazalwa noma osanda kuzalwa <u>athelele</u> umntwana wakhe			1,882	est~1.txt	40
9	ongakazalwa ma osanda kuzalwa <u>athelele</u> umntwana wakhe (			4,457	est~1.txt	95
10	yiTB yamaphaphu, futhi lesi sifo <u>siyathelelana</u>. IMBANGEL			1,462	est~1.txt	32

iii) **Verbal multi-terms**

Just like nouns, there are also verbal forms which form the basis for multi-terms, e.g. **-ya ocansini** (have sexual intercourse, literally *go to the sleeping mat*). The problem of extraction for the terminologist is that the words contained in such multi-terms are listed as separate words in the *wordlist*, each with its own frequency. Since the *wordlist* will not isolate multi-terms, the tool of *concordance* can perform this task with some manipulation. First of all the stem of the verb, **-ya** which lacks inflection, is isolated. Thereafter a word search is conducted by entering the multi-term ***ya ocansini** to reveal a frequency of 21 in the medical corpus. See Table 9 below.

Through experimenting with the *concordance* tool, by conducting a single word search, entering the manipulated form **-lalan-**, for instance, the following verbal multi-terms (not reflected here in a table) can be discovered, e.g. **-lalana ngendlela ephephile** (practise safe sex), **-lalana ngendlela engaphephile** (practise unsafe sex) and **-lalana ngendlela engavikelekile** (practise unprotected sex).

TABLE 9

Concordance: verbal multi-terms

WordSmith Tools -- 2002/09/11 12:06:10

xYA OCANSINI: 21 entries (sort:5L,5L)

N	Concordance	SetTag	Word No.	File	%
1	amakhondomu noma bayeke <u>ukuya</u> <u>ocansini</u>. Akulona iqiniso		2,619	est~1.txt	55
2	ngaso sonke isikhathi uma <u>uya</u> <u>ocansini</u> ngokubuza kosebenz		4,602	est~1.txt	98
3	ngadluliselwa komunye umuntu <u>ngokuya</u> <u>ocansini</u>, uma igazi		2,808	est~1.txt	59
4	Yiba nesiqiniseko esikhulu <u>ngokuya</u> <u>ocansini</u> ucophelele uku		3,002	est~1.txt	64
5	nomnane ongenaso lesi sifo. <u>Ngokuya</u> <u>ocansini</u> nezintombi n		3,040	est~1.txt	65
6	wane lengculaza landa kuphela <u>ngokuya</u> <u>ocansini</u>. Ingculaza		2,907	est~1.txt	61
7	sebenzisa ikhondomu. Yenza <u>ukuya</u> <u>ocansini</u> okuphephile. Y		2,546	est~1.txt	54
8	Ingculazi ingathelelana <u>ngokuya</u> <u>ocansini</u> ungazivikele nang		2,757	est~1.txt	58
9	ntu o-HIV positive ngaphandle <u>kokuya</u> <u>ocansini</u> naye ungaziv	2,295	est~1.txt	49
10	thengisa ngomzimba wakhe. <u>Ukuya</u> <u>ocansini</u> nomuntu osebe		3,095	est~1.txt	66
11	dakamizwa ngokujova njalo. <u>Ukuya</u> <u>ocansini</u> okungafanele o		3,102	est~1.txt	66
12	elela ukuthi umhlobo wakho <u>uya</u> <u>ocansini</u> nawe kuphela. No		2,509	est~1.txt	53
13	nobakho. Ubutabani busho <u>ukuya</u> <u>ocansini</u> nomuntu owubuli		2,671	est~1.txt	56
14	isilisa nesifazane busho <u>ukuya</u> <u>ocansini</u> nomuntu owubulili o		2,663	est~1.txt	56
15	njengokulalana ngasemuva <u>Ukuya</u> <u>ocansini</u> nowesilisa noma		3,087	est~1.txt	66
16	nokuzilethela amadlingozi <u>kokuya</u> <u>ocansini</u> uze ufeze uwedw		3,112	est~1.txt	67
17	ukuthi awuyitholi ingculazi...<u>Iya</u> <u>ocansini</u> ngendlela ephephile		2,539	est~1.txt	54
18	mu. Le zindlela ezilandelayo <u>zokuya</u> <u>ocansini</u> ziyingozi: Ok		3,074	est~1.txt	66
19	sengculaza. Kuphela uma <u>uya</u> <u>ocansini</u> nomuntu oyedwa loyo		3,054	est~1.txt	65
20	wukuphi ukugula. Yenza <u>ukuya</u> <u>ocansini</u> okuphephile. Qikele		2,502	est~1.txt	53
21	ziyingozi: Okunye nokunye <u>ukuya</u> <u>ocansini</u> okungenza ukuthi		3,079	est~1.txt	66

5.5.1.3 The qualificative/adjectival category

Adjective stems like **-ncane** (small), -**bili** (two) are commonly used as qualificatives (qualifiers) in Zulu but not generally in a terminological context. To a vowel verbal root such as **-elaph-** (cure/treat) the neuter verbal extension, **-ek-** is added to form **-elaphek-**, conveying the notion of 'curable'. Although **-elaphek-** is an entirely verbal formation, it can also be used qualificatively in Zulu and has an adjectival notion in English. Thus, in order to capture both the verbal origin and semantic context of a term such as **-elapheka**, it can be called a *verbo-qualificative* (own designation).

The practice of adding the word category such as 'adj.' next to the term, is followed in national medical terminology lists mainly to guide the user to distinguish between terms, e.g.

fat *adj*   **-nonile**, **-khuluphele** (DACST 1997a:77).

Likewise **-elapheka** would then be entered as follows:

curable/treatable *adj*   **-elapheka**.

The application of the *concordance* tool will trace all the realisations of a qualificative/adjectival term (inflected or uninflected), indicating its actual frequency which the *wordlist* fails to do. The word search would be morphologically manipulated to suit the terminologist's needs, so as not to search for the root only but for the root **-elaph-** plus the neuter verbal extension **-ek-** to form

-elaphek-. This word search is entered as ***elaphek***,   the * indicating where the term is broken down. Combinational patterns of terms as displayed through *concordance* can help the terminologist in making predictions about the general use of a term in context, i.e. initiate some form of corpus annotation. From Table 10 below, for instance, it is clear that the verbal stem **-elapheka** is mostly preceded by a subject that denotes an illness such as N1 **ingculazi - yingculazi** (AIDS), N2 N3 N4 N6 **i-TB** (TB) and N5 **isifo sofuba** (tuberculosis).

Other examples of verbs with a qualificative notion in Zulu (but an adjectival notion in English) which are

strictly speaking Zulu verbal stative forms (adding the perfect verbal extension **-ile**) that should be listed because they enrich technical terminology are, for instance:

-**hlanzekile**   hygienic/clean

and

-nukubezekile   contaminated.

TABLE 10

Concordance: verbo-qualificative

WordSmith Tools -- 2002/09/12 13:19:03
xELAPHEKx: 7 entries (sort:5L,5L)

N	Concordance	Set Tag	Word No.	File	%
1	Ukugula okuningi okudalwa yingculazi <u>kuyelapheka</u> emitho	2,123	est~1.txt	45	
2	ukukwelapha ngaleso sikhathi I-TB <u>IYELAPHEKA</u> Kufane		1,788	est~1.txt	38
3	LWA EKLINIKHI ESEDUZANE I-TB <u>IYELAPHEKA</u> noma		1,657	est~1.txt	36
4	ngubani angayithola i-TB futhi umuntu <u>angelapheka</u> ngoku	1,664	est~1.txt	36	
5	oma ngaphezulu ukuthi isifo sofuba <u>selapheke</u> ngokuphelele		1,240	est~1.txt	27
6	i ngabe unayo yini noma awunayo I-TB <u>iyelapheka</u> noma nga		1,963	est~1.txt	42
7	ede onke amaphilisi owanikiwe ukuze <u>welapheke</u> Amaphilisi		1,795	est~1.txt	39

5.5.2   Technical aspects

In the extraction of terms from a medical corpus one does not only need linguistic analytical knowledge of the Zulu language as discussed above, but one should also consider some related technical aspects such as how to deal with high and low frequency, sub-corpora, loan terms, acronyms and additional information.

5.5.2.1 High and low frequency

The words with the highest frequency in this medical corpus are conjunctions, e.g **noma** (107), **uma** (94), **ukuthi** (67), **futhi** (28), **ukuba** (20), etc. (see 5.4.1.1 Table 2). As could be established, this is generally the case in most African language corpora, regardless of language. In Zulu most of these conjunctions are all morphologically free morphemes with no inflection since they can stand on their own and need not be lemmatised.   However, for the purpose of compiling a technical terminology list, they are not terms, irrelevant and can therefore be discarded.

It must be remembered that high frequency is not the ultimate criterion for lemmatising/listing although it is a good indication of the popularity of a term. The word **uketshezi** (body fluid), for instance, has a frequency of only 2 in the main corpus, but may be regarded a better and shorter term than the longer equivalent **amanzi omzimba** with a frequency of 5 since the former word is already popularly used in the context of AIDS education.

5.5.2.2 Sub-corpora

With synonymous or variant terms it is often very difficult to decide which term is the most representative. It is interesting to note that a variant term for **ingculaza** (AIDS) appears in the form of **ingculazi**. In the *wordlist* **ingculazi** would appear with a number of 15 occurrences, which is quite a high frequency in a corpus of 2187 words. Yet its frequency would even be much higher if all inflected forms of this same term are also considered. Such inflected forms, however, are listed as words separately from **ingculazi**, e.g. **abanengculazi** (2) **lengculazi** (2), **ngengculazi** (3), **yingculazi** (5), etc. Yet, all the latter words have a similar stem, namely **-ngculazi**. In order to trace all the realisations of this stem (inflected and uninflected) the application of the *concordance* tool is of great help when the word search *ngculazi is conducted in the same manner as for ***mpawu** (see 5.4.2 Table 4), indicating that the actual frequency is 34.

What complicates the issue is to decide which term to list when both have a high frequency in the *wordlist*: **ingculaza** (16) and **ingculazi** (15) as well as a high actual frequency, as established through *concordance*: **ingculaza** (34) and **ingculazi** (34) respectively. Ideally one term would suffice. The next step now would be to consult the sub-corpora, dealing with *HIV/AIDS*, *tuberculosis* and *cholera*. In the *HIV/AIDS corpus* the frequency is: **ingculaza** (15) and **ingculazi** (14) and in the *tuberculosis corpus* the frequency is: **ingculaza** (1) and **ingculazi** (1). It has thus been established that both the terms are used interchangeably and are equally representative since they have the same frequencies in one *non-HIV/AIDS sub-corpus* (*tuberculosis)* and almost the same in the *HIV/AIDS sub-corpus*. The solution would then be that both terms should be listed as variants in a terminology list. In Zulu there are other variant nouns that are also used interchangeably such as **intombazana** and **intombazane** (girl) and **umngani** and **umngane** (friend).

Furthermore, if a word search is conducted, revealing the use of the word in context, a terminologist would sometimes unexpectedly discover a new term by linguistic intuition. Such an example is the nominal term **ukwelashwa** (cure), derived from the passive stem **-elashwa**, discovered when entering -**elashwa** as a word search. The term **ukwelashwa** has a frequency of 8 in the corpus and is much more popular than its alternative synonymous nominal term **ukwelapha** (cure) with a frequency of 1. The term **ukwelapha** appears in only one sub-corpus of *HIV/AIDS* while **ukwelashwa** is used in another *HIV/AIDS sub-corpus* with a frequency of 7 and in one *tuberculosis sub-corpus* with a frequency of 1. It is thus clear that the term containing the passive verb **ukwelashwa** is the more representative of the two in the medical corpus as a whole and should therefore be listed.

5.5.2.3 The role of acronyms

Common acronyms such as **HIV**, **TB** and **AIDS** generally have a high frequency of use in medical texts. These acronyms are loans from English and are adjusted to suit the Zulu morphological word structure, i.e. by adding a noun class prefix e.g. **i-HIV**. Most acronyms also have a high frequency in the medical Zulu corpus **i-HIV** (35), **uHIV**(5), **i-TB** (23), **i-AIDS** (2) and **i-DOTS** (2). However, even if the prefix **i-** is not always used in these terms some will still have a high frequency such as **AIDS** (21). Another problem prior to the listing of terms is an orthographical one, i.e. whether to list these acronyms with or without the hyphen, e.g. **iTB** or **i-TB**. Since the latest orthographical rules have not been updated as regards the use of a hyphen in loan acronyms, the common use of such a term in written texts could be the determining factor. The use of the hyphen is quite popular as reflected in the frequency data e.g .**i-HIV**(35) as opposed to **iHIV**(2); **i-TB** (23) as opposed to **iTB** (4). In an alphabetical *wordlist* acronyms containing a hyphen are listed first in a particular letter before the other words follow in strict alphabetical order. For acronyms it can thus be recommended that the hyphen be retained in the listing of terms. See also 3.2.2.3 in this regard.

5.5.2.4 Loan terms

Medical multi-terms are very often indigenous terms in Zulu. The question that now arises is whether it is better to use the indigenous term than the loan term. The term **isifo sofuba** (tuberculosis, literally, the *illness of the chest*) is the indigenous multi-term equivalent for the acronym **i-TB** (TB) or the loan **ituberculosis** (tuberculosis). The loan acronym **i-TB** has a frequency of 23 in the *wordlist*. The purer multi-term **isifo sofuba**, on the other hand, has a frequency of 10 as established through *concordance*. It thus proves that the purer term is not necessarily the most popular. The loan term **ituberculosis** has a frequency of only 1 in the corpus. It would be logical then to list only the two terms with the highest frequency while discarding the latter term with a very low frequency in the following manner:

i-TB/**isifo sofuba**   TB/tuberculosis

The frequency of the alternative single loan terms **icholera** (9) and **ikholera** (2), is easy to establish by means of the *wordlist*. It is interesting to note that the term **ikholera** is spelt according to the orthographical rules of Zulu, adapting the English **c** to the aspirated Zulu plosive **kh** while its alternative **icholera** is simply taken over from English and also pronounced in the same manner, merely adding a class prefix in Zulu. One cannot argue that **icholera** is not a Zulu term since it does not adhere to orthographical rules; but rather argue that it is kept as close to the well-known English medical term in order to prevent confusion and maintain similarity. It is logical then to list **icholera** (8) as a term since it has a much higher frequency than **ikholera** (2) and perhaps keep the synonymous indigenous alternative multi-term **isifo sohudo** (4), with the second highest frequency, for more clarity, (especially for use in the rural areas) in the following manner:

icholera /**isifo sohudo**   cholera.

Yet, a terminologist can never make generalisations about the spelling of terms, since each example is different and it is sometimes impossible to dictate phonological conditions. If the previous example **icholera**

versus **ikholera** is considered it would follow logically that the loan term **icondom** (condom) which is very close to English would be more popular than the Zuluised term **ikhondomu** (condom). However, it is not the case, since **ikhondomu** has a frequency of 4 which is slightly higher than the frequency of 1 for **icondom** in the corpus. It is thus advisable to list the term **ikhondomu** with the highest frequency of use.

5.5.2.5 Additional information

In a terminology list it would be useful to include additional information for the user in order to complement a specific lemma on the conditions of its use in context. For instance, if the actual frequencies of -**vikel**-, established through concordance, are considered, it is clear that the form
-**zivikel**- occurs in 16 out of a total of 33 frequencies, e.g. **yokuzivikela**, **kokuzivikela**, **singazivikela**, etc. (see 5.5.1.1 iii Table 6). The popularly used reflexive prefix **-zi-** in the latter examples means 'to protect oneself', a very important notion that will be captured in medical publications discussing contagious or preventable illnesses such as TB, HIV/AIDS and cholera.

Since this reflexive **-zi-** appears to be quite significant in the applications of the verb stem **-vikela**, the terminologist should add additional information about -**zi**- in the front or back matter of such a terminology list, marking **-vikela** with an * in the alphabetical entry, e.g.

>   lemma:
>
>   **-vikela*** (protect/prevent, in relation to contagious or preventable illnesses such as TB and HIV/AIDS)
>
>   Front or back matter:
>
>   **-vikela*** Very often the reflexive -**zi**- is used together with -**vikela** in order to express the notion of oneself, i.e. myself, ourselves, themselves, himself, herself, e. g. **Singazivikela kanjani kwingculazi?** (How can we protect ourselves against AIDS?)

It is advisable to differentiate between word categories by adding the additional *v* for 'verb', *n* for 'noun' and *adj*. for 'adjective' so as not to confuse the uninformed user in a terminology list:

-elapha   cure (v)

ukwelashwa   cure (n)

-**elapheka**   curable (adj.)

Also see 5.5.1.3 for the marking of word categories in terminology.

The multi-term, **-ya ocansini** (have sexual intercourse, literally *go to the reeds sleeping mat*), can, for instance, be listed as follows:

-ya ocansini ◇   have sexual intercourse.

The ◇ sign appearing after the term is added, for instance, to explain conditions of the use of this specific term in the front or back matter of a terminology list in order to inform users as follows:

◇ This term has a sexual connotation and therefore it conforms, in its use, to the social practice of avoidance or taboo. The term **-ya ocansini** is therefore referred to in an indirect fashion, i.e. literally meaning *to go to the reeds sleeping mat*.

5.6   Exemplifying a lemmatised terminology list

After manual and computerised lemmatisation procedures applying analytical and technical language skills to a medical corpus have been executed, an alphabetical lemmatised terminology list can be compiled (although it was not a specific aim of this chapter). The eleven medical pamphlets which formed the medical corpus were scanned, edited and medical terms extracted both manually and automatically. The use of computerised tools such as WST, i.e. by applying the *wordlist* tool and the *concordance* tool, thus formed the basis for semi-automatic term extraction.

The following term list is only exemplary of what can be achieved in written medical term extraction and does not claim to be fully representative. This term list practically exemplifies the result of not only manual,

but also manipulated computerised term extraction.

UHLA LWAMAGAMA APHATHELENE NEZEMPILO / A LIST OF MEDICAL TERMS

amajaqamba	cramps
amanzi aphuma esifazaneni	vaginal fluids
amanzi wesilisa	semen
-bhebhetheka	spread fast
-elapha	cure/treat
-elapheka	curable/treatable
emphakathini	(in the) community
-hlasela	attack (the immune system)
-hlanzekile	hygienic/clean
-hlola igazi	have blood tested
icholera / isifo sohudo	cholera
i-DOTS	DOTS - Directly Observed Treatment Shortcourse
igciwane	germ/virus
igonorrhea	gonorrhea
i-HIV	HIV (human immunodeficiency virus)
i-HIV positive	HIV-positive
i-HIV negative	HIV-negative
ikhondomu	condom
iliflethi	leaflet
imiphumela	results
ingculazi / ingculaza / i-AIDS	acquired immune deficiency syndrome/AIDS
inyumonia	pneumonia
isampula	sample
isibulalamagciwane	disinfectant
isidoda	sperm
isifilisi	syphilis
isifo	illness/disease
isifo esithathelwanayo	infectious disease
isikhwehlela	sputum
isitabani	gay man/person
i-TB / isifo sofuba	TB/tuberculosis
ivayirasi	virus
izeluleko	counseling/advice
izidakamizwa	drugs
izimpawu	symptoms/signs
izinhlungu esifubeni	chest pain
izivikelamzimba / amasotsha omzimba	body's immunity system
i-x-ray	x-ray
-khathazeka	feel depressed
-khipha isisu	have an abortion

-lalana ngendlela ephephile	practice safe sex
-lalana ngendlela engaphephile	practice unsafe sex
-lethela amadlingozi kokuya ocansini	masturbate
-lwa nezifo	combat disease
-nikela igazi	donate blood
-nokolotha	innoculate
-nukubeza	contaminate/abuse sexually
-**nukubezekile**	contaminated
-phelelwa ngamanzi	dehydrate
-thelela / **-thelelana** / **-thathela**	spread/infect
-thinta isilisa nesifazane	be heterosexual
-**thuma**	make a stool
ubutabani	homosexualism
ucansi oluphephile	safe sex
ucansi olungavikelekile	unprotected sex/unsafe sex
uhudo	diarrhoea
uketshezi / amanzi asemzimbeni	body fluid
ukudla okunomsoco	nutritious food
ukugoma	immunisation
ukugonywa	vaccination
ukuhlanzeka	hygiene
ukuhlolwa kwegazi	blood test (n)
ukukhubaza	complications
ukukhwehlela	coughing
ukulahlekelwa ngamanzi emzimbeni	dehydration
ukuncipha komzimba	weight loss
ukuqubuka	rash
ukuvuvukala	swelling
ukwelashwa / **ukwelapha**	cure (n)/treatment
umdlavuza wamaphaphu	lung cancer
umeluleki	councillor
-vikela	protect against disease
-ya ocansini	have sexual intercourse
-**xilonga**	examine - by medical staff

5.7   Conclusion

The lack of terminology in Zulu can be overcome if action is taken towards effective practical elaboration

through the collection of suitable written corpora (texts), i.e. by applying the theory and practice of universal

corpus linguistics.

A corpus is a body of written or spoken language data, which can be used as a basis for linguistic research. In this chapter the Zulu medical pamphlets dealing with the aspects of primary health care constitute the written (published) corpus. In order to compile a structured corpus specific texts were collected dealing with three aspects of primary health care: cholera, tuberculosis and HIV/AIDS which can each be seen as sub-corpora of the main medical corpus. The aim of the compilation of this small corpus (a total of 2187 words) was to exemplify a methodology in the extraction of terms from a written Zulu corpus to facilitate lemmatisation for eventual standardisation purposes. The applications and insights gained can thus be made applicable to other corpora. The selected texts were electronically scanned, edited, and stored as computer files. Finally the corpus was encoded by means of computerised query tools aimed at enabling the researcher to manipulate the data and supplying the researcher with additional information about the text.

Through computerised query tools, here the *WordSmith Tools* (WST), frequency word counts and coordinating alphabetical word lists, called *wordlist*, can be applied in order to isolate and extract words, some of which are terms, e.g. **izimpawu** (symptoms). This eases the task of the terminologist in the compilation of a terminology list. S/he can now immediately have access to the ranked frequency of a word/term in the medical corpus, e.g. **izimpawu** has a frequency (occurrence) of 15, ranked in the 33rd. position. In the alphabetical list **izimpawu** is the 773rd alphabetical word with a frequency of 15 in the corpus. To grasp a term such as **izimpawu** in the context of its use, i.e. in the environment of immediate preceding and following words, it can be entered as a search word by making use of another tool, the *concordance* query tool.

When extracting terms from a written corpus linguistic analytical aspects and lemmatisation have to be taken into account. Before a term can be lemmatised, all the realisations of a term, both uninflected and inflected, have to be considered to determine its representativeness or frequency. To get an overview of **izimpawu** (symptoms), the noun stem -**mpawu** is isolated through morphological manipulation and entered as a word search through the application of the *concordance* tool. Eventually it is revealed that **izimpawu**

has a high frequency of 23 in the corpus, which makes the term a representative one. It is also noted that the initial frequency of the term **izimpawu** of 15 in the *wordlist*, has increased to actual frequency 23 by the application of the *concordance* tool.

It is noted that it is particularly difficult to lemmatise inflected verbs since prefixes such as concords can be added to the root to form different tense or modal forms; affixes can be added to the root such as verbal extensions and endings. The verb **ikuvikela** (it protects you), for instance**,** comprises the morphemes **i-** (subject concord class 9) **-ku-** (object concord 2nd person singular)-**vikel-**(verb root) and -**a** (verbal terminative). The basic forms of terms, e.g. the root **-vikel-** should first be isolated for the purpose of lemmatisation (i.e. the listing of basic nouns and verb (also other) stems in a terminology list). Manual morphological manipulation is needed before *concordance* can be used effectively in Zulu, since these tools were particularly designed to suit the structure of European languages and not the complex structure of the Zulu language (or the structure of any of the African languages for that matter).

Concordance is not only a sufficient tool to reveal an overview of all the actual frequencies of nominal forms such as **izimpawu**, but also of basic verbal forms such as **-vikel-** and more complex verbal forms with extensions such as **-thelel-**:

Once the verbal root **-vikel-** (protect) has been identified and entered in the form of a search word *****vikel***** (***** to indicate the morpheme boundary), *concordance* reveals a high frequency of 33 in the corpus, indicating a representative term. It is noted that the initial frequency of 2 in the *wordlist* has increased to the actual frequency of 33 by the application of the *concordance* tool. C*oncordance* is also used to extract more complex verbal terms such **-thelel-** containing verbal extensions. The word search is conducted by entering the root plus the extension, i.e. ***thelel***. The *concordance* result is that the verbal term -**thelela** has an actual frequency of 10 in the corpus.

A corpus of lemmatised medical terms, does not only contain single terms such as **isifo** (illness) but also contains multi-terms such as the nominal **isifo sofuba** (tuberculosis, literally *the illness of the chest*) and

the verbal **-ya ocansini** (have sexual intercourse - literally, *go to the reeds sleeping mat*), which consist of two or more words. Since the *wordlist* cannot isolate multi-terms *concordance* can perform this task through the application of a word search. Entering the multi-term **isifo sofuba** results is an actual frequency of 10 in the corpus. What has become clear thus far is that lemmatising terminology in Zulu, is a combination of manual and computerised processes.

The *wordlist* tool does not have much application in semi-automatic term extraction in Zulu especially since it does not facilitate lemmatisation. The *concordance* tool complements the *wordlist* tool since it can establish the actual frequency instead of the initial frequency of a term, especially for inflected nominal and verbal terms and multi-terms, but also since it reveals how a certain term is used in context - in relation to other words in its immediate environment. Furthermore, it helps in determining the frequency of multi-terms. Although not discussed in this chapter, another WST tool, the *keyword* function, can complement the *concordance* tool, especially when the researcher has access to a large computerised language corpus as exemplified by Taljard & De Schryver (2002).

When extracting terms from a written corpus by means of text encoding not only statistical information has to be taken into account but also related additional information and conditions of the use of a term, for instance:

*	In the 33 occurrences of -**vikel**- in the corpus established through concordance, the form -**<u>z</u>ivikel**- employing the reflexive prefix **-zi-**, in examples such as **uku<u>z</u>ivikela** is used in 16 cases and means 'oneself', a very important notion in the combating of contagious preventable illnesses. Such additional information (notion) about this reflexive **-zi-** can thus be added in the front matter or back matter of a terminology list.

*	The term **isifo sofuba** is the indigenous multi-term for the acronym **i-TB** or the loan **ituberculosis**, all synonymous terms for 'tuberculosis'. The acronym **i-TB**, the multi-term **isifo sofuba**, and the loan term **ituberculosis**, as established through a word search by using the *concordance* tool,

have frequencies of 23, 10 and 1 in the corpus respectively. It thus proves that the shorter acronym is the most popular, a universal tendency in terminological practice.

* Through *concordance* the actual frequencies of the term variants for 'cholera' **icholera** (9), **ikholera** (2) and **isifo sohudo** (4) were established. Although the term **icholera** is almost entirely taken over from English - without orthographical adaptation, it has a much higher frequency than **ikholera** which adheres to the orthographical rules of Zulu. **Icholera** is thus a more representative term. However, the indigenous alternative multi-term **isifo sohudo** can also be listed since it can be used for more clarity, especially in the rural areas.

Atkins & Ostler (1992) in De Schryver & Prinsloo (2000a:92) make it clear that text encoding is a continual process: After a specific representative corpus has been built and its strengths and weaknesses analysed through text encoding, some material would be added or deleted to enhance the corpus. Thereafter the whole process will start all over again.

It is recommended, however, that such application of corpus linguistics, i.e. the use of corpus query tools, should be coordinated among the African languages in South Africa. In some respect this is already true in the functioning of the NLUs and in the establishment of corpora for the eleven official languages. However, terminological practice is not that far removed from lexicographical practice and the NLUs should coordinate their work with the terminological practices of the NLS and all the NLBs under the auspices of PanSALB.

For the proper elaboration and standardisation of terminology, the starting point is term extraction from the written living language, in the context of its use, and not merely the random intuitive creation of terms by the terminologist. However, terminologists must have practical training in the groundwork of term extraction. Corpus linguistics has a major role to play in technical elaboration and modernisation: Corpus query tools can enhance term extraction and ease lemmatisation for the terminologist, but never replace human involvement. In Zulu with its complex agglutinating structure in particular, manipulation of the morphology

is needed in order to extract or isolate terms for lemmatisation purposes. Such manipulation would also enable the terminologist to determine the actual frequency of words to decide on their inclusion or exclusion in a terminology list, alongside other synonymous or variant terms.

CHAPTER 6

THE VALUE OF ORAL CORPUS ANNOTATION FOR IMPROVING THE ACCEPTABILITY OF TECHNICAL TERMINOLOGY IN ZULU

6.1 Introduction

In this thesis, broadly speaking, the emphasis lies in establishing to what extent corpus planning, particularly as regards technical elaboration and related standardisation in Zulu, has been acceptable. This process naturally involves identifying the problem areas in the technical elaboration process. The problem area identified for this chapter can be phrased in the following questions:

Where are the oral sources of existing standard technical terminology?

How do these sources reflect the living language in the workplace?

Is this standard oral based and does it serve the consumer market (is it actually used

on a daily basis amongst medical staff and also between medical staff and their

patients)?

In the South African context, the standardisation of terminology in the African languages has not been realistically verified by, for instance, comparing oral terms with existing written standard terms. The use of oral corpora as a basis for research in practical elaboration and standardisation in the African languages, Zulu in particular, has generally not been exploited.

Written corpora should not be the only source for the elaboration and standardisation of technical terminology but also oral corpora which reflect the living language in the workplace. According to Calteaux (1996:42) the evidence of standard should also be oral-aural based. Terminologists should use terms that already exist in society, thereby promoting natural term development (Jafta 1987:127). The use of oral corpora as a basis for research in practical elaboration and standardisation in the

African languages is thus of the utmost importance.

In order to gather the latest existing oral spoken corpora available in the specific technical (medical) field, professional mother-tongue speakers who are medical workers in hospitals or clinics were interviewed in their work domain by the use of questionnaires designed to get an intuitive natural response on terminology (see Addendum 1). By comparing oral terms with existing written standard ones, the success of the elaboration and standardisation process can be evaluated, at least to some extent. In this manner it can be determined what technical terms have been incorporated into the Zulu lexicon and to what extent they have been accepted and used by the broad Zulu speech community. Manual frequency counts form the basis for such comparison, although high frequency is not the sole criterion for acceptability. See 5.4.1.1 for 'the frequency *wordlist*'.

Corpus annotation is based on an analysis of term equivalents for the same concept as preferred and/or additionally rendered by the informants. Frequency counts based on the informants' natural response play a decisive role in corpus annotation. Such counts and comparisons are manifested in corpus annotation and can indicate the acceptability of terminology and consequent validity of standardisation, i.e. what terms are actually used and accepted in the field and which ones should be included in a modern terminology list and which ones should be discarded.

The standard written Corpus A is based on an extraction of terms from lists compiled by the previous *Zulu Language Committee/Board* (ZLC/B) and the previous *National Terminology Services* (NTS). Although criticised by scholars, the efforts of these previous structures cannot be underestimated and thus need to be acknowledged in such a corpus. The oral Corpus B is based on fieldwork conducted at hospitals or clinics, i.e. the response of informants interviewed via a questionnaire. After comparing Corpus A (written) and Corpus B (oral) with the emphasis on the difference between the two, one cannot really determine the outcome of the standardisation process

of medical terminology thus far, since relatively small corpora were used which cannot claim full representativeness of medical corpora in Zulu. However, the basis of this oral corpus annotation is in the form of a practical approach which can contribute positively towards the acceptability of terms. This practical oral corpus annotation is based on some recognised tendencies in the comparison of the written and spoken corpora and can throw some light on the terms that are actually being used in the practical environment, some of which are more popular, trendy and accurate than the standardised ones. For the sake of systematisation, these tendencies are arbitrarily named 'indigenous coinage', 'accurate designation', 'phonological adaptational trends', 'semantic shift alternative' and 'taboo preference'.

To put the value of oral corpus annotation in the elaboration and standardisation of technical (here medical) terminology into perspective, it is important to refer to Chapter 5 (5.1-5.3) where the written corpus was dealt with and general aspects of corpus linguistics such as *corpus*, *text encoding* and *lemmatisation* were discussed in detail. Therefore, for the purposes of this chapter, similar and sometimes overlapping aspects are only referred to briefly as they become applicable to the oral corpus.

It must be noted that the research methodology followed for fieldwork procedures in the obtaining of oral data (also see 1.6) including the utilised questionnaire, is contained in Addendum 1. Corpus statistics in the form of a frequency profile is contained in Addendum 2.

In order to understand corpus linguistics to its fullest extent in language elaboration, after having discussed the written corpus, it is only appropriate to include the oral corpus.

6.2   The concept 'oral corpus'

A *corpus* is a body of language data, which can be used as a basis for linguistic research. This language data is obtainable in a written or oral format. See the definitions of *corpus* in 5.2.1. Whereas the written corpus was the topic of discussion in Chapter 5, the oral/spoken corpus forms the point of departure for this chapter. Written corpora are easier to collect and compile than corpora of oral discourse (Leech 1997:8). Whereas the written corpus can be regarded as a 'given', the oral corpus can be regarded as a 'second given' since it is a transcript of what was said (Leech 1997:4).   It therefore follows logically then that in spoken discourse it may sometimes be difficult to distinguish between representative and interpretative information (Leech 1997:3).

For this chapter a standard written medical term list was compiled. This term list constitutes a summary of the most general medical terms standardised by the ZLC/B and the NTS   - see Addendum 1. Despite the shortcomings of these term lists, they at least bear the evidence that some progress has been made in the field of medical terminology. However, the purpose of listing these existing standard medical terms that constitute <u>Corpus A</u>, is to use them as a prompt for informants to initiate the rendering of original oral equivalents.

The terms contained in Corpus A can then be compared with the oral terms (gathered through interviews via questionnaires during fieldwork), which constitute <u>Corpus B</u> in order to improve the future elaboration and eventual standardisation process of technical terminology in Zulu. This means to produce acceptable terms by making use of corpus annotation. It is commonly known that description and comparison play a major role in corpus linguistics, or at least in its application in the form of corpus annotation.

The aim is to make the small corpora A and B exemplary of the type of research that can be done

in the field of oral and/or standard technical terminology. Furthermore, only a small corpus can be dealt with within the scope of a single chapter. However, the method followed for corpus annotation in a small oral corpus can very well be applied to a more substantial corpus where research is conducted on a much bigger scale, on a national level, for instance. Thus, after a specific representative oral corpus has been built, it must be analysed (annotated) and some material added or deleted to enhance such a corpus. Thereafter the whole process will start all over again.

6.2.1   The concept 'oral corpus annotation'

When dealing with a corpus, one should also understand a basic concept that goes hand in hand with it, namely *corpus annotation*, also known as *encoding*. In Chapter 5 where computerised corpus query tools were used to encode the written corpus, the term 'text encoding' was used. However, in this chapter which deals with oral corpus, the term 'annotation' is used in order to distinguish between the two types of corpora. Furthermore, in this chapter, the corpus is not annotated (encoded) electronically but manually.

It must be clear that corpus annotation is not found in the original (also called *raw*) text, in this case the 'given' (Leech1997:3) but is information derived from the text. Corpus annotation is thus the process of adding 'enriching information' to the text (Leech et al. 1995:2). Leech (1997:3) explains that corpus annotation has interpretive value in that it tells us 'about the language of the text'. The corpus is thus to be analysed linguistically at one or more levels in order to annotate or label the text with information thus obtained (Leech 1997:2). This information is used for a multitude of purposes such as to understand and manipulate the language more successfully (Garside et al. 1997:viii). For instance, if one extracts information from an existing oral corpus in order to create more acceptable standard technical terms from 'real language data', one applies corpus annotation. In this chapter the purpose of corpus annotation is to determine which terms are actually utilised in the oral medical

discourse in the work environment and therefore worth lemmatising (listing in their most basic form) and eventually standardising.

In short, the corpus compiler makes use of a corpus and applies its annotation to a specialised area of research with practical benefits gained from it (Leech et al. 1995:2). The kind of corpora will determine a specific type of annotation to be followed. The medical corpora in this chapter will, for instance, determine what type of annotation is to be followed. It must be remembered, however, that no annotation scheme can claim authority as absolute standard since no corpus can be representative of all language specific contexts (Leech 1997:7). However, an annotated corpus is valuable because it is a reusable resource (Leech 1997:5). In this chapter, for instance, corpus annotation has value as resource for improving the elaboration process in order to produce acceptable standard terms.

6.3   The compilation of a structured oral corpus

In this discussion of a structured corpus compilation two steps, namely corpus design and text collection, are applied to the field of research done for this chapter.

6.3.1   Oral corpus design

Many different types of corpora can be distinguished which will eventually determine the design of a corpus. In this instance the corpus design is directed towards two small corpora, namely one written standard medical corpus and one oral medical corpus (where terms were obtained through the interviewing of medical workers). The standard terms were selected because they represent an existing medical written Zulu terminology. From these officially standard term lists the most general terms were extracted in order to compile a fairly representative term list or corpus of 350 standard

medical terms. However, it was, after careful scrutiny of terms, decided to limit the written standard to a core corpus of 145 terms (225 including equivalents) in order to deal with corpus annotation in the scope of a single chapter. This obviously also limits the comparative oral corpus to a core corpus of 145 (823 including equivalents). See Addendum 1 for detailed information on corpora.

The oral corpus is based on terms obtained from interviews via questionnaires in which only health workers who were also Zulu mother-tongue speakers took part. This was done in order to obtain a natural intuitive response. The format of the questionnaire was such that *informants* (a term henceforth used to refer to persons who took part in the survey) who were interviewed had to either approve or disapprove of the given standard Zulu medical terms. In cases where they disapproved of terms, they had to supply the more acceptable or popularly used terms (for detailed research methodology consult Addendum 1).

However, what ultimately determines corpus design is the specific purpose for which it will be used. In this chapter the purpose is to find practical means of improving the elaboration process and eventual acceptability of standard technical medical Zulu terminology. This purpose was initiated by comparing standard written medical terms with oral medical terms (gathered through research).

For practical reasons it is at this stage important to render the structural design of both the written and oral corpora used for comparison and eventual annotation purposes. It must be understood that both these corpora contain medical terms in isolation and not in discourse. The written standard medical corpus, is called *Corpus A* and the oral medical corpus based on the responses of informants interviewed, is called *Corpus B*. Corpus A consists of a single corpus while the oral corpus comprises six sub-corpora, each subdivided according to the region where the hospital is situated:

CORPUS A

Main written standard medical corpus

containing standard terms from terminology lists as sources

Total of 225 terms (145 core terms including equivalents)

CORPUS B

Main oral medical corpus (containing different responses of informants as sources)

Total of 823 terms

Total of 100 questionnaires (informants) out of 120 questionnaires used.

sub-corpora	number of terms rendered	number of informants /questionnaires used
Amajuba Memorial Hospital, Volksrust	426	17
King George V Jubilee Hospital, Durban	419	17
McCord Zulu Hospital, Durban	330	10
Newcastle Provincial Hospital, Newcastle	476	24
Prince Mshiyeni Hospital, Umlazi	266	18
Vryheid Hospital, Vryheid	428	14

6.3.2 Oral text collection

Written standard Zulu medical terms were collected. These terms are the ones standardised thus far by official language authorities such as ZLC/B and the NTS (see Addendum 1). From these official standard term lists the most general terms were extracted in order to compile a fairly representative term list of standard medical terms - Corpus A. In this chapter, for instance, the purpose of compiling this standardised medical corpus is to eventually compare it with the corpus of terms extracted from the oral responses of medical workers - Corpus B.

The relevant collected written standard medical terms (Corpus A) were entered into computer files by means of term extraction. From terms standardised thus far by the official language authorities mentioned above (see Addendum 1 for detail), the most general medical terms were entered in order to compile a fairly representative term list (corpus) of standard medical terms. These terms from Corpus A were not only used for comparison but also <u>to serve as prompts for informants</u> in order to initiate the rendering of original oral terms used in the workplace.

The relevant collected oral terms (Corpus B) were also entered into computer files after interviewing and transcription. The questionnaires were completed during interviews by the researcher or by trained assistants. In some cases the informants completed the questionnaires themselves. The transcriptions were based on the oral response of medical workers, in other words, spoken terms were recorded by writing them down. The transcription of oral terms posed a problem in that different spellings occurred for the same term. Therefore it was decided to make the spelling of terms as uniform as possible in accordance with the Zulu orthography. These orally rendered terms form Corpus B.

Once a structured corpus has been compiled by means of corpus design and text collection, the annotation of the corpus is the next logical step. However, one of the cornerstones of corpus annotation is frequency count.

6.4   Frequency count in relation to corpus annotation

Conducting a frequency count can be regarded as an important and valuable tool for the annotation of a corpus or corpora for satisfactory elaboration and standardisation procedures in terminological work. By means of a frequency count, for instance, one can determine the following:

*   The popularity of a term or its functional use in a specific subject field by frequency of occurrence:

>   Such popularity is determined by the rank of items in frequency lists, overall counts in the entire corpus and distribution of the term across the different sub-corpora.

*   How many equivalent or variant terms exist for the same concept.

*   Which terms to list as lemmas in a terminology list and eventually standardise:

    i)   An item with a substantial overall count and with a reasonable spread across sub-corpora, will be likely to be included in lemmatised format and will be likely to influence corpus annotation.

    ii)   The absence of commonly used words in African language terminologies and dictionaries, could have been avoided if the compilers utilised frequency counts even with regard to a small corpus. Although De Schryver and Prinsloo (2000b:302) claim that lemmata which show low and uneven distribution do not qualify to be

included amongst the frequently used words, this principle cannot be applied as a rule in the annotation of terms. High frequency is not the ultimate criterion for the listing of a term. It can, for instance, be that a term with a low frequency is a more suitable or correct term than the most popular term with the highest frequency. Such judgements can be made by a terminologist who is equipped with intuitive linguistic knowledge together with an expert equipped with scientific field related knowledge.

*   To what extent corpus planning, i.e. language development and elaboration and eventually standardisation, can be improved:

i)   One can start by comparing the oral medical terms rendered in the questionnaires with the existing standard medical terms.

ii)   The aim is that elaboration be successful and that standard terms are actually utilised in the public and private sector.

To eventually come up with a realistic frequency count the spelling of terms must be standardised. If the spelling of terms differs with one letter only, these words will appear as different lemmas which will obviously influence the frequency count considerably. Out of the 100 usable questionnaires 145 core medical terms (823 including equivalents) were selected as basis for an oral corpus annotation. Since this is regarded as a relatively small corpus, it was decided to conduct the frequency count manually, i.e. supplying statistics in terms of frequency counts. See Addendum 2 for the frequency profile of discussed terms. The reason for doing it this way is that single and multi-terms are already in the extracted format and not in discourse. A computerised frequency count, for instance, would be problematic and time consuming since multi-terms consist of phrases which can be divided into several words, each word of which has its own frequency count. However, with research projects of greater magnitude such a frequency count would definitely be conducted electronically via corpus querying tools such as *WordSmith Tools* (see Chapter 5) to facilitate the task of the terminologist.

After a successful frequency count, corpus annotation takes place when the raw terms are encoded by corpus processing annotations. Making use of such corpus annotation can enhance the development of the Zulu language.

6.5   The value of oral corpus annotation for improving the acceptability of technical (medical) terminology

As is known by now, the 'standard' Corpus A and 'oral' Corpus B were designed, compiled and eventually compared in order to meet the compiler's needs, in this instance, to come up with a corpus annotation that is aimed at improving elaboration and standardisation processes. It must be noted that terms could only be compared on the basis of concept equivalence. This corpus annotation is roughly based on the difference between the standard written corpus and oral corpus and the motivation of such differences.

To verify the success of elaboration and standardisation processes of technical (medical) terminology extensive research on a much bigger scale is required than the scope of this chapter allows. The aim of the annotation of this corpus is therefore rather aimed at improving the elaboration process and eventual acceptability of standard technical (medical) Zulu terminology by applying a practical approach. Although negative perceptions of standardisation have developed globally, standard terms are in many instances acceptable terms which are being used as proven by informants' responses.

The value of this oral corpus annotation is that some suggestions forthcoming from it can contribute positively towards the acceptability of terms. The concept 'acceptability ' is a relative concept, rather than an absolute one, especially in this chapter, since the responses of only 100 informants from different regions of Kwa-Zulu Natal, and from Volksrust in Mpumalanga, and not of the whole speech community per se, were taken into account. Acceptability is indeed a complex issue, since

for every new term that is coined, even if it has a high frequency, new equivalents may spring up due to new trends and change in the language. The standard written terms from Corpus A **ingculazi** (AIDS) and **isidakwa** (alcoholic), for instance, are highly acceptable according to informant responses as reflected in high frequency count. Yet, the more informal harsher term equivalents for **ingculazi** being **umashaya bhuqe** (literally, the one who wipes out), **unogawula** (the one who fells trees - also the Xhosa term), and **ikhodi** (apparently referring to the acronym, AIDS) have also entered the language, although with a much lower frequency. Similarly, the more informal term equivalents for **isidakwa** being **impuzi tshwala** (the liquor drinker), **indlamanzi** (the water eater) and **unsutha** (the saturated one who does not eat), are also in use even though they have much lower frequencies. These latter three terms remind of the informal *isiHosi* (hospital language) recorded by Zungu (1995). See Addendum 2 'per example' for the frequency profile of the terms discussed above.

In this chapter acceptability is based on high frequency of use, i.e. a high frequency count is also imbedded in tendencies on which corpus annotation is based. However, high frequency is not the ultimate criterion for the listing of a term. It can, for instance, according to sound linguistic intuitive judgements made by a terminologist and a scientific field specialist, be that a term with a lower frequency is a more suitable or correct term than the most popular term with the highest frequency. In Addendum 2 a frequency profile (in table format) of the oral medical corpus is given. This profile contains the frequency count of all the written standard terms (Corpus A) and their oral equivalents (Corpus B) according to hospital regions listed in 6.3.1. Only the first four equivalents, starting with the highest frequency are included in this profile. The profile follows the chronological order of numbering in this chapter, e.g. paragraph 6.5.1 corresponds with Table 6.5.1 in Addendum 2.

The oral corpus annotation in this chapter is based on some identified tendencies in the comparison of the written and spoken corpora. Such annotation can throw some light on the oral terms that are actually being used in the practical environment, some of which are more popular, trendy and

accurate than the standard ones. Furthermore, annotation can also indicate which standardised terms are not popularly used and as a result have become outdated. It can also eventually determine which terms merit listing in a terminological list for standardisation purposes. These tendencies are arbitrarily named 'indigenous coinage', 'accurate designation', 'phonological adaptational trends', 'semantic shift alternative' and 'taboo preference' for the sake of systematisation. The 'methods of word-formation' in Chapter 4 also have significant relevance to oral corpus annotation in this chapter. However, it must be emphasised that these tendencies cannot be entirely separated since they do overlap to some extent.

6.5.1   Indigenous coinage

Indigenous coinage occurs when a Zulu coinage is preferred to the phonologically adapted loan term, e.g. a very generally used compound term **umtholampilo** (place where one can find (**-thola**) health (**impilo**) is preferred to the loan term **ikliniki** (clinic) for 'clinic'.

Other examples of indigenous coinage that were found are:

umkhuhlane wamaphaphu (literally, the fever of the lungs) with a significant high frequency, is preferred to the loan term **ibhroyinkhanthisi** (bronchitis) or the other indigenous term **izilonda emaphashini** (literally, *wounds in the lungs*) for 'bronchitis';

isihlungu (one of its meanings, 'nettle-rash') is preferred to the loan term **i-alegi** or **ukungezwani nomzimba** (literally, *not to agree with the body*) for 'allergy';

igciwane (literally, *small light floating particle*) is a very popular medical term in Zulu forming the basis for a variety of terms. The term **igciwane** is preferred to the standard loan terms **ibhakthiriya** and **ibhakithiriya** for 'bacteria'. It follows logically then that indigenous coinage in the form of a

paraphrase, **isibulala** <u>**magciwane**</u> (the killer of germs) was generally preferred to the standard **isibulala mabhakithiriya** for 'antibiotic'. Even in the term **isilwa magciwane** which has a low frequency, the part **magciwane** is preferred to **mabhakithiriya**. The term **igciwane** is a case of semantic shift in that the smallness of a particle correlates to the smallness of a germ. **Igciwane** is also the most popular Zulu term for 'germ' with a higher frequency than the equivalents **imbewu yokufa** (literally seed of death) and the loan **ijemu** (germ). See Addendum 2 Table 6.5.1.

However, a terminologist should, when examining the acceptability of a term, in order to conclude with an objective corpus annotation, not only consider high frequency but also the distribution of the term across the different sub-corpora. The loan term **imayigreni** (migraine), for instance, has a higher frequency than any of its indigenous equivalents **ikhanda elibuhlungu kakhulu** (very painful headache), **ikhanda elibuhlungu elingapheli masinyane** (literally, *a head that is painful which does not stop quickly*) and **ukuphathwa ikhanda kakhulu** (have a severe headache). However, the distribution of these terms across the different sub-corpora (different regional hospitals) is evidenced by the reoccurrence of the word **ikhanda**. This is quite significant since the overall count of these three indigenous terms matches the frequency of the loan term **imayigreni**. Furthermore, these indigenous terms capture the concept of a migraine, being an intense headache by the use of the qualificatives **elibuhlungu** (which is painful); **elingapheli** (which does not stop) and the adverb **kakhulu** (much/a great deal). Above all, indigenous terms, exploiting the internal resources of the language, are generally more understandable and descriptive, particularly to the rural community. Needless to say, the term **ikhanda elibuhlungu elingapheli masinyane** is the most explicit since it is in the form of a definition. However, such terms are long and cumbersome which would make a term like **ikhanda elibuhlungu kakhulu** more practical. Indigenous terms have not all necessarily become documented but some have become established terms by continual oral use as evidenced by many similar informant responses.

The internal resources of the language are also used in the following example where it becomes evident that terminologists need not always search beyond their language for suitable terms, since they

are imbedded in the language and they are revealed intuitively by the users of the language:

ukhakhayi (mid-part of the head) is a pure indigenous term preferred by the informants to the standard loan term **ifontaneli**. (fontanelle). The term **isikhala sokhakhayini** (opening in the mid part of the head - of a baby) is also quite descriptive, but unfortunately has a very low frequency in comparison with the previously mentioned shorter term. The other indigenous equivalent **ufokothi** for 'fontanelle' is proof that there are sufficient terms available in the language to opt for outright coinage instead of borrowing. Furthermore, as in the previous example, there are three indigenous terms compared to one loan term, which shows that indigenous coinage is alive and well and is intuitively reverted to by informants. Indigenous coinage is a positive sign that people still take pride in their language and that Zulu has enough elaboration capacity.

6.5.2 Accurate designation

Some medical terms rendered by informants are far more accurate and more concept related than the standard ones. These terms do not necessarily have the highest frequency, but should be considered because of accuracy in concept designation. The health workers who provided these accurate terms are not linguists, but intuitively use the elaboration mechanisms of the language to the fullest. The examples that follow can illustrate how the elaboration accuracy of the Zulu lexicon can be improved in order to lead to successful standardisation by making use of oral corpus material. See Addendum 2 Table 6.5.2.

Instead of the official term **umdlavuza wesifuba** (breast cancer - literally *cancer of the chest*), the vast majority of informants rendered the term **umdlavuza webele** (literally *cancer of the breast*). The latter term is exclusive, accurate and in line with the concept and English equivalent 'breast cancer'. The term **umdlavuza wesifuba** (chest cancer), on the other hand, is less accurate and caters for a far more generic and inclusive concept (chest). The standard Corpus A is inconsistent

in the use of **umdlavuza** for 'cancer'. In other instances the synonym **umhlaza** is used, e.g. in **umhlaza wesikhumba** (skin cancer). The informants generally prefer the term **umdlavuza wesikhumba**. The use of **umdlavuza** for cancer in general is a form of intuitive consistent accuracy.

Similarly, instead of the standard term **izinga lokushisa <u>kwegazi</u>** (literally *temperature of the <u>blood</u>*) the rendered term **izinga lokushisa <u>komzimba</u>** (literally *temperature of the <u>body</u>*) is the correct term for accurate designation of 'body temperature', even if it has a lower frequency than the former. However, the correctness of the latter term is confirmed by the fact that the other two equivalents, i.e. **ukushisha <u>komzimba</u>** (the heating of the body) and **izinga lokushisa <u>emzimbeni</u>** (measuring the heat in the body) also include reference to **umzimba** (body) instead of **igazi** (blood).

The Zulu term for both 'homosexual male' and 'lesbian' is a single general term in the standard Corpus A, namely **isitabani**. Terms like the latter are not strictly speaking medical terms but are considered important in relation to sex and AIDS education and are therefore dealt with here. In English it also seems to be a problem since 'homosexual' is also a generic term referring to both men and women who are attracted to the same sex, but generally it is used to refer to a homosexual male. A few informants intuitively solved this problem by rendering two differentiating terms for 'homosexual', namely **isitabani sesilisa** (homosexual male) and **isitabani sowesifazane** (lesbian). The concluding possessive parts of the latter two terms **sesilisa** and **sowesifazane** indicate masculine and feminine gender respectively which obviously make these two terms more descriptive and accurate in concept designation.

The informants rendered the oral term **ungqingili** for both 'homosexual male' and 'lesbian'. Obviously this is also inaccurate since it is a generic term for 'homosexual' according to Nyembezi's (1992) monolingual Zulu dictionary which happens to be the only dictionary which lists this term. The other alternatives **inkonkoni** ('blue wilde beeste' for male homosexual) and **uncukumbili** ('double-sexed being'/'hermaphrodite'/'bisexual' for lesbian) are at least distinctive in words (sex).

However, the latter term is not suitable since it deals with both the concepts of heterosexuality and of homosexuality (lesbian). What has come to the fore is that the general speech community (here represented by the informants) have not yet conceptualised the differentiating concepts of the generic terms 'homosexual' and 'heterosexual' or the sexually distinctive concepts of 'homosexual male' and 'lesbian'. These related terminological problems can only be solved by revisiting these closely related concepts and then designating them accurately. Ideally this should happen before standardisation. However, there is nothing wrong with improving existing standard terms or even creating new ones if required for accurate and distinctive designation of concepts.

To be HIV positive means to be infected with the virus that causes AIDS. The standard Zulu term **ukuba nesandulela ngculazi** (HIV positive) with the highest frequency by far, is rather descriptive, literally meaning *to be with the proceeder of AIDS*, implying the virus that causes AIDS. However, the other three equivalents, although with a much lower frequency, explicitly include the significant term **igciwane** 'virus' in their paraphrases, i.e. **ukuba negciwane lengculazi** (to have the virus of AIDS), **ukuba ne-HIV gciwane egazini** (to have the HIV virus in the blood) and **ukutholakala negciwane lengculazi** (to contract the virus of AIDS). Of the three terms **ukuba negciwane lengculazi** (to have, literally *to be with the virus of AIDS*) seems to be the most accurate since it designates the concept 'to be infected with the virus that causes AIDS'. The second term **ukuba ne-HIV gciwane egazini** (to have the HIV virus in the blood) does not include the concept of 'the virus that causes AIDS' as reflected in the possessive construction **igciwane lengculazi** (the virus of AIDS) contained in two of the other equivalents, but directly mentions the 'HIV virus'. The problem with the third term **ukutholakala negciwane lengculazi** (to contract the virus of AIDS) lies in the infinitive verb **ukutholakala** (to get/contact the virus of AIDS) which implies a possibility of contacting it instead of **ukuba na-**(being with it/having 'the virus that causes AIDS') in **ukuba negciwane lengculazi**.

Another standard verbal term which lacks accuracy is - **khipha iqanda** (ovulate). This term is very

general since it literally means 'produce an egg' which does not have any link with the concept of fertility. On the other hand, the term -**khipha iqanda lenzalo** with a much lower frequency, which literally means *to produce an egg of offspring* (**inzalo**) does correlate with the concept of fertility and is therefore a much better descriptive term. Although the other terms also deal with the concept of fertility in that -**akhela iqanda** literally means *build up an egg* implying fertilisation and -**khipha iqanda lowesifazane** literally means *produce an egg of feminine gender*, they are less accurate than the term -**khipha iqanda lenzalo** which indicates the result of ovulation, namely generating an **inzalo** (offspring).

6.5.3   Phonological adaptational trends

Phonological adaptational trends, which are initiated by making use of external resources of the language, like making use of borrowing from other languages, can be seen as an opposite trend to that of indigenous coinage which utilises the internal resources of the language. However, this does not mean that the one should be preferred to the other, but rather that they exist alongside each other. Which one to use depends entirely on the language situation and the preference of the language users. Purist coinage cannot be striven for at all costs, for instance, if a loan term is more suitable. See Addendum 2 Table 6.5.3.

Phonological adaptation occurs spontaneously whereby English terms are Zuluized, i.e. phonologically adapted to suit the Zulu language system, e.g. **ibhulima** for 'bulimia'. See 4. 4.6 'borrowing' for a detailed explanation. The informants' response on the term **ibhulima** (bulimia) was poor since it is not yet a well known state of illness in the Zulu health community and it is generally considered a mental rather than a physical illness. However, it was the more popular alongside indigenous alternatives **isifo sokudla** (literally, *the illness of food*) **ukuminza** (to gulp down which relates to the eating habits of bulimia sufferers) - see 'semantic shift alternative' 6.5.4 below - and **isifo sokuhlanza ukudla** (literally, *the illness of bringing up food*).

The standard term **iganjirini** (gangrene) has a higher frequency than the other equivalents **igangrini, igangirini** and **ukubola** (literally, *to rot*). However, the former three phonologically adapted equivalents are spelt differently. The term **iganjirini** (gangrene) is closer to the Zulu sound system, using **j** instead of **g**, adapting an open syllabic system in which a syllable ends in a vowel. The term **igangirini** (gangrene), on the other hand, is closer to the English in keeping **g** but closer to Zulu, retaining the open syllabic system. The term **igangrini** (gangrene), however, is closest to the English proving the acceptance of the inadmissible sound sequence **ngr** and thus also making the word shorter. The three different spellings of the same adapted term is thus an indication that Zulu orthographical rules of allowing only certain sound combinations have changed; facts that have to be noticed by terminologists. Since English is such a dominant language it will exert an influence on the phonology and spelling of the Zulu language. See 3.2.5.1 for new phonological trends in Zulu.

Some phonological adaptations can be considered trendy since foreign sounds are adapted or elided in such a way that the term of the source language can hardly be recognised as in the variant terms **ukunokolota** and **umnokoloto** for 'inoculation'. These terms are significant since they are derivative nouns from the same loan verb stem - **nokolota** (inoculate) represented in two out of four equivalents although **umjovo** (injection) has a higher frequency. However, **umjovo** (injection) and **umgcabo** (incision) are generically inclusive whereas the former two adoptive terms are more specific, denoting a specific type of injection or incision, namely an inoculation. To some extent **ukunokolota** and **umnokoloto** could also be classified under 'accurate designation' above. It becomes clear then that oral corpus annotation cannot always be strictly classified as a specific type and that overlapping annotations do occur.

As is the case in other languages, it is also the trend in Zulu to use shortenings and abbreviations which really catch on, such as the following examples:

The term **ikhwashi** for 'kwashiorkor' was preferred by the informants to the more indigenous terms

isifo sendlala (illness of hunger) and **isifo sokungondleki** (illness of malnutrition - not being nourished). It is interesting to note that **ikhwashi** is a phonological adaptation of the scientific term 'kwashiorkor' but at the same time a shortened trendy form. The other trendy equivalent **umTopia** is an informal *IsiHosi* (hospital language) example as recorded by Zungu (1995) but is derived from the word 'Ethiopia', a place associated with famine. The term **iskizo** (schizophrenic) which is also quite common characterises a similar type of phonological trend (shortening) as **ikhwashi**. Also see 'clipping' in 4.4.7.2.

The acronym **i-AZT** (AZT) for 'AIDS medication known as zidovudine' was preferred to other equivalents such as the phonologically adapted **azathi** (converting the acronym AZT into a catchy word), **isithibingculazi** (restrainer of AIDS) and the indigenous **iphilisi elapha ingculazi** (the pill which cures AIDS). Just like other international acronyms such as **i-TB** (TB), **i-HIV** (HIV) and **i-IUD** (IUD), **i-AZT** is popularly used. Even informal acronyms like **i-BP** (blood pressure) and **i-MC** (mental case - mentally handicapped person) are used in hospitals and are likely to increase.

For the term 'sonar scan' three of the Zulu equivalents are phonological adaptations, i.e. the preferred standard term **isona**, **isikeni** and **iscan**. The other equivalent is an indigenous example of semantic shift, namely **emafutheni** (literally, *at the place of fat*) - the gel rubbed on to body parts to conduct the scan reminds of animal fat **amafutha**. The three mentioned phonological adaptations can also be regarded as shortenings since part of the term is omitted in each case. In the first term **isona** 'scan' is left out, in the second **isikeni** and third **iscan** 'sonar'. The first two have clearly been phonologically adapted to Zulu and are short and catchy. The third term **iscan** can hardly be called an adaptation (except for the **i-** class prefix) since it is taken over from English as it is.

For the term 'yellow fever' one equivalent is outright indigenous, i.e. **umkhuhlane** (fever/flu), two equivalents are partially indigenous coinages, i.e. **imfiva encombo** and **imfiva ephuzi**. The part **imfiva** is phonologically adapted from ' fever' while both -**ncombo** and -**phuzi** are Zulu relative stems

denoting 'yellow'.   However, the preferred term is the standard **iyelo fiva** almost taken over from English as it is, sounding like the English term.

Preferred terms like **isona** and **iyelo fiva** remind of other trendy adopted terms such as **ireshi** (rash) and **idragi** (drug) which are more likely to be used by the youth than the purer Zulu terms since they can immediately be associated with the English terms. There seems to be a trend that a near English term like **iyelo fiva** is to be regarded better since it is associated with social advancement. See 4.4.6 'borrowing' in this regard.

6.5.4   Semantic shift alternative

Semantic shift is a means of word formation whereby the meaning of a word is extended or modified to suit another, mostly related meaning. Semantic shift is an interesting and popular alternative commonly found in the terms rendered by informants. A perfect example of meaning extension is found in **isithombe** (picture), a popularly used term in favour of the standard term **iX-reyi** or any of the other rendered terms **isithombe seX-rayi** (literally, *picture of x-ray*) and **igesi** (apparatus working with electricity) for 'X-ray '. It is clear that the meanings of **isithombe** (picture) and even **igesi** (gas/electricity) have been extended to include the related meaning (taking of a picture by using an electrically driven apparatus) of 'X-ray'. The use of **isithombe** and **esithombeni** (literally, *place of pictures*) for 'X-ray room' is quite common in the hospitals visited and can be associated with *isiHosi* (urban Zulu hospital language used in the Durban area) as documented by Zungu (1995). See Addendum 2 Table 6.5.4.

Another example of semantic shift is the term **ungwengwezi emehlweni** which is commonly preferred to the standard **ikhatharakhithi** for 'cataract'. The association **ungwengwezi**, literally meaning *thin layer such as dust on object or scum on water* (Doke &Vilakazi 1972), has with the hindrance cataracts cause, is obvious. Since a cataract can also be associated with 'a spider

(web) over the eye' and a 'transparent object' the respective Zulu terms **ulwembu nasemehlweni** and **untwentwezi** are also perfect examples of semantic shift which exemplify the creativity of speakers of the language.

The popular term **unhlangothi** indicated by a significantly high frequency, is preferred to the standard **isitrokhi** for 'stroke'. The semantic shift is evident in the association **unhlangothi**, meaning 'tree trunk that is charred by lightening' (Doke & Vilakazi 1972), has with the devastating mental effect of a stroke on the human mind. The term **unhlangothi** is quite prominent as it also appears in the paraphrase equivalent **isifo sonhlangothi** (stroke illness). Semantic shift is further evident in the association the idiomatic expression **ukushaywa yinyoni**, meaning 'to be hit by a bird (in flight)', has with the sudden and devastating impact a stroke has on the human mind.

Another popular term **uklilo** with quite a high frequency is preferred to the standard terms **uxhilo** and **idifuteria** for 'diphtheria'. Since **uklilo** and not **uxhilo** could be found as a dictionary entry, an assumption can be made that the only difference between the two is pronunciation. The semantic shift can probably be explained in the cultural association **uklilo** has with **iklilo** meaning ' beast with a white marked throat'. The term **uklilo** is an acceptable term which fits the latter reference to a beast since diphtheria is a serious infection of the throat making it difficult to breathe. The other indigenous alternative **umphimbo omhlophe**, literally meaning *the white throat* - caused by white sores, also links up with the reference to a beast with a white mark but is less acceptable since it has a low frequency.

It has become clear yet again that oral corpus annotations cannot strictly be categorised since elaboration by means of semantic shift is actually also indigenous coinage whereby the internal resources of the language are utilised to the fullest as exemplified in **uklilo** above.

6.5.5 Taboo preference

Zulu language elaboration would not be complete without mentioning the way in which such elaboration is linked to Zulu culture. Taboo, for instance, is a culture-related aspect which should be taken into account in terminological development. See 4.5.2 for 'taboo'. Taboo is an umbrella term used to refer to terms that are unsuitable for use in a specific social context. Taboo also deals with acting respectfully or modestly and using language of avoidance. A good example to illustrate taboo is the preferred standard term **isifo socansi esithathelanayo** (sexually transmitted disease), literally translated as *the illness of the sleeping mat that is contageous*. In Zulu culture it is usually taboo to refer to terms with a sexual connotation in a direct manner. A culturally bound word such as **ucansi** a 'reeds sleeping mat', is therefore used to indirectly refer to 'sex.' However, many oral responses by informants for the same term also adhere to cultural taboo in the form of **isifo samasoka** (literally, *illness of the young men*) and **ukubhajwa** (to be entrapped entangled - indicating the difficulty of cure). The other alternative **isipatsholo**, however, cannot be considered a taboo term since it appears in the lexicon as the generic term for venereal disease, i.e. syphilis and gonorrhoea.

Although the loan term **ivasekhithomi** for 'vasectomy' was preferred, some informants rendered the euphemistic taboo alternatives **ukuvala inzalo kwabesilisa** (literally, *to close the reproduction of the male*) and **ukuqeda inzalo kumuntu wesilisa** (literally, *to end the reproduction at the male person*). These latter two terms could be regarded more closely related to Zulu cultural taboo than the former loan term, **ukuvala inzalo kwabesilisa** being the shorter more condensed better term. The other alternative **ukuthena**, with the second highest frequency, however, can be considered the complete opposite of a taboo term since it is gross and actually means 'to castrate and emasculate (animals)'. Yet, this is an exaggeration (hyperbole) since a vasectomy is a procedure whereby only the tube carrying semen is cut; ruling out **ukuthena** as an acceptable term. See Addendum 2 Table 6.5.5.

It is interesting to note that there is no evidence of taboo in the alternatives rendered for 'bisexual'.

It is significant that the term rendered by informants **uncukumbili** (literally *double-sexed being*/hermaphrodite) is by far the most popular. However, the entry found in Doke & Vilakazi (1972), namely **uncukubili** is spelt without the **m**. The standard term **ukuba uncumbili** with the second highest frequency seems to be a derived shortened form, especially the concluding part **uncumbili**, or even a linguistically concealed taboo form of the term **uncukumbili** since no entry of **uncumbili** could be found in dictionaries. Even the two other rendered terms**, isitabani** (previously rendered for both 'homosexual male' and 'lesbian') and **ungqingili** (also previously rendered for both 'homosexual male' and 'lesbian' but actually meaning homosexual according to Nyembezi (1992)) seem to bear little evidence of taboo. Also rendered terms such as **isipatsholo** (sexually transmitted disease), the previously mentioned **ukuthena** (vasectomy, literally to castrate) and **uncukumbili** (bisexual) are examples where the taboo custom is deviated from.   It is furthermore observed that terms for sexual diseases are sometimes used interchangeably such as **isipatsholo** and **ukubhajwa** for the generic term 'sexually transmitted disease' and specifically 'venereal diseases' such as 'syphilis' and 'gonorrhoea'. This shows that there is still much confusion amongst speakers as far as terminology denoting sexually transmitted diseases is concerned. There is thus a need to revisit such terms and eventually distinguish them by means of accurate concept designation - indeed a formidable task for the terminologist also involving cultural aspects such as taboo preference (or not).

However, the following are some responses by informants strictly adhering to cultural taboo. A good example to illustrate taboo preference is the very popular preferred standard term **ijazi lomkhwenyane** (condom), literally translated as 'the coat of the young man'. The shortened standard taboo form of the latter term **ijazi** is also quite popular, compared to the less popular rendered taboo form **iglavu** (condom, a derived loan from 'glove'). The words **ijazi** and **iglavu**, exemplifying semantic shift, clearly fulfil the purpose of taboo, namely to avoid direct reference to the sexual object, 'condom'.

Another good example to illustrate taboo preference is the standard term with quite a high frequency

ukukhipha isisu (abortion), literally translated as 'to take out the stomach'. The other rendered oral equivalents **ukuhushula isisu** (to slip/draw out of the stomach), **ukuphuphuma kwesisu** (overflowing of the stomach) and **ukuchiteka kwesisu** (spilling out of the stomach), although less popular, are also clearly taboo forms. It should be noted that the latter two terms can also be regarded as equivalents for 'miscarriage'. The word **isisu** (stomach), another example of semantic shift, clearly fulfils the purpose of taboo, namely to avoid direct reference to objects with a sexual connotation such as **isibeletho** (womb) and **umbungu** (foetus, here being aborted). It has become clear yet again that oral corpus annotations cannot strictly be categorised since elaboration by means of taboo techniques actually also involves semantic shift whereby the internal resources of the language are used to the fullest.

One can conclude that taboo preference is not adhered to in all situations. Sometimes the loan terms or the gross terms are preferred to the taboo ones. It could be an indication that linguistic taboo is on the decline or that deviation from taboo has to do with linguistic intuition of the speakers of the language and cannot always be accounted for. That is why research in this regard is important; to determine for which terms taboo is preferred and for which ones direct reference is acceptable.

6.6 Conclusion

After comparing Corpus A (written) and Corpus B (oral) with emphasis on the difference between the two, one cannot say that the standardisation process of medical terminology was unsuccessful, since a fair amount of the standard terms are being used in the health environment. The efforts of the ZLC/B and the NTS cannot be underestimated; however, some practical suggestions can be made towards the improvement and acceptability of the elaboration and eventual standardisation process. These suggestions are forthcoming from an analysis based on some identified tendencies in the comparison of the written and spoken corpora, called *oral corpus annotation*. This oral corpus annotation throws some light on the terms that are actually being used in the practical environment,

some of which are more popular, trendy, accurate and in line with linguistic and cultural practices than the written standard ones.

Although the frequency of a term plays a role in recognised tendencies it should not be the ultimate criterion for corpus annotation purposes. It can, for instance, be that a term with a lower frequency, according to logical linguistic judgement, is a more suitable or correct term than the most popular term with the highest frequency. Practically, annotation can indicate which standard terms have become outdated or which new oral terms merit listing in a terminological list for standardisation purposes. These corpus annotation tendencies are arbitrarily named 'indigenous coinage', 'accurate designation', 'phonological adaptational trends', 'semantic shift alternative' and 'taboo preference' for the sake of systematisation.

The term **ukhakhayi** (mid-part of the head) is a popular indigenous term preferred to the standard loan term **ifontaneli** (fontanelle). The other indigenous equivalent **ufokothi** is further proof that there are sufficient terms available in the language to opt for outright coinage instead of borrowing. Indigenous coinage is mostly more understandable to language users since the internal resources of the language are used.

The standard term **umdlavuza wesifuba** for 'breast cancer', literally,cancer of the <u>chest</u>, is inaccurate since it caters for a far more generic and inclusive concept (chest) instead of an exclusive concept (breast). However, the oral term **umdlavuza <u>webele</u>** (literally 'cancer of the <u>breast</u>') for 'breast cancer' is far more popular and appropriate since it exemplifies accurate concept designation.

Preferred standard terms such as **isona** (sonar scan), **iyelo fiva** (yellow fever) and the acronym **i-AZT** (AZT) exemplify phonological adaptational trends in the Zulu language. These terms are more likely to be used rather than purer Zulu terms by the youth since they can immediately be associated with the English terms; they catch on; they represent phonological change in the language.

Nevertheless, the terminologist should take notice of such popular phonological trends in the language even if it seems a blatant taking over of English terms, resisting purist coinage.

A perfect example of the semantic shift alternative is found in **isithombe** (picture) a popularly used term in favour of the standard term **iX-reyi** for 'X-ray'. It is clear that the meaning of **isithombe** (picture) has been extended to include the related meaning of X-ray, i.e. the taking of a picture/photo.

Taboo is an avoidance phenomenon in the language that should be taken into account in the African languages when terms for sex education, also in relation to AIDS, have to be coined.
A good example to illustrate taboo preference is the preferred standard term **isifo socansi esithathelanayo** (sexually transmitted disease), literally translated as *the illness of the reeds mat /sleeping mat that takes on*. Since it is culturally taboo to refer to terms with a sexual connotation in a direct fashion, a culturally bound word such as a ' the reed mat/sleeping mat' **ucansi** is used to indirectly refer to 'sex'. However, not all terms with a sexual connotation are rendered in taboo form such as **uncukumbili** (bisexual, literally *double-sexed being*). This tendency cannot be explained other than attributing it to speaker intuition.

It has become clear in the discussion that oral corpus annotations cannot strictly be categorised since elaboration by means of semantic shift is actually also indigenous coinage and even taboo preference. A good example to illustrate this is the term **isifo socansi esithathelanayo** (sexually transmitted disease). The meaning of a cultural object such as a ' the reeds mat/ sleeping mat' **ucansi**, is semantically extended to refer to 'sex.' Yet, this is done indirectly, thus practising taboo, simultaneously practising indigenous coinage by using a pure Zulu word.

Massive work has already been done in the field of corpus building in the African languages by De Schryver and Prinsloo (2000a,b,c). However, the ultimate aim of any language or term elaboration process is not merely the compilation of organic (growing) corpora but also of acceptable corpora

representative of the living language. This implies that terminology be evaluated in terms of acceptability to speakers - and it is here that oral corpus annotation proves valuable.

Dissemination of draft or standardised existing term lists to professional persons working in the medical domain like nurses, nurse aids and doctors for their input prior to final standardisation, is thus of utmost importance. Only in this manner will terms be listed or standardised that are actually utilised, alive and acceptable to the consumer (language user) in the public and private sector.

Nevertheless, the lack of (medical) terminology in Zulu can be overcome if action is taken towards effective standardisation by means of proper corpus annotation, including oral corpus annotation. This means that terminologists should always annotate the corpora they work with, be they written or oral, organic or small, in order to improve the elaboration and related standardisation process and above all, to render acceptable terms.

CHAPTER 7

CONCLUSION

7.1 Introduction

In this chapter the main findings of this study are reviewed, stemming from two background perspectives, i.e. a national language planning perspective and a practical approach perspective. The first perspective concludes with general observations, based on Chapters 1 and 2. The second perspective concludes with major findings of this study originating from Chapters 3, 4, 5 and 6. Also the contribution this study has made to the standardisation and elaboration of Zulu as a technical language is examined. Although some limitations of this study are pointed out, the significance this study has for future application, training and possible research is also highlighted.

7.2 Background perspectives

The main findings of this study are reviewed, stemming from two background perspectives, i.e. a national language planning perspective and a practical approach perspective.

7.2.1 Standardisation and elaboration in the African languages within a language planning perspective

Mainly, the lack of terminological development in the African languages can be attributed to the lack of language and educational policy implementation, but also to the lack of coordination in the national language standardisation and elaboration processes. Considering a national language planning perspective general problems (observations) are stated and possible solutions are presented in Table 11 below.

However, although solutions to language policy problems may be integrated, they cannot be regarded as absolute but as an attempt to systematise the South African language planning scenario. The following identified problems and solutions are mainly based on findings of the (LANGTAG) Report on language services in DACST (1996) but also include the researcher's own observations after the completion of this study:

TABLE 11

Language (policy) problems in South Africa and possible solutions, specifically as regards the African languages

Problem	Solution
i) Bamgbose (1991) recognises that African language policies are generally characterised by avoidance, vagueness, arbitrariness and fluctuation, especially as far as implementation is concerned. Also in South Africa there is a gap between the language policy adopted by government and its implementation.	The government should properly implement its language policy by promoting language equality and practising multilingualism.
ii) Language services lack an adequate infrastructure and language workers enjoy a very low status.	The proper sophisticated technical infrastructure for language development and documentation should be put in place. Language workers (including teachers) should enjoy a higher professional status facilitated by a proper job description with corresponding higher compensation.

iii)	Owing to a legacy of colonial educational policies in Africa, unilingualism (English) instead of multilingualism is favoured in education.	Material must be made available for tuition in the African languages, even at the secondary level. African languages should increasingly be offered as languages of both learning and teaching. Language equity should be encouraged through active language awareness campaigns in all educational institutions.
iv)	In the African languages there is a lack of trained language workers such as translators, teachers and interpreters.	Language workers need proper training in the field of language teaching, teachers' training and literacy programmes. Training methods should be regularly updated, be more Afrocentric and more needs based in order to provide a meaningful language service, also in education. Linguistic training should be combined with computational training, for instance. Such training efforts in the African languages though, should be initiated and coordinated by language authorities and language interest structures in South Africa.
v)	Standard-setting structures do not function or coordinate properly.	Standard-setting structures such as the NLBs, the PLCs and the NLS, especially in the African languages, where there is a lack of terminology, should coordinate their efforts in order to ultimately develop and modernise these languages.

vi)	Because of the widespread use of English in trade and administration, language equity is lacking at the private, provincial and governmental sectors	The dignity of all languages in South Africa, including the Khoe and San, Nama and Sign Languages and even the more marginalised African languages Tsonga and Venda, must be respected in all sectors of society. Furthermore, the non-standard language varieties in the African languages can no longer be ignored. "African languages must be given high user status in society, beyond a symbolic official status" (De Klerk 1995b:33). Symbolism can only be overcome with a concerted drive, from the government and speakers of the African languages themselves, to develop and promote their languages.

7.2.2   A practical approach to the standardisation and elaboration of Zulu as a technical language

Although the African languages possess the basic tools that are necessary for their development such as orthographical standards, terminology lists, dictionaries, grammars and publications, these tools have some serious deficiencies that get in the way of effective technical elaboration and standardisation. In order to overcome such deficiencies, they need to be approached from a practical perspective.

When official and existing grammatical sources prove inadequate, research data in the form of written texts and oral responses should also be incorporated. Sager (1990) actually also implies that guidelines for language development should be laid down since national standardisation bodies, unlike international standardisation bodies such as the *International Organisation for Standardisation*

(ISO), rarely lay down standards and guidelines for the naming, compilation, selection and publication of terminology.

A practical approach does not claim to be absolute or prescriptive but it will at least serve as an example for those concerned with elaboration and standardisation and pave the way forward. Its aim is to guide language planners and terminologists, or to some extent even train them, to follow certain practical guidelines towards realistic technical standardisation and elaboration.

The language of exemplification for this method is Zulu and the field mainly medical terminology which may also include other fields to offer a general perspective. Four main deficiencies (problem areas) underlying the lack of language policy implementation and consequent lack of coordinating corpus planning, in this case elaboration and standardisation in Zulu, have been identified. These deficiencies mainly deal with:

i)   Inconsistency in the formulation, application and exemplification of orthographical (terminological) rules.

ii)   The lack of documentation and standardisation of the methods of word-formation, also in relation to culture.

iii)   Overlooking the value of written sources for the purpose of technical elaboration and standardisation for the purpose of term extraction, for instance.

iv)   Overlooking the value of oral sources for the purpose of improving the acceptability of technical terminology, for instance.

In the following section (see 7.3) the main findings of this study are reviewed against the background of the main aims set out for this study (see also 1.4). The first aim towards proper standardisation (here documentation) in relation to the applicability of the Zulu orthography and elaboration methods, i.e. word-formation, has been achieved. The second aim of providing a methodology to illustrate how

the real valid written and oral sources of the Zulu language in a specific technical (medical) field can be utilised for terminological work, has also been achieved. This second aim is closely linked to the application of corpus linguistics, especially the development of corpus linguistics methodology in Zulu. Considering a practical-approach-perspective, the main findings of this study are presented in terms of identified deficiencies and possible solutions.

7.3 **Orthographical terminological standardisation problems in Zulu**.

Deficiency 1: Inconsistency in the formulation, application and exemplification of Zulu orthographical (terminological) rules

After existing developmental tools such as orthographies compiled by the previous *Language Committee/Boards*, official terminology lists, grammars, dictionaries, published literature and technical (medical) leaflets distributed for primary health care, were investigated, it became apparent that there are still some deficiencies as far as orthographical standardisation is concerned. It was generally found that there are serious inconsistencies in the application of orthographical rules in the Zulu language in general and specifically as far as the latest developments in technical language and terminology are concerned. Another related problem is a lack of proper exemplification in juxtaposition to the standard formulated rules.

In particular, inconsistencies mainly relate to old versus new roman orthography, writing disjunctively or conjunctively, lack of accuracy in morphological notation, capitalisation and changing linguistic trends in the language which are not reflected in the orthography. There is thus an urgency that the work done on orthographical standard by the previous *Language Committees/Boards* be looked at critically.

7.3.1 **Solution 1: Solving orthographical terminological standardisation problems regarding Zulu**

The theory and issues of standardisation, in particular as far as the orthography is concerned, need to be addressed. The present ZNLB should be made aware of the mentioned problems. Poor formulation leads to problems in both the interpretation and application of the latest orthographical rules. These problems have been identified by questioning the logic behind some rules formulated by the previous ZLC/B. Problems have not only to be identified but also to be addressed by a workable practical approach, that will simultaneously enhance development and standardisation in the language. Adjustments and change in the orthography of Zulu have to be discussed by language interest structures, first of all on national level by the ZNLB, and also on the provincial level by the PLC for Zulu. In order to solve the five identified problematic issues concerning Zulu (technical) orthographical standardisation, the following recommendations are made:

7.3.1.1 Old versus new Zulu orthography

Traces of old Zulu orthography used in earlier Zulu publications should consistently be replaced by new orthography in, for instance, Zulu surnames, e.g. **dhl** in **Dhlomo** should be replaced by **dl** in **Dlomo**. The **h** in -**hahama** should be replaced by **hh** as in -**hhahhama** (growl), in the case of the voiced glottal fricative sound.

Aspiration (**h**) should follow the plosives **k**, **p** and **t** where required after a period during which it was omitted or used inconsistently. Since the lack of aspiration can bring about change in meaning, e.g. **-teta** (carry on back) and **-thetha** (reprimand) its inclusion has become imperative and entrenched by grammarians such as Doke (1945). However, lack of aspiration that is still found in place names should be changed to new orthography, e.g. **Tokoza** should become **Thokoza** (including aspiration).

7.3.1.2 Writing disjunctively or conjunctively

Writing disjunctively or conjunctively deals with the writing system of a language (the written word):

In following (in simplified terms) the disjunctive writing system morphemes are written separately to form a word/phrase, e.g. **si ya cul a** (we sing), whereas in the conjunctive system morphemes are combined to form a word /phrase, e.g. **siyacula** (we sing). However, Zulu gradually developed from a language with a disjunctive writing system to a language with a conjunctive writing system although both systems were earlier arbitrarily used by grammarians. Even though the conjunctive writing system is established in Zulu, the disjunctive writing system is evident in the writing of the demonstrative.   The previous ZLC/B constantly varied the rules for the writing of the demonstrative from conjunctive to disjunctive and back again.   Even today it is common to find the demonstrative written conjunctively as part of the following noun, e.g. **lesisitsha** (this plate - one word) instead of **lesi sitsha** or **isitsha lesi** (two words) in accordance with the latest Zulu orthography.

The use of the apostrophe, like the hyphen also deals to some extent with the writing system. Their use in words can be regarded as part of the conjunctive writing system since they prevent lexical items or sounds from being entirely separated. According to the Zulu orthography (DET 1993) the apostrophe should only be used to indicate elision, e.g. **angesabi 'nja** (I do not fear any dog) where the **i** in **inja** is elided. However, in examples **iyul'sa** (ulcer) and **iyul'na** (ulna) in DACST (1997a) its use to indicate elision is not altogether clear since they can, without further ado, be written without the apostrophe as **iyulsa** and **iyulna** without any apparent pronunciation problem.

According to the Zulu orthography (DET 1993) the hyphen is mainly used to join concords to numerals, e.g. **zingu-9** (they are 9); to separate two vowels, e.g. **ama-apula** (apples); in enclitics, e.g. **woza-ke** (come then) and for 'practical reasons' in lengthy compounds, e.g. **ikhaboni-dayoksayidi** (carbon dioxide - *own example*). The vague phrase 'for practical reasons' does not explain via examples on what grounds a compound can be considered lengthy enough to merit the use of a hyphen. For instance, should a lengthy compound like **ukhandalimtshelokwakhe** (wayward person) simply be written conjunctively or should it be hyphenated? To solve this confusion the rule to include the hyphen in lengthy compounds should be revisited and properly exemplified by the

ZNLB, taking conjunctiveness as a given.

Furthermore, the hyphen will undoubtedly appear more and more in technical acronyms and can therefore not be ignored by the present ZNLB. A proposed rule is that the hyphen be inserted between the initial vowel prefix and the capitalised acronym, e.g. **i-AZT** (zidovudine - drug used for the treatment of AIDS). However, in order to maintain consistent conjunctivism, the use of the hyphen and apostrophe in words/ terms should be minimised.

7.3.1.3 The lack of accuracy in morphological notation

Notation specifically deals with the way in which lemmas (basic forms) or terms are listed in terminology lists. Inconsistencies in morphological notation may be attributed to terminologists lacking sufficient linguistic knowledge or training. Specific official terminology lists were investigated and the main notational morphological inaccuracy that could be found is the inconsistent use of the prefix **uku-** (class15) in the notation of verbs, specifically in relation to verb stems, infinitive verbs and deverbative nouns. The prefix **uku-** can precede any Zulu verb stem, forming an infinitive implying the meaning of 'to' or bringing about change in the word category and eventually in the meaning, e.g.

> **gaya** (digest) can form an infinitive verb **ukugaya** (to digest) or a deverbative noun **ukugaya** (digestion).

Morphologically the most accurate and least confusing manner in which to notate the verb, is to consistently list the stem, e.g. **gaya** (digest). This is a simple rule which can be applied in all future official publications to solve inconsistencies in the listing of lemmas which are not complete words. This form of notation also implies that the **uku-** (class 15 prefix) be omitted so as to avoid confusion by the uninformed user as to whether **ukugaya** , is a verb (to digest) or a noun (digestion).

7.3.1.4 Capitalisation

Capitalisation in terminology practice has not been discussed orthographically by the previous ZLC/B. In relevant official and commercial publications it was found that capitalisation as far as terminology is concerned can roughly be based on capitalisation in English, adhering to the linguistic structure of Zulu. The following rules could be formulated in order to deal with capitalisation in terminology:

* The first letter after the initial vowel in the name of a product or name of an international technical (medical) term is capitalised; an easy rule to apply since the first letter of such a name is capitalised in English too, e.g.

  iPurity (Purity - commercial baby food) and **ukubakhona kwefekitha yeRhesus** (Rhesus positive).

* In technical capitalised acronyms the initial vowel preceding it is in the lower case, e.g. **i-HIV** (HIV - human immunodeficiency virus).

7.3.1.5 Changing linguistic trends in the language which are not reflected in the orthography

Phonologically, the traditional open syllabic system of consonant vowel (CV) in a syllable, e.g. **ikilasi** (classroom) does not necessarily apply any longer. It is already obvious that the Zulu syllables are becoming more closed, not necessarily ending in vowels, especially in modern loans, e.g. **istradi** (street) instead of **isitaladi**. It is noted that the **r** is increasingly being used where it was previously replaced by **l**.

Morphological changes are also quite common as exemplified by (Koopman1994), e.g. the use of the verbal extension **-ish-** in adopted verbs such as **-filisha** (fill out a form).

It is important that recent phonological and morphological changes or trends be discussed and evaluated. If these are continually being used, they should be taken up in the orthography of the language so as to represent the living language, a task to be facilitated by the present ZNLB.

7.3.2   Main findings concerning orthographical standardisation

The first practical step towards effective practical standardisation occurs on the most basic level, i.e. on the orthographic level. This level is the easiest to attain as put forward by Sager (1990) who reasons that standardisation can be successful on at least the levels of spelling, pronunciation, morphology and syntax. Scholars such as Thipa (1989) and Mathumba (1993) recommend that the composition of the *Language Boards/Committees* (now known as *National Language Bodies*) be changed to include more members who are qualified in linguistics and language planning. Such composition will ease the task of terminologists since inconsistencies in the linguistic formulation and exemplification of Zulu orthographical rules and in morphological notation will be reduced.

7.4   Deficiency 2: The lack of standardisation of the methods of word-formation that facilitate language and technical elaboration in Zulu

For a long time African language development was approached from a Eurocentric perspective (LANGTAG in DACST 1996): Initially such development focused on the establishment of an orthography, Christian terminology and basic school terminology. Furthermore, the indigenous word-formation patterns in the African languages (also Zulu) by means of which technical elaboration is achieved, have thus far not been properly documented or standardised. Also very little attention has thus far been given to extra- linguistic factors such as culture in relation to these word-formation patterns. Since comprehensive works on word-formation patterns is largely lacking in the grammars of the African languages, including Zulu, this is another problem to be addressed by the ZNLB.

7.4.1   Solution 2: Towards the standardisation of the methods of word-formation that facilitate language and technical elaboration in Zulu

Language elaboration in the African languages, including Zulu, is achieved by means of linguistic tools

being indigenous word-formation methods. However, these word-formation methods still have to be properly documented and standardised, specifically by the NLBs (Alberts 2003). Different types of word-formation methods in Zulu were discussed and exemplified in order to give an overview of all possible types. This was done since the documentation of such word-formation patterns will provide guidelines not only to ease the task of terminologists but also help in their linguistic training. In a way it is a practical approach towards standardisation, at least to some extent, of the word-formation methods of Zulu. For effective terminological practice, terminologists must have the necessary linguistic insight in the formation strategies of words or terms. The proper documentation of word-formation methods in Zulu can lead to the possible publication of a style or reference manual for training in terminological practice.

Corpus planning deals to a great extent with language elaboration (more specifically word- /term-formation) which may require the use of technical language. Technical language is more formal, logical and more inclined to show a correlation between the concept and the term than ordinary language. It has been found that the methods of word-formation that facilitate (technical) language elaboration are derivation, semantic shift, compounding, loan-translation, deideophonisation, borrowing and abbreviation. These methods of word-formation draw either on the internal resources of the language or on the external resources which are borrowings from other languages.

7.4.1.1 Derivation

When terminologists want to coin terms, one of the very basic methods to expand the lexicon is through derivation. Derivation is word-formation through affixation, i.e. creating a word or term from indigenous roots. To some extent, affixation (the adding of prefixes or suffixes or both) is evidenced in all the mentioned methods of word-formation but particularly strongly in derivation. The verb root **-nyel-** (to sprain), for instance, derives the noun **isinyelo** (sprain) by adding the class prefix **isi-** and the impersonal nominal suffix **-o**.

7.4.1.2 Semantic shift

Like derivation, semantic shift is also a method that utilises the internal resources available in the language. Semantic shift occurs when an existing meaning of a word is extended to name a new related concept, e.g. **isithombe** (photograph) now also names a new concept **isithombe** (x-ray) because of parallels in meaning between the former and latter terms.

7.4.1.3 Compounding

It has been found that the method of compounding, the formation of one word out of two or more words, produces some of the most original purist coinage, e.g. **um**- (class prefix)
- thola (obtain) + **impilo** (health) > **umtholampilo** (clinic). The latter example is a typical combination of a verb phrase (VP) plus a noun phrase (NP) commonly found in Zulu compounds.

7.4.1.4 Deideophonisation

Deideophonisation is a word-formation method of coining new terms by means of ideophones (onomatopoeic words).   This method exemplifies unique purist coinage in that a term is formed on the basis of its onomatopoeic resemblance to a specific sound.   An example is **isithuthuthu** (motorcycle) formed on the basis of its resemblance to the sound of a running engine **thuthuthu**.

7.4.1.5 Loan-translation / calquing

Unlike the previously discussed methods loan-translation is a method of word-formation that draws on foreign language resources since it entails translating the meaning of a new term from the donor language into the target language, e.g. 'embryo' is translated as **isibindi sembewu** (literally, *the core of the seed*).   Such Zulu loan-translations are mostly morphologically speaking possessive constructions and semantically speaking conceptual translations.

7.4.1.6 Borrowing

The method of word-formation that draws exclusively on foreign resources in terms of loans from other languages, is borrowing. Borrowed terms from a donor language are usually phonologically and morphologically adapted to suit the target language structure. For instance, **ijondisi** (jaundice), is morphologically adapted by prefixing the class prefix **i-** and orthographically adapted by introducing a Zulu vowel **o,** the consonant **s** and finally concluding the syllable with the vowel **i** .

Many new sound combinations occur in borrowings which were previously uncommon in Zulu, such as **thr** in **ibhethri** (battery). According to Koopman (1994) and Hlongwane (1995) these sounds should be incorporated into the language.

7.4.1.7 Abbreviation

Abbreviation is another method of word-formation that occurs in the blending and clipping of words and in acronyms. Blending occurs when the parts of two or more words merge and can therefore also be regarded as part of compounding, e.g. as in the case of **isidaka imizwa** (literally, *intoxication of the senses*) > **isidakamizwa** (narcotic). Clipping also occurs when a term is reduced to one of its parts, e.g. **iskizo** for 'schizophrenic patient'.   Acronyms occur universally and are initial words of an expression which are abbreviated in a sequence of capital letters in order to form a term, e.g. **i-HIV** (human immunodeficiency virus - HIV).

7.4.2   Main findings on elaboration methods, also in relation to culture

After having discussed the seven identified methods of word-formation in Zulu it became evident that no clear-cut division exists between these methods since a great deal of overlapping occurs. The

classification of such word-formation methods is thus merely pursued for the sake of systematisation. Compounding, for instance, is a method of word-formation that not only conforms to derivation but also to semantic shift and abbreviation (phonological blending), e.g. **izi-akha** (build) + **umzimba** (body) > **izakhamzimba**(energy food). Furthermore, the divisions between compounding and loan-translation are far from clear-cut. Had the loan-translation **isibulala magciwane** (literally, *killer of germs* - antibiotic), for instance, been written as one word or hyphenated, it could have been considered a compound.

Furthermore, most purists are not in favour of indiscriminate borrowing as they feel that it may eventually suppress the source language and pollute it. In this regard Ohly (1987) refers to the suppression of kinship terms in the African languages by English. In Zulu, for instance, the following specific kinship terms have been suppressed, namely **udadewethu** (my/our sister), **udadewenu** (your sister) and **udadewabo** (his/her/their sister) by an only Afrikaans /English borrowing **usisi** (sussie/sister). The African languages are not simply recipients of foreign terms. On the contrary, they supply the Western world with some unique, almost untranslatable terms, e.g. **indaba** (story, case) and **imamba** (snake species) (Ohly 1987). On the other hand, scholars also mention the advantages of borrowing being faster and cheaper than purist coinage (Fourie 1993). Mochaba (1987) also points out that near-similarity in technical terms would promote acceptability. However, borrowing will not pose a threat to purist coinage if the African languages are maintained and if borrowing is pursued within the framework of the existing resources of the language, such as derivation, semantic shift and compounding.

Besides the effective use of the discussed word-formation patterns, the study of Zulu language elaboration will not be complete without mentioning the way in which such elaboration is linked to extra-linguistic factors such as culture. ' World view' and 'taboo', for instance, are two culture-related aspects which should be considered in the terminological development of the language.

The world view hypothesis is based on the theory that a person's mother-tongue offers him /her a

framework for his/her perception of the world. A term that, for instance, through semantic shift, correlates with a certain cultural world view is **ukubeletha** (literally, *to carry on the back* - labour).

The connotation with world view here is that in African culture labour is associated with the action of carrying a child on the back after the birth. Taboo, on the other hand, applies to terms that are not allowed to be used in a specific social context because they would be considered offensive or vulgar such as **isifo socansi** (sexually transmitted disease, literally *the illness of the reeds mat/sleeping mat*). A cultural object such as **ucansi** (the reeds mat/sleeping mat), is therefore used to avoid direct reference to a term with a sexual connotation. However, the utmost form of taboo occurs when loan terms are used to replace the Zulu taboo terms, e.g. **ukuba phregi** (be pregnant) replaces **ukuba nesisu** (literally, *to be with a stomach*). The English term is preferred in order to avoid the use of the target language - Zulu itself! Nevertheless, the proper understanding and application of taboo have become imperative for terminologists who have to devise terms for important current health issues such as sex education and AIDS.

7.5 Deficiency 3: Overlooking the value of written Zulu sources in term expansion for elaboration and standardisation purposes

A question from the STANON-Report (Calteaux 1996) that relates to standardisation in the African languages thus far, is whose standard was used and where the sources of this standard were. Could it be that this standard was based on the standard of the previous *Language Boards/Committees* or even the standard of certain terminologists involved in the task of official standardisation? Basically there are two main sources of language, namely the written and the oral version. Firstly the written

source of a language such as Zulu should be taken as point of departure; a source mostly overlooked until recently. In relation to the latest developments in corpus linguistics it is appropriate to use written sources for term expansion, i.e. to extract terms by means of electronic query tools from published text corpora. Term extraction can, for instance, establish which technical terms have been incorporated into the Zulu lexicon to such an extent that they are used in the written (published) format.

Pioneering work has been done in the field of term extraction from written corpora in the African languages by De Schryver and Prinsloo (2000a,b,c). However, automatic term extraction is not possible in the African languages since most computerised query tools were developed for European languages such as English which is morphologically a simple language in comparison to Zulu which is an agglutinative complex language with a conjunctive writing system. Thus, the only option left is to resort to a method of semi-automatic term extraction.

7.5.1 Solution 3: Utilising written corpora for semi-automatic term extraction for elaboration and standardisation purposes

For the proper elaboration and standardisation of terminology, the point of departure should always be the language, be it the written or oral format. In this case it is using the written living language in the context of its use as source for terminology. Term extraction from valid written sources can contribute to the documentation and standardisation of representative and acceptable terms, a process that earlier depended to a great extent on the random intuitive creation of terms by the terminologist. The lack of terminology in the African languages can be overcome if action is taken towards effective practical standardisation through the collection of suitable written text sources to which the theory of corpus linguistics can be applied.

A corpus is a body of written or spoken language data, which can be used as a basis for linguistic research. In this thesis the Zulu medical pamphlets dealing with the aspects of primary health care

constitute the written (published) corpus. Specific texts were collected dealing with three aspects of primary health care; three texts for cholera, four texts for tuberculosis and four texts for HIV/AIDS which are the sub-corpora of the main medical corpus of eleven texts comprising a total of 2187 words. Since this structured corpus is very small it is by no means representative of the medical field, per se. The aim of the compilation of this small corpus is to exemplify methodology and insights gained in the extraction of terms and make them applicable to other more representative or bigger corpora. The selected texts were electronically scanned, edited, and entered into computer files. Eventually the terms were extracted and encoded according to frequency count and concordance (the use of a term in context) by means of computerised query tools.

Corpus query tools can ease term extraction for the terminologist, but never replace human involvement. In Zulu with its complex conjunctive structure in particular, manual manipulation of the morphology is needed by the terminologist in order to use the *concordance* tool effectively to extract terms for lemmatisation (entry) purposes. Such skill will further enable the terminologist to determine the actual frequency of a term in order to decide, alongside other synonymous terms, on its inclusion in a terminology list. What follows is a methodology on term extraction, for instance, how simple nominal terms, complex verbal terms and complex multi-terms are extracted.

7.5.1.1 The extraction of simple nominal terms from a written corpus

Through computerised query tools, called *WordSmith Tools* (WST), frequency word counts and coordinating alphabetical word lists, called *wordlist*, were conducted in order to isolate and extract words, some of which are terms, e.g. **izimpawu** (symptoms). Obviously these tools facilitate the task of the terminologist since s/he has immediate access to the ranked frequency of a word or term in the medical corpus, e.g. **izimpawu** has a frequency (occurrence) of 15, ranked in the 33rd position, being the 773rd alphabetical word. In this manner the terminologist is supplied with additional information about the text to decide which terms can be included in a term list. To grasp a term such as **izimpawu**

in the context of preceding and following words, it can be entered as a search word by making use of another tool, the *concordance* query tool.

When extracting terms from a written Zulu corpus, morphological analysis goes hand-in-hand with lemmatisation, i.e. establishing a suitable entry in a term list. Before a term can be lemmatised an overview of all its realisations, both uninflected and inflected, is needed to judge how representative it is, or simply how high its frequency is. Manual morphological manipulation is particularly needed before the *concordance* query tool can be used effectively in the African languages (Zulu), since these tools were particularly designed to suit the structure of European languages and not the complex conjunctive structure of the Zulu language, for instance.

To get an overview of the nominal term, **izimpawu** (symptoms), for instance, its basic noun stem -**mpawu** is isolated through sensible morphological manipulation by the terminologist and is entered as a word search through the application of the *concordance* tool. All realisations of **-mpawu**, e.g. **zimpawu** (3) and **benezimpawu** (6) are listed as separate words with different frequencies in the initial *wordlist*. It is also noted that the *initial frequency* of **izimpawu** of 15 in the *wordlist*, has increased to 23 after the application of the *concordance* tool. This increased frequency of 23 can thus be regarded as the *actual frequency* of the representative term **izimpawu** in the corpus.

7.5.1.2 The extraction of complex verbal terms from a written corpus

The verbal category is particularly difficult to lemmatise since the verb root is usually inflected - different prefixes and suffixes can be added to it to indicate different tense forms/moods. The verb **ikuvikela**, for instance**,** comprises the morphemes **i-** (present tense subject concord class 9) **-ku-** (object concord 2nd person singular) -**vikel-** (verb root) and -**a** (positive verbal ending). The most

basic form of this verbal term, the root, **-vikel-** is established through manual morphological manipulation. Such manipulation is needed for the purposes of lemmatisation, which involves the listing of basic uninflected nouns such as **izimpawu** (symptoms) and basic verb stems such as **-vikela** (prevent) in a Zulu terminology list.

Once the basic form has been identified, here being the verbal root **-vikel-** (protect), and entered in the form of a search word *****vikel*** (* being used to indicate where the word is broken), the *concordance* tool reveals a high frequency of 33 in the corpus, indicating a representative term. All realisations of **-vikel-** (inflected and uninflected) are listed as separate words with different frequencies in the initial *wordlist*, e.g. **ikuvikela** (2) , **ukuvikela** (6), etc. It is noted that the *initial frequency* of the term **ikuvikela** of (2) in the *wordlist* has increased to the actual frequency of 33 (for the basic stem **-vikela**) by the application of the *concordance* tool.

The *concordance* tool is also quite handy to extract more complex verbal terms such as **-thelel-** containing verbal extensions. In the wordlist each realisation of **-thelel-** is listed separately with its own frequency, e.g. **ingathelelana** (1), **kungakuthelela** (1), etc. When a word search through the *concordance* tool is applied, not the root -**thel**- but rather the root plus the extension, i.e. -**thelel**- (spread disease), thus ***thelel*** should be searched for. The *concordance* result is that the verbal term -**thelela** has an actual frequency of 10 in the corpus.

7.5.1.3 The extraction of complex multi-terms from a written corpus

A corpus of medical texts, does not only contain single terms such as **isifo** (illness) but also multi-terms such as **isifo sofuba** (tuberculosis, literally *the illness of the chest*), which consist of two or more words. These multi-terms are listed as separate words with different frequencies in the same

wordlist, e.g. **isifo** (25) and **sofuba** (18). Since the *wordlist* tool cannot isolate multi-terms, the *concordance* tool can perform this task through the application of a word search. Entering the multi-term **isifo sofuba** as a word search results in an actual frequency of 10 in the corpus.

7.5.2 Main findings concerning semi-automatic term extraction from written corpora

The *concordance* tool complements the *wordlist* tool since it can establish the actual frequency instead of the initial frequency of a term, especially in inflected nominal and verbal terms, but also since it reveals how a certain term is used in context.

What has become clear thus far is that extracting (and eventually lemmatising) terminology in the African languages, here Zulu, is a combination of manual and computerised processes. Applying the *wordlist* and *concordance* functions of the corpus query tool *WordSmith Tools* is actually a form of text encoding that can be used as a basis for recommendations and predictions to be made for more effective practical term extraction, lemmatisation and eventual standardisation as regards terminology. Text encoding reveals additional information and conditions of the use of a term, for instance:

* In the 33 occurrences of -**vikel**- (protect) in the corpus established through *concordance*, the form -**zivikel-** employing the reflexive prefix **-zi-** (oneself), a very important notion in the combatting of serious illnesses, is used in 16 of these occurrences. Information that complements a lemma such as **-vikela** and the conditions of its use such as the notion of the reflexive **-zi-**, should thus be added in the front matter or back matter of a terminology list to the benefit of the user.

* The term **isifo sofuba** is the indigenous multi-term for the acronym **i-TB** or the loan **ituberculosis**, all synonymous terms for 'tuberculosis'. Through a word search via the *concordance* query tool, these terms have respective frequencies of 23, 10 and 1 in the corpus, thus proving that the shorter acronym is the most popular, in accordance with universal

tendencies in terminology.

However, after a specific representative corpus, like this medical corpus, has been built and encoded by means of corpus query tools, some material may be added or deleted to enhance the corpus. Thereafter the whole process will start all over again. It is recommended, however, that the application of corpus linguistics, here the use of computerised corpus query tools, should be coordinated among the African languages in South Africa. In some respect this is already done in the functioning of the NLUs and in the establishment of corpora for eleven official languages thus far. Since terminological practice is not that far removed from lexicographical practice other language interest structures such as the NLS and all the NLBs should coordinate their terminological practices with those of the NLUs. Obviously such coordinating efforts should be supported by PanSALB whose operations include lexicography and terminology development, according to Alberts (2003).

7.6   Deficiency 4: Overlooking the value of oral sources for improving the acceptability of technical terminology in Zulu

Not only written corpora should be the source for the elaboration and standardisation of technical terminology but also oral corpora which reflect the living language in the workplace. In other words oral corpora should serve the consumer market and be used in general technical (medical) discourse. However, the use of oral corpora as a basis for research in practical elaboration and standardisation in the African languages, Zulu in particular, has thus far been avoided. The reason perhaps is that it is more complex to compile a structured oral corpus which involves transcriptions of interviews and recordings, than a structured published written corpus. What should happen is that terminologists should use terms that already exist in society, thereby promoting natural term development (Jafta 1987).

Thus far the standardisation and elaboration of terminology has not been realistically verified by, for

instance, comparing oral terms with existing written standard terms. By this comparison the acceptability of terms can be improved and the success of the process can be verified, at least to some extent. In this manner it can be determined which technical terms have been incorporated into the Zulu lexicon and to what extent they have been accepted and used by the Zulu-speaking community in actual oral use as evidenced by everyday communication. The use of oral corpora as a basis for research in practical elaboration and standardisation in the African languages is thus of the utmost importance. For this to happen the oral corpus should be analysed in order to find out why certain terms are more acceptable than others.

7.6.1 Solution 4: Utilising oral corpora and their annotation for improving the acceptability of technical terminology in Zulu

Not only written corpora should be the source for the elaboration and standardisation of technical terminology but also oral corpora which reflect the living language in the context of its use, for instance, in the workplace. The use of terms from oral sources can contribute to the documentation and standardisation of representative and acceptable terms. However, term elaboration and standardisation earlier depended to a great extent on terms created by the previous *Language Committees/Boards* or by the random intuition of terminologists. According to Calteaux (1996) the evidence of standard should also be oral-aural based. The use of oral corpora as a basis for research in practical elaboration and standardisation in the African languages will promote natural term development by using terms that already exist in the workplace, for instance. Questionnaires were distributed to health workers in six (provincial) hospitals, five in KwaZulu-Natal and one in Mpumalanga. Professional medical mother-tongue speakers were interviewed in their practical work domain in order to get an intuitive natural response on terminology. See Addendum 1. In this indirect manner the viewpoints of speakers on term standardisation and elaboration, also implying the acceptability of technical health terms, were determined.

Selected existing written standard medical terminology (Corpus A) was compared with equivalent oral medical terminology (Corpus B), with emphasis on the difference between the two. This comparison was conducted to determine the acceptability of standardisation, i.e. to give an indication of which terms are actually used in the field to the extent that they can be listed as entries in a valid medical terminology list. After such comparison it was concluded that the efforts of the ZLC/B and the NLS cannot be underestimated since a fair amount of the standard terms are being used in the health environment and their documentation of medical terminology was contributing to language development. However, some practical suggestions can be made towards the improvement of the acceptability of terms in the elaboration and eventual standardisation process. These suggestions are based on some identified tendencies mainly determined by frequency count, while comparing the standard written and oral corpora. These tendencies were arbitrarily named 'indigenous coinage', 'accurate designation', 'phonological adaptational trends', 'semantic shift alternative' and 'taboo preference' for the sake of systematisation. See Addendum 2. This analysis of common tendencies is called *oral corpus annotation* and indicates which terms are more popular, trendy, accurate and in line with cultural practices in the practical environment than the written standard ones or vice versa. However, high frequency of a term is not the ultimate criterion for corpus annotation purposes and does not merit the listing of a term in a terminology list. Sometimes a term with a lower frequency is a more suitable or accurate term.

7.6.1.1 Indigenous coinage

The term **ukhakhayi** (mid-part of the head) is a popular indigenous term preferred to the standard loan term **ifontaneli** (fontanelle) and the other indigenous equivalent **ufokothi**. The two indigenous terms prove that Zulu has sufficient internal elaboration capacity to avoid borrowing and in so doing making terms more understandable to language users.

7.6.1.2 Accurate designation

The standard term **umdlavuza wesifuba** for 'breast cancer' is inaccurate since it caters for a far more generic and inclusive concept (**isifuba -** chest) instead of an exclusive concept (**ibele** - breast). The very popular oral term **umdlavuza webele** (literally, *cancer of the breast*) exemplifies accurate concept designation and thus merits listing above the former term.

7.6.1.3 Phonological adaptational trends

Phonological adaptational trends in the Zulu language are exemplified in preferred written standard terms like the shortened **isona** (sonar scan), **iyelo fiva** (yellow fever) and the acronym **i-AZT** (AZT). Since these terms can immediately be associated with English terms they catch on in youth circles; they represent change and have an influence on the standardisation of terminology.

7.6.1.4 Semantic shift alternative

In the popularly used term **isithombe** (picture) in favour of the standard term **iX-reyi** for 'X-ray', the meaning of **isithombe** (picture) has clearly been extended to include the related meaning of X-ray, i.e. the taking of a photograph. The use of **isithombe** (X-ray) and **esithombeni** (literally, *the place of pictures*) for X-ray room are perfect examples of using the semantic shift alternative in order to find a suitable term.

7.6.1.5 Taboo preference

The concept of taboo deals with acting respectfully in a specific social context including using language of avoidance. The preferred standard term **isifo socansi esithathelanayo** (sexually transmitted disease), literally translated as *the illness of the reeds mat/sleeping mat that takes on* illustrates such taboo preference. A culturally bound word such as a ' the reeds mat/ sleeping mat' **ucansi** is used to avoid direct reference to 'sex'. However, not all terms with a sexual connotation are orally

rendered in taboo form such as **uncukumbili** (bisexual - literally *double-sexed being*), a tendency that cannot be explained other than attributing it to speaker intuition.

7.6.2   Main findings concerning oral corpus annotation

It has become evident that oral corpus annotations cannot strictly be categorised since elaboration by means of the semantic shift alternative is actually also indigenous coinage and even taboo preference, e.g. in the term **isifo socansi esithathelanayo** (sexually transmitted disease).   The meaning of a cultural object such as a 'reeds mat' **ucansi**, is semantically extended to indirectly refer to 'sex' , thus practising taboo and indigenous coinage by using a pure Zulu word.

The ultimate aim of any (technical) language elaboration process is not merely the compilation of organic (growing) corpora but also of acceptable corpora representative of the living language. This implies that terminology be evaluated in terms of acceptability to speakers - and it is here that oral corpus annotation proves valuable.

Proper dissemination of draft term lists to professional persons working in the technical (medical) domain like nurses, nurse aids and doctors for their input prior to final standardisation, is thus a priority. This will ensure that only those terms be listed or standardised that stem from the living language, those are that are actually used and acceptable.

Nevertheless, the acceptability of (medical) terminology in the African languages, and Zulu, can be improved if action is taken towards effective elaboration and standardisation by means of proper corpus annotation, including oral corpus annotation. This means that terminologists should always annotate the corpora they work with, be they written or oral, organic or small and above all, make their findings known to fellow terminologists working in other (related) African languages and even to language authorities. Only in this manner can term elaboration increase (more terms become available) and acceptability improve.

7.7   The contributions of this study

The main contribution of this study is the adding of a practical approach to the standardisation and elaboration of Zulu as a technical language, specifically in the medical field. Practical solutions to standardisation and elaboration problems are offered through proper methodology and exemplification, the point of departure being the real written and spoken Zulu language in the professional (medical) domain.

The exploratory part of this study was able to determine that the lack of terminology in the African languages, Zulu in particular, could firstly be attributed to lack of implementation of language policy and secondly to lack of proper coordination between language interested structures such as the NLS and the NLBs, for instance.

In addition, this thesis provides a practical methodological approach as basis on which future research on technical elaboration and standardisation in Zulu can be built. This approach specifically considers the orthography, the methods of word-formation and the utilisation of relevant written sources to extract terminology by means of corpus query tools, and the utilisation of relevant oral sources for term comparison in order to arrive at corpus annotation (analysis).

This study attempts the proper documentation of the orthography and methods of word-formation that facilitate (technical) language elaboration in Zulu with the aim of standardisation. The practical approach proves valuable since terminologists now have a designed methodology at their disposal to ease their task towards more effective elaboration and standardisation. The exemplification of these methods can form a basis for similar applications in term creation in other fields. The practical methodology promotes the utilisation of relevant real language sources, be they written or oral, in terminological work. These sources can, for instance, be used for term extraction and corpus annotation which is closely linked to the application of corpus linguistics, thus providing corpus

linguistics methodology specifically to an African language such as Zulu.

This study emphasises the utilisation of the real relevant sources of the living language in terminological work instead of the intuition of terminologists. Another implication of this study is that it can contribute to the practical training of terminologists through proper language exemplification of the methodology.

7.8   The limitations of this study

This study does not attempt to be prescriptive in its practical approach to the standardisation and elaboration of Zulu as a technical language. Rather, it provides some guidelines towards an understanding of national standardisation and elaboration of the African languages, in particular Zulu. The methodology focuses on the application of corpus linguistics to a specific set of data. This data is mainly selected written and oral medical corpora which cannot claim representation of the entire Zulu medical corpus.

This study suggests how the orthographical rules can be improved to ease their application in the Zulu language, also in relation to terminology. But, although this study can make the ZNLB aware of orthographical problems, and of the application and exemplification, for instance, of helpfully improved orthographical rules, it cannot see to it that these rules are duly adjusted.   Changes can only be effected if they go through the official channels of the national language authority PanSALB. Thereafter official documentation and dissemination of the latest orthographical rules to all the public and educational institutions should be administered.

It will require more than a single study to contain sufficient motivation to adjust orthographical rules. Suggested changes will have to be reviewed by linguistic experts of Zulu as well as members of the ZNLB alike.   To effect orthographical changes has been a long and cumbersome process in well established languages such as German and English since changes have not always been readily

accepted by speech communities.

This study provides an overview of the methods of word-formation that can be (re)used for elaboration purposes, having proper documentation and eventual standardisation in mind. It also emphasises that a great deal of overlapping occurs in these methods and that they cannot clearly be divided. It is in this respect that more research, based on sound linguistic motivation, is needed towards a proper less overlapping division of these word-formation methods. This thesis also makes an attempt to provide, through this overview, for the training of terminologists, even if it is just the initiation of further more advanced training.

In this thesis relatively small written and oral technical corpora were compiled. However, the aim is to make insights and methods gained from these small corpora applicable to bigger growing corpora, such as those in national research projects. The application of methods includes the utilisation of both valid written and oral sources in terminology by applying corpus linguistics. It shows how technical written texts can be utilised for semi-automatic term extraction towards lemmatisation and also how a type of corpus annotation can be developed by comparing written and oral corpora. It also shows how corpus annotation can improve the acceptability of terms.

Although only Zulu speakers who are also medical workers were interviewed by means of a Zulu questionnaire in order to gather oral terms, there is a still a slight possibility that some informants could misinterpret some procedures, or that there was a misunderstanding between the interviewer and the interviewee.

Yet, all the parameters of corpus linguistics and its application are far from exhausted. The application of corpus query tools should be exploited further, especially in the African languages, also Zulu, where automatic term extraction seems a very distant goal and where advanced text tagging has not been achieved.

It is not possible to cover all aspects of standardisation and elaboration of Zulu as a technical language in this study. Rather it is possible, considering the lack of previous guidelines, to provide a methodological practical foundation with sufficient exemplification in order to ease application in terminological work.

7.9 The way forward: applications and research

This study paved the way forward in the direction of practical applications and future research in terminological elaboration and standardisation.

Rather than covering most aspects of standardisation and elaboration of Zulu as a technical language, this study provides a methodological practical foundation with sufficient exemplification in order to encourage further research. The methodology and insights gained in this study can be made available to other technical fields where elaboration is needed, such as the legal and linguistic terminological fields in Zulu. This methodology can even be made applicable to another African language be it related such as Swazi or unrelated such as Tswana. However, this proposed practical methodology can change by adding new parameters of standardisation and elaboration. Terminologists can evaluate the methodology and make further contributions based on their own experiences in the terminological domain.

The proposed adjustments made in this study to the Zulu orthography (also in relation to terminology) can be tested by terminologists . They should then, once consensus has been reached about these adjustments, make concrete proposals towards officialisation by approaching the ZNLB so that existing rules can be reformulated. Later, during the final standardisation process, adjustments will finally have to be approved by PanSALB.

Although an attempt was made to document the methods of word-formation, the process may still need

further research input before standardisation can be achieved. An aspect that needs further research is the categorisation of word-formation methods in order to overcome the overlap between categories such as compounding, loan-translation and semantic shift.

The insights gained from the compilation and the encoding (by means of query tools) of a relatively small written medical corpus can be made applicable to much bigger growing corpora or corpora of another technological field. The more one could experiment with corpus query tool applications such as WST, the closer one would get to automatic term extraction and the easier lemmatisation would become. Already the research input of Bosch and Pretorius (2002) towards a computational morphological analysis of Zulu has made automatisation a more feasible reality.

The oral corpus, based on responses by the informants, provides a basis for comparison with written standard terms. This comparison forms the basis for a type of corpus annotation (analysis) which can prove valuable in terminological work. However, this corpus annotation is exemplary of what can be achieved. Although every corpus will initiate its own type of annotation, the oral corpus annotation arrived at for this study may direct further insights in corpus annotation. Corpus annotation is a must since it can serve as a reusable resource in terminological work, especially at improving the acceptability of terms. The input of oral corpus annotation in this study can encourage the utilisation of oral sources in the professional work domain in terminological work. This study may prompt more involvement of researchers in obtaining oral corpora, which indicates trends in the real language, instead of concentrating on written corpora which is relatively easier to obtain.

Yet, all the parameters of corpus linguistics and its application are far from exhausted. The application of corpus query tools, such as WST *keyword* application should be exploited further, especially in the African languages, also Zulu, where automatic term extraction and advanced text tagging have not been fully realised. Not only research in the field of term extraction, tagging and corpus annotation should be encouraged, but also African language specific training in the application of query tools.

7.10   Conclusion

The proposed practical approach to the elaboration and standardisation of Zulu as a technical language can in broad terms be captured into four suggested methods:

*   Improve the understanding of orthographical rules by means of accurate formulation with appropriate exemplification. Modify the orthography to ease its application to the latest trends in terminology development in Zulu. Make sure that (amended) orthographical rules for Zulu, finalised by the ZNLB, reach all the relevant language interest groups in the public and educational sectors.

*   Gain a comprehensive overview of the word-formation methods in Zulu, to the extent of standardising them, also bearing in mind cultural factors such as *taboo*. Acquired insight within a structured reference framework of word-formation patterns can only enhance the term creativity of the terminologist and language development in the long term.

*   Use relevant written sources in a specific technical field to compile corpora which can be used for the purpose of term extraction and eventual standardisation. Where possible, make use of computerised corpus query tools to facilitate the task of lemmatising (listing).

*   Use relevant oral sources in a specific technical field to compile corpora which can be used for the purpose of term verification in the standardisation process by, for instance, comparing existing standard terms to oral terms. In the comparison of terms, make use of corpus annotation to find reasons why some terms are more acceptable, suitable and popular than others.

However, these methods can only become feasible through proper training. As could be established

in terminological practice, the emphasis up to now has fallen on computer training without combining it with language specific linguistic training. What is really needed is training both in language specific linguistic analysis as well as in the effective and accurate application of computerised tools to a corpus.

Within the national perspective, standardisation was discussed with reference to the aspects of definition, models, norm, stages, purpose, agents, limitations and problems with specific linguistic exemplification in Zulu. National language standardisation was found to be quite a vague process which generally lacks coordination between the different language interest structures, a situation that should be rectified by national language authorities such as the NLBs and PanSALB. Standardisation, although recently negatively perceived by many scholars, and ongoing documentation are still necessary to develop a language to capacity.

For natural language elaboration and the proper standardisation of terminology, the starting point is term extraction from the written or oral living language, and not merely the random intuitive creation of terms by terminologists appointed by language authorities. Standardisation can only be sucessful if technical terms are evaluated in terms of acceptability to speakers since partially acceptable terms cannot be afforded by any language. Therefore, oral terms have to be established and popularised, by for instance the media, before they become standardised. Terminologists should use terms that already exist in society, thereby promoting natural term development (Jafta 1987). According to Thipa (1989) the standardisation of language deals with the functional efficiency of such a language, i.e. that everyone should be able to understand and use the language with the least misunderstanding.

Furthermore, these standard technical terms have to reflect current developments and change (in the phonology and lexicon) in the real language. The terminologist should take notice of popular phonological trends in the language; they flow on the tongue and catch on; they represent change and creativity in the language and may influence the standardisation process, even if it seems a blatant taking over of English terms, resisting purist coinage. Where needed, spelling should be adjusted to suit the phonology of language change.

The only fully-developed languages in South Africa are English and, to a lesser extent, Afrikaans. However, this lack of terminology in the African languages can be overcome only if action is taken towards effective computerised standardisation (including text encoding and corpus annotation) and if coordination in the term-creating activities among the African languages is promoted.

An idealistic language policy alone cannot change the developmental dilemma the African languages find themselves in. Language policy cannot be fully implemented by language interest structures or even the government of the day if the speakers of the African languages themselves do not do their part. Modernisation (per implication also standardisation) will only be successful if the African languages are maintained and their status thus improved by their speakers.

To sum up, for Zulu to elaborate to its capacity, a proper modern discourse dealing with technical and scientific topics needs to develop in the language. Modernisation occurs when a language becomes an appropriate medium of communication for modern discourse (Cooper 1989). Unfortunately this 'modern discourse' does not yet exist in any of the African languages, including Zulu. Such discourse can only be achieved through the continual development of the African languages in all sectors, including the advanced technological sector. Yet, if given the opportunity, the Zulu language has sufficient elaboration capacity to develop a technological and scientific discourse in order to function as a proper technical language.

BIBLIOGRAPHY

Abdulaziz, M.H.1989. Development of scientific and technical terminology with special reference to African languages. *Kiswahili* 56: 32-49.

Adam, H. M. & Geshekter, C. L. 1980. *The revolutionary development of the Somali language (Occasional Paper No. 20)*. Los Angeles (CA): UCLA African Studies Centre.

Alberts, M.1997. Legal terminology in African languages. In: Du Plessis, J.C.M.D. (ed.)   *Lexikos*. Stellenbosch: Buro van die WAT. Afrilex-Series 7: 179-191.

Alberts, M.1999(a). Terminology and definitions. Tutorial on Principles, Procedures and Practice of Terminology and Terminography presented by the African Association for Lexicography (AFRILEX) at the University of Pretoria, Economic and Management Sciences Building, 29-30 November 1999. Notes 1-10.

Alberts, M.1999(b). Theoretical principles of terminology and terminography. Tutorial on Principles, Procedures and Practice of Terminology and Terminography presented by the African Association for Lexicography (AFRILEX) at the University of Pretoria, Economic and Management Sciences Building, 29-30 November 1999. Notes p1-28.

Alberts, M. 2003. Collaboration between PanSALB and terminology structures. In: De Schryver, G. (ed.) *TAMA 2003 South Africa. Terminology in advanced management applications*. 6th International TAMA Conference Proceedings. Pretoria: TermNet SF2: 40-44.

Alexander, N. 1991. *Language policy and national unity in South Africa/Azania*. Cape Town: Buchu Books.

Alexander, N. 1995. Multilingualism for empowerment. In: Heugh, K. Siegrühn, A. & Plüddemann, P. (eds) *Multilingual education for South Africa. For the project for the study of alternative education in South Africa and the National Language Project*. Johannesburg: Heinemann: 37-41.

Andrzejewski, B. W. 1979. The development of Somali as a national medium of education and literature. *African languages* 5, 21: 1-9.

Ansre, G. 1971. Language standardisation in sub-Saharan Africa. In: Sebeok, T.A. (ed.) *Current trends in linguistics*. Vol. 7: Linguistics in sub-Saharan Africa. The Hague: Mouton: 680-698.

Baker, M. 1987. Review of methods used for coining new terms in Arabic. *Meta* 32, 2: 186 -188.

Bamgbose, A.1991. *Language and the nation - the language question in sub-Saharan Africa*. Edinburgh: Edinburgh University Press.

Batibo, H. 1992. Term development in Tanzania. In: Crawhall, N. T. (ed.)*Democratically speaking*: *international perspectives on language planning*. Salt River: National Language Project: 92-99.

Bible Society of Southern Africa. 1986.*Ibhayibheli elingcwele*. Cape Town: National Book Printers.

Bogdan, R. C. & Biklen, S. K. 1992. *Qualitative research for education*: *an introduction to theory and methods*. Boston:Allyn & Bacon.

Bokamba, E. G. 1981. Language and national development in Sub-Saharan Africa: A progress report. *Studies in the linguistic sciences*. 11, 1: 1-25.

Bokamba, E. G. 1991. French colonial policies and their legacies. In: Marshall, D. F. (ed.) *Language planning. Focusschrift in honor of Joshua A. Fishman on the occasion of his 65th birthday*. Vol. 3. Amsterdam: John Benjamins: 175-215.

Bosch, S. E. & Pretorius, L. 2002. The significance of computational morphological analysis for Zulu lexicography. *South African Journal of African languages* 22, 1: 11-20.

Breton, R. 1991. The handicaps of language planning in Africa. In: Marshall, D. F. (ed.) *Language planning*: *Focusschrift in honor of Joshua A. Fishman on the occasion of his 65th birthday*. Vol. 3. Amsterdam : John Benjamins: 153-174.

Bryant, A. T. 1905. *A Zulu-English Dictionary*. Pinetown: Marianhill Mission Press.

Buthelezi, J. C. 1993. *Kushaywa edonsayo*. Goodwood, Cape Town: Maskew Miller Longman.

Calteaux, K. V. 1996. (ed.) *Standard and non-standard African language varieties in the urban areas of South Africa - Main Report for the STANON research programme*. Pretoria: HSRC Publishers.

Calteaux, K. V. 1994. A sociolinguistic analysis of a multilingual community. Johannesburg: Unpublished D Litt et Phil thesis, Rand Afrikaans University.

Chimhundu, H. 1992. Zimbabwe - Standard Shona: myth and reality. In: Crawhall, N. T. (ed.) *Democratically speaking - international perspectives on language planning*. Salt River: National Language Project: 69-76.

Chiwome, E. 1992. Zimbabwe: Term creation: The case of Shona. In: Crawhall, N. T. (ed.) *Democratically speaking - international perspectives on language planning*. Salt River: National Language Project: 89-91.

Cluver, A. D. de V. 1989. *A manual of terminography*. Pretoria: Human Sciences Research Council.

Cluver, A. D. de V. 1993. Towards a democratic language policy for South Africa. In: Von Staden, P.M.S. (ed.) *Linguistica. Festschrift E. B. van Wyk*: *'n Huldeblyk*. Pretoria: J. L. Van Schaik: 26-44.

Cluver, A. D. de V. 1994. Preconditions for language unification. *South African Journal of linguistics*. Supplement, 20: 168-194.

Colenso, J. W. 1882. *First steps in Zulu*. Pietermaritzburg & Durban: Davies & Sons.

Colenso, J. W. 1905. *Zulu-English Dictionary*. Pietermaritzburg: Shuter & Shooter.

Cooper, R. L. 1989. *Language planning and social change*. Cambridge: Cambridge University Press.

Crawhall, N. T. (ed.) 1993. *Negotiations and language policy options in South Africa*: *The National Language Project Report to the National Educational Policy Investigation Sub-Committee on Articulating Language Policy*. Salt River: National Language Project.

Crystal, D. 1993. *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.

Davey, A. & Koopman, A. 2000. Adulphe Delegorgue's Vocabulaire de la Langue Zoulouse. *South African Journal of African languages* 20, 2: 134-147.

De Klerk, G. 1995(a). Slaves of English. In: Heugh, K. Siegrühn, A. & Plüddemann, P. (eds) *Multilingual education for South Africa. For the project for the study of alternative education in South Africa and the National Language Project*. Johannesburg: Heinemann: 8-14.

De Klerk, G. 1995(b). Three languages in one school: a multilingual exploration in a primary school. In: Heugh, K. Siegrühn, A. & Plüddemann, P. (eds) *Multilingual education for South Africa. For the project for the study of alternative education in South Africa and the National Language Project*. Johannesburg: Heinemann: 53-62.

Department of Arts and Culture. 13 November 2002. Final Draft: *National Language Policy Framework* (NLPF). Pretoria: Department of Arts and Culture.

Department of Arts, Culture, Science and Technology. 1996. *Towards a national language plan for South Africa. Final Report of the Language Plan Task Group (LANGTAG)*. Pretoria: The Government Printer.

Department of Arts, Culture, Science and Technology - National Terminology Services: 1997(a). *Draft List*: *Basic Health Terms.* Pretoria: The Government Printer.

Department of Arts, Culture, Science and Technology - National Terminology Services: 1997(b). *Draft List*: *Sex Education*. Pretoria: The Government Printer.
Department of Bantu Education.1962. *Zulu terminology and orthography No. 2*. Pretoria: The Government Printer.

Department of Bantu Education.1976. *Zulu terminology and orthography No. 3*. Pretoria: The Government Printer.

Department of Education. 1997(a). *Language-in-Education Implementation Plan (LIEIP).* Pretoria: The Government Printer.

Department of Education.1997(b). *Language-in-Education Policy (LIEP).* http://www.polity.org.za/govdocs/politcy.html// (accessed 26/04/2000).

Department of Education and Training. 1993.*IsiZulu terminology and orthography No. 4*. Pretoria: The Government Printer.

Department of Native Affairs. 1957. *Zulu-Xhosa terminology and spelling No. 1*. Pretoria: The Government Printer.

De Schryver, G. & Prinsloo, D. J. 2000(a). The compilation of electronic corpora with special reference to the African languages.*Southern African linguistics and applied language studies* 18, 89-106.

De Schryver, G. & Prinsloo, D. J. 2000(b). Electronic corpora as a basis for the compilation of African-language dictionaries. Part 1 : The macrostructure. *South African Journal of African languages* 20, 4: 291-309.

De Schryver, G. & Prinsloo, D. J. 2000(c). Electronic corpora as a basis for the compilation of African-language dictionaries. Part 2 : The microstructure. *South African Journal of African languages* 20, 4: 310-330.

Dhlomo, R. R. R. 1961. *UDingane kaSenzangakhona*. Pietermaritzburg: Shuter & Shooter.

Döhne, J. L. 1857. *Zulu-Kafir Dictionary*. Cape Town: G. J. Pike's Machine Printing Office.

Doke, C. M. 1945. *Text-Book of Zulu Grammar*. Johannesburg: Longmans, Green & Co.

Doke, C. M. 1984. *Textbook of Zulu Grammar*. Johannesburg: Maskew, Miller & Longman.

Doke, C. M. & Cole, D. T. 1961. *Contributions to the history of Bantu linguistics*. Johannesburg: Witwatersrand University Press.

Doke, C. M. & Vilakazi, B. W. 1949. *Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.

Doke, C. M. & Vilakazi, B. W. 1972. *Zulu-English Dictionary*. Johannesburg: Witwatersrand University Press.

Dore, W. & Silva, W. (eds) 1996. *A Dictionary of South African English on historical principles*. Oxford: Oxford University Press.

Drozd, L. & Roudny, M. 1980. Language planning and standardization of terminology in Czechoslovakia. *International Journal of the sociology of language* 23: 29-41.

Eastman, C. M. 1983. *Language planning*. San Fransisco: Chandler & Sharp.

Ekwelie, S. 1971. Swahili as a lingua franca: a study in language development. *Geneve-Afrique* 10, 2: 86-99.

Felber, H. 1982. Some basic issues of terminology. *The Incorporated Linguist* 21, 1 Winter: 12-24.

Fellmann, J. 1979. Language development in Tigrinya. *Language problems and language planning* 3: 25-27.

Ferguson, C. A. 1977. Sociolinguistic settings of language planning. In: Ruben, J. Jernudd, B. H. Das Gupta, J. Fishman, J. A. & Ferguson, C. A.(eds) *Language planning processes*. The Hague: Mouton: 10-29.

Fishman, J. A 1974. Language modernization and planning in comparison with other types of national modernization and planning. In: Fishman, J. A. (ed.) *Advances in language planning*. The Hague: Mouton: 79-102.

Fishman, J. A 1977. Comparative study of language planning: Introducing a survey. In: Ruben, J. Jernudd, B. H. Das Gupta, J. Fishman, J. A. & Ferguson, C. A. (eds) *Language planning processes*. The Hague: Mouton: 31-39.

Fourie, D.1993. Reflections on African languages and technical translation theory. *South African Journal of African languages* 13, 3: 81-85.

Galinski, C. 1982. Standardisation in terminology. An overview. In: Nedobity, R. (ed.) *Terminologies for the eighties. With a special section: 10 years of Infoterm*. München: K. G. Saur Verlag: 186-226.

Garside, R. Leech, G. & McHenry, T. (eds) 1997. *Corpus annotation - linguistic information from computer text corpora*. New York: Longman Inc.: 1-15.

Garvin, P. L. 1993. A conceptual framework for the study of language standardization. *International Journal for the sociology of language*, 37-54.

Gonzalez, A. 1979. Language and social development in the Pacific area. *Philippine Journal of linguistics* 10, 1/2: 21-44.

Gonzalez, A. 1993. An overview of language and development. *Journal of multilingual and multicutural development* 14, 1/2: 5-23.

Gregerson, E. A. 1977. *Language in Africa. An introductory survey*. New York: Gordon & Breach.

Grout, L. 1893. *A grammar of the Zulu language*. London: Kegan Paul, Trench, Trübner & Co.

Hanks, P. McLeod, W. T. & Urdang, L. (eds) 1986. *Collins dictionary of the English Language*. Glasgow: Collins.

Haugen, E. 1966. Language standardization. In: Coupland, N. & Jaworski, A. (eds) 1997. *Sociolinguistics. A Reader and Coursebook*. London: Macmillan Press Ltd.: 341-352.

Heugh, K. Siegrühn, A. & Plüddemann, P. (eds) 1995. *Multilingual education for South Africa. For the project for the study of alternative education in South Africa and the National Language Project*. Johannesburg: Heinemann.

Heugh, K. 1995. The multilingual school: modified dual medium. In: Heugh, K. Siegrühn, A. & Plüddemann, P. (eds) *Multilingual education for South Africa. For the project for the study of alternative education in South Africa and the National Language Project*. Johannesburg: Heinemann: 83-88.

Hlongwane, J. B. 1995. Growth of the Zulu Language and its structural changes. *South African Journal of African languages* 15, 2: 60-65.

Hudson, R. A. 1980. *Sociolinguistics*. Cambridge: Cambridge University Press.

Insight: A way with words. 1999. *Daily News.* 23 August 1999: 9.

Jafta, N. 1987. The development of terminology in Xhosa - a case study. *Logos* 7, 2: 127-138.

Kaschula, R. H. & Anthonissen, C. 1995. *Communicating across cultures in South Africa*: *Toward a critical language awareness*. Johannesburg: Holder & Stroughton Witwatersrand University Press.

Kennedy, G. 1998. *An introduction to corpus linguistics*. London: Longman.

Khumalo, N. H. E. 1995. *The language contact situation in Daveyton*. Johannesburg, Soweto: Unpublished M. A. dissertation, Vista University.

Koopman, A. 1992. Zulu and English adoptives: Morphological and phonological interference. *South African Journal of African languages* 12, Supplement 1: 105-115.

Koopman, A. 1994. Lexical adoptives in Zulu. Pietermaritzburg: Unpublished D Litt et Phil thesis, University of Natal.

Kubeka, I. S. 1979. A preliminary survey of Zulu dialects in Natal and Zululand. Durban: Unpublished M. A. dissertation, University of Natal.

Kumalo, M. B. 1987(a). Current and proposed new terminology in language description and Zulu literature. *Educamus* 33, 4: 21-24.

Kumalo, M. B. 1987(b). Revised current and proposed new terminology in language description and isiZulu literature. *Logos* 7, 2: 147-165.

Leech, G. 1987. General introduction. In: Garside, R. Leech, G. & Sampson, G.(eds) *The computational analysis of English*: *a corpus-based approach*. New York: Longman Inc.: 1-15.

Leech, G. 1997. Introducing corpus annotation. In: Garside, R. Leech, G. & McHenry, T.(eds) *Corpus annotation - linguistic information from computer text corpora*. New York: Longman Inc.: 1-18.

Leech, G. Myers, G. & Thomas, J. (eds) 1995. *Spoken English on computer-transcription, mark-up and application*. New York: Longman Publishing.

Leedy, P. D. 1993. *Practical Research*: *planning and Design*. New York: Macmillan Publishing Company.

Louw, J. A. 1983. The development of Xhosa and Zulu as languages. In: Fodor, I. & Hagege, C. (eds) *Language reform*: *History and future*. Hamburg: Helmut Buske Verlag: 371-92.

Louwrens, L. J. 1993. Semantic change in loan words. *South African Journal of African languages* 13, 1: 8-16.

Madiba, M. R. 2000. Strategies in the modernisation of Venda. Pretoria: Unpublished D Litt et Phil thesis, University of South Africa.

Malcolm, D. McK. 1966. *A new Zulu manual*. Cape Town: Longmans Southern Africa.

Malimabe, R. M. 1990. The influence of non-standard varieties on the standard Setswana of High School Pupils. Johannesburg: Unpublished M. A. dissertation, Rand Afrikaans University.

Marivate, C. N. 1992. The evaluation of language policy for Africans in South Africa 1948-1989. Harvard: Unpublished D Litt et Phil thesis, Harvard University.

Mathumba, D. I. 1993. A comparative study of selected phonetic, phonological and lexical aspects of some major dialects of Tsonga in the Republic of South Africa, and their impact on the standard language. Pretoria: Unpublished D Litt et Phil thesis, University of South Africa.

Matšela, Z. A. 1987. The problems of modernizing the development of Sesotho scientific technical terminologies. *Logos* 7, 2: 79-90.

Matšela, Z. A. & Mochaba, M. B. 1986. Development of new terminology in Sesotho. *South African Journal of African languages* 6, Supplement: 136-148.

Matthews, P. H. 1997. *Oxford Concise Dictionary of Linguistics*. Oxford: Oxford University Press.

Mkhulisi, N. 1996. Rural versus urban language usage. Paper read at the conference: The feasibility of technical language development in the African languages, by the Department of Arts, Culture, Science and Technology (The National Terminology Services in collaboration with the State Language Services). 8 March 1996 in Forum 150, HSRC Building, Pretoria.

Mochaba, M. B. 1987. Shift of meaning in Sesotho modern terminology. *Logos* 7, 2: 139-146.

Msimang, C. T. 1989. *Some phonological aspects of the Tekela Nguni dialects*. Pretoria: D Litt et Phil thesis, University of South Africa.

Msimang, C. T. 1992. The future status and function of Zulu in the new South Africa. *South African Journal of African languages* 12, 4: 139-143.

Mthembu, P. 1996. Accommodating changes in the phonology and vocabulary of the African languages. Paper read at the conference: The feasibility of technical language development in the African languages.

Presented by the Department of Arts, Culture, Science and Technology (The National Terminology Services in collaboration with the State Language Services). 8 March 1996 in Forum 150, HSRC, Pretoria.

Mtintsilana, P. N. & Morris, R. 1988. Terminography in African languages in South Africa. *South African Journal of African languages* 8,4: 109-113.

National Education Policy Investigation (NEPI). 1992. Language policies for medium of instruction: an information document for discussion by parents. In: Stevenson, I . A. & Papo, F. M.(eds) *Sociolinguistics for applied linguistics - Reader for HSOPER-U*. Pretoria: University of South Africa: 214-226.

National Place Names Committee.1951. *Official Place Names in the Union and South-West Africa*. Pretoria: Government Printer.

Neustuphy, J. V. 1974. Basic types of treatment of language problems. In: Fishman, J. A. (ed.) *Advances in language planning*. The Hague: Mouton: 37-48.

Nhlapo, J. M. 1945. *Nguni and Sotho, a practical plan for the unification of the South African Bantu languages*. Cape Town: The African Bookman.

Njogu, K. 1992. Kenya - Grassroots standardisation of Swahili. In: Crawhall, N. T. (ed.) *Democratically speaking - international perspectives on language planning*. Salt River: National Language Project: 69-76.

Nkabinde, A. C. 1968. Some aspects of foreign words in Zulu. *Communications of the University of South Africa*, c 59.

Nkabinde, A. C. 1982. *Isichazamazwi 1*. Pietermaritzburg: Shuter & Shooter.

Nkabinde, A. C. 1985. *Isichazamazwi 2*. Pietermaritzburg: Shuter & Shooter.

Nkondo, C. P. N. 1987. Problems of terminology in African languages with special reference to Xitsonga. *Logos* 7, 2 69-78.

Ntuli, D. B. Z. 1994. *Isibhakabhaka.* Pretoria: Aktua Press.

Nyembezi, S. 1992. *Isichazimazwi sanamuhla nangomuso*. Pietermaritzburg: Reach out Publishers.

Ohly, R. 1977. Swahili to be a world-language. *Swahili Studies, a supplement to Kiswahili* 47, 1:

119-128.

Ohly, R. 1981. The silent absence of term. *Babel* 27, 2: 100-103.

Ohly, R. 1987. Corpus planning, glottoeconomics and terminography. *Logos* 7, 2: 55-67.

Oostdijk, N. 1991. *Corpus linguistics and the automatic analysis of English*. Amsterdam: Rodopi.

Poulos, G. & Msimang, C. T. 1998. *A linguistic analysis of Zulu*. Pretoria: Via Afrika.

Prah, K. K. 1998. *Between distinction and extinction*. *The harmonisation and standardisation of African languages*. Johannesburg: Witwatersrand University Press.

Richards, J. C. Platt, J. & Platt, H. 1992. *Longman dictionary of language teaching and applied linguistics*. London: Longman.

Roberts, C. 1899. *The Zulu-Kafir Language*. London: Kegan Paul, Trench, Trübner & Co.

Roberts, C. 1915. *An English-Zulu Dictionary*. London: Kegan Paul, Trench, Trübner & Co.

Rubin, J. 1977. Language standardization in Indonesia. In: Ruben, J. Jernudd, B. H. Das Gupta, J. Fishman, J. A. & Ferguson, C. A. (eds) *Language planning processes*. The Hague: Mouton: 157-179.

Sager, J. C.1990. *A practical course in terminology processing*. Amsterdam: John Benjamins Publishing Company.

Samuelson, R. C. A. 1925. *Zulu Grammar*. Durban: Knox.

Scott, M. 1999. *English language teaching*. *ELT Catalogue Multimedia*. *WordSmith Tools*. Oxford: Oxford University Press. Distributed by the World Wide Web. http://www1.oup.co.uk/elt/catalogu/multimed/4589846/4589846.html// (accessed on 06/8/2002).

Seymour-Smith, C. 1986. *Macmillan Dictionary of Anthropology*. London: The Macmillan Press Ltd.

Shabangu, S. S. 1987. *Isichazamazwi samagama amqondo ofanayo*. Pietermaritzburg: Shuter & Shooter.

South African Government of National Unity (SAGNU Index). 1996. *Act 106 of 1996*. http://www.polity.org.za// (accessed on 05 /11/ 2001).

Taljaard, P. C. & Bosch, S. E. 1988. *Handbook of IsiZulu*. Pretoria: Van Schaik.

Taljard, E. & De Schryver, G. 2002. Semi-automatic term extraction for the African Languages, with special reference to Northern Sotho. *Lexikos* 12: 44-74.

Temu, C. W. 1984. Kiswahili terminology: Principles adopted for the enrichment of the Kiswahili language. *Kiswahili* 51, 1/2: 112-127.

TermNet. [CD-ROM]. Also available: http// www.termnet.at//.

The IsiZulu Language Board. 1990. The orthographical rules for IsiZulu as approved by the isiZulu Language Board on 19 September 1989. *Educamus* 36, 3: 24-26.

Thipa, H. M. 1989. The difference between rural and urban Xhosa varieties: A sociolinguistic study. Pietermaritzburg: Unpublished D Litt et Phil thesis, University of Natal.

Thipa, H. M. 1992. The difference between rural and urban Xhosa varieties. *South African Journal of African languages* 12, Supplement 1: 77 - 90.

Toffelson, J. W. 1991. *Planning language, planning inequality.* London: Longman.

Tumbo, Z. 1982. Towards a systematic terminology development in Kiswahili. *Kiswahili* 49, 1: 87-93.

Ungerer, H. J. 1983. Komposita in Zulu. Johannesburg: Unpublished D Litt et Ph Roberts, C. 1899. *The Zulu-Kafir Language*. London: Kegan Paul, Trench, Trübner & Co. il thesis, Rand Afrikaans University.

Van Eeden, B. I. C. 1956. *Zoeloe-Grammatika.* Stellenbosch: Universiteitsuitgewers en Boekhandelaars.

Van Huyssteen, L. 1999. Problems regarding term creation in the South African African languages, with special reference to Zulu. *South African Journal of African languages* 19, 3: 179-187.

Van Huyssteen, L. 2001. The value of oral corpora in the development of standardised medical corpora in Zulu. Paper read at the 6th International Conference for the African Association for Lexicography (AFRILEX) held at Oasis Conference Centre Pietersburg: University of the North on 4-6 July 2001.

Van Huyssteen, L. 2002. The role of culture in Zulu language development. In: LeBeau, D. & Gordon, J. (eds) *Challenges for Anthropology in the 'African Renaissance' - A Southern African Contribution*. Windhoek: University of Namibia Press Publication Number 1: 217-224.

Van Wyk, E. B. 1958. Woordverdeling in Noord Sotho en Zoeloe: 'n bydrae tot die vraagstuk van woordidentifikasie in die Bantoetale. Pretoria: Unpublished D Litt et Phil thesis, University of Pretoria.

Van Wyk, E. B. 1992. The concept standard language. *South African Journal of African languages* 12, Supplement: 23-34.

Wanger, W. 1917. *Konversationsgrammatik der Zulu-Sprache*. Marianhill: St. Thomas Aquins Druckerei.

Wilkes, A. 1985. Words and word division: A study of some orthographical problems in the writing systems of the Nguni and Sotho languages. *South African Journal of African languages* 5, 4: 148-153.

Ziervogel, D. Louw, J. A. & Taljaard, P .C. 1981. *Handbook of the Zulu Language*. Pretoria: Van Schaik.

Zungu, P. J. 1995. Language variation in Zulu: A case study of contemporary codes and registers in the greater Durban area. Durban: Unpublished D Litt et Phil thesis, University of Durban- Westville.

MEDICAL PAMPHLETS USED AS TEXT SOURCES

Abesifazane, abantwana neHIV. Pietermaritzburg AIDS Action Group:Department of National Health and Population Development.

Amaphuzu abalulekile nge-HIV /AIDS. Pretoria: Department of Health, AIDS Helpline.

Cholera isifo sohudo - yazi amaqiniso ukuze uphile. Gauteng Provincial Government: Department of Health.

Hlolelwa i-TB mahhala. Gauteng Provincial Government: Department of Health.

ICholera isifo esingabulala abantu. Pretoria:Department of National Health and Population Development.

Ingculazi emphakathini (p24-31). Johannesburg: National HIV/ AIDS Programme Department of Health, sponsored by Old Mutual, BP, AIDS Helpline, European Union, the Open Society Foundation for South Africa and UNAIDS.

Isifo sofuba i-TB kanye ne-HIV / AIDS. Pretoria: Department of Health, AIDS Helpline.

Isifo sofuba siyalapheka. Pretoria: City Council of Pretoria - TB Services.

Isifo sohudo ikholera. Ukuhlanzeka kuyasiza ekuvikeleni ikholera. Durban Local Councils: Health /Ezempilo KwaZulu Natal.

Konke okufanele ukwazi ngesifo sofuba noma i-TB. Pretoria: Department of Health.

Ukukhipha isisu - zikhethele. Reproductive Health Materials Package by PHRU, SPH and AMREP.

Ukuqala umntwana wakho nokudla. Gerber Purity.

Ukuvalwa kwenzalo yowesilisa - zikhethele. Reproductive Health Materials Package by PHRU, SPH and AMREP.

Uthando...ukuvikela umndeni wakho kungculaza. Yazi amaqiniso. Pretoria: Department of National Health and Population Development.

ADDENDUM 1

RESEARCH METHODOLOGY FOR ORAL CORPUS ANNOTATION IN ORDER TO IMPROVE THE ACCEPTABILITY OF ZULU TECHNICAL TERMS

1 Introduction to and aims of research

Besides the viewpoints of scholars on term elaboration, the acceptability of technical (medical) terms to speakers will also have an impact on the Zulu language and therefore also has to be taken into consideration. Such acceptability was tested by means of Zulu medical-term questionnaires that were distributed to medical staff in several (provincial) hospitals, mainly in KwaZulu-Natal and one in Mpumalanga.

A survey on the acceptability of Zulu health terms was conducted and only health workers who were also Zulu mother-tongue speakers took part in it in order to obtain a natural intuitive response.

2 The questionnaire used for research purposes

The questionnaire consisted of an introduction, stating the conditions of confidentiality pertaining to persons taking part in the research followed by a list of English health terms and their existing Zulu equivalents. This list of health terms contains the standard written Corpus A. Corpus B is not contained here since it represents the feedback of the questionnaires, i.e. the oral responses from all the informants for each term (concept). Following the list of health terms, questions were asked in order to establish the region of origin and age of the informants as well as the section of the hospital

where they worked.

The given list of basic Zulu medical terms is an excerpt from both the 1997 official draft lists, *Basic Health Terms* and *Sex Education*. These two lists were both prepared by the *National Terminology Services* (NTS) of the *Department of Arts, Culture, Science and Technology* (DACST). Now the NTS is known as the *National Language Service*, abbreviated (NLS) of the *Department of Arts and Culture* (DAC). Also included are some general health terms appearing in *IsiZulu Terminology and Orthography No 3 (1976)* and *No 4 (1993)* prepared by the *Zulu Language Committee/Board* (ZLC/B). From these officially standardised term lists the most general terms were extracted in order to compile a fairly representative term list - corpus - of 350 standardised medical terms. However, it was, after careful scrutiny of terms, decided to limit the written standard terms, called Corpus A, to a core corpus of 145 terms (225 including equivalents) in order to deal with corpus annotation in the scope of a single chapter. This obviously also limits the comparative oral corpus based on informants' response, called Corpus B, to a core corpus of 145 (823 including equivalents).

The Zulu phrasing of the questionnaire was such that *informants* (a term henceforth used to refer to persons who took part in the survey) had to either approve (using a ✓) or disapprove of the given Zulu health terms. In cases where informants disapproved of terms, they had to supply the more acceptable or popularly used terms which they use in their work environment.

Out of the 120 questionnaires, based on interviews with health workers, which were returned, only 100 were found to be suitable for research purposes. Questionnaires which were just completed for the sake of completion (only ticking ✓ the given terms instead of supplying own equivalents) and incomplete questionnaires were not considered part of the survey and discarded.

The structural design of both the written and oral corpora for comparison and eventual annotation purposes is given below. Medical terms contained in these corpora appear in isolation and not in discourse. The written standard medical corpus is called **CORPUS A** and the oral medical corpus, based on the responses from informants interviewed, is called **CORPUS B**. Corpus A consists of a single corpus while Corpus B comprises six sub-corpora, each subdivided according to the region where the hospital is situated:

CORPUS A

Main written standard medical corpus

An original corpus containing a total of 350 standard terms from terminology lists as sources which was reduced to a total of 225 terms (145 core terms including equivalents).

CORPUS B

Main oral medical corpus (containing different responses of informants as sources)

Total of 823 terms (145 core terms including equivalents from informants from different hospitals).

Total of 100 informants out of 100 questionnaires used.

sub-corpora	number of terms rendered	number of informants /questionnaires used
Amajuba Memorial Hospital, Volksrust	426	17
King George V Jubilee Hospital, Durban	419	17
McCord Zulu Hospital, Durban	330	10
Newcastle Provincial Hospital, Newcastle	476	24

Prince Mshiyeni Hospital, Umlazi	266	18
Vryheid Hospital, Vryheid	428	14

3 The format of the questionnaire

The questionnaire, containing conditions, instructions and a list of terms was available in Zulu only since

the informants were all Zulu speakers. However, the English translation follows for the sake of clarity.

Uhla lwemibuzo ngamagama aphathelene nezempilo

Qaphela ukuthi uhlanganyela NGOKUTHANDA KWAKHO futhi NANGESIKHATHI SAKHO kulolu cwaningo. Igama lomuntu alizufakwa kulolu hla lwemibuzo ukuqinisekisa ukuthi imininigwane etholakele kulolu cwaningo igcineka iyimfihlo. Usizo lwakho njengomuntu okhuluma isiZulu kanye nosebenza ngezempilo luzothokozelwa kakhulu ngoba imininingwane ozosinikeza yona izokwelekelela ekuthuthukiseni amagama esiZulu aphathelene nezempilo. Uma uthanda ukuhlanganyela kulolu cwaningo, **yenza okushiwoyo** usebenzisa uhla lwamagama olulandelayo:

Yenza okushiwoyo ngohla lwamagama aphathelene nezempilo anikeziwe

Uzonikezwa uhla oluphathelene namagama ezempilo. Kukhona amagama esiNgisi aphathelene nezempilo kanye nahambelana nawo esiZulwini. Sebenzisa lolu hla lwamagama ukwenza okushiwo ku-a) ukuya ku-c) ngezansi:

a)   Uma ucabanga ukuthi igama lesiZulu elinikeziwe ohleni, yilona elihambelana nelesiNgisi, faka uphawu ngaphansi kwalo, isibonelo:

<div align="center">

cancer   *ikhensa*

✓
</div>

b)   Uma ucabanga ukuthi lelo gama elinikeziwe akulona, ungalufaki uphawu, kodwa nikeza ngaphansi kwalelo gama, igama ocabanga ukuthi yilo elingcono noma elisetshenziswa ngokujwayelekile esiZulwini, isibonelo:

<div align="center">

cancer   *ikhensa*

isimila
</div>

c)   Lapho kunamagama amaningana esiZulu (amabili noma ngaphezulu), nakhona yenza njengoba kuchazwe ku-a) no-b) ngenhla. Khumbula ukuthi ungawakhetha womabili noma womathathu amabili ezinhlotsheni ezintathu njengamagama esiZulu afanele ukusetshenziswa, isibonelo:

cancer *ikhensa / isimila / umdlavuza*

✓ ✓ ✓

Ungakhohlwa ukunikeza igama elingcono lesiZulu ngaphansi kwalelo elingemukeleki kahle, isibonelo: cancer *ikhensa / isimila*

i-ayidsi

UHLA LWAMAGAMA APHATHELENE NEZEMPILO
A LIST OF MEDICAL TERMS

ENGLISH   **ISIZULU**

A

English	Isizulu
abortion *	ukukhipha isisu
abscess	ithumba
acquired immune deficiency syndrome (AIDS) *	ingculazi
addicted *	-huhekile
adenoids	amankanka / i-adenoyidi
aids-related complex (ARC) *	izimpawu ezihlobene nengculazi
alcoholic *	isidakwa
alcoholism *	ubudakwa
allergy *	i-aleji
anaemia *	ukuphaphatheka kwegazi / i-anemiya
anaesthetic / anaesthesia (n) *	isidakamizwa / ukudaka imizwa
anorexia nervosa *	anoresiya nevosiya
antacid	isihlambululi sisu
anthrax	undicosho
antibiotic *	isibulala mabhakithiriya
antibody *	amasotsha omzimba / isivikela-mzimba
antidote	isibiba / isibulala buthi
antidepressant *	isiqeda kucobeka
antiseptic (n)	isinqandakuvunda / isibulala magciwane
appendix *	ithunjana / i-aphendisi
appetite	inhliziyo yokudla
arthritis *	isifo samathambo
artificial insemination	ukukhulelisa ngokujova / ukuhlanganiswa kwembewu ngaphandle kokuya

	ocansini / ukutshalwa kwembewu yowesilisa kowesifazane ngaphandle kokuya ocansini
artificial respiration	ukuphefumulisa / ukuphefumuliswa
asthma *	umbefu / isifuba somoya / i-asimu

*   indicates a term (including equivalents) in the core corpus that was selected for corpus annotation purposes

B

baby formula	ifomula / indlela yokuthaka
bacteria *	ibhakthiriya / ibhakhithiriya
balanced diet	ukudla okukhethwe ngendlela
	kwapheleliswa ngohlelo / ukudla okunomsoco
bile (gal)	inyongo
Bilharzia (red water)	isichenene segazi / isichenene (kubantu) / umbendeni
birth control *	ukuhlela umndeni
bisexual *	ukuba uncumbili
blister	intshabusuku
blood alcohol level	izinga lotshwala egazini
blood bank	indawo yokubeka igazi
blood cholesterol level	izinga lezinto ezingamafutha egazini
blood-clot	ihlule / ihlulu legazi
blood donor *	onikela ngegazi
blood-poisoning	ukonakala kwegazi ngesihlungu
blood pressure *	umfutho wegazi / ukucindezeleka kwegazi
blood smear	ukugcoba ngegazi
blood test *	ukuhlolwa kwegazi
blood transfusion *	ukuthasiselwa igazi
body temperature *	izinga lokushisa kwegazi
boil (tumour)	ithumba
bottle-feed	-ncelisa ibhodlela
botulism	ibotulism / ibhothulisimu
bowel movement	ukuya ngaphandle
breast milk	ubisi lwebele
breast cancer *	umdlavuza wesifuba
breast-feeding	ukuncelisa

bronchitis *	ibhroyinikhanthisi / umkhuhlane wamaphaphu
building food	izakhamzimba
bulimia *	ibhulima
bunion	isiqagalane / isitshophi
burn	isilonda sokusha
burns and scalds	isilonda somlilo / samanzi ashisayo

C

caesarian section *	ukubeletha ngokuhlinza
caffeine	ikhafeni
calorie	ikhalori
calcium	ikhalsiyumu / ikhalisiyamu
cancer *	ikhensa / isimila / umhlaza / umdlavuza
carbohydrate	ikhabonhayidrethi / ikhabhohayidrade
carbon dioxide	isikhutha /ikhaboni-dayoksayidi ikhabhonidayokisayidi
casualty	umuntu olimele noma oshonile engozini
cataract *	ikhatharakhithi
chemotherapy *	ikhemotirapi
chickenpox	ixhikhinipokisi / inqubulunjwana
chilblain	amabhamuza amakhaza
child abuse *	ukunukubezwa kwengane
child neglect	ukunganakekelwa kwengane
choke	-miwa / -binda / -bindwa
cholera *	ikholera / uhudo
cholesterol *	ikholestaroli
circulation (blood) *	ukuzungeleza
circumcision *	ukusoka
clinic *	iklinikhi / umtholampilo
colic	ikholikhi
coma *	ukuquleka
concussion	ukuxukuzeka kobuchopho /ukuguquka komhwamuko / ukulahlekelwa umqondo
condition (state)	isimo / isimiso
condom *	ikhondomu / ijazi lomkhwenyana / ijazi

conjunctivitis *	ikhonjangithivithisi / isifo samehlo / ikhonjakhithivithisi
constipation	ukusongeleka / ukuqumbelane
contagious disease *	isifo esisuselwanayo / isifo esingathathelana ngokuthintana kwabantu
contamination	ukungcola / ukonakala
contraceptive *	isivikela kukhulelwa
contraceptive (by injection)	*isitofu / ukutofa*
convulsions *	ukuqhashaqhasha / amafithi
coronary thrombosis	ihlulu legazi elisenhliziyweni
cut (wound)(n)	inxeba
croup	isifo sezingane lapho ingane iphathwa umphimbo ihluleke ukuphefumula ikhwehlele futhi
cure (n)	ikhambi

D

dairy produce	okuvela obisini
dehydration *	ukuphela kwamanzi emzimbeni
Department of Health	Umnyango wezeMpilo
depression *	ukukhathazeka
diabetes *	isifo sikashukela
diabetic (n)	umuntu onesifo sikashukela
diagnose	ukuthola nokukhomba ukuthi iyini inkinga esigulini
diarrhoea *	isihudo / uhudo
diet	ukudla okukhethwe ngohlelo / idayethi
digestion	ukugayeka / ukugayekokudla
diphtheria *	uxhilo / idifuteria
disinfect *	-bulala imbewu yokufa
disinfectant *	isihlanzisisi
disinfection	ukubulala amagciwane
dislocate *	-bhinyilika / -bhinyika
dislocation (of joint)	isenyelo / ukubhinyilika / ukubhonculuka kwethambo / ukugudluka kwethambo endaweni yawo
Down syndrome	idawuni sinidromu
dressing (bandages)	okokubopha inxeba imidweshu / ukumbozwa kwesilonda

drug (habit-forming) *	isidakamizwa
drug (medication)	okokwelapha / umuthi wokwelapha / ikhubalo
drug addict *	umuntu ophila ngezidakamizwa
drug addiction *	ukuhuheka
due-date	usuku olunqunyiwe

E

eczema	umuna
embryo *	isibindi sembewu / ihlule / i-embriyo / umbungu
emergency (n)	okuphuthumayo / isigubhukane
energy food	isinikamandla
enteric fever	intheriki
enzyme	i-enzayimi
epilepsy	isithuthwane
epileptic (n)	onesithuthwane
epileptic fit *	ukuphathwa isithuthwane
eye inflammation	ihlo elibomvu / ukuthinteka kwehlo

F

faint	-quleka
family planning *	ukuhlela umndeni / ukuhlelwa komndeni
family planning clinic	umtholampilo wokuhlela umndeni
fatal injury	ulimale wafa
fat content	ubungako bamafutha
feeding	ukondla / ukudlisa
fertility *	imvundo / ukuba sesimeni sokukwazi ukwenza ingane
fetus / foetus *	umbungu / umutwana osakhula esisweni sikamama
fever	imfiva
fibre	inxoza / uzi / okumahadlahadla / ukudla okungagayekile kahle
first-aid	usizo lokuqala
fluoride	ifulorayidi

fontanelle *	ifontaneli
food poisoning	ukubakhona kobuthi ekudleni / ubuthi obusekudleni
fracture (v)	-phoqoka / -feceza
frost-bite	umshazo / umonakalo ezinyaweni odalwe amakhaza
full-cream milk	ubisi olunokhirimu
fungus	isikhunta

G

gall	inyongo
gallstone *	iqondo
gangrene *	iganjirini
gargle	-hahaza
gastric juice	ujengezi lwasesiswini / igastrikhi jusi
gastroenteritis *	ukuphazamiseka kwesisu okuhlanzisayo
gauze	igozi
germ *	imbewu yokufa / ijemu / igciwane
German measles	isimungumungwane (sesiJalimane)
glucose	iglukhozi / iglukhosi
goitre *	isibobo / igoyitri
gonorrhoea / DROP *	idropha / idrophu / igonoriya
gout *	igawuthi

H

haemorrhage	umopho / ukopha kakhulu
haemorrhoids (piles) *	amathunjana emdidini
hare-lip	inhlewuka
hay fever	ukuthimula okubangwa izinto ozihogelile
health education	imfundo yezempilo
health services	imisebenzi yezempilo
hearing aid *	izinsizakuzwa

heart attack	isifo senhliziyo
heartburn	isilungulela / isilungulelo
hepatitis *	ihephathithisi
herpes *	ihephisi
HIV antibody test *	ukuhlolo-ngculazi
HIV disease	isifo sesandulela ngculazi
HIV-infected person *	umuntu onesandulela ngculazi
HIV negative *	ukungabi naso isandulela ngculazi
HIV positive *	ukuba nesandulela ngculazi
homosexual *	isitabani
hookworm	isikelemu
hormone	ihomoni
human immunodeficiency virus (HIV) *	i-HIV
hygiene	inhlanzeko
hygienic	ukuphatheka nenhlanzeko
hyperventilation *	ihayihayi
hysterectomy	ukususwa kwesibeletho

I

immune system *	isivikela zifo / izivikeli-mzimba
immunisation	ukugonywa
immunise *	-goma / -holoya
immunity	ukugomeka / ukuholoywa
incubator *	isichamiseleli / isifukameli
incurable disease	isifo esingelapheki
infect *	-thelela
infection *	ukuthathelwana / i-infeshini / i-infekishini / umhabulo
infertile *	ukungabi nanzalo
inflammation	ukudumba / ukuvuvuka / ukuqubuka
influenza / flu	imfluyenza / imfuluyeza / i-influenza / umkhuhlane
inoculation *	umjovo
insulin *	i-insulin
intestinal worms	izikelemu / izilo

intrauterine device/contraceptive (IUD) *	iluphu
intravenous	intravenasi / -ngemithambo
itch (n) *	utwayi

J

jaundice *	ijondisi

K

kidney failure	ukungasebenzi kahle kwenso
kidney stone *	iqhutshana elisezinsweni elibanga ubuhlungu
kwashiorkor *	ikhwashi

L

lactose	ilakhithozi
laryngitis	ilarijithisi
laxative	umuthi ohambisayo / uhlambululayo /okuhambisayo
leprosy	uchoko / ubulephero
lesbian *	isitabani
listlessness	ukucobeka
local anaesthesia	ukubulala imizwa elungeni elithile lomzimba
loss of appetite	ukungakuthandi ukudla
loss of co-ordination	ukungakwazi ukuma uqonde
low blood-pressure *	umfutho wegazi omncane
lump	isigaxa

M

magnesium	imagineziyamu
malaria	uqhuqho / umalaleveva
malnutrition	ukungondleki / ukungondleki komzimba
measles	isimungumungwane / isamungu

meningitis *	imenigithisi
menopause	ukuphela kwenzalo
mental	-kwekhanda / -kwengqondo
mental disease *	isifo senqondo
mentally deranged	ukuphambana mgqondo
mentally handicapped patient / mental case	iklabishi / i-M.C.
mercy killing	ukubulala ngoba wenza isihe
metabolism	imethabholisimu
migraine *	imagreyini
mineral	iminirali
miners' phthisis	ithayisisi
miscarriage *	ukucithekelwa isisu
mumps	uzagiga
muscle-cramp *	amajaqamba
muscular development	ukukhula kwemisipha / ukuqina kwezinyama
muscular dystrophy *	imasikhula dayistrofi

N

nappy rash	ukuqubuka okubangelwa inabukeni elimanziswe umchamo
narcotic (n)	isilalisa-mizwa / isidakamizwa
natural childbirth	ukubeletha ngokujwayelekile
nephritis	ukuvuvuka kwenso
nervous disturbance	iziphazamiso zemizwa
nervous system *	uhlelo lwemizwa
nicotine	inikotini / inikhothini
nose-bleeding	umongozima
nutrient	umsoco
nutrition	ukudla ukudla okunomsoco / ukondleka komzimba / isondla somzimba

O

obesity	okunona
oestrogen *	ukukhipha (izinyo) / i-estrogen

ointment	isigcobo
operation (surgical)	ukuhlinzwa / umthungu
osteoporosis *	isifo samathambo
ovulate *	ukukhipha iqanda
ovulation	ukuphuma kweqanda esizalweni
oxygen	i-oksijini

P

pain-killer*	okuqeda izinhlungu / isibulala zinhlungu
Pap smear	ucwephana lomlomu wesibeletho
paralysis *	uvendle
paraplegia *	ukufa esingezansi
parasite	ipharasayithi / isiphilangokunye / okunxibayo
pasteurised milk	ubisi olungenawo amabhakithiriya
penicillin *	iphenisilini
perspiration *	ukujuluka / umjuluko
physically disabled	ukukhubazeka emzimbeni
pill	iphilisi
plaster cast *	ukhonkolo
pleurisy	iplurasi / iplorisi / amangwe
pneumonia *	inyumoniya / umkhuhlane wamaphaphu
polio	uvendle
pregnancy test *	ukuhlolwa kokukhulelwa
pregnant *	ukubanzima / ukukhulelwa
prescription	isiyalo sokudliwa kwemithi
primary health care	unakekelo lwempilo
processed food	ukudla okuhlelwe ngandlela thile ukuze kugcineke
progesterone	iprogesterone
protein *	amaprotheni / amaphrotheni
public health service	umsebenzi wempilo kawonkewonke
pulse beat / pulse rate	ukushaya kwenhliziyo / ukushaya kwegazi
pyorrhea	isifo sezinsini

Q

quadriplegia *	ukukhubazeka isingenhla nesingezansi
quadriplegic *	isigoga
quarantine	ukuvalelwa / ukuvalelwa ngenxa yesifo

R

rabies	amarabi
radiation *	ukushisa ngemisebe
rape (n) *	ukudlwengula
rape (v) *	-dlwengula
rash	umqubuko / ukuqubuka
refined food	ukudla okucosakele
reflex action	okuzenzekalelayo
rehabilitation	ukuhlenga / ukubuyisela
remedy	iselapho
rheumatism *	irumathizimu / ikhunkulo / ubuhlungu ezinyameni nasemajoyintini
ringworm	umbandamu
roughage	umhadlahadliso
roundworm	izikelemu

S

scab (skin disease)	ukhwekhwe
scabies *	utwayi
scansion	isilinganisosigqi
scar	isibazi
scarlet fever	imfiva yengqubukane ebomvu / imfiva eyenza isikhumba sibe bomvu / isinkalethi fiva
schizophrenia *	iskizo
sclerosis	ukuqina kwesitho somzimba esithambile
sedative	isedathivu
septic sore	isilonda esibhibhayo
sex education *	imfundo kwezocansi
sexually transmitted disease (STD) *	isifo socansi (esithathelanayo)

side effect	imiphumela embi
sinusitis *	isayinasithasi
skin cancer *	umhlaza wesikhumba
smallpox	ingxibongo
smoking habit	umkhuba wokubhema
smoking risk	ingozi yokubhema
sonar scan *	isona
sprain *	isipreyini
sterilise (render germ-free) *	-bulala imbewu yokufa / ukubulala amagciwane
sterilise (render unproductive) *	-thena
stimulant	isikhuthaza / isihlumelelisi
stroke *	isitrokhi
sunstroke	ukuwiswa isango selanga
syphilis *	ugcunsula
symptom	izimpawu zesifo esithile

T

thermometer	ithemomitha
tonic	umuthi wokuqinisa igazi
toxin	ithoksini / isihlungu / ubuthi obukhiqizwa mabhakhithiria
tranquilliser	itrankwilaza
treat*	-phatha / -lapha
tuberculosis *	isifuba sexhwala / iT.B. / isifo sofuba
tumour	yinoma ikuphi ukuvuvuka okungajwajekile

U

ulcer *	isilonda / uzokozela / iyul'sa / i-alsa
ultra violet rays	umbala i-ultra violet
ultrasound clinic *	emafutheni
unconscious state *	isihlwathi / umukiswe ingqondo / isimo sokuquleka
unhealthy	-ngaphilile / -xhwalile / -ngafanele
unhygienic	-ngenanhlanzeko

V

vaccinate *   -gcaba / -goma
vaccination *   umgcabo
vaccine   umuthi wokugcaba / umgcabo / umgomo
vasectomy *   ivasekhithomi
virus *   ivirusi
vitamin *   ivithamini / uvithamini
vulgar fraction   iqhezungqo / ingxenyengqo

W

wart   insumpa / ukuvuvuka
weight   isisindo
weight gain   ukwanda kwesisindo
weight loss   ukwehla kwesisindo
whiplash injury   ukulimala entanyeni okubangwa ingozi yemoto / ukulimala
                           kwemisipha / isingezansi intamo okubangelwa ingozi yemoto
whooping-cough *   umkhuhlane / amankonkonko / ugonqogonqo /
                                   ugonklo / umpenge

X

X-ray *   iX-reyi
X-rays (venue)   imisebe yeX-reyi / i-x-reyi
Y

yellow fever *   imfiva encombo / iyelo fiva

Z

zidovudine (AZT) *   i-AZT

Manje phendula imibuzo elandelayo

a) Uhlala kuphi?

....

b) Usebenza kuyiphi ingxenye yesibhedlela noma yomtholampilo?

....

c) Shono ubudala bakho ngeminyaka

....

Sibonga kakhulu ngokunikela kwakho kulolu cwaningo. Uma ucabanga ukuthi kukhona okuthile esikushiyile, sicela ukuba usitshele noma usazise ngakho ngencwadi noma ngocingo.

Thintana noNkk. L van Huyssteen
University of South Africa
Department of African Languages
PO Box 392
Pretoria
0003
Tel: 012 4298258   Fax: 012 4293221
e-mail: vhuysl@.unisa.ac.za

The English translation of the questionnaire follows. The terminology list is not repeated, since it is the same for both the Zulu and English questionnaires.

Questionnaire on Zulu health terms

Please note that you take part in this survey on a VOLUNTARY basis in your OWN TIME. No name of a person is attached to the questionnaire to ensure the strictest confidentiality. Your help as a Zulu speaker and health worker is most appreciated as the information you supply can aid in developing more Zulu health terms. If you want to take part in this survey you should **follow the procedures below**:

Use the given list of health terms and follow the procedures

You will be given a list containing English health terms with their corresponding equivalent Zulu health terms. Keep the term list at hand and follow procedures a) to c):

a)   If you think that the given Zulu term in the list is the correct equivalent, make a tick underneath it, e.g.

<div align="center">

cancer   *ikhensa*
✓
</div>

b)   If you think that the given term is incorrect, do not make a tick, but supply the term you think is a 'better', more common, Zulu equivalent underneath it, e.g.

<div align="center">

cancer   *ikhensa*
isimila
</div>

c)   In cases where there is a variety of Zulu terms (two or more), apply the same procedures as explained in a) and b) above. Remember that you may choose both, all three or two of the three varieties as suitable Zulu equivalents, e.g.

<div align="center">

cancer   *ikhensa / isimila / umdlavuza*
✓   ✓   ✓
</div>

Do not forget to supply the 'better' Zulu term underneath the 'wrong' ones, e.g.

<div align="center">

cancer   *ikhensa / isimila*
i-ayidsi
</div>

Previous term list is repeated here

...

<div style="border:1px solid black">

Now answer the following questions

a)   Where do you live?

        ....

b)   In which section of this hospital or clinic do you work?

        ....

c)   What is your age?

        ....

Thank you very much for contributing towards my research. If you think that there is something important we have missed, please tell us or let us know by telephone or mail:

      Mrs. L. van Huyssteen
      University of South Africa
      Department of African Languages
      PO Box 392
      Pretoria
      0003
      Tel: 012 4298258 Fax: 012 4293221
      e-mail: vhuysl@.unisa.ac.za

</div>

4   Research methodology followed

For research purposes I paid visits to four provincial hospitals, one semi-private hospital in KwaZulu-Natal and one provincial hospital on the border of Mpumalanga and KwaZulu-Natal. These six

hospitals (see specifics below) were chosen because the majority of health workers (nursing sisters, medical consultants and nursing assistants) employed and most of the patients admitted were Zulu speakers. Before visiting these hospitals, I had to consult with hospital authorities, who were either superintendents or matrons, communicating via fax, e-mail or telephone, to obtain (official) permission to conduct the research. At some hospitals it was extremely difficult to obtain such permission. After a successful application the Research Committee of the University of South Africa put funds at my disposal to enable me to undertake the research. Thereafter specific appointments were made with regard to dates of my research undertaking and interviews with medical staff. The consultation and planning were done months in advance. The detailed research methodology followed will be outlined as far as the actual programme and the role played by the informants (medical staff) are concerned:

4.1   Research programme

The first research undertaking took place from 30 March 1998 to 7 April 1998. This undertaking was meant to be a pilot study. If the initial undertaking proved to be successful, the next study would be undertaken. The hospitals visited are situated in urban areas. During this period I visited the following three hospitals while consulting with their respective authorities and staff:

a)   King George V Jubilee Hospital in Durban (KwaZulu-Natal)

b)   McCord Zulu Hospital in Durban (KwaZulu-Natal)

c)   Prince Mshiyeni Hospital in Umlazi (KwaZulu-Natal).

The second research undertaking took place from 22 June 1998 to 30 June 1998. This undertaking was a follow-up of the pilot study since the first undertaking had proved to be both successful and worthwhile. These hospitals visited were chosen as research venues because they are situated in (semi-) rural areas. This was planned with a view to include the Zulu medical terminology used in (semi-) rural areas with that used in urban areas in order to make terms more representative. During this period I visited the following three hospitals while consulting with their respective authorities and

staff in the same manner as during the initial research undertaking:

a)   Newcastle Provincial Hospital in Newcastle (KwaZulu-Natal)

b)   Vryheid Hospital in Vryheid (KwaZulu-Natal)

c)   Amajuba Memorial Hospital in Volksrust (Mpumalanga).

Although the venues are important, the role played by informants in the actual conducting of the research is of utmost importance.

The questionnaires were either collected by myself or mailed to me as late as February 1999.

4.2   The role played by informants

To fulfil the aims of this research project (see 1 earlier) it was important to investigate the extent of the acceptability of the existing documented Zulu health terms. Ultimately the elaboration capacity of the Zulu language also had to be investigated. I therefore had to visit some hospitals and clinics where the health workers employed could assist me to put the extent of the acceptability of Zulu health terms and the elaboration capacity of the Zulu language into practical perspective.

At the hospitals (some of which included clinics), the hospital authorities arranged for me to interview Zulu health workers who dealt with Zulu patients on a daily basis. All informants who took part in this survey did so voluntarily in their own time, with the strictest confidentiality and with compensation from the research fund. The interviews were not interviews in the true sense of the word but took the form of an informal conversation, either with appointed assistants or myself, during which the procedures

were explained and the methodology clarified.

In the questionnaire the informants had to compare the standard written terms on the given standard Zulu medical term list with the Zulu medical terms they actually use. Appropriate terms had to be marked with a tick (✓). If the informant did not agree with the given Zulu medical term s/he had to record the 'better', 'more appropriate' or 'more popular' term instead of (underneath) the given one; in other words, they had to then supply the Zulu terms that are actually used in their daily encounter with medical colleagues and patients. The response of the informants would be influenced by the areas from which they originated, i.e. either urban or (semi-) rural. However, this aspect of origin was not included for discussion in this research project.

In the concluding section of the questionnaire the informants had to answer short Zulu questions about their origin, the section of the hospital where they worked and their age.

At the first three hospitals visited, 56 questionnaires were completed and returned, while at the last three hospitals visited, 64 questionnaires were completed and returned, giving a total of 120 questionnaires that were completed and returned, some via mail.

5   Conclusion

On the whole it can be said that a number of African language scholars were dissatisfied with the official elaboration process mainly undertaken by the previous African Language Committees/ Boards in that they saw it as a stigmatised process forced by legislation, rather than as a testing of the acceptability of terms by documenting the oral terms rendered by mother-tongue speakers in their work environment.

Eventually the completed questionnaire data had to be analysed in order to reach certain conclusions concerning the acceptability of given Zulu health terms to Zulu health workers, specifically the oral

terms that are used on a daily basis. The responses of the informants had to be analysed in order to design a practical oral corpus annotation, mainly based on frequency of occurrence of terms. The aim of this type of annotation is to enhance the elaboration and eventual standardisation process. Eventually oral corpus annotation will also have significance as far as the general elaboration capacity and status of the Zulu language is concerned.

ADDENDUM 2

FREQUENCY PROFILE OF RESPONSES FOR ORAL CORPUS ANNOTATION

Abbreviations for hospitals and regions

A	Amajuba Memorial Hospital in Volksrust (Mpumalanga)
K	King George V Jubilee Hospital in Durban (KwaZulu-Natal)
M	McCord Zulu Hospital in Durban (KwaZulu-Natal)
N	Newcastle Provincial Hospital in Newcastle (KwaZulu-Natal)
P	Prince Mshiyeni Hospital in Umlazi (KwaZulu-Natal)
V	Vryheid Hospital in Vryheid (KwaZulu-Natal)
( )	Total responses of regional hospitals per term
[ ]	Sum total of responses of regional hospitals per term

Per example

WRITTEN CORPUS A	ORAL CORPUS B			
Given standard term(s):	Equivalent 1	Equivalent 2	Equivalent 3	Equivalent 4
ingculazi (AIDS)	**ingculazi** [85] A(12) K(13) M(9) N(20) P(18) V(13)	**umashaya bhuqe** [8] A(2) K(1) N(5)	**unogawula** [2] A(2)	**ikhodi** [1] K(1)
isidakwa (alcoholic)	**isidakwa** [77] A(15) K(12) M(9) N(14) P(17) V(10)	**impuzi tshwala** [10] A(2) K(4) N(4)	**indlamanzi** [8] K(3) N(4) V(1)	**unsutha** [2] N(2)

INDIGENOUS COINAGE (CF. 6.5.1)

WRITTEN CORPUS A	ORAL CORPUS B			
Given standard term(s):	Equivalent 1	Equivalent 2	Equivalent 3	Equivalent 4
ikliniki / **umtholampilo** (clinic)	**umtholampilo** [80] A(14) K(11) M(10) N(19) P(14) V(12)	**ikliniki** [48] A(4) K(7) M(5) N(11) P(17) V(4)		
ibhroyinkhanthisi / **umkhuhlane wamaphaphu** (bronchitis)	**umkhuhlane wamaphaphu** [61] A(4) K(11) M(7) N(14) P(14) V(11)	**ibhroyinkhanthisi** [16] A(3) K(3) M(1) N(2) P(6) V(1)	**izilonda emaphashini** [1] N(1)	
i-alegi (allergy)	**isihlungu** [50] A(6) K(6) M(10) N(14) P(3) V(11)	**i-alergy** [32] A(4) K(8) M(1) N(1) P(14) V(4)	**ukungezwani nomzimba** [4] K(2) N(2)	**ukuqubuka** [2] A(1) N(1)
ibhakthiriya / **ibhakithiriya** (bacteria)	**igciwane** [55] A(10) K(8) M(9) N(19) V(9)	**ibhakthiriya** [23] A(5) K(1) M(1) N(1) P(13) V(2)	**ibhakithiriya** [21] A(2) K(2) M(1) N(1) P(13) V(2)	

WRITTEN CORPUS A	ORAL CORPUS B

Given standard term(s):	Equivalent 1	Equivalent 2	Equivalent 3	Equivalent 4
isibulala mabhakithiriya (antibiotic)	**isibulala magciwane** [44] A(5) K(7) M(7) N(13) V(12)	**isibulala mabhakithiriya** [34] A(7) K(6) M(1) P(13) V(1)	**uzifozonke** [5] A(1) N(4)	**isilwa magciwane** [2] N(2)
imbewu yokufa / **ijemu** / **igciwane** (germ)	**igciwane** [80] A(12) K(9) M(10) N(24) P(15) V(10)	**imbewu yokufa** [27] A(3) K(6) M(1) N(13) P(1) V(3)	**ijemu** [12] A(2) K(1) N(6) V(3)	
imagreyini (migraine)	**imagreyini** [38] A(11) K(5) M(1) N(3) P(16) V(2)	**ikhanda elibuhlungu kakhulu** [26] A(4) K(4) M(3) N(10) P(1) V(4)	**ikhanda elibuhlungu elingapheli masinyane** [9] A(2) M(3) N(2) V(2)	**ukuphathwa ikhanda kakhulu** [2] M(1) V(1)
ifontaneli (fontanelle)	**ukhakhayi** [34] A(8) K(7) M(10) N(17) P(1) V(1)	**ifontaneli** [27] A(6) K(2) P(14) V(5)	**ufokothi** [5] K(2) N(3)	**isikhala sokhakhayini** [1] N(1)

ACCURATE DESIGNATION (CF. 6.5.2)

WRITTEN CORPUS A	ORAL CORPUS B			
Given standard term(s):	Equivalent 1	Equivalent 2	Equivalent 3	Equivalent 4
umdlavuza / **isimila** / **umhlaza** / **umdlavuza** (cancer)	**umdlavuza** [82] A(15) K(9) M(10) N(19) P(17) V(12)	**ikhensa** [48] A(4) K(4) M(5) N(15) P(16) V(4)	**isimila** [41] A(10) K(5) M(4) N(15) P(1) V(6)	**umhlaza** [39] A(6) K(5) M(3) N(16) P(4) V(5)
umdlavuza wesifuba (breast cancer)	**umdlavuza webele** [84] A(12) K(12) M(9) N(21) P(16) V(14)	**umdlavuza wesifuba** [17] A(6) K(4) M(1) N(3) P(1) V(2)		
umhlaza wesikhumba (skin cancer)	**umhlaza wesikhumba** [56] A(11) K(11) M(7) N(13) P(10) V(4)	**umdlavuza wesikhumba** [37] A(6) K(1) M(4) N(10) P(5) V(11)		
izinga lokushisa kwegazi (body temperature)	**izinga lokushisa kwegazi** [41] A(10) K(4) M(3) N(3) P(14) V(7)	**izinga lokushisa komzimba** [32] A(6) K(3) M(8) N(8) P(2) V(5)	**ukushisa komzimba** [7] K(1) N(6)	**izinga lokushisa emzimbeni** [4] K(1) V(3)

WRITTEN CORPUS A	ORAL CORPUS B

Given standard term(s):	Equivalent 1	Equivalent 2	Equivalent 3	Equivalent 4
isitabani (homosexual male)	**isitabani** [68] A(14) K(11) M(8) N(20) P(4) V(11)	**isitabani sesilisa** [13] P(13)	**inkonkoni** [6] K(1) N(2) V(3)	**ungqingili** [6] K(1) N(2) V(3)
isitabani (lesbian)	**isitabani** [68] A(13) K(9) M(7) N(15) P(4) V(13)	**isitabani sowesifazane** [19] K(2) N(3) P(14)	**incukumbili** [7] A(7)	**ungqingili** [4] A(2) K(1) M(1)
ukuba nesandulela ingculazi (HIV positive)	**ukuba nesandulela ingculazi** [78] A(17) K(9) M(6) N(19) P(15) V(12)	**ukuba negciwane lengculazi** [4] A(1) K(2) M(1)	**ukuba ne-HIV gciwane egazini** [1] K(1)	**ukutholakala negciwane lengculazi** [2] M(2)
-khipha iqanda (ovulate)	**-khipha iqanda** [75] A(14) K(12) M(4) N(20) P(15) V(10)	**-khipha iqanda lenzalo** [6] M(3) N(3)	**-akhela iqanda** [2] V(2)	**-khipha iqanda lowesifazane** [1] K(1)

PHONOLOGICAL ADAPTATIONAL TRENDS (CF. 6.5.3)

WRITTEN CORPUS A	ORAL CORPUS B

Given standard term(s):	Equivalent 1	Equivalent 2	Equivalent 3	Equivalent 4
ibhulima (bulimia)	**ibhulima** [45] A(10) K(4) N(6) P(17) V(8)	**ukuminza** [5] A(2) V(3)	**isifo sokudla** [2] K(1) V(1)	**isifo sokuhlanza ukudla** [1] A(1)
iganjirini (gangrene)	**iganjirini** [22] A(5) M(1) N(13) V(3)	**igangrini** [9] P(9)	**ukubola** [7] A(2) N(4) P(1)	**igangirini** [4] P(4)
ikhwashi (kwashiokor)	**ikhwashi** [47] A(9) K(6) M(5) N(7) P(13) V(7)	**isifo sendlala** [40] A(7) K(4) M(6) N(14) V(9)	**umTopia** [13] A(2) K(3) M(1) N(1) P (4) V(2)	**isifo sokungondleki** [5] K(1) M(3) N(1)
umjovo (inoculation)	**umjovo** [51] A(12) K(12) M(5) N(12) P(3) V(7)	**ukunokolota** [25] M(6) N(1) P(15) V(3)	**umnokoloto** [4] K(2) V(2)	**umgcabo** [2] M(1) N(1)

WRITTEN CORPUS A	ORAL CORPUS B

Given standard term(s):	Equivalent 1	Equivalent 2	Equivalent 3	Equivalent 4
i-AZT (zidovudine / AZT)	**i-AZT** [58] A(12) K(15) M(6) N(6) P(15) V(4)	**azathi** [2] V(2)	**isithibingculazi** [2] N(2)	**iphilisi elapha ingculazi**[2] K(1) M(1)
isona (sonar scan)	**isona** [47] A(9) K(5) M(2) N(12) P(13) V(6)	**emafutheni** [6] A(1) N(2) V(3)	**isikeni** [4] V(2)	**iscan** [2] V(2)
imfiva encombo / iyelo fiva (yellow fever)	**iyelo fiva** [46] A(11) K(4) M(3) N(8) P(13) V(7)	**imfiva encombo** [32] A(4) K(9) M(2) N(9) P(3) V(4)	**umkhuhlane** [3] N(1) V(2)	**imfiva ephuzi** [2] N(2)

SEMANTIC SHIFT ALTERNATIVE (CF. 6.5.4)			
WRITTEN CORPUS A	**ORAL CORPUS B**		
Given standard term(s):	Equivalent 1	Equivalent 2	Equivalent 3	Equivalent 4
iX-reyi (X-ray)	**isithombe** [40] A(6) K(6) M(8) N(10) V(10)	**iX-reyi** [35] A(7) K(8) M(2) N(6) P (9) V(3)	**isithombe seX-reyi** [4] A(1) N(3)	**igesi** [2] K(1) V(1)
ikhatharakithi (cataract)	**ungwengwezi (emehlweni)** [38] A(3) K(2) M(9) N(14) P(1) V(9)	**ikhatharakhithi** [32] A(3) K(4) N(5) P(15) V(5)	**untwentwezi** [4] K(2) N(2)	**ulwembu nasemehlweni** [3] A(2) K(1)
isitrokhi (stroke)	**isitrokhi** [36] A(5) K(5) M(2) N(7) P(13) V(4)	**unhlangothi** [23] A(4) K(4) N(7) V(8)	**isifo sonhlangothi** [9] A(2) M(2) N(2) P(1) V(2)	**ukushaywa yinyoni** [8] A(2) K(3) N (1) V(2)
uxhilo / **idifuteria** (diphtheria)	**uklilo** [49] A(2) K(3) M(4) N(12) P(14) V(14)	**uxhilo** [34] A(6) K(7) M(3) N(11) P (3) V(4)	**idifuteria** [15] A(5) K(4) N(4) P(1) V(1)	**umphimbo omhlophe** [2] A(1) N(1)

TABOO PREFERENCE (CF. 6.5.5)				
WRITTEN CORPUS A	**ORAL CORPUS B**			
Given standard term(s):	Equivalent 1	Equivalent 2	Equivalent 3	Equivalent 4
isifo socansi esithathelanayo (sexually transmitted disease)	**isifo socansi** [81] **esithathelanayo** A(14) K(11) M(10) N(19) P(17) V(10)	**ukubhajwa** [5] A(2) M(1) V(2)	**isifo samasoka** [3] A(1) V(2)	**isipatsholo** [1] K(1)
ivasekhithomi (vasectomy)	**ivasekhithomi** [44] A(7) K(3) M(11) N(8) P(11) V(4)	**ukuthena** [23] A(3) K(4) M(4) N(8) V(4)	**ukuvala inzalo kwabesilisa** [21] A (2) K(2) M(6) N(8) P(1) V(2)	**ukuqeda inzalo kumuntu wesilisa** [4] K(1) N(1)
ukuba uncumbili (bisexual)	**uncukumbili** [43] A(4) K(1) M(5) N(11) P(13) V(9)	**ukuba uncumbili** [27] A(8) K(1) M(4) N(8) P(3) V(3)	**isitabani** [9] A(2) K(2) M(1) N(2) V(2)	**ungqingili** [7] A(2) K(4) N(1)

WRITTEN CORPUS A	ORAL CORPUS B			
Given standard term(s):	Equivalent 1	Equivalent 2	Equivalent 3	Equivalent 4
ikhondomu / **ijazi lomkhwenyana** / **ijazi** (condom)	**ijazi lomkhwenyana** [77] A(13) K(7) M(9) N(19) P(15) V(14)	**ikhondomu** [46] A(8) K(5) M(3) N(10) P(17) V(4)	**ijazi** [33] A(5) K(2) M(3) N(7) P(13) V(3)	**iglavu** [2] V(2)
ukukhipha isisu (abortion)	**ukukhipha isisu** [55] A(6) K(11) M(8) N(10) P(16) V(4)	**ukuhushula isisu** [25] K(5) N(8) P(1) V(11)	**ukuphuphuma kwesisu** [13] A(2) K(5) M(1) N(2) V(3)	**ukuchitheka kwesisu** [7] A(2) N(4) V(1)