

**THE USE OF CORPORA IN THE COMPILATION OF A SPECIALISED
ENGLISH-ISINDEBELE GLOSSARY OF MEDICAL TERMS**

by

Nomsebenzi Johanna Malele

submitted in accordance with the requirements for the degree of

DOCTOR OF LITERATURE AND PHILOSOPHY

in the subject

AFRICAN LANGUAGES

at the

UNIVERSITY OF SOUTH AFRICA (UNISA)

SUPERVISOR: Prof Sonja Bosch

CO-SUPERVISOR: Dr Gertrud Faass

February 2021

DECLARATION

Student Number: 08126259

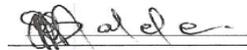
I declare that this thesis entitled: **The use of corpora in the compilation of a specialised English-isiNdebele glossary of medical terms** is my own work. All the sources that I have used or quoted have been indicated and acknowledged by means of complete references in the text and in the list of sources.

I further declare that I submitted the thesis to originality checking software and that it falls within the accepted requirements for originality.

I further declare that I have not previously submitted this work, or part of it, for examination at Unisa for another qualification or at any other higher education institution.

24 February 2021

Date

A handwritten signature in black ink, appearing to read 'D. de V.' or similar, written over a horizontal line.

Signature

DEDICATION

To my husband Mbuso and my daughters Simphiweyinkosi and Lesedi, this degree belongs to you! Your support and love kept me going even when my spiritual legs were weary.

To my late mom NaNdala, *wakhamba ungakaboni bonyana ubelethe udorhodere.*

To my late Dad, Skosana sikaMusi, you really believed in me. If you were still around you would say: *Angirareki!*

To my late brother Aaron, how I wish you could be here and witness the faithfulness of the Lord.

To all my brothers, sisters, nephews and nieces, *ngiyathokoza maJali!*

To Reverend Sydney Chami Ntuli, *Mhlalukwana qala umsebenzakho!*

To Dr Ntswaki Masithela, Dr Bongsi Mashiane-Nkabinde, Mrs Bongsi Mhlanga, Mrs Nadia Davis and all my prayer partners, thank you all for walking this journey with me. Thank you for encouraging and also praying for me when discouragement set in.

To Ms Thabiso Monakgisi, God brought you in my life 'for such a time as this!'

ACKNOWLEDGEMENTS

Firstly, I would love to thank God who enabled me to complete this project. If it had not been for the Lord, this would not have been attainable. *Ngiyathokoza Zimu!*

To my promoter Prof Sonja Bosch, thank you for your fruitful guidance, words of encouragement and understanding.

To my co-promoter Dr Gertrud Faass, whose speciality in corpus linguistics and invaluable criticism has contributed towards the improvement of this study.

To Prof Inge Kosch, for editing this thesis.

To Ms Sizakele Selatlhe, for formatting this thesis.

To Ms Refilwe Matatiele, thank you for being an angel that works at Unisa Library.

To Ikwewezi FM for having allowed me to record the health program, *ngiyathokoza!*

To the presenter of the programme Mrs Busisiwe (Bussai) Mathebula from Ikwewezi FM, 'thank you for having ensured that I do not miss the health program. Even when I missed it, you would always email me the topics that were presented so that I can get them.' *Ngiyathokoza naNtuli!*

To the Department of Health, Mpumalanga for granting me permission to use their health facilities and also to interview the nurses for this study.

To Nomusa Skhosana-Hlatshwayo, thank you for ensuring that the CEO for hospitals and the CEO for clinics attend to my applications. *Ngithokoza khulu naThamana!*

To the following medical doctors who participated in this study: Dr Nomthandazo Tjiyane, Dr Paul Sibanyoni, Dr Vusi Masombuka and Dr Mahlangu Sfiso. I thank you all!

Special thanks to Prof Stanley Madonsela the CoD, for words of encouragement and genuine concern.

To all members of the Department of African Languages at the University of South Africa (Unisa) thank you so much. Prof Mojapelo Mampaka 'You do not know the role you played in my journey. You always knew what to say!'

To Dr Kina Malobola-Ndlovu, 'Thank you for your continuous support Dr Sesi'.

To Ms Marissa Griesel for an introduction to WordSmith Tools that I have so much fallen in love with.

To Mr Bethuel Malele for always addressing my IT challenges whilst busy with this project.

To Mr Mothusi Moswele, thank you for assisting with the lay-out of this thesis.

To my IT neighbour Mr Costa Martinis, thank you for always having been willing to assist me with the graphs and the lay-out of this thesis.

ABSTRACT

The Constitution of South Africa legislated that South African languages should be developed to the status of languages such as English and Afrikaans. This necessitated translations in specialised fields such as technology, science, justice and finance. The shortage of terminology and resources such as bilingual glossaries in African languages had a negative impact on the development and intellectualisation of languages such as isiNdebele in particular. Most researchers and scholars are of the opinion that specialised lexicography is significant in the development of language terminology so that African languages can be usable in specialised fields.

Based on the above background, this study explored how corpora can be interrogated using WordSmith Tools to extract both monolingual terminology from the spoken corpus and bilingual terminology from the written medical corpus. The purpose of the extraction was to compile a specialised English-isiNdebele glossary of medical terms.

A corpus-driven approach was utilised in this study. This approach was selected because through it, every piece of information is drawn from the corpus without pre-conceived ideas. Both corpus linguistics theory and the Grounded Theory were implemented in this study. The two theories complement each other well. Whilst corpus linguistics studies real examples of what people have said, the Grounded Theory on the other hand (through its method of agreement and the method of difference) enabled the researcher to determine what is common and uncommon in what has been said (through corpus linguistics theory).

The analysis and extraction of both the monolingual and the bilingual terminology for glossary making was successful through the use of the KeyWords function and the Concordance function of the WordSmith Tools. Terms that resulted from the KeyWords function had to be validated manually. This is because the KeyWords function cannot sort the variants of the term candidate list, neither can it identify the multiword-term candidates. The researcher therefore had to sort the variants and also identify multiword-term candidates manually. Through the Concordance function, the researcher was able to extract multiword-terms. Finally, an English-isiNdebele glossary of medical terms was compiled. The study therefore contributes to corpus-driven lexicography as well as standardisation of medical terminology.

Key words: Corpus, parallel corpora, bilingual corpora, corpus linguistics, Grounded Theory, WordSmith Tools, language for special purposes, language for general purposes, terminology, written corpus, spoken corpus.

IRHUNYEZORHUBHULULO

UmThethosisekelo weSewula Afrika wagunyaza ukobana umbuso uthathe amagadango wokuthuthukisa nokusetjenziswa kwamalimi womdabu. Lokhu kwenza ukobana kube netlhogeko ekulu yokutjhugululwa kwemitlolo okukhethekileko, emikhakheni efana netheknoloji, isayensi, umthetho kanye nezeemali. Ukutlhogeka kwetheminoloji neensetjenziswa ezifana namadlhozari amalimimbili emalimini womdabu, khulukhulu emalimini anjengesiNdebele kunomthelela omumbi ekutjhugululweni nekusetjenzisweni kwesiNdebele emikhakheni ethileko. Abarhubhululi abanengi abasemalimini wabomdabu bayavumelana ukobana ilekzikhografi ekhethekileko iqakathekile ekuthuthukisweni kwetheminoloji yelimi ukuze amalimi womdabu asebenziseke emikhakheni ekhethekileko.

Ukuya ngesendlalelo esinikelwe ngehlesi, lomsebenzi urhubhulule indlela ikhophora engazizingwa ngayo ngokusebenzisa iWordSmith Tools ekutsomuleni itheminoloji elimilinye ekhutjwe ekhophasini ekhulunywako kanye netheminoloji emalimimbili ekhutjwe ekhophasini yezamaphilo etlolweko. Ihloso yokukhutjwa kwetheminoloji bekukubuthelela idlhozari ekhethekileko elimimbili (isiNgisi nesiNdebele).

Indlela esunduzwa yikhophasi (corpus-driven approach) isetjenzisiwe. Lendlela ikhethwe ngebanga lokobana ngayo, lelo nalelo lwazi elisetjenziswako lisuselwa kukhophasi ngaphandle kwemicabango ehleliweko yangaphambili. Amathiyori amabili, i-corpus linguistics neGrounded Theory asetjenziswe womabili kilomsebenzi. Adlhegana kuhle lamathiyori ngombana i-corpus linguistics isebenzisa iimbonelo zamambala zalokho okutjhiwo babantu. IGrounded Theory yona ngakwelinye ihlangothi, ngebanga leendlela zayo zokuvumelana nokuphikisana, ivumela umrhubhululi ukobana akwazi ukulemuka okufanako nalokho okuhlukileko kilokho okutjhiwiweko ngebanga lecorpus linguistics.

Ukutsengwa nokukhutjwa kwetheminoloji elimilinye naleyo elimimbili ibe yipumelelo ngebanga lokusetjenziswa kweWordSmith Tools, khulukhulu ngokusebenzisa iKeyWords kanye neConcordance. Amathemu akhiqizwa ngeKeyWords aqinisekiswa ngezandla. Lokhu kungebanga lokobana iKeyWords ayikwazi ukusota amathemu begodu ayikwazi nokulemuka amathemumvango. Umrhubhululi wakateleleka ukobana akwenze lokho ngezandla. IConcordance, ngilo ithulusi umrhubhululi akwazi ukulisebenzisa ukukhupha amathemumvango. Ekugcineni, idlhozari yezamaPhilo, elimimbili (isiNgisi nesiNdebele) yabuthelelwa. Lomsebenzi udlala indima ekulu kulekzikhografi esunduzwa yikhophasi kanye nekulinganisweni kwetheminoloji yezamaPhilo.

OPSOMMING

Die Grondwet van Suid-Afrika het deur wetgewing bepaal dat die Afrikatale van Suid-Afrika ontwikkel moet word tot die status van tale soos Engels en Afrikaans. Dit het gelei tot die noodsaaklikheid van vertalings in vakgebiede soos tegnologie, wetenskap, justisie en finansies. Die tekort aan terminologie en hulpbronne soos tweetalige woordelyste in Afrikatale het 'n negatiewe impak gehad op die ontwikkeling en intellektualisasie van tale soos isiNdebele in die besonder. Die meeste navorsers en vakkundiges is van mening dat vakgerigte leksikografie beduidend is in die ontwikkeling van taalterminologie sodat Afrikatale in gespesialiseerde navorsingsgebiede gebruik kan word.

In die lig van bogenoemde agtergrond, ondersoek hierdie studie hoe korpora onder die loep geneem kan word deur van WordSmith Tools gebruik te maak om beide eentalige terminologie van die gesproke korpus en tweetalige terminologie van die geskrewe mediese korpus te onttrek. Die doel van die ekstraksie was om 'n gespesialiseerde Engels-isiNdebele woordelys van mediese terme saam te stel.

'n Korpus-gedrewe benadering is in hierdie studie gevolg. Dié benadering is gekies omdat dit daardeur moontlik was om elke stukkie inligting uit die korpus te onttrek sonder vooropgestelde idees. Beide die 'Corpus linguistics' teorie en die 'Grounded' teorie is in hierdie studie aangewend. Die twee teorieë vul mekaar goed aan. Terwyl die 'Corpus linguistics' teorie daadwerklike voorbeelde van wat mense gesê het, ondersoek, het die 'Grounded' teorie aan die ander kant (deur sy metodes van ooreenkoms en verskil) dit vir die navorser moontlik gemaak om te bepaal wat is alledaags en wat is ongewoon in wat gesê is (deur die 'Corpus linguistics' teorie)

Die ontleding en ekstraksie van beide die eentalige en tweetalige terminologie vir die samestelling van 'n woordelys was suksesvol deur die aanwending van die KeyWords funksie en die Concordance funksie van die WordSmith Tools. Terme wat uit die KeyWords funksie gespruit het moes nie-meganies geverifieer word. Die rede hiervoor is dat die KeyWords funksie nie die variante van die term kandidatelys kan sorteer en ook nie die multiwoord term kandidate kan identifiseer nie. Die navorser moes gevolglik die variante en die multiwoord terme per hand identifiseer. Deur die Concordance funksie kon die navorser die multiwoord terme onttrek. Ten slotte is 'n Engels-isiNdebele woordelys van mediese terme saamgestel. Die studie dra dus by tot korpus-gedrewe leksikografie sowel as standardisering van mediese terminologie.

ABBREVIATIONS AND ACRONYMS

The following abbreviations and acronyms are used in this study:

AC:	Analysis corpus
AFRILEX:	African Association for Lexicography
ALLEX:	The African Languages Lexical Project
ANC:	American National Corpus
ATE:	Automatic Term Extraction
BNC:	British National Corpus
CDA:	Corpus-Driven Approach
CI:	Curriculum Implementer
CLTAL:	Centre for Legal Terminology in African Languages
DAC:	Department of Arts and Culture
DBE:	Department of Basic Education
EFL:	English as a Foreign Language
FET:	Further Education and Training
GEMS:	Government Employees Medical Scheme
KWIC:	Key Word In Context (KWIC)
LGP:	Language for general purposes
LSP:	Language for special purposes
MTE:	Manual Term Extraction
NLB:	National Language Board
NLS:	National Language Services
NLU:	National Lexicography Unit
OCR:	Optical Character Recognition-tool
OPM:	Operational Manager
PanSALB:	Pan South African Language Board
PHC:	Primary Health Care
PLC:	Provincial Language Committee
RC:	Reference corpus
SABC:	South African Broadcasting Corporation
SADiLaR:	South African Centre for Digital Language Resource
SANTED:	South Africa Norway Tertiary Education Development
SL:	Source Language
SLT:	Second Language Teaching

SPeLCAL: Special Language Corpora for African Languages
TC: Term Candidates
TE: Term Extraction
TL: Target Language
ULPDO: University Language Planning and Development Office
WST: WordSmith Tools

TABLE OF CONTENTS

DECLARATION.....	ii
DEDICATION	iii
ACKNOWLEDGEMENTS.....	iv
ABSTRACT.....	vi
IRHUNYEZORHUBHULULO	vii
OPSOMMING	viii
ABBREVIATIONS AND ACRONYMS	ix
LIST OF TABLES.....	xvi
LIST OF FIGURES	xvii
CHAPTER 1	1
INTRODUCTION	1
1.1 BACKGROUND TO AND RATIONALE BEHIND THE RESEARCH PROBLEM	1
1.2 STATEMENT OF THE PROBLEM AND RESEARCH QUESTIONS.....	5
1.3 AIM AND OBJECTIVES OF THE STUDY	5
1.4 METHODOLOGY	6
1.5 THEORETICAL FRAMEWORK	6
1.6 DEFINING THE KEY TERMS AND CONCEPTS.....	8
1.6.1 Corpus	9
1.6.2 Corpus-driven approach	9
1.6.3 The monolingual corpus	9
1.6.4 An English-isiNdebele parallel corpus	9
1.6.5 A specialised corpus	10
1.6.6 The WordSmith Tools	10
1.6.7 Defining a glossary	11
1.7 DATA COLLECTION TOOLS AND PROCEDURES.....	11
1.7.1 Data analysis and interpretation.....	11
1.7.2 Ethical consideration	12
1.8 STRUCTURE OF THE THESIS	12
CHAPTER 2	15
LITERATURE REVIEW.....	15
2.1 INTRODUCTION.....	15
2.2 THEORETICAL OVERVIEW	16
2.2.1 Corpus linguistics theory	16
2.2.1.1 <i>Defining corpus linguistics</i>	24
2.2.1.2 <i>Advantages of corpus linguistics</i>	25

2.2.1.3	<i>Disadvantages of corpus linguistics</i>	25
2.2.1.4	<i>Computers and corpus linguistics</i>	25
2.2.2	The Grounded Theory	27
2.2.2.1	<i>Defining the Grounded Theory</i>	27
2.2.2.2	<i>Advantages of the Grounded Theory</i>	29
2.2.2.3	<i>Disadvantages of the Grounded Theory</i>	29
2.3	METHODOLOGIES BASED ON THE THEORIES DISCUSSED	29
2.3.1	LSP Corpus design	30
2.3.2	Compilation of corpora for terminology and glossary	30
2.3.2.1	<i>Defining ‘term’</i>	30
2.3.2.2	<i>Defining ‘terminology’</i>	31
2.4	PROJECTS BASED ON THE THEORIES DISCUSSED	33
2.4.1	ALLEX project	33
2.4.2	UKZN project	34
2.4.3	CLTAL project	36
2.4.4	SANTED Multilingualism Project	37
2.4.5	SPeLCAL project	38
2.5	AN OVERVIEW OF GLOSSARY COMPILATION	40
2.5.1	English-Portuguese glossary of cooking terms	40
2.5.2	Glossary on multilingual concept literacy terms	41
2.6	STUDIES CONDUCTED ON COMPILATION OF MEDICAL TERMS	42
2.6.1	IsiZulu-English medical terms	42
2.6.2	English-Shona biomedical terminology	43
2.7	TERMINOLOGY DEVELOPMENT	44
2.7.1	Challenges faced when developing new terms in African languages	45
2.7.2	Guidelines for the development of new terms for African languages	46
2.8	TERM-FORMATION PROCESSES	47
2.8.1	Primary term-formation strategies (foreign)	50
2.8.1.1	<i>Paraphrasing</i>	50
2.8.1.2	<i>Borrowing</i>	50
2.8.1.3	<i>Derivation</i>	53
2.8.1.4	<i>Deideophonisation</i>	53
2.8.2	Secondary term formation strategy (internal resources)	56
2.9	TABOO IN TERM-FORMATION PROCESS	56
2.10	COPYRIGHT AND PERMISSIONS	58
2.11	CONCLUSION	59

CHAPTER 3	61
RESEARCH METHODOLOGY	61
3.1 INTRODUCTION	61
3.2 RESEARCH DESIGN	61
3.2.1 Research methods	61
3.2.1.1 <i>The qualitative research method</i>	61
3.2.1.2 <i>The quantitative research method</i>	62
3.2.2 Data collection instruments	63
3.2.2.1 <i>Questionnaires</i>	63
3.2.2.2 <i>Interviews</i>	68
3.2.3 Collection of written and spoken corpora	68
3.2.3.1 <i>The written corpus</i>	68
3.2.3.2 <i>The spoken corpus</i>	69
3.3 THEORIES FOR DATA ANALYSIS	69
3.3.1 The corpus-driven approach (CDA)	69
3.3.2 The Grounded Theory	69
3.3.3 The corpus linguistics theory	70
3.4 CONCLUSION	71
CHAPTER 4	74
PRESENTATION AND ANALYSIS OF DATA COLLECTED THROUGH QUESTIONNAIRES	74
4.1 INTRODUCTION	74
4.2 QUESTIONNAIRES	74
4.2.1 Summary of tables of respondents	78
4.2.2 Analysis of questionnaires	78
4.2.3 Theory used in analysing questionnaires	78
4.2.3.1 <i>Open coding</i>	79
4.2.3.2 <i>Axial coding</i>	92
4.2.3.3 <i>Selective coding</i>	96
4.3 TERM-FORMATION PROCESSES	96
4.3.1 Primary term-formation strategies (foreign) in isiNdebele	96
4.3.1.1 <i>Paraphrasing</i>	97
4.3.1.2 <i>Borrowing</i>	97
4.3.1.3 <i>Derivation</i>	98
4.3.1.4 <i>Deideophonisation</i>	99
4.3.1.5 <i>Abbreviation</i>	99
4.3.2 Secondary term-formation strategy (internal resources) in isiNdebele ... 100	

4.3.3	Overview of term-formation strategies	100
4.4	TABOO AND TERM-FORMATION PROCESSES IN ISINDEBELE	103
4.5	CONCLUSION.....	104
CHAPTER 5		106
DATA ANALYSIS OF THE WRITTEN AND THE SPOKEN CORPUS		106
5.1	INTRODUCTION	106
5.1.1	The spoken corpus	106
5.1.2	The written corpus.....	115
5.2	SUMMARY OF ALL THE DATA COLLECTED (SPOKEN AND WRITTEN ISINDEBELE).....	117
5.3	ALIGNMENT OF TEXTS	117
5.4	TERM EXTRACTION	122
5.4.1	Term extraction through the WordSmith Tools	125
5.4.2	The manual extraction of terms	138
5.4.2.1	<i>The manual extraction of terms from the corpora.....</i>	139
5.4.2.2	<i>Comparison between terms manually extracted from the written and the spoken (literacised) corpus</i>	142
5.5	COMPARISON BETWEEN THE SPOKEN (LITERACISED) AND WRITTEN CORPUS OF THIS STUDY	146
5.5.1	Term-formation processes.....	147
5.5.2	Corpus size.....	151
5.5.3	WordSmith Tools and corpora (written and spoken)	151
5.6	CONCLUSION.....	151
CHAPTER 6		154
FINDINGS AND INTERPRETATIONS		154
6.1	INTRODUCTION	154
6.2	AIM, OBJECTIVES AND RESEARCH QUESTIONS OF THE STUDY	154
6.3	OVERVIEW OF CHAPTERS	156
6.4	CONTRIBUTION OF THE STUDY.....	161
6.5	LIMITATIONS OF THE PRESENT STUDY	163
6.6	RECOMMENDATIONS	164
6.7	FUTURE RESEARCH.....	166
BIBLIOGRAPHY		168

APPENDICES	182
APPENDIX A: ETHICS CERTIFICATE FROM THE DEPARTMENT OF AFRICAN LANGUAGES, UNISA	182
APPENDIX B: LETTER TO MPUMALANGA REQUESTING ACCESS TO HEALTH FACILITIES.....	184
APPENDIX C: APPLICATION FOR RESEARCH & ETHICS APPROVAL.....	186
APPENDIX D: LETTER OF INFORMED CONSENT FOR PARTICIPANTS IN A SEMI-STRUCTURED INTERVIEW.....	187
APPENDIX E: QUESTIONNAIRES.....	189
APPENDIX F: IRHELO LAMAGAMA APHATHELENE NEZAMAPHILO: A LIST OF MEDICAL TERMS FROM THE WRITTEN AND THE SPOKEN CORPUS	201
APPENDIX G: ENGLISH-ISINDEBELE GLOSSARY OF MEDICAL TERMS.....	204
APPENDIX H: TURN-IT-IN DIGITAL RECEIPT.....	235
APPENDIX I: TURN-IT-IN SIMILARITY INDEX.....	236

LIST OF TABLES

Table 4.1: <i>Metadata of medical pamphlets from which questionnaire terms were sourced</i>	75
Table 4.2: <i>Details of doctors who participated in the study</i>	76
Table 4.3: <i>Details of hospitals and clinics where nurses participated</i>	77
Table 4.4: <i>Details of Curriculum Implementers who participated in the study</i>	77
Table 4.5: <i>Details of isiNdebele translators who participated in the study</i>	78
Table 4.6: <i>Order of preference of the use of isiNdebele medical terms (Open & Axial Coding)</i>	94
Table 4.7: <i>Examples of paraphrasing</i>	97
Table 4.8: <i>Example of direct loan</i>	97
Table 4.9: <i>Examples of coinage</i>	97
Table 4.10: <i>Examples of transliteration</i>	98
Table 4.11: <i>Examples of derivation</i>	98
Table 4.12: <i>Example of deideophonisation</i>	99
Table 4.13: <i>Examples of acronyms</i>	99
Table 4.14: <i>Examples of clipping</i>	99
Table 4.15: <i>Examples of semantic transfer</i>	100
Table 4.16: <i>Summary of term-formation strategies</i>	101
Table 4.17: <i>Examples of taboo</i>	104
Table 5.1: <i>Comparison between the written and spoken corpus</i>	108
Table 5.2: <i>Summary of the spoken corpus from Ikwewezi FM</i>	109
Table 5.3: <i>Summary of doctors who participated in interviews</i>	113
Table 5.4: <i>An extract from the spoken corpus showing sentences with and without fillers</i> ..	114
Table 5.5: <i>Metadata of written corpus</i>	118
Table 5.6: <i>Summary of isiNdebele general corpus</i>	123
Table 5.7: <i>isiNdebele most frequent first 100 types</i>	127
Table 5.8: <i>Doctors' interviews frequency-ranked word list without function words</i>	129
Table 5.9: <i>The resulting KeyWords function terms</i>	131
Table 5.10: <i>Manually identified terms from the corpora of this study</i>	139

LIST OF FIGURES

<i>Figure 5.1: Summary of the written isiNdebele corpus</i>	116
<i>Figure 5.2: Un-aligned parallel English-isiNdebele texts</i>	120
<i>Figure 5.3: Aligned parallel English-isiNdebele text</i>	121
<i>Figure 5.4: Wordsmith Tools functions</i>	126
<i>Figure 5.5: Multi-words on cancer</i>	134
<i>Figure 5.6: Display of collocates for the search pattern <i>ikhenza</i></i>	137
<i>Figure 5.7: Collocates of the term <i>iingazi</i></i>	138
<i>Figure 5.8: Verbal root <i>-beleth-</i></i>	146

CHAPTER 1 INTRODUCTION

1.1 BACKGROUND TO AND RATIONALE BEHIND THE RESEARCH PROBLEM

The Constitution of the Republic of South Africa makes it clear that one of the fundamental rights of a human being is a language. This calls for the development and promotion of all languages, particularly the official ones, so that they can also be used in the same environments the other languages already enjoy. IsiNdebele (ISO 639-3: ndl) is one of the languages of South Africa that has been granted official status.

IsiNdebele is amongst the nine official African languages of South Africa and it is spoken predominantly in Mpumalanga, in former kwaNdebele. It has been a spoken language for many years, it was only made a formal, written language in 1985, when it was introduced in schools. Before 1985, the language of instruction for the children of amaNdebele (Ndebele speakers/people) was isiZulu, instead of isiNdebele. It was examined for the first time as a matriculation subject in 1996 (Jiyane, 1994:1).

According to the multilingual policy that South Africa has adopted, all languages should reach modernity. This would be achieved by developing their technical terminology commensurate with English and Afrikaans. It is in response to this call and the afore-mentioned background that the researcher is compiling an English-isiNdebele glossary of medical terms.

IsiNdebele is one of the South African languages that has been marginalised for some time. There is therefore, a greater need of ensuring that it is developed and elevated to the level of other African languages like isiXhosa and isiZulu, and ultimately to the required status of South African official languages like English and Afrikaans. This will make isiNdebele to be part of those languages that are highly recognised within the economic and educational spheres.

The language under study is under-developed. It has a scarcity of resources as well as a shortage of terminology. It does not compare well with other official languages, particularly those of the Nguni language group to which it belongs. The language has no dictionary for language specific purposes (LSP). Dictionaries that are available in isiNdebele are dictionaries for general purposes (LGP). A general or reference language corpus (RC) is necessary to extract terms with the KeyWords function of the WordSmith Tools. It was thus necessary for the researcher to compile a general language corpus of isiNdebele. Bowker and Pearson

(2002:45) maintain that the LSP corpus focuses on an aspect of a language. It could be confined to the LSP of a subject field, a language variety or specific text type. One cannot use this kind of corpus in making observations about a language in general. In this regard, the present study focuses on the medical domain.

The LGP or RC corpus is non-technical whilst the LSP or analysis corpus (AC) is domain specific and technical (Khumalo, 2015:497). In previously conducted and current researches in computational terminology, a variety of techniques have been and are used to extract and recognise terms. Through these techniques the identified terms can be made public knowledge and accessible through terminological databases to end-users.

A glossary of medical terms that was compiled in this study will be of great use to the medical fraternity, particularly doctors. It will also benefit patients who sometimes become the victims of misdiagnoses because of miscommunication caused by the lack of terminology. The glossary will also satisfy the needs of isiNdebele translators, educators and end users. It will contribute towards the standardisation of medical terms and the growth of this language, in turn.

Although this is the first study about the compilation of a medical corpus in isiNdebele, similar works have been undertaken in other languages. Through the University of Kwazulu-Natal (UKZN) project, the isiZulu National Corpus of 31,775,784 tokens was collected. Through the African Languages Lexical (ALLEX) project, the Zimbabwean Ndebele corpus of 691,268 tokens was collected. The South Africa Norway Tertiary Education Development (SANTED) project, although it did not mention the size of the corpus collected, compiled a booklet with 150 computer science terms. Projects such as the Centre for Legal Terminology in African Languages (CLTAL) and the Special Language Corpora for African Languages (SPeLCAL) also did not mention the size of the corpus they collected.

Mpofu and Mangoya (2005) also collected corpora. They did not mention the size of corpora they collected either. Van Huyssteen (2003) collected isiZulu medical corpora of 3,010 tokens. This includes 2,187 written and 823 spoken tokens.

The researcher also reviewed literature on other languages which are not in the Nguni group. For instance, Prinsloo (2015) did a study on Sepedi, which also had limited corpora. This is a confirmation that it is possible to utilise small and unbalanced corpora for lexicographic purposes. Mabasa (2005) conducted a study the aim of which was to verify the Xitsonga medical and health terminology of the Department of Sports and Culture. This study is relevant

to the present one because the medical terms collected for the present study, also had to be verified, before they could be included in the glossary.

IsiNdebele as a Nguni language, is a language that has an agglutinating structure that presents some challenges. For instance, one isiNdebele orthographic word corresponds to one linguistic word. For example, the word *uyagula* 'he/she is sick' is classified as a verb in isiNdebele, whereas its English translation 'he/she is sick' consists of three orthographic words, each of which is also a linguistic word, belonging to a different word category.

The shortage of linguistic resources for isiNdebele implies a shortage of terminology in the language. Examination of both spoken and written isiNdebele texts shows a high degree of borrowing from Afrikaans and from two of the Sotho languages, namely Sesotho sa Leboa and Setswana. This is evident in the large collection of vocabulary which originates from these languages, for instance, the term 'hypodermic syringe' has been translated as *isipeyidi* in isiNdebele, a borrowing from the Afrikaans word 'spuit'. In the Nguni languages, to which isiNdebele belongs, the same noun features as *umjovo* in isiZulu and *sifutsamjovo* in Siswati (DAC, 2013a:49).

Regarding borrowing from Sotho languages, Mashiyane (2002:67) says that the two languages (isiNdebele and Sesotho sa Leboa) were once in close proximity. Because of this proximity, a relationship of free-giving and accepting developed. Mashiyane (ibid.) further maintains that in such a relationship, one language may accept more than the other one. In this case it is isiNdebele that freely accepted more items from Sesotho sa Leboa than vice versa. This is evident in the large collection of vocabulary which originates from Sesotho sa Leboa.

Pretorius and Bosch (2012) maintain that for an under-resourced language such as isiNdebele (ISO 639-3: ndl) to be developed, the basic tools and technologies for processing the language should be developed, be made accessible and available for future use. This means that even if the formats, platforms and software systems on which they were developed become outdated or be no more available, the resources themselves should remain available and accessible.

In South Africa, the national Department of Arts and Culture (DAC) is amongst the departments that play a significant role in the development of the eleven official languages of the country. DAC has, for instance, developed the following multilingual terminology lists and a dictionary. Multilingual financial terminology list (2017a), multilingual human, social,

economic and management sciences (2017b), multilingual natural sciences and technology terminology list (2013d), multilingual mathematics R-6 (2013c), multilingual ICT dictionary (2013b), multilingual parliamentary/political term list (2013e), multilingual HIV/AIDS terminology (2013a) and multilingual soccer terminology list (2013f).

Khumalo (2015:21-24) maintains that only few African languages have really developed corpora. This scarcity of corpora causes the development of African languages to be slower as compared to the development of European languages. This situation in African languages is aggravated by the fact that some of the corpora that have been developed for African languages are not easily accessible. Permission is always required before one can access them.

Moors, Wilken, Calteaux and Gumede (2018:297-299), did an audit regarding the state of Human Language Technology. The focus of their audit was on the following areas: An overview of text and speech resources, an increase in both the speech and text data as well as the increase in both text and speech software. According to their findings, there is an improvement in percentages regarding isiNdebele. In the data/text category for instance, isiNdebele showed an increase of 71%. This percentage is higher compared to languages such as Sesotho sa Leboa, isiZulu and the South African English and Afrikaans. Whilst there is a notable increase in text resources, there is none in the case of speech resources. As far as the soft-ware text is concerned, isiNdebele is at 92%, this being higher than English, Afrikaans, isiZulu and Sesotho sa Leboa. Moors et al. (ibid.) discovered that despite an improvement in percentages as indicated above regarding isiNdebele, this language is, however, still among the least maturely developed languages. Its resources are also least accessible to the end-users. This is an indication that although there are some developments as far as the isiNdebele texts and software category are concerned, isiNdebele is not developing, and is still under-resourced.

The researcher used questionnaires and interviews as well as semi-automatic term extraction methods in order to extract terms from texts, with the purpose of compiling a glossary of medical terms. The glossary will be made available on the Mpumalanga Department of Health and the South African Centre for Digital Language Resources (SADiLaR)' websites.

1.2 STATEMENT OF THE PROBLEM AND RESEARCH QUESTIONS

The dictionaries that are presently available in isiNdebele (of which there are four) are dictionaries for general purposes. This means that even if their development was corpus-driven, the corpora utilised were collected for general purposes. The first isiNdebele dictionary was compiled by Shabangu and Swanepoel (1989). The second isiNdebele dictionary was published by the isiNdebele National Lexicography Unit (2006). The third one was the monolingual dictionary published by the isiNdebele Lexicography Unit (2014) followed by the English-isiNdebele-Afrikaans trilingual dictionary, which was also published by the isiNdebele Lexicography Unit (2015).

Nkomo and Madiba (2011:150) state that 'most of the available dictionaries have major limitations and barely satisfy the people they are purposed to serve and because of that, glossaries are alternatives for dictionaries'. Owing to this finding, the development and the production of glossaries is justifiable because glossaries address the needs which dictionaries could not address, due to their capacity.

The current study therefore, aims to explore how corpora can be used as a basis for the compilation of a specialised English-isiNdebele glossary of medical terms. The study therefore addresses the following research questions:

- How does the paucity of standardised terminology in specialised subject fields negatively impact the development of isiNdebele?
- What role can corpus-driven lexicography play in filling the gap caused by the shortage of terminology in isiNdebele?
- How can existing methods of producing medical terminologies be transferred to isiNdebele where only sparse resources are available?
- How can a bilingual glossary be designed and represented in a user-friendly way in the case of isiNdebele, and how can such a glossary be made available to its intended users, inter alia, medical fraternity, translators of medical texts, lexicographers and other end-users?

1.3 AIM AND OBJECTIVES OF THE STUDY

The main aim of this study is to investigate how specialised corpora can be used in the compilation of an English-isiNdebele bilingual glossary of medical terms. This is in line with the above stated questions of the present research.

The **objectives** of this study therefore are:

- To create an English-isiNdebele Parallel Corpus of medical texts
- To extract terminology for glossary making, from both the monolingual corpus (spoken corpus of this study) and parallel corpus (written corpus of this study)

The researcher devised a record sheet with the purpose of recording information as she found it. The researcher was aware of available standards and devices of storing terminological data. She explored those standards with the purpose of choosing the ones relevant for this study. Storing her data on the Unisa's OneDrive, proved to be optimal as it was a cheaper and the safest way of storing the present study's data. There are however, limitations to the amount that one can store.

1.4 METHODOLOGY

The research data was collected using a variety of methods. Both the quantitative and the qualitative methods of collecting data were employed in this study. Dissertations, articles and papers were used also, as data sources.

Both open-ended questionnaires and semi-structured interviews were utilised as tools for data collection. The researcher used questionnaires in testing the acceptability of isiNdebele medical terms. Terms approved or suggested through the questionnaires were then included in the interview questions for verification by medical doctors. Two theories were applied in this study, namely the corpus-linguistics and the Grounded Theory.

1.5 THEORETICAL FRAMEWORK

The theoretical framework is significant in that it forms the basis of the research problem, it explains the path of a research and also grounds the research firmly in theoretical constructs. It is through the theoretical framework that the research findings are more meaningful and acceptable. It further stimulates research while ensuring the extension of knowledge. It does this by providing both direction and momentum to the research inquiry.

It is a framework based on an existing theory in a field of inquiry that is related and/or reflects the hypothesis of a study. The theoretical framework provides a general or broader set of ideas within which a study belongs. It is based on existing theory/theories in the literature which has been tested and validated by other scholars. It offers a focal point for approaching

the unknown research in a specific field of inquiry. The theoretical framework consists of theories that seem interrelated with their propositions deduced. It is used to predict and control the situations within the context of a research inquiry and also to test theories (Adom, Hussein and Agyem, 2018: 438 and 440).

Kerlinger (1986:9) defines a theory as a 'a set of interrelated constructs, definitions, and propositions that present a systematic view of phenomena by specifying relations among variables with the purpose of explaining and predicting phenomena'.

This study is about terminology development in a language that has a scarcity of terminology and resources. Through corpora, the researcher came to understand the medical terms used on a day to day basis in medical institutions. The study uses corpora for the compilation of an English-isiNdebele glossary of medical terms. Corpus linguistics is therefore relevant as the theoretical framework of this study.

There is an ongoing debate as to whether corpus linguistics is a theory or a methodology. In response to this debate, Kübler and Zinsmeister (2015:14) acknowledge that corpus linguistics is viewed by some linguists as a research tool or methodology and by others as a discipline or theory in its own right.

The researcher therefore, aligns herself with the linguists that refer to corpus linguistics as a theory. Through corpus linguistics, the researcher was able to explain observed evidence. Corpus linguistics further guided the researcher in highlighting relevant aspects that needed to be considered. It also guided the researcher in predicting appropriate steps in an analysis. Kübler and Zinsmeister(ibid.) further emphasise that corpus linguistics can be both the theory or a method, it depends on how it is applied.

In response to the debate whether corpus linguistics is a theory or methodology, McEnery, Xiao and Tono (2006:7-8) maintain that corpus linguistics is actually a whole system of methods and principles of how to apply corpora in language studies and teaching or learning. In other words, they are convinced that corpus linguistics should be considered a methodology with a wide range of applications. This means it can be applied across many areas and theories of linguistics.

Baker (2010:93) describes corpus linguistics as a popular field of linguistics which involves the analysis of very large collections of electronically stored texts, aided by computer software. The corpus linguistics theory enabled the researcher to get access to real-world examples

from the medical environment, rather than relying on hypothesis. This means that the English-isiNdebele glossary in this study was compiled from authentic terminology. In the present study, the researcher collected and analysed both the bilingual written corpora and the monolingual spoken medical corpora and they were manipulated easily by the WordSmith Tools.

Corpus linguistics displays salient features that were beneficial for the present study. Teurbet (2001:129) observes that corpus linguistics has combined three different approaches, namely, the identification of language data, the correlation of data by statistical methods and, finally, interpretation of results. This approach was applicable in all the phases of this research, such as selection of text types to be included in the corpus, analysis tools and the interpretation of corpus evidence.

Based on the background given above, corpus linguistics and a corpus-driven approach are highly relevant to this study. Corpus linguistics enabled the researcher to understand the usage of medical terms in their context. Through the corpus-driven approach, the researcher was able to use an underlying corpus as a source from which data was extracted. In addition, the Grounded Theory was also used in this study. Smith and Davies (2010:150) explain the Grounded Theory as 'a theory in which observations are constantly compared to each other and the emerging patterns are noted and coded'. The combination of corpus linguistics and the Grounded Theory is of great value for the present study. A detailed explanation regarding the two theories follows in Chapter 2.

This study is methodological in approach in that it explores how WordSmith Tools can be used to extract information from corpora for the compilation of a bilingual glossary.

1.6 DEFINING THE KEY TERMS AND CONCEPTS

It is necessary that terminology used is defined so that it is understood in the context of this study.

1.6.1 Corpus

Bowker and Pearson (2002:9) define the term 'corpus' as a 'body of text which comprises a large collection of authentic texts that have been gathered in an electronic form, according to a specific set of criteria'. 'Authentic' here means that a corpus is a sample of live, real language and that it consists of genuine communication between people going about their normal business. This means that the text occurs naturally. 'Authentic' further means that the text occurs naturally, it is not created for a specific purpose. A corpus is usually machine-readable, that is, the data is in the format that can be processed by a computer.

1.6.2 Corpus-driven approach

According to Tognini-Bonelli (2001:84) 'corpus-driven approach is an approach where the corpus is seen as more than a repository of examples to back pre-existing theories. In this approach, a corpus is seen as more than a probabilistic extension to an already well-defined system. The theoretical statements are fully consistent with, and reflect directly, the evidence provided by the corpus'.

1.6.3 The monolingual corpus

One or more languages may be covered by a corpus. A corpus is said to be monolingual when it contains texts in one language. Monolingual corpora are further divided by Kenny (2001) into two namely, the single translational as well as the non-translational. Non-translational corpora consist of texts that are original texts in one language and single translational corpora, on the other hand, consist of texts which are translated in one language only. The most familiar monolingual corpus that comprises English original texts which are not translations, is the British National Corpus (BNC).

1.6.4 An English-isiNdebele parallel corpus

The written corpus of this study is bilingual (it consists of two languages, namely English and isiNdebele) and parallel. Corpora are referred to as parallel when they comprise a collection of texts, each translated into one or more languages, other than the original. Parallel corpora, like any monolingual corpora, can either be bilingual (with an original text and its translated version), or multilingual (where corpora contain translations of several target languages of the same source language) (Tognini-Bonelli, 2001:6).

1.6.5 A specialised corpus

Bowker and Pearson (2002:22) emphasise that parallel corpora can either be general or specialised. They can also be written or spoken. They are specialised when focusing on a subject, exemplifying an aspect of a language.

To emphasise the difference between the general and language for special purpose (LSP) corpora, Baker (2009:241-242) claims that 'general and specialised corpora differ in terms of coverage, that is, the range of genres and domains they are supposed to represent. General corpora tend to cover as many genres and domains as possible to ensure maximum balance and representativeness. LSP corpora, on the other hand, tend to be composed of texts from a specific domain'.

1.6.6 The WordSmith Tools

Scott (2010) defines the WordSmith Tools as an integrated suite of programs used for corpus analyses. It includes three text analysis tools: a monolingual Concordance (Concord), and two-word list extractors (WordList and KeyWords). The Concordance gives one a chance to see any word or phrase in context so that one can see what sort of company it keeps. The WordList enables one to see in a text, a list of all the word clusters or just words in a text, which are set out in alphabetical or frequency order. Through the KeyWords function, one can see the key words in a text. Key words are defined by Prinsloo and Prinsloo (2011:101-102) as 'words that appear either as substantially more frequently than expected (positive keys) or substantially less than expected (negative keys) in an LSP corpus than in a comparable general corpus.'

Key words are those words whose occurrence is unusually frequent in comparison to other words in the corpus. Next to each key word, there are various numbers indicating how frequently each word appeared in the source text(s) and how that compares with its frequency in the reference corpus. Terminologists have discovered that the KeyWords function is helpful in extracting terms for the compilation of language for specific purpose (LSP) dictionaries (Scott, 2010:35).

1.6.7 Defining a glossary

A glossary according to Merriam-Webster's online dictionary (2021) is defined as follows:

1. a collection of textual glosses or of specialised terms with their meanings
2. a list that gives definitions of the hard or unusual words found in a book
3. a dictionary of the special terms in a field or job.

A glossary is a collection or list of terms in a special subject, field, or area of usage, with accompanying definitions. It only concerns itself with the terms that will enhance one's comprehension about a certain topic. It is an alphabetical list of terms in a domain of knowledge with the definitions for those terms. Traditionally, a glossary appears at the end of a book and includes terms within that book that are either newly introduced, uncommon, or specialised.

1.7 DATA COLLECTION TOOLS AND PROCEDURES

Data collected in this study comprises both the written and spoken corpora. Written corpora are made up of a pamphlet collected from a clinic, the internet and the SADiLaR website (2021). Spoken corpora are made up of data from Ikwewezi FM health program as well as from interviews with the doctors.

1.7.1 Data analysis and interpretation

After data collection, the next step that followed was data analysis. The parallel corpora collected were analysed through the WordSmith Tools. The WordSmith Tools has functions or programmes which were utilised in analysing the present research's corpora. The first function is the WordLister or WordList tool; this is the function that helped to create the alphabetical and frequency lists of words. This function assisted the researcher in determining consistent words. The second function is the KeyWords. Through this tool, key words used in the medical corpus were determined.

The other function the researcher used is the Concordance. This helped with the retrieving of all occurrences of a word as well as its immediate contexts. The Concordance also helped with the retrieving of multi-words.

It also needs to be mentioned that the manual aspect cannot be discarded in lexicography. The terms that were identified through this tool were verified manually.

1.7.2 Ethical consideration

Owing to the fact that the researcher had to collect data from hospitals and clinics, it was necessary to observe ethical requirements. The researcher was granted ethical clearance by The Research Ethics Review Committee of the Department of African Languages, University of South Africa who granted the researcher ethical clearance to proceed with the research study (See **Appendix A**).

A letter was sent to the Mpumalanga Department of Health requesting permission to access their health facilities (See **Appendix B**).

A letter of approval/permission was obtained from the Mpumalanga Department of Health (See **Appendix C**). This permission letter granted the researcher access to selected clinics and hospitals.

The researcher met face-to-face with the CEOs of both clinics and hospitals in Mpumalanga, Nkangala district. The purpose of the meetings was to verbally explain the researcher's intentions for visiting health facilities (clinics and hospitals). All participants signed participation and consent forms. They were also informed that their participation was voluntary, they could withdraw at any time they felt like withdrawing (See **Appendix D**).

1.8 STRUCTURE OF THE THESIS

This thesis is divided into six chapters:

CHAPTER 1: INTRODUCTION

This chapter presents the background to the proposed study. It introduces the study by outlining the historical background of isiNdebele (ISO 639-3: ndl), as well as the challenge posed by the scarcity of resources. The chapter also gives the reasons and rationale for undertaking the study. It outlines the aim, the research problem, the objectives, the theoretical framework and the methodological approaches used in this study. Finally, key concepts are defined.

CHAPTER 2: LITERATURE REVIEW

This chapter presents an overview of the corpus-driven approach (CDA), followed by theoretical frameworks that underpin this research, namely corpus linguistics and the Grounded Theory. Methodologies on which these theories are based, are also discussed. The chapter further discusses the development of isiNdebele terminology and the challenges experienced in terminology development. It also reviews projects undertaken by different linguists and terminologists.

CHAPTER 3: RESEARCH METHODOLOGY

This chapter discusses the approach the researcher followed in collecting and analysing data for this study. This includes methods and instruments used for data collection. It describes the compilation and the analyses of both the written and spoken corpus and describes the theories and tools for term extraction and analysis. All these are explained with reference to the aim of this study, which is the compilation of an English-isiNdebele glossary of medical terms.

CHAPTER 4: DATA COLLECTION AND ANALYSIS

This chapter presents the interpretation and the discussion of data collected only through questionnaires. It also presents term-formation processes applicable in the formation of isiNdebele medical terms of this study. The researcher compares the word-formation processes of this study and that of the projects discussed in Chapter 2. The development of isiNdebele terminology and the challenges experienced are also discussed in this chapter.

CHAPTER 5: DATA ANALYSIS AND CORPUS ANALYSIS

This chapter discusses the design, the compilation as well as the analysis of both the written and the spoken corpora. The analysis of corpora is done semi-automatically. It is done with the purpose of answering the question on the role corpus-driven lexicography can play in filling the gap caused by the shortage of terminology and the question on ways in which existing methods of producing terminologies can be transferred to isiNdebele where only sparse resources are available.

CHAPTER 6: FINDINGS AND INTERPRETATIONS

This chapter summarises, interprets and concludes the findings of the study. It also presents the recommendations for further research.

CHAPTER 2

LITERATURE REVIEW

2.1 INTRODUCTION

This chapter presents an overview of the corpus-driven approach (CDA) followed by theoretical frameworks that underpin this research, namely Corpus linguistics and the Grounded Theory. Methodologies on which these theories are based, are discussed. The development of isiNdebele terminology and the challenges experienced are also discussed in this chapter. At the end of the chapter, projects undertaken by different linguists and terminologists are reviewed.

As mentioned in Chapter 1, language is one of the basic rights of human beings. All languages therefore, need to be promoted and developed so that they are at the same level as those languages that have already reached modernity. To achieve this, the technical terminology of the indigenous, under-developed languages will have to be developed.

It is in response to this need for development that this study explores the use of corpora in the compilation of a specialised English-isiNdebele glossary of medical terms. The literary works consulted during the study are able to illustrate how corpora are used in the compilation of lexicographic tools such as glossaries, dictionaries and terminology lists.

The corpus-driven approach (CDA) is an approach in which every piece of information is drawn from the corpus, without pre-conceived ideas. This approach does not use carefully selected examples from the corpus just to support pre-conceived theories, but it seeks to formulate hypotheses and theoretical positions based on the evidence found (Hanks, 2012:417).

The corpus-driven approach emphasises that corpus is an only source of the hypotheses about language. According to this approach, it is only from the analyses of the corpus that linguistic constructs can emerge. A corpus is seen in this approach as more than a repository of examples to back pre-existing theories and systems. This means that corpus is not used to find examples that fit new or pre-existing entries. Theoretical statements directly reflect and are also fully consistent with the evidence provided by the corpus, since evidence plays a significant role in this approach. Theories are built step by step in the presence of the evidence. Examples are normally taken verbatim. The observation of certain patterns leads to a hypothesis. The hypothesis in turn leads to generalisations in terms of rules of usage and findings, on the unification within a theoretical statement (Tognini-Bonelli, 2001:84-85).

Baker (2010:95) confirms Tognini-Bonelli's (ibid.) statement by mentioning that a corpus is used by corpus-driven linguists in an inductive way. They use the corpus to form hypotheses about language, without referring to linguistic frameworks that are existing.

The use of corpora in developed languages such as English and Afrikaans is at an advanced stage. This is due to the availability of rich, accessible resources such as term-banks, a variety of dictionaries, glossaries, etcetera, in those languages. In African languages, however, such resources are scarce. In the languages where they are available, they are at an infancy stage.

2.2 THEORETICAL OVERVIEW

Theories that underpin the present study are discussed below, namely the corpus linguistics theory and the Grounded Theory.

2.2.1 Corpus linguistics theory

Bowker and Pearson (2002:9) regard corpus linguistics as a relevant approach for studying language.

Although many scholars have contributed to the development of modern-day corpus linguistics, John Sinclair is considered the most influential scholar in this field. He observes that a word on its own does not carry meaning, meaning only comes through several words in sequence. This means that the meaning of the word is determined by the context in which it occurs. This hypothesis forms the backbone of corpus linguistics.

John Sinclair believes that a description should be 'driven' by the corpus. In order to fulfill this, he, on several cases attempted to utilise the emergent computer technologies for the purpose of automating the initial stages of analysis and also to defer the intuitive, interpretative contributions of linguists to increasingly later stages in the process. John Sinclair's model of corpus-driven lexicography has spread to English as a Foreign Language (EFL) dictionaries, native-speaker dictionaries (e.g. the New Oxford Dictionary of English, and many national language dictionaries in emerging or re-emerging speech communities) and bilingual dictionaries (e.g. Collins, Oxford-Hachette). This shows that it has spread far beyond its initial implementation at Cobuild (Krishnamurthy, 2008:1).

Sinclair maintained that the entry point to phraseological meaning, is a lexical item. He saw that the meaning of utterances is flexible and probabilistic, often arising out of a dynamic

interpretation of collocations and not just a list of dictionary 'meanings'. He discovered this at a time when linguists were using theories of speculation, and their focus was not on meaning and usage but on logical form of sentences that were idealised. He was unimpressed by the fact that linguists were not interested in looking for what was going on in a sentence or in a word. Instead they were using corpora as 'fish ponds' to fish examples from, just for supporting their theories.

At the centre of the term corpus linguistics, is the term corpus (which has been defined in Chapter 1). It is however important to mention that nowadays the term corpus means a collection of texts held in electronic form, capable of being analysed automatically or semi-automatically rather than manually only. It contains both spoken and written texts (Kruger, 2002:71).

To elaborate on what the above linguists said, Bennett (2010) defines a corpus as a principled collection of authentic texts stored electronically. Bennett (ibid.) therefore suggests that when creating a corpus, the focus should be on the following three factors:

- **Principle:** This means that there are explicit linguistic criteria used to select and order corpora. The collection of corpora is governed by particular principles or criteria. It is the intended use of the corpus that determines the criterion or principle to be used. This confirms the fact that a corpus is not a random selection of texts, but a selection of texts based on particular principles (McEnery, Xiao and Tono, 2006:13).
- **Specific characteristics:** The language comprising the corpus must be chosen according to specific characteristics.
- **Authentic texts:** The term authentic here means that the text chosen is an example of a real, live language and that it consists of genuine communication between people going about their normal business. The term authentic further means that the text occurs naturally. It has not been created for a particular purpose and it can be computer-processed.

A corpus must therefore include authentic texts. The main idea behind the authenticity of the texts is that the language it contains must not be made up for the sole purpose of creating the corpus (Bennett, 2010:3).

There have been developments in the manner of storage, design and analysis of the corpus. The term corpus has therefore expanded to include these developments. In modern linguistics, the term corpus has become almost synonymous with the term 'machine-readable'. The term 'machine readable' refers to how easy or difficult it is for the machine to read a text. It also refers to how visible the text is, for the machine to be able to read it. McEnery and Wilson (2001:17) say that it is only when the content of the document can be readily processed by a computer that the document can be said to be 'machine-readable'. This is when the computer program can read and interpret data without the need for manual, human intervention. All the above linguists emphasise that texts should be 'real'; this means that texts should not be created for the purpose of a particular research. Texts to be used should also be 'authentic' and 'natural'. This means that they should consist of genuine communication between people under real circumstances.

After having looked at the definition of the term corpus, it is also important to look into its classifications. Corpora may either be written or spoken. Written corpora are based on formal written texts such as newspapers, school books, etcetera. They can also be based on fiction. Written language corpora have already been collected by certain institutions for some indigenous languages, including isiNdebele. The South African Centre for Digital Language Resources (SADiLaR, 2021) has for instance collected both isiNdebele language for special purposes (LSP) and language for general purposes (LGP). The LSP obtained from the SADiLaR web page only, is 2,068,648 tokens and the LGP (from SADiLaR only), is 1,186,015 tokens. <https://repo.sadilar.org/handle/20.500.12185/272>.

This study is about language for special purposes (LSP). Khumalo (2015:498), when emphasising the importance of LSP corpora, maintains that these kinds of corpora comprise of texts that have been written by subject field experts. There will also be evidence regarding expressions, functions and usage of words in the LSP.

Spoken corpora are also important. It is important to mention at this stage that authors such as Van Huyssteen (2003) and others refer to spoken corpora as oral corpora, the term spoken corpora will however be used in this study.

Spoken corpora are corpora that consist of transcripts of spoken material. Spoken materials include conversations such as radio and television broadcasts, interviews, etcetera. Spoken corpora are discourses that can only be listened to and not read. Emphasising the importance of spoken corpora, Allwood and Hendrikse (2003:190) say that it is through spoken corpora that language used in real life can be observed. Spoken corpora present varieties of languages

with regard to the vocabulary of those languages. This includes the pronunciation of words, the grammar of various languages as well as their communicative function. To add on what Allwood and Hendrikse (*ibid.*) say, Newman (2008:29) maintains that these kinds of corpora have become a significant genre. They represent day to day basic human communication and also present language as used in speech communities.

Spoken language corpora are relatively unexplored whilst it is actually spoken corpora (as compared to their written counterpart) that contribute towards the promotion of the languages' status. Amongst other reasons why spoken corpora are unexplored is that their compilation is expensive and also demanding. That is why they are so scarce and as such small. Spoken language corpora need to be explored because for many languages, they are the dominant modalities (a system of linguistic options), and the only modalities for certain genres for instance, activities involving traditional knowledge. SADiLaR (2021) has also collected spoken corpora for isiNdebele. The file size of isiNdebele spoken corpora on the SADiLaR website (2021) is 4.3GB.

After the collection of spoken corpora, transcription follows. Bailey (2008:128) and Davidson (2009:38) maintain that transcription is an important first step in data analysis. Transcription is the transformation of sound or image from recordings into text. It involves close observation of data through repeated, careful listening and/or watching. Loubere (2017:1) says that transcription is a necessary and an important methodological tool, to the researchers focusing on discourse. Transcription plays an important role in qualitative research and improves the thoroughness of qualitative research.

Types of transcription

A distinction is made between various types of transcription.

Oliver, Serovich and Mason (2005:1273) confirm that transcription is central in qualitative data collection. It is a pivotal aspect of qualitative inquiry. Oliver et al. (*ibid.*) maintain that transcription should be an object of study in its own right and not be viewed as a behind-the-scenes aspect of data management, as is usually the case.

Transcription practices can be thought of in terms of a continuum with two dominant modes, namely naturalism and denaturalism.

- Naturalism transcription

Bucholtz (2000:1461) describes this transcription type as a word-for-word transcription. It is also known as verbatim transcription. In this transcription type, the verbal and non-verbal components of the discussion are transcribed. This means that all the details omitted in the denaturalism or intelligent transcription, are retained in naturalism or verbatim transcription. In this transcription type (naturalism), features of oral language such as “ums” and “ers” are preserved. The more a transcript preserves or retains the features of spoken language, ‘the less transparent it becomes for readers unaccustomed to encountering oral features in written texts’. Readers used to denaturalised transcription practices might find the transcripts odd looking and difficult to read.

- Denaturalism transcription

This type of transcription is also referred to as Intelligent verbatim or smart transcription. This is the type of transcription in which idiosyncratic elements of speech such as stutter and pauses are removed. Things such as fillers and purposeless repetitions are deliberately excluded. They are excluded because they may distract one from focusing on the content of the speech.

Recorded speech is converted into text format and any redundant phrases and words are cut out. The focus in this transcription is to capture the meaning of what was said and not how it was said. How it was said is not essential. The result of this transcription is that it is straightforward, and the final written product reads intelligently. What is important to note is that regardless of the fillers that are excluded, the essence and the whole idea of the text is still maintained. The removal of the fillers does not in any way influence the content of the recordings. It does not make the content to be unnatural because sentences are not altered or edited. Although softwares for transcriptions are now available as compared to previous times, it must be noted that such softwares are not available for a language such as isiNdebele. This type of transcription is however time-consuming, this is because the transcriber must be able to differentiate between what is important and what is not important in the audio or video file. This includes cases such as laughter, mannerisms and telephone ringing.

Davidson (2009:38) maintains that transcription is selective, irrespective of its nature or type, and emphasises that it is impossible to transcribe all features of talk and interaction from recordings. Selectivity needs to be acknowledged and explained in relation to the goals of a study. This is applicable in this study because the researcher transcribed the recordings

herself and she has acknowledged that fillers, purposeless repetitions and telephone ringing were for instance not transcribed. This is because they would distract the reader from focusing on the content of the speech.

Duranti (2007:302) maintains that naturally, transcription is a selective process. It relies on conventions and it reflects theoretical goals and definitions. Oliver et al. (2005) maintain that there is a relationship between the Grounded Theory and the denaturalised transcription style. This means that the Grounded Theory by its nature, channels one to employ the denaturalised transcription.

Having described the two transcription types, it needs to be mentioned that Davidson (2009:38) discovered another type of transcription that he (Davidson, *ibid.*), refers to as the literacised transcription.

- Literacised transcription

This is the type of transcription in which spoken utterances are taken and written down as if they were written. In this type of transcription, everything that makes a 'spoken' corpus gets removed. The spoken corpus therefore becomes literacised. The language appearing in it becomes identical to written language.

For this study, the denaturalised-literacised transcription was implemented. It is denaturalised because fillers were removed; it is on the other hand literacised because the spoken corpus then became a written corpus. The word 'spoken' in this study is therefore used to refer to the spoken material that was literacised.

Owing to the fact that transcription is a practice central to qualitative research, it is therefore important to understand what qualitative research is. Qualitative research is the research type that requires the collection of qualitative data in order to answer the research questions. This research type is expressed in words. It is used to understand concepts, thoughts or experiences. Qualitative research enables one to gather in-depth insights on topics that are not well understood. It is therefore suitable for research which is exploratory in nature. It enables one to gain an in-depth understanding of individual experiences, opinions, trends and also to dig deeper into the problem at hand. The perception and the beliefs of participants about a preference is cleared through the qualitative research tool (Greeff, 2005:296).

Interviews with open-ended questions are common methods for qualitative research. In open-ended questions, participants are to choose between the given responses and also qualify their choices or express a more complex or subtle meaning (Siniscalco and Auriat, 2005:23-24).

On the other hand, quantitative research is the research type that uses mathematical models and statistics for analysis, providing numerical results. This research type is expressed in numbers and graphs. It is therefore said to be more objective as compared to its qualitative counterpart. It is used to test or confirm theories and assumptions. This type of research can be used to establish generalisable facts about a topic. It is used to quantify opinions, attitudes, behaviours, and other defined variables with the goal to support or refute hypotheses about a specific phenomenon, and potentially contextualise the results from the study sample in a wider population or specific groups (McMillan and Schumacher, 2006:23).

A study may use both qualitative and quantitative methods either concurrently or sequentially. Such a research type is called a mixed methods research. The advantage of using both methods (qualitative and quantitative) is that one method compensates for weaknesses of the other method and the results are good, than in research that uses only one method. A mixed methods study could lead to better and more accurate inferences because of the depth of review. There may be the discovery of conflicting or divergent findings (Moore, 2016:2-3).

A distinction is further made between generalised and specialised corpora. General corpora are ones that aim to be representative of a particular language, such as The British National Corpora. Corpora of this type include various types of texts, both written or spoken (and transcribed), on a variety of subjects. A corpus is specialised when it focuses on a particular subject, for example, when it exemplifies a particular aspect of a language (Bowker and Pearson, 2002:22).

According to Bennett (2010:13), specialised corpora can be large or small and are often compiled to answer very specific questions. Examples of specialised corpora include:

- Michigan Corpus of Academic Spoken Language (MICASE), which contains only spoken language;
- CHILDES corpus, which contains language used by children;
- Michigan Corpus of Upper-Level Student Papers, a collection of papers from a range of University disciplines, and a medical corpus containing language used by nurses and hospital staff (MICUSP).

In discussing the classification of different types of corpora we also need to look at the language or languages in which a corpus is encoded. It is called a *monolingual corpus* when it contains texts in a single language and a *bilingual corpus* when it contains text data in two languages. When it contains multiple languages, it is called a *multilingual corpus*.

A distinction is also made between comparable and parallel corpora. In the case of comparable corpora, there are sets of texts in different languages. The choice of texts is determined by some features or characteristics that they have in common; the main feature that distinguishes one set of texts from another is the language in which the texts are written. Parallel corpora on the other hand, comprise of texts and their translations into one or more languages. Parallel corpora can be bilingual; this is when they comprise of original texts and their translated versions in another language. They can also be multilingual. It needs to be emphasised that a multilingual corpus, may also be comparable even if it does not contain translations (Bowker and Pearson, 2002:92-93). In order to facilitate comparison between the two texts, original texts are paralleled to their translated versions at paragraph, sentence and word levels. This is done through a semi-automatic alignment facility. Olohan (2004:25-26) describes alignment as 'linking a unit of text in one language with a unit of text in another language'.

It is necessary when working with a corpus to know how it will be queried and interpreted semi-automatically. Taljard and De Schryver (2002:46) explain semi-automatic term extraction as a process whereby the computer software is used to detect and extract potential terms from electronic corpora. Owing to the fact that the issue of human intervention in term extraction cannot be ruled out, human beings will always have to decide if what has been picked up by the software is really a term or not. Humans are therefore, final arbiters.

The process is said to be semi-automatic when both the computer tool and human being are utilised. Taljard and De Schryver (2002:58) as well as Bowker and Pearson (2002:145) emphasise that in terminological practice, the human factor cannot be discarded. In any terminological activity, human beings will always remain the final judges irrespective of whether the endeavour is manual or computational.

Taljard and De Schryver (ibid.) did a research on the feasibility of semi-automatic term extraction in African languages. The research was motivated by the fact that texts of a technical nature were beginning to appear on the internet. They explained factors that have an impact on the compilation of corpora. At the time of their research, African languages were historically so disadvantaged that subject specific texts to be used in building electronic databases, were not available. They also discovered that terminologists from African

languages had little or no access to special field texts. To evaluate the success rate of semi-automatic term extraction, Northern Sotho was used as their pilot. The manual term excerption was compared with the computational extraction of terms. The manual success rate served as a bench-mark for the computational success rate. The results of computational processing were compared to the results of manual excerption.

To realise the aim of their research, the WordSmith Tools was utilised. The WordList, KeyWords and Concord functions of the WordSmith Tools were applied. Through the WordList, they discovered that only the 20 items appearing in the top 100 of manually excerpted terms could be regarded as terms. This means that, out of the 335 lemmatised terms, 285 or 85% were excerpted manually, and 237 or 71% were extracted computationally. The difference between the two approaches (which is 14%) is smaller than the number of items not retrieved in either approach. When looking at the end product, it was discovered that extracting terminology using the semi-automatic approach is indeed feasible for and in the African languages. The researcher is yet to prove this point in the current study.

2.2.1.1 Defining corpus linguistics

Bowker and Pearson (2002:9) define corpus linguistics as a theory for studying language use. Corpus linguistics is empirical, as it involves the study of samples of what people have actually said, rather than hypothesising about what they might or should say. Litosseliti (2010) confirms Bowker and Pearsons' (2002) definition of corpus linguistics by saying that corpus linguistics is about studying real-world language use.

Baker (2010:93) also emphasises that 'corpus linguistics' is firmly rooted in empirical and inductive forms of analysis, relying on real-world instances of language use. There are no models of made-up examples or introspection in corpus linguistics. Corpus linguistics involves analysing large collections of texts which are stored electronically. This is done by the computer software.

Corpus linguistics is firmly based on the integration of four interdependent elements, namely data, description, theory and methodology. Data plays a significant role in every research, it is the driving force of research. Without data, there is no research. In terms of description, it means that language is organised systematically in new descriptions of language behaviour. Through the theory, concepts and language models are created with the purpose of explaining the phenomena which are empirically observed. Hypotheses are put forward for further testing.

To examine and process a corpus, a specific methodology is used. These elements are interrelated for the process of creation, evaluation and testing of the corpus (Laviosa, 2002:8). In the present study, the researcher collected medical corpora and analysed them using the WordSmith Tools. There are both advantages and disadvantages of corpus linguistics as discussed below.

2.2.1.2 Advantages of corpus linguistics

McEnery and Wilson (2001:103) state that corpus linguistics makes the analysis of corpora more objective. McEnery et al. (2006:7) confirm this by saying that corpus linguistics is not restricted to a particular aspect of language. Khumalo (2007:53-54) also agrees with this point by saying that corpus linguistics may be used in almost any area of linguistic research. McEnery et al. (ibid.) maintain that through the corpus, the linguist can discover what intuition alone cannot find. Baker (2010:94) says that corpus linguistics enables researchers to quantify linguistic patterns and also to reach more solid conclusions.

2.2.1.3 Disadvantages of corpus linguistics

Owing to the fact that language is dynamic, it is not possible to collect all the evidence needed through corpus linguistics. Another fact is that corpora are unable to filter mistakes and errors present in speech; it then becomes the responsibility of the researcher to carefully select evidence appropriate for the study. The amount of evidence that is available is limited although it does not mean that it actually does not exist. This makes the use of corpora less valid for research purposes (Andersen, 2010:555). McEnery et al. (2006:7) argue that not all research questions can be addressed by the corpus. It needs to be emphasised that although there are disadvantages, the advantages outweigh the disadvantages.

2.2.1.4 Computers and corpus linguistics

The major features of corpus linguistics are computer technology, corpora and analysis tools. Computers have played significant roles in corpus linguistics because they have enabled more complex forms of analysis, that were previously only manually possible.

Exploration of text is only possible if one is using a computer. In this study the automated ways were not fully applied when extracting terminology for questionnaires. The researcher identified term candidates manually. The core vocabulary of a specialised domain is identified through Term Extraction (TE). Terminologists who list potential Term Candidates (TC) apply

the Manual Term Extraction (MTE). The purpose is to consult with a domain expert in order to arrive at a final list of validated terms. The Manual Term Extraction is a labour-intensive enterprise. In the long run, Automatic Term Extraction (ATE) might replace MTE completely (Foo, 2012:16).

Computers facilitate the study of registers, author style, dialects, lexico-grammatical aspects of a language and regional differences. The results of these empirical investigations are not only beneficial to lexicographers and linguists, but also to teachers, users of dictionaries, translators, language students, second language learners, amongst others.

McEnery and Wilson (2001:18) mention that computers are important because they have the ability to store, search for, retrieve, sort and calculate linguistic data. This can be done in a quicker and more reliable manner than was previously possible, that is, when this was done manually.

Computers are utilised to compile corpora for linguistic research purposes (Van Huyssteen, 2003:134). Through computers, it is possible to identify and analyse complex patterns of language use. They also make it possible to store and analyse larger databases of natural language than could be dealt with by hand. Emphasising the significance of computers, Moropa (2007:183) states that computer tools have contributed to the compilation and in turn the development of technical terminology of isiXhosa. They have provided constant and reliable analyses.

The beginning of modern corpus linguistics was marked by computers. This also contributed to a change in the way corpora were viewed. According to Bennett (2010:2), computers have played a very important role in the creation of modern-day corpora. However, the beginning of modern-day corpus linguistics, also known as the second (or computational) era of the discipline, goes back to the early 1960s when the 'first generation' of one-million-word computer-readable corpora were first compiled.

To emphasise the role played by computers, Bosch and Pretorius (2011:138) state that towards the end of the previous millennium, corpus linguistics and electronic corpora started to increase in relevance and importance in the case of language studies and linguistic research. This is because of technological advances that were experienced during this time.

The first computer-based corpus, the Brown corpus, was created in 1961 which comprised of one million words. It was regarded as a reference corpus. The Brown corpus consisted of a

written corpus of press reports, fiction and government documents. A counterpart to the Brown corpus, is the Lancaster-Oslo-Bergen (LOB) corpus which was compiled in Britain. It also consisted of one million words of British-English texts. These early attempts at corpus design provided a foundation for later researches in corpus studies. The mid-1970s to 1980s saw the compilation of the 500,000-word London-Lund corpus, which was a spoken component of the Survey on English Usage. The spoken English corpora comprised of various genres, conversations and radio broadcasts. In their day, such corpora were considered as large-scale, but by today's standards they would be judged as relatively small-scale (Bowker and Pearson, 2002:48).

The current study uses both corpus linguistics theory as well as the Grounded Theory. There is a relationship between the two theories. Corpus linguistics studies real examples of what people have said. The Grounded Theory on the other hand, through its method of agreement and the method of difference, enables the researcher to determine what is common and uncommon in what has been said (through corpus linguistics). In this study, it is data collected through questionnaires which was analysed through the Grounded Theory. This theory enabled the researcher to establish agreements and disagreements about an isiNdebele term.

2.2.2 The Grounded Theory

In this section the definition of the term Grounded Theory, its advantages and disadvantages are discussed.

2.2.2.1 Defining the Grounded Theory

Grounded Theory is defined by Charmaz (2000:509) as 'systematic guidelines for collecting and analysing data to build middle-range theoretical frameworks that explain the collected data'. Smith and Davies (2010:150), on the other hand, define the Grounded Theory as the theory in which observations are constantly compared to each other and any emerging patterns are noted and coded. Categories are then generated.

The researcher is aware of the debates on whether Grounded Theory is a theory or a method. Neuman (1997:428-429) refers to it as the theory of analytic comparison. This is because this theory is made up of the method of agreement as well as the method of disagreement(difference). The method of agreement focuses a researcher's attention on what is common across cases – the common terms in the case of the present study. The researcher is able to locate cases that are similar in many respects yet differ in a few crucial ways, through the method of difference.

Coding is an important aspect of the Grounded Theory. It is the categorisation of data. Smith and Davies (2010:155) reason that coding does not constitute the totality of data analysis, but it is a method of organising data so that underlying messages portrayed by the data may become clearer to the researcher. Coding shapes an analytic frame from which the analysis is built. It also serves as a pivotal link between the collection and the explanation of the meaning of data. Coding is therefore an interpretive activity. Through it, various codes are attributed to the same data by various researchers. A 'code' can be a word or a short phrase that represents a theme or an idea (Charmaz, 2006:45-46).

Three coding techniques are distinguished in the Grounded Theory, namely the open coding, the axial coding as well as the selective or theoretical coding (Strauss and Corbin, 1996:40).

- *Open coding:* According to Goulding (1999), it is in open coding that raw data is organised by breaking it down into distinct parts. The researcher then examines these parts closely and compares them for similarities and differences. Open coding is about labelling and categorising data. Open coding is regarded by some scholars as a process in which data is broken down into separate units of meaning. Open coding aims at conceptualising and labelling of data. Open coding will be used in this study as the first stage of interpreting data. It is in this stage that all written data from questionnaires will be *Axial coding:* This is where data that was split during the open coding gets re-assembled, it gets put into categories. Corbin and Strauss (1990:13) explain axial coding as the process in which sub-categories are related to a category. It is a process in which elements that are linked to each other, are examined. Connections or links that were identified between categories or sub-categories in the open coding, are drawn. Neuman (1997:443) explains axial coding as the process of putting together data into categories after it has been fractured through open coding. Creswell (1998) says that the purpose of axial coding is to sort, synthesise and organise large amounts of data and re-assemble them in new ways.

- *Theoretical or selective coding*: This relates to all categories and sub-categories. It addresses the *how* and *why* questions, to explain the phenomena. When no further new codes or categories can be identified, then *theoretical saturation* is reached (Theron, 2015:5-6).

2.2.2.2 *Advantages of the Grounded Theory*

The Grounded Theory avoids making assumptions and instead adopts a more neutral view of human action, in a social context. This theory provides a methodology to develop an understanding of social phenomena that is not pre-formed or pre-theoretically developed with existing theories and paradigms. The Grounded Theory is exploratory in nature. It is then suitable for investigating social processes that have attracted little, prior research attention.

The Grounded Theory can identify the situated nature of knowledge, as well as the contingent nature of practice. It produces a description that acknowledges areas of conflict and contradiction. This theory is better at determining what happens. As a general theory, the Grounded Theory adapts readily to studies of diverse phenomena. It can respond and change as conditions that affect behaviour, change.

2.2.2.3 *Disadvantages of the Grounded Theory*

The Grounded Theory fails to recognise the embeddedness of the researcher and thus obscures the researcher's considerable agency in data construction and interpretation. This theory tends to produce large amounts of data, often difficult to manage. Researchers therefore need to be skillful in using this theory. There are no standard rules to follow for the identification of categories as yet.

2.3 METHODOLOGIES BASED ON THE THEORIES DISCUSSED

Methodologies include LSP corpus design and the compilation of corpora for terminology, glossary, projects based on the theories discussed, an overview of glossary compilation in African languages, studies conducted on compilation of medical terms as well as copyright and permissions.

2.3.1 LSP Corpus design

The aim of the present study is to compile an English-isiNdebele glossary of medical terms. This means that a special corpus will have to be compiled. According to Bowker and Pearson (2002:45), there are matters one needs to pay attention to when designing an LSP corpus. The matters include size, number of texts, medium, subject, text type, authorship, language as well as the date of publication.

About the size of the corpus, Litosseliti (2010:95) maintains that it is possible to carry out corpus analysis on very small texts. The corpus size should be big enough to reveal something about frequencies of certain linguistic phenomena, enabling researchers to examine what is typical and what is rare in a language. There are no hard and fast rules regarding how large a corpus ought to be, instead size is dictated by several criteria. One of those criteria concerns the purpose for which the corpus is intended. Litosseliti (2010:96) says that there may be more pragmatic reasons for building a corpus of a small size. It mainly depends on the availability of texts and how much money or time must be devoted to a project.

The purpose for which the corpus is compiled also determines the method to be followed when designing corpora. The aim of this study is to design a parallel, LSP corpus. When designing a parallel LSP-corpus, several issues must also be taken into consideration. The issues include: the medium (written or spoken text or both), the text type (fiction, non-fiction, scientific/technical writing, newspaper articles, etc.) and time (period of production or publication of the text).

2.3.2 Compilation of corpora for terminology and glossary

In this section a distinction will firstly be made between terms and terminology. Methods used for compiling terminology and the types of compilation, will also be discussed in this section.

2.3.2.1 *Defining 'term'*

Sager (1990:76) defines a term as the linguistic representation of a mental construct and also a lexical element used in specialised fields (subjects or their branches). This means that terms are generated from specialised fields or modified from elements already existing in other fields. A term is created when various linguistic labels are used to describe or name a specific object or concept. Therefore, a term refers to a definite concept which is clearly defined within specific parameters.

Alberts (2017:65 and 70) also confirms that a term refers to a linguistic representation of a mental concept and also that the concept term can be a single or compound word, phrase, collocation as well as acronym or abbreviation. Terms are formed after a new concept has been created which means that a new concept necessitated the creation of a new term. A term should have only one meaning and should not have emotional connotations attached to it. When a term has more than one meaning it loses its term-hood and becomes a mere word and therefore forms part of the general vocabulary.

Sager (1990:57) further explains the special interrelationship between the symbol, the concept (that is, its mental representation in one's brain) and the various linguistic labels used in different languages to describe the object and concept. He emphasises that there should be a relationship between the object and the concept. If the relationship does not exist between the two, a misunderstanding or miscommunication will result. A term should conform to the morphological, spelling and pronunciation rules of the receiving language.

2.3.2.2 *Defining 'terminology'*

Terminology is 'a representation of an equally coherent, but possibly differently structured system of concepts' (Sager, 1990:114). Terminology is further defined as 'the study and use of the systems of symbols and linguistic signs employed for human communication in specialised areas of knowledge and activities' (Sager, 1990:4). Terminology ensures a meaningful, effective communication between different language groups (for example, specialists and lay persons).

Alberts (2017:64), on the other hand, defines terminology as the collection of terms of a specific subject field, discipline or domain. Terminology forms part of an LSP of a particular subject field. It is the vocabulary of a domain, it plays a pivotal role in language development and also in the promotion of multilingualism (Alberts, 2017:148).

If the correct terminology is used, the technical and scientific communication skills of all citizens will be developed. Terminology must have characteristics such as specialisation, technicality, single acceptance as well as conciseness (Alberts, 2017:64 and 67). Alberts (2014:19) explains terminology as a polysemic word referring to three different aspects related to compilation, description and presentation of terms.

- **Compilation of terms** refers to the preparation of lists with terms belonging to a certain subject, according to a previously established methodology.

- **Description of terms** refers to the definition of each term (semantic focus) and the description of elements composing the term and its generation process (morphological focus).
- **Presentation of terms** refers to the preparation of dictionaries and glossaries. The researcher in the present study followed the processes mentioned above. The researcher started with the compilation of terms, followed by the description (this was done during the analyses of questionnaires and also through the usage of WordSmith Tools). Finally, a glossary of English-isiNdebele medical terms was compiled.

(a) Methods of compiling terminology

As seen in the previous paragraph, Alberts (ibid.) explains compilation of terms as the preparation of lists with terms belonging to a certain subject, according to a previously established methodology. The move towards automation is of overwhelming importance. Techniques and methods are based on the assumption that now and in future, all aspects of terminology compilation are being assisted or directly carried out by a computer. Principles of compilation are different now from the ones that were used in pre-automation days. Machines have influenced the way compilation was done initially.

It is, however, important to note that automatic terminology extraction tools are never perfect. Through them, issues of 'noise' and 'silence' are experienced. 'Noise' refers to invalid term candidates whereas 'silence' refers to the missing legitimate term candidates. Clean up work is therefore required to be able to identify 'noise' and 'silence' before converting the texts to txt. and also, before using a particular software.

(b) Types of compilation

According to Otlogetswe (2007:122), there is a variety of terminology compilation methods. The variety is brought by the fact that some compilers prioritise quality over quantity and vice versa.

Nowadays corpus plays a significant role in systematic terminology compilation. Terminology is no longer extracted from previous lists or by individual searches but from a corpus of texts. The compilation of terminology is an ongoing and repeated activity. The principles of compilation are affected by the following.

- the raw data as found in the corpus;

- the database which contains all the information that is collected in suitable structured form;
- the various sub-sets of information which are created for specific purposes.

Decisions must therefore be made on the nature of the texts: Where the texts are stored and the format in which they are stored, the method used to extract them (how to go from texts to terms) as well as how the compilation process is organised. Terms are not created in a haphazard way. There are specific ways of supplying term equivalents in various languages. Those ways include paraphrasing, borrowing, transliteration, semantic transfer etc. (This is discussed in detail in 2.8).

Projects undertaken with regard to terminology compilation in African languages are discussed in the next section.

2.4 PROJECTS BASED ON THE THEORIES DISCUSSED

Having explained what a term and what terminology is, it is important to also look at the relevant projects undertaken, projects in which terminology compilation in African languages took place. In reviewing these projects, the researcher needed to establish amongst others, the relevance of each project to the project under study, the purpose of the project, that is, the lexicographic product (glossary or dictionary etc.) the project aimed to produce and how the product would be made available. The researcher was also interested in the term extraction tool used as well as the size of the corpus collected of each project. Finally, key insights of each project will be weighed against each other.

2.4.1 ALLEX project

The African Languages Lexical Project, known by its acronym as the ALLEX Project started in 1992. It was a joint co-operative project between the Universities of Oslo and Zimbabwe. It was organised and financed under the Norwegian Council of Universities' Programme for Development Research and Education (NUFU) agreement. The University of Gothenburg also participated in this project.

The objectives of this project were to develop the African languages of the country through research to encourage the general public to use African languages in all its diverse affairs of life and also to produce corpus-based products. Corpus-based products included the general and specialised monolingual dictionaries, terminological glossaries and other information and

communication technology (ICT) language products for the main Zimbabwean languages, that is, Shona and Ndebele. The Ndebele corpus consisted broadly of spoken texts (interviews, radio and television recordings from The Zimbabwe Broadcasting Corporation) and written texts (consisting of publications such as novels, dramas and textbooks as well as manuscripts which included unpublished dissertations).

This project is similar to the present one because a corpus-driven product (which is the glossary of medical terms) will also be produced in the present project.

Both written and spoken texts were converted to machine-readable texts before a tool could be used. The total size of the Ndebele corpus (both written and spoken) was 691,268 tokens (Khumalo, 2007:45-47 and 103). The collection of both the written and spoken texts in this project, makes it relevant to the project under study in which the written bilingual texts and the monolingual texts from Ikwewezi FM were collected.

2.4.2 UKZN project

Since 2015 the University of KwaZulu-Natal (UKZN) has been compiling a series of language for special purpose (LSP) dictionaries for various specialised subject domains. This makes the UKZN project relevant to the present study which also aims to produce an LSP, medical glossary. Khumalo (2015) focused on term extraction for words in the linguistics subject domain. Through this project, isiZulu corpora and terminology have been developed. Both the corpus and terminology have contributed towards the building of the isiZulu National Corpus (INC) which started in 2014. The INC was piloted in November 2014 and the numbers of tokens were under 2 million, then. This terminology development project took place under a serious shortage of resources. The shortage of resources made all the processes of corpus and terminology building difficult.

By the end of the project, a total of 1,863 terms were in the isiZulu Term Bank (Khumalo, 2015:495-499). To analyse corpora, the WordSmith Tools (version 6) was used. Only the frequency analyser and the KeyWords function were used. As is known, for KeyWords analysis to be able to take place, a reference corpus is needed. In this project, the isiZulu National Corpus (INC) of about 1,22 million tokens was used as a reference corpus. The LSP corpus of about 100,000 tokens was used as a study corpus (Khumalo, 2015:495). The use of WordSmith Tools in this project makes this project relevant to the project under study, which also intends using the same tool.

This project at UKZN was further developed. The purpose was to intellectualise isiZulu to such an extent that it (isiZulu) functions at the same level as English, in all domains of the University. A terminology development process with 5 stages was designed. The first stage involved the harvesting of terms, after terms were harvested, they would be sent to the University Language Planning and Development Office (ULPDO). The second stage was the description and translation of terminology. The ULPDO would organise a workshop of experts for the description and the translation of terms.

Then the consultation and verification of terms with end-users followed. Workshops were organised with the purpose of verifying terms. The terms agreed upon in the verification workshop were submitted to the isiZulu national language body. The body was expected to consider and approve the verified terms. The fourth stage was authentication and standardisation of terms through the official national structures. The final stage was the listing of terms through the terminology databases. The authenticated terms were stored in a database (Khumalo, 2017:256-258).

All these stages are relevant to the present study. The researcher learnt from these stages that for instance terms need to be verified by end-users. That is why in the case of the present study, the isiNdebele speaking medical doctors have played a role in verifying terms.

Khumalo (2018:26-28), in a paper he presented at the 23rd International Conference of the African Association for Lexicography (AFRILEX), explained the growth of the IsiZulu National Corpus. The size of the corpus had moved from just under 2 million tokens to a total of 31,775,784 tokens. Instead of the WordSmith Tools that had been used in the former project, the Sketch Engine was used this time for the creation, management and analysis of corpora. The Sketch Engine has corpora in ninety (90) languages, with some languages like English having multiple corpora (Khumalo, 2018:18-26). This project is also relevant to the present study because of the usage of the WordSmith Tools for term extraction.

In this UKZN project, an explanation of how the compiled (LSP) dictionaries would be made available, was not given. This is a gap which the present study had to close.

2.4.3 CLTAL project

The research done by the Centre for Legal Terminology in African Languages (CLTAL) is presented by Alberts and Mollema (2013:51-52). CLTAL compiled various dictionaries for legal terms. The dictionaries were targeted to benefit legal practitioners as well as members of the community who at some stage would have to go to court or deal with law enforcement issues.

The process of compiling dictionaries on legal terms was divided into four phases. In the first phase, term extraction from legal texts took place. This was done by subject specialists from various legal fields. It is however not mentioned how term extraction was done, whether it was done manually or through the computer.

In the second phase of the project, data was computerised, and the Tshwane Term database was created. In the third phase, translation and definitions of terms took place. The first African languages to be included in the term base were Sesotho sa Leboa and isiZulu. Languages such as Tshivenda, Setswana, Sesotho, Siswati and isiXhosa were to be included in the term base, in the next project.

Various stakeholders played an important role in the translation process. This included language and subject specialists whose tasks were to translate terms and definitions accurately. Terminologists, researchers and anthropologists were also involved in this project. Problematic terms were discussed at quarterly meetings with representatives from different disciplines, for example lawyers, magistrates, court interpreters, members of the Justice Training College, the Department of Justice, the SAPS (South African Police Service), members of NLBs (National Language Boards), members of NLUs (National Lexicography Units) as well as members of PLCs (Provincial Language Committees), the SABC (South African Broadcasting Corporation), the Department of Education, the Tshwane Municipality, the SABS (South African Bureau of Standards), AFRILEX and academics from legal and African language departments.

The final and fourth phase comprised the verification and authentication by the National Language Boards (NLBs) of the relevant target languages. Before this stage, other language specialists and experts in law were appointed for editing and quality assurance purposes. Lastly, the dictionaries were published and disseminated. It is however not indicated where they were published.

Also, in this research, no software tool was used to extract and analyse data. The compilation of data was done manually. This makes this project different from the present one, which aims to use a term-extraction tool. Quality assurance of terminology was also done manually.

Although the purpose of this project was to compile dictionaries and not glossaries as is the aim of the current study, the project is however relevant to the present study. This is because it also compiled language for special purposes, in the legal domain. The present study is compiling language for special purposes in the medical domain.

2.4.4 SANTED Multilingualism Project

Mawonga, Maseko and Nkomo (2014:66-68) explain the role played by the South Africa Norway Tertiary Education Development (SANTED) project in the development of African languages, in different institutions of higher learning. This project focused on the promotion of multilingualism and the development of indigenous African languages for use in South African higher education. The project enabled terminology work to thrive at some institutions, namely the University of KwaZulu-Natal (UKZN) in collaboration with Durban University of Technology (DUT), Rhodes University (RU) and the University of Cape Town (UCT).

SANTED also enabled the translation of discipline-specific glossaries. UCT developed multilingual (English-isiXhosa-Afrikaans) glossaries in Statistics and Mathematics, Economics and Law. UKZN-DUT developed an English-isiZulu terminology list and glossaries in Education, Nursing (midwifery), Psychology and Dental Assisting. RU developed various multilingual resources in Political Philosophy, Computer Science, Geography, Law, Pharmaceuticals, Education and Journalism.

Although these institutions collaborated (as inspired by the SANTED Programme), they implemented different approaches to terminology development. They also used different approaches for publishing the glossaries online. A glossary booklet for Computer Science was compiled, it consisted of a total of 150 Computer Science terms and definitions in isiXhosa. This makes this project similar to the present one in that the glossary the researcher intends to compile, will also consist of terms from two languages (English and isiNdebele) as well as their definitions. The booklet that SANTED compiled was made available for isiXhosa-speaking beneficiaries. The Computer Science bilingual glossary is available for students in a hardcopy (as a booklet) and also online through the RU e-learning platform, RUconnected. This also makes this project relevant to the present one which aims to make the glossary available online (both on the Mpumalanga Department of Health and the SADiLaR's website).

As indicated above, the SANTED Multilingualism Project at RU accelerated the compilation of glossaries (that were published), bilingual terminology lists and phrasebooks. The bilingual texts were also developed through translation and also published online. In this project, it is not mentioned as to which tool was used to extract and analyse terms.

The School of Nursing of the University of KwaZulu-Natal also participated in the SANTED Multilingualism Project. The aim of the school of nursing was to develop scientific nursing terminology for isiZulu. Terminology development was one of the core activities that involved both the subject specialists and language departments. This they did with the purpose of enhancing effective communication between health professionals and patients in the health care services of KwaZulu-Natal. The health professionals speak 'medical' language in English while the patients speak 'every day' language in isiZulu and this leads to miscommunication. As isiZulu is one of the official languages in KwaZulu-Natal, patients in this province have the right to be served in their own language which often puts the nursing students in difficult situations.

The extraction of terms for this project was done manually. Subject specialists played an important role in the project. Ultimately, 1,400 terms were developed, and they were made available to nursing and midwifery (Engelbrecht, Shangase, Majeke, Mthembu and Zondi, 2010:249-267).

Through the SANTED project, the discipline-specific glossaries were compiled and published.

2.4.5 SPeLCAL project

According to Madiba (2004:135-136), the Special Language Corpora for African Languages (SPeLCAL) project was born out of the need for language resources. The language resources would support the implementation of South Africa's multilingual language policy adopted after the democratic changes of 1994. SPeLCAL would also provide a language resource for the compilation of specialised dictionaries, terminology lists and glossaries in the official African languages of South Africa as well as a resource for research in linguistic fields such as terminology, terminography, translation, language for special purposes (LSP) and second language teaching (SLT). This project illustrated how parallel corpora can be used as tools for developing the indigenous languages of South Africa.

Through SPeLCAL, only written texts were compiled. Oral texts were not compiled. This is what makes this project to differ from the present one. The project under study aims to compile both written and spoken texts. Only texts for specialised purposes and from specialised subject fields, such as the political domain (government and administration), health, education, law, science and technology were collected.

For the European Union's Lingua project (No. 36860-CP-1-96-1 FR-LINGUA-LD), the Multiconcord software was used in order to deal with parallel texts. The program ensures that an indexed corpus of texts together with their translations are automatically compared. It enables the comparison of up to ten European languages. The programme has been found to be highly flexible and fast and can achieve a high degree of alignment accuracy with any pair of bilingual texts. The Multiconcord software also has a 'Minmark program' which is very useful in encoding texts. It automatically inserts the markers for the body text, paragraphs and sentences. It proved to be very useful in the processing of parallel corpora. Although Multiconcord is excellent in aligning source and target texts, it has been discovered that this exercise requires pre-processing that can be done by means of Microsoft Word. Once this process has been completed, the aligned documents appear in the programme and are ready for processing (Madiba, 2004:141). A Multiconcord software was at least used here, which makes this project relevant to the present one which will use the Wordsmith Tools.

Bilingual parallel corpora are resources that have not yet been fully exploited, as far as the development of the indigenous languages of South Africa is concerned. For instance, regardless of how small the bilingual corpora are, such corpora are necessary for yielding translation equivalents in a variety of contexts. Unlike large corpora, bilingual corpora can be analysed electronically or manually using standard tools such as Multiconcords. Furthermore, small corpora as compared to their large counterparts, allow early human intervention. Large corpora, on the other hand, enable late human intervention. 'Where small corpora are concerned, a researcher can begin searching while the corpus is still under construction. SPeLCAL corpora can be useful for terminographers, translators and lexicographers' (Madiba, 2004:146). This project is relevant to the present study because it aimed at providing resources for the compilation of glossaries, amongst others. The only challenge is that no mention is made as to where these resources are available; one may thus not know where to access them. This is also a gap that the present study had to close.

Finally, the above projects had to be analysed in terms of their differences and similarities. What was discovered is that the ALLEX and the UKZN projects are similar in that written and spoken texts were collected in both projects. The SANTED and the SPeLCAL projects on the other hand, only collected written corpora.

Regarding the size of corpora collected, it is only the ALLEX, UKZN and SANTED projects that mentioned the size of corpora they collected. The SPeLCAL and CLTAL projects did not mention the size of the corpora they collected. Whilst the CLTAL and SANTED projects did not use any term extraction tool, the ALLEX, UKZN and SPeLCAL projects did use the softwares for the extraction and the analysis of their corpora.

Finally, whilst SANTED and UKZN mentioned the websites on which their products were made available, CLTAL, SANTED and SPeLCAL did not.

2.5 AN OVERVIEW OF GLOSSARY COMPILATION

According to Nkomo and Madiba (2011:149-150), glossaries are meant to serve certain needs which dictionaries have failed or do not seem to have the capacity to address. They state that most of the available African language dictionaries have major limitations and barely satisfy the people they are purposed to serve and because of that, glossaries are short term alternatives for dictionaries. Owing to the above findings, the production of glossaries is justifiable.

In the next sections, as part of the literature review on glossary compilation, the researcher aims to establish the following: types of glossaries, the tool used for extraction of terms, the number of terms compiled and how the glossaries were made available.

2.5.1 English-Portuguese glossary of cooking terms

This glossary is not on African languages, but the researcher included it because of the corpus-driven approach employed in the compilation of this English-Portuguese glossary. Tagnin and Teixeira (2012:51) compiled this bilingual, mono-directional English-Portuguese glossary of cooking terms. Tagnin and Teixeira (ibid.) stated the significance of compiling such a glossary by saying that 'although there are many English-Portuguese technical glossaries on the market, very few are designed to meet the specific needs of translators'.

This is an indication that they targeted to meet the needs of translators with their glossary. The present study also aims to meet the needs of isiNdebele translators. Two sub-corpora consisting of cooking recipes were gathered, one in English and one in Brazilian Portuguese. The corpora were gathered only from the internet as they were able to evaluate websites that contained adequate texts. The number of recipes they collected for Portuguese was 8,300 (1,520,864 tokens) and 7,400 (1,578,125 tokens) for English. The fact that in this project, corpora were only gathered from the internet makes this project different from the present one. In the present project, corpora were gathered from many sources, including the internet.

To analyse these corpora, a corpus query software, the WordSmith Tools was used. The first function (in the WordSmith Tools) they used is the WordList function. Through this function, a word list was generated (that is, a list of all the words in all the texts of the corpus with their frequencies). After the word list, a key word list was generated. For generating the key word list, a reference or general corpus is necessary. The researchers used the American National Corpus (ANC) as reference corpus. It is important to note that the KeyWords function is monolingual. Through the key word list, the researchers were able to get a more accurate picture of the vocabulary, specific to the cooking area. This made it easy to highlight the vocabulary of cooking recipes (ibid.).

After the key word list was generated, the WordSmith Tools' Concordance function was used. This function allows the lines containing the search word in the centre to be sorted according to the words at its left and/ or right. Through this function it became easy to visualise recurrent patterns. All recurrent collocations and phraseologies that contained one of the 300 key words analysed, were then included in the glossary.

The glossary therefore, consisted of the 300 most probable words (as well as the multi-word combinations they are part of) that occur in recipes (Tagnin and Teixeira, 2012:59). Also, here, it is not indicated how the glossary was made available.

2.5.2 Glossary on multilingual concept literacy terms

Nkomo and Madiba (2011:149-150) conducted research on the compilation of multilingual concept literacy glossaries at the University of Cape Town. Glossaries resulted from the collection as well as descriptions and presentation of terms. The glossaries were for various subject fields.

The purpose of the glossaries was to assist those students whose first language was not English. The size of the economics corpus was 70,000 tokens. This is a small size but Nkomo and Madiba (ibid.) maintain that the size contained relevant key concepts and contexts. The WordSmith Tools was used for term extraction. The completed glossaries were then uploaded on Vula.

Vula is the online collaboration and learning environment of the University of Cape Town. It is used to support UCT courses as well as other UCT-related groups and communities. The word "Vula" means "open", and refers to the many possibilities provided by Vula, as well as its Open Source origins: <https://vula.uct.ac.za/portal/directtool/177933a3-8af8-44c1-8d0c-e40a202f2ed2/>.

2.6 STUDIES CONDUCTED ON COMPILATION OF MEDICAL TERMS

This section is about studies conducted on the compilation of medical terms for various indigenous languages.

2.6.1 IsiZulu-English medical terms

Van Huyssteen (2003) compiled an isiZulu-English list of medical terms. The list was compiled from both written and spoken corpora. The written corpus comprised of printed medical pamphlets whereas the spoken one comprised of interviews with medical staff. As far as the written corpus is concerned, terms were extracted from the medical pamphlets distributed to hospitals and clinics by the Department of Health. Only 11 medical pamphlets covering health care issues such as tuberculosis, cholera, and HIV/AIDS were used to build the medical corpus of this study. The researcher emphasises that these pamphlets were produced by different institutions and translators. According to the researcher, they represented an existing medical written isiZulu discourse, which reveals terms in contexts.

The medical pamphlets collected were entered into computer files through re-keyboarding and scanning. Through re-keyboarding, printed medical texts were typed into computer files, while scanning was done by means of an *Optical Character Recognition*-tool (OCR). Manual correction of errors took place after re-keyboarding and scanning had been completed. No medical texts were downloaded from the internet, in this study.

The spoken corpus comprised of interviews with medical staff, after which transcription took place. After transcription, the WordSmith Tools (WST) was used to extract terms. In order to establish terms that were actually used in the practical medical field of health care and the ones that had to be lemmatised and then later taken up in a terminology list, a semi-automatic term extraction method was used. Van Huyssteen (2003:150) emphasises that computerised lemmatisation was preceded by manual lemmatisation. This is proof that manual intervention cannot be ignored.

In the WordSmith Tools, the researcher used the WordList function (to produce frequency and alphabetical list) as well as the Concordance function. The resulting written medical corpus collected was a total of 2,187 tokens. The spoken corpus contained a total of 823 tokens. Van Huyssteen points out that although the corpus she used was small, the purpose was to make it exemplary of the type of research that can be done in the medical terminology field.

Although this study mentions that the spoken corpus collected was transcribed, it is not mentioned as to which type of transcription was employed.

2.6.2 English-Shona biomedical terminology

Mpofu and Mangoya (2005) compiled an English-Shona dictionary of biomedical terms. Mpofu and Mangoya's (ibid.) study is relevant to this present study. This is because whilst the aim of their study was to collect biomedical terms in order to compile a dictionary, the present study's aim is to collect medical terms for compiling a glossary. The purpose of the dictionary compiled by Mpofu and Mangoya (ibid.) was to improve the efficiency of communication between doctors and patients. The dictionary is composed of terms from both modern and traditional medicinal practices from major areas of medicine such as general medicine, gynecology, ophthalmology, pediatrics, surgery, urology, etcetera. One of the challenges they faced was to maintain a balance as far as terms to be included in the dictionary, is concerned. Preference was therefore given to those terms used in everyday medical practice and consultation.

Their process of compiling terms for the dictionary started with field-work. Four people in the nursing field conducted field research. The researchers' focus was on diseases, their symptoms and treatments. Interviews were conducted with the medical personnel and the general public. Those interviews were recorded and later transcribed. For the collection of traditional medical terms, four traditional healers from the Zimbabwe Traditional Healers' Association (ZITHA) were interviewed.

It is only after the field work process was completed that the headword selection process took place. To support this process, library research was undertaken. Both the English and Shona dictionaries for general purposes were consulted. Medical books were also consulted (Mpofu and Mangoya, 2005:119).

No tool was used to extract and analyse data in this research. The head word selection was done manually. In corpus linguistics, the manual identification of terms is allowed on condition that those terms will be validated by a language expert.

Having discussed the concept terms and terminology as well as projects in which terminology in various languages was compiled, it is also important to pay attention to the concepts terminology development as well as term-formation processes.

2.7 TERMINOLOGY DEVELOPMENT

In this section, the development of isiNdebele terminology is investigated. The investigation is necessary in order to try and understand how the paucity of standardised terminology in specialised subject fields negatively impacts the development of isiNdebele. According to Finlayson and Madiba (2002:53), terms need to be developed for the purpose of intellectualisation. Intellectualisation is a process that ensures that the terminology of the language carries the full weight of scientific rigor and precision. This is done through cultivating, developing, elaborating and modernising the terminology of a particular language. Intellectualisation further ensures that the language is able to function in all domains and has the capacity to discuss any issue regardless of its complexity. This process helps the previously under-developed African languages to grow faster. It further ensures that the language transforms in such a way that it has the capacity to carry and also to convey all forms of knowledge, in all spheres of life.

The translation of the terms Covid-19 as *i-Covid 19* in isiNdebele and 'corona virus' as *ingogwana ye-corona* and *i-corona* in isiNdebele (Vukuzenzele, 2020:3) is an indication that isiNdebele terms on 'Covid-19' are being intellectualised.

Van Huyssteen (2003:93) asserts that the methods of word-formation play a very important role in the elaboration of language because they are the very linguistic tools that make the expansion of the lexicon and the technical modernisation possible. The concept language elaboration refers to the creation of new terms in order to meet the educational, scientific and technical demands of a language. When one looks at the above terms (that is, corona virus and Covid-19), one realises that these are terms we did not have prior to the break-out of the pandemic.

Mnguni (2004:7) says that when it comes to the creation of technological terms, African languages are faced with a serious challenge. This is because as technology advances, there is always a need for the creation of new terms.

To emphasise the views of the above scholars, Madonsela (2017:202) says that the widespread use of the internet, cell phones and other technological devices necessitated the formation of new terms. Owing to the fact that the process of inventions and interventions is ongoing, there is also an ongoing need to create new words in order to convey the concepts accurately. This is necessary because African languages in general do not easily assimilate words of foreign origin.

According to Madzimbamuto (2012:135), terms need to be developed in such a way that they can be used in all domains. Madzimbamuto (ibid.) further says that terms should be used in tertiary environments for 'dual' language teaching. A space should be developed where the use of African languages is promoted in publications as well as in scientific meetings. The use of African languages in medical education should be increased in order to lessen the marginalisation of the African culture.

2.7.1 Challenges faced when developing new terms in African languages

The development of new terms is always characterised by challenges. Term-creation processes are always complicated. It is because of these complications that translators resort to different methods and strategies of formulating new terms. The process of creating terms is even worse when one is translating scientific terms (Ndhlovu, 2012:329). Van Huyssteen (2003:173) says that the complications are brought about by the fact that unlike in the case of ordinary language, technical language shows a one-to-one correlation between a concept and a term.

There is a lack of insight and standardisation in the term-creation mechanisms especially when it comes to African culture. The development of terminology in African languages (isiNdebele included) is unfortunately characterised by compilers who possess little or no knowledge in the term development theory. There is also a lack of documented terms. This on its own contributes towards all these challenges (Van Huyssteen, 2003:58).

2.7.2 Guidelines for the development of new terms for African languages

Taljard (2008:90) created guidelines that must be taken into cognisance when creating term equivalents that will be acceptable to both linguists and special field subjects. The guidelines are as follows:

- There should be a one-to-one relation between any given term and the concept it represents, that is, any term should ideally refer to one concept and one concept only. This implies that there should be no synonyms and no morphological or spelling variants for any specific term.
- Terms referring to closely related concepts should also be similar in some way, in order to reflect the similarity between the related concepts. Conversely, concepts not closely related should be expressed by terms that differ markedly in appearance and sound.
- Terms should conform to spelling, pronunciation conventions and morphology of the language for which it is intended. Terms should be concise and not contain unnecessary information. This should happen without sacrificing precision.
- A term should be more or less self-explanatory, which means it should be transparent.
- The meaning of a term should be independent of context.
- Terms should be capable of providing derivatives.
- Once a term has gained general acceptance, it should not be changed without compelling reasons and a strong certainty that the new term will be accepted as a full substitute (Taljard, 2008:90).

Term-formation processes in general are discussed below. Term-formation processes as applicable in isiNdebele are discussed in Chapter 4.

2.8 TERM-FORMATION PROCESSES

According to Madonsela (2017:201), term-formation strategies refer to all the processes used to create new lexical units. When people are familiar with the processes underlying the creation or formation of new words, it becomes easier for them to understand and assimilate these new terms into their languages.

Because of new inventions and interventions, there is an ongoing need to create new words to accurately convey these new concepts. Although the process of word-formation generally addresses morphological issues, in its wider sense it also deals with the processes of creating new lexical units. It became apparent that in some instances, a number of different strategies had to be employed in order to version the same word in different languages.

Madzimbamuto (2012:132) says that from the medical perspective, patients prefer information in their own languages. In Africa, however, the development of such technical language has been neglected. Most citizens do not speak English as their first language and may not be able to adequately express their illnesses. On the other hand, most health professionals lack the terminology for the information being delivered. Therefore, developing African languages so that they can be able to function at a technical level is a necessity. The development of African languages includes the dispersion of technical information to a lay level, thereby improving access to knowledge. This has become an important area of language advocacy.

To develop ChiShona medical terms, new words in Chishona have been assimilated from English, Nguni, Afrikaans and Portuguese. Word borrowing is one of the methods used in terminology development.

Mabasa (2005) just like Madzimbamuto (2012), did a study on translation equivalents for health/medical terminology, with the focus on Xitsonga. According to Mabasa (ibid.), miscommunication of health information has a devastating effect on South African health providers and health consumers (community). The curing of infections is dependent on communication of information. Now, if there is a communication breakdown, a serious challenge can be envisaged. Mabasa also did research on the term-formation processes mostly applied in Xitsonga when translating health/medical terms.

Mabasa discovered that over and above all term-creation strategies applied, transliteration occurs frequently in Xitsonga. It is actually one of the most frequently used methods for creating Xitsonga translation equivalents for English and Afrikaans terms, particularly in a subject such as medicine (Mabasa, 2005:16).

Mtintsilana and Morris (1988:111) caution that 'although transliteration seems to be the most productive method of developing terminology, a language may run the risk of losing its character if it allows transliteration to fill lexical gaps. Transliteration should therefore perhaps not be used as a shortcut or first aid, but rather as a last resort, for instance in the fields of medicine and the natural sciences'. Mphahlele (2004:341-342) on the other hand, maintains that transliteration is nothing but a mere duplication of a source and that does not provide a best solution for the language that has a shortage of terminology.

Ndhlovu (2014:327-330) also did a study on Zimbabwean Ndebele health terms. Just like in the case of other indigenous languages, there is a scarcity of terminology for special language purposes in Zimbabwean Ndebele. The language has three dictionaries which are dictionaries for general purposes. General purpose dictionaries are of little use in the translation of health terms.

Ndhlovu (ibid.) explores how English-Ndebele translators deal with the challenge of term scarcity when translating specialised terms in the health sector. The focus of her study was on term-creation strategies used by Zimbabwean Ndebele translators.

Ndhlovu (ibid.) explored an English-Ndebele Parallel Corpus by means of a parallel concordance, ParaConc. Through ParaConc, the researcher was able to identify source terms and their equivalent translations in the Ndebele language. She also managed to draw frequencies, word counts, alphabetical lists and other possible translations and "hot words".

Ndhlovu (ibid.) noted that pure borrowing is the most commonly used strategy. This was supported by word counts and frequency of terms she discovered in the Ndebele corpus.

Borrowing involves various strategies such as the use of pure loan words, pure loan words provided with a preceding explanation, acronyms and abbreviations, paraphrased acronyms and abbreviations as well as indigenised loan words. It was discovered that the lesser the frequency of a term, the higher the chances of the term being retained in its pure form.

Most translators rely on pure loan words. The reason for this is that pure loan is an easy option, as most translators have not received any formal training. Because English is one of the official languages and a dominant language in Zimbabwe, it seems as if translators assume that their readers will understand borrowed terms even though the structure of these terms is alien to the Ndebele language. The Zimbabwean situation is contrary to the South African linguistic environment, where the Constitution stipulates that all 11 languages are equal and where some of these 11 languages are in fact minority languages. The languages in Zimbabwe do not enjoy the same support for development as is the case in the 11 officialised languages of South Africa. This finding points to a dire need to train translators in Zimbabwe and also to develop a language policy that takes cognisance of all the languages spoken in Zimbabwe. The least used strategies are paraphrasing, acronyms and abbreviations, coinage and compounding (in this order). There is a need for terminographers to continue developing terms, working hand-in-hand with other language experts and researchers in order to fill the gap of scarcity of terminology in African languages (Ndhlovu, 2012:342-343).

One of the questions this study seeks to answer is how existing methods of producing medical terminologies can be transferred to isiNdebele where only sparse resources are available. Different term-formation processes used in forming isiNdebele medical terms are therefore discussed. Hadebe (2002:143) maintains that term-formation processes play a significant role in developing indigenous languages. Mtintsilana and Morris (1988:110) say that languages develop terminologies through term-formation processes that are both internal and external to the language.

Term-formation is about terminology development. Sager (1990:2) defines terminology as the study and the field concerned with the collection, description, processing as well as the presentation of terms, that is, lexical items belonging to specialised areas of usage of one or more languages.

Term-formation processes are classified differently by different scholars; however, this study aligns itself with Sager's classification. Sager (1990:80) distinguishes between two types of term-formation processes, namely the primary term-formation and the secondary term-formation processes. Primary term-formation processes involve the usage of new forms or new lexical entities that did not exist in the lexicon before.

Van Huyssteen (2003:93) refers to this approach as that of forming words from foreign resources. Hadebe (2000:229) refers to the primary term-formation as borrowing from other languages. Some of the term-formation mechanisms which are used are: compounding, derivation and neologism or coinages as well as semantic expansion.

Secondary term-formation processes occur when a new term is created for a concept which is in existence. In other words, there is already an existing term in the source language. The term can serve as the basis for the secondary term formation. Secondary term-formation processes therefore require that the terminographer should have a thorough understanding of linguistic mechanisms. In that way, the expansion of terminology will be ensured. Van Huyssteen (ibid.) refers to this as internal resources of term-formation and on the other hand, Hadebe (2000:229) refers to secondary term-formation processes as processes from within the language.

2.8.1 Primary term-formation strategies (foreign)

2.8.1.1 Paraphrasing

Paraphrasing is a way of explaining or describing a concept by making use of a phrase or even a sentence. This is regarded as a productive way of extending vocabulary in indigenous languages. When term developers are confronted with new concepts they are unable to express, they resort to paraphrasing (Madiba, 2000:219-220). Van Huyssteen (2003:112) is amongst the scholars who refer to this strategy as loan translation or calquing. With this strategy, a word is borrowed from a source language and its meaning gets rendered in the recipient language through paraphrasing it. By so doing the message becomes clearer, as foreign concepts are explained. According to Madiba (2000:200), loan translation strategy (paraphrasing) is a productive way of describing foreign concepts and engineering knowledge with ease.

2.8.1.2 Borrowing

Borrowing is the process whereby linguistic elements are taken over from one language to another. There are different terms that are used interchangeably with the term borrowing, namely 'loaning' and 'adoption'. According to Mahlangu (2007:1), the definition of these terms means one and the same linguistic feature. Therefore, all three of these terms will be used in this study to refer to all words and phonemes that have been introduced to isiNdebele vocabulary from other languages.

Van Huyssteen (2003:115) asserts that no language is self-sufficient. This is because no perfectly homogeneous language group exists. This means that all languages are therefore bound to borrowing as long as they are in contact with one another. Borrowing is regarded as the least strategy of term-formation because a word is taken from one language into one's own language. With borrowing, the word retains its original meaning that it had in the original language, although the orthographic form of the word might change.

Borrowing occurs when a new term is created due to transferring knowledge to another linguistic community in which a corresponding term needs to be created. It should be noted that borrowing is the only term-formation strategy that brings in the lexicon of any given speech community, the foreign lexical elements (Sager, 1990:80).

According to Moropa (2007:191), borrowing occurs in areas where two languages are unequal. In this kind of a situation, the less developed technical language takes over concepts and terms from the more developed one.

Using a loan word is not new in the indigenous languages. Whenever a language is reduced to writing, such a language needs to receive items which are foreign to the indigenous language's culture. Translators use different loan words, such as indigenised loan words, which is where a loan word becomes slightly modified so as to remove some foreignness in it. In this case, the loan word is spelled according to the orthography of the borrowing language (Wallmach and Kruger, 1999:281).

Sometimes an indigenised loan word may be preceded by the explanation. In such a case the loan word is put in brackets. The explanation helps to enhance understanding. A pure loan word may also be used, in this case source language words remain the same in the target text. Translators often use this strategy to deal with culture-specific items, modern concepts and buzz words. Acronyms and abbreviations are retained in their original form in the target text, but the necessary prefixes should be added (Moropa, 2007:191-192).

Borrowing is classified into direct loan, coinage and transliteration.

(a) Direct loan

Direct loan is whereby a term is taken over into the target language, from a source language. The term is taken over as it is. That is, it is taken into the target language without changing its source language's morphological structure. It retains its spelling. This is resorted to when the target language has no equivalent for source language units (Sager, 1990:90).

(b) Coinage

Coinage is the process of inventing totally new words. Kalima (2007:5) makes a distinction between borrowing and coinage by saying that borrowing simply copies the word and meaning from one language to another. Coinage, on the other hand, is the process of adding a word to a language by giving an object or an idea a new name. Mahlangu (2015:90) maintains that coinage as a term-formation strategy is still problematic in isiNdebele because when coining terms, terminologists are trying to bring forth all the resemblances that are found in the source language (SL) term. The longer the term, the more ambiguous it becomes. For instance, the term AIDS is translated as *umbulalasihlangu* in isiNdebele. This term is a compound noun; it is made up of *bulala* '(to) kill' and *isihlangu* which is 'a shield'. This means that AIDS is the killer of the shield. The shield refers to the 'immune system'. This means that, 'AIDS' destroys one's 'immune system'. Mahlangu (2015:82) discovered that in isiNdebele, coined terms are usually compounded to express the meaning of the English phrase.

(c) Transliteration

Regmi, Naidoo and Pilkington (2010:18) define the concept transliteration as 'a process of replacing or complementing the meanings of words in one language with meanings of another as sometimes the exact equivalence or exact meaning might not exist. The important aspect of transliteration is an unavoidable loss of meaning that occurs in everyday language, which helps to set the context in which cross-cultural translation can be better understood'.

Mtintsilana and Morris (1988:111) confirm that transliteration is the most productive method of developing terms, however, it is easy for a language to lose its character when it heavily depends on transliteration.

IsiNdebele is one of the South African indigenous languages that has been in close contact with Afrikaans and English-speaking communities for many years and a part of its vocabulary has been drawn from these non-African indigenous languages, particularly Afrikaans (Mahlangu, 2007:1).

2.8.1.3 Derivation

Derivation is the term-creation strategy in which new terms are coined from roots by adding affixes. Madonsela (2017:203) confirms this by saying that in indigenous South African languages, prefixes and suffixes are often added to stems to create new words.

Derivation, according to Van Huyssteen (2003:218), is a basic method of term-formation since it draws on the internal resources of the language.

2.8.1.4 Deideophonisation

Deideophonisation is the term-creation strategy in which a term is formed when a prefix is added to an ideophone. It deals with the coinage of terms from sounds that can be associated with the action or object that must be named. The process of deideophonisation shows parallels with the English onomatopoeia. In isiNdebele the process involves the prefixing of a noun class prefix to the grammatical term.

2.8.1.5 Abbreviation

Sager (1990:263) refers to abbreviation as a process of compression and further defines this as a term-formation strategy where complex terms are reduced or compressed to a more compact, standardised form. Abbreviation has sub-categories of clipping, acronyms and blending.

(a) Acronyms

An acronym is a term-formation strategy in which a word is made up from initial letters of several words which make up the name of something.. This strategy uses fewer words with the purpose of designating the same concept, for example, 'AIDS' instead of 'Acquired Immune Deficiency Syndrome'. Acronyms have an advantage in that shortening of terms serves the purpose of creating more concise forms, which speakers prefer. This is especially for frequently used terms. In other words, this process creates terms that are easier to

remember than lengthy ones which may be difficult to remember. Acronyms have an advantage in that after some time, they can be integrated into a language as fully-fledged words. In this case they become words and part of the lexis of the language concerned (Van Huyssteen, 2003:124).

(b) Clipping/Shortening

Clipping or Shortening is a term-formation process that creates new words by reducing the length of the existing words. The sense of a term is reduced to only one of its parts.

(c) Blending/Compounding

Blending and compounding are two term-formation processes which overlap. It is for this reason that they will be discussed together. In compounding, two independent (autonomous) words that are mostly members of a word group, are combined to form a compound (Krüger, 2006:41). In this process a new single term or lexical unit is formed by combining two or more free morphemes (Sager, 1990:76-77).

Blending on the other hand is the term-formation process in which parts of two or more words combine to create a new word whose meaning is often a combination of the original words.

Cluver (1989) in Moropa (2007:195) claims that to create a new technical term in any language, compounding is the most important mechanism. Compounding involves combining two or more words (two nouns, a noun and a verb, etcetera) into one unit. The meaning of compounds is often dependent on the context in which they are used and the underlying relations between the constituting elements.

It is also observed that translators create terms by using a loan word, compounding, derivation and paraphrasing. Four types of loan words, i.e. an indigenised loan word, an indigenised loan word preceded by an explanation, a pure loan word and a pure loan word preceded by an explanation, were identified. Loan words were used mainly for modern computer terms as well as for financial and other technical terms. Moropa (ibid.) maintains that the successful ways of creating technical terms in isiXhosa are through loan words and compounding.

The process of intellectualisation of indigenous languages has contributed towards the relevancy of compounding as a word formation process. Finlayson and Madiba (2002:53) confirm that it is a tendency in many countries that the creation of new terminologies is based on the deliberate and conscious use of word-formation patterns or methods such as borrowing, compounding, derivation, loan translation or calquing, semantic shift, blending, clipping, etcetera.

Pretorius, Viljoen, Pretorius and Berg (2008:5) did a research on computational morphological analysis of Setswana compounds. They developed a computational morphological analyser for Setswana. The purpose was to model the accurate implementation of compounding as a term-formation process. Their focus was on the formation of noun + noun compounds by computational morphological means in order to understand how this process should be formalised, modelled and subsequently implemented.

Pretorius et al. (2008:18-19) wanted to find a way in which compounds can be formalised, modelled and implemented so that compounding in Setswana could be handled as accurately as possible by the morphological analyser which was under development. The findings were that when they applied the basic computational assumption (compound = word + word) to the analysis of compounds, it transpired that certain word-formation rules do not strictly apply in compounding. They then had to relax these rules in order to gain insight into the processes that govern the formation of noun + noun compounds.

Pretorius et al. (ibid.) discovered that in the formation of compounds consisting of noun plus noun, other than those having reduplicated stems, the prefix of the second noun gets omitted. Another insight gained from this study is that the rules governing deverbative morphology are not adequate for a complete description of the formation of noun + noun compounds in which the second or both constituents are deverbatives. They maintain that from a linguistic perspective the status of the constituents of compounds warrants further investigation.

It is easy to confuse blends with compounds. This is because blends are also formed by combining two different words. The distinction is that blends, as their name suggests, blend two words into one unique entity. It is easy to separate the two components of the compound but with blending, it is impossible because it is only parts of the words that are combined to form a new word. For example, a word such as **smog** is made up of two words that is, **smoke + fog** (Kalima, 2007:6).

When one looks at the term **smog**, its different components are not easily identifiable. In compounding, one can easily determine the different components and separate the two. See more examples in Chapter 4.

Madiba (2000:214-215) regards compounding as another productive term-formation strategy that has been used in most languages of the world. Although the term compound has been defined in various ways in literature, Madiba (ibid.) advises that the new entity created must represent the original concept.

2.8.2 Secondary term formation strategy (internal resources)

Semantic transfer is a secondary term formation strategy. It is a process whereby a new meaning is attached to an existing word by modifying its semantic content. In this process the existing word and the new term are used side by side, one in ordinary speech and the other as a term in a special field. In this case, a word that comes from the general vocabulary acquires a more technical meaning. The meaning gets narrowed down and becomes more specialised (Sager, 1990:71). The existing meaning of a word acquires an expanded or modified meaning in order to name a new, generally related concept (Van Huyssteen, 2003:10).

2.9 TABOO IN TERM-FORMATION PROCESS

Taboo is a type of avoidance that refers to terms that are not permissible in a specific social environment. This is because such terms are considered as vulgar, offensive and disrespectful. When one discusses terminology, especially terminology that belongs to a particular domain, one cannot avoid the issue of taboo. There are euphemisms and certain expressions that are created in order to express the unspeakable, pertaining for example to parts of the body that perform (1) reproductive tasks, and (2) 'private' functions such as urinating and passing stool. These body parts are considered as 'private' because in the anatomy of the human being they are not exposed. The 'private' functions are also performed in private. De Klerk (1988) in Mfazwe (2003:17) states that 'such words are avoided, considered inappropriate and loaded with affective meaning'.

Yamanti, Brata and Sedeng (2016:162) define taboo as a cultural custom that forbids talking about certain things. This is done so that social cohesion and social order can be preserved. Wardhaugh (1992:239) says that the reason why certain things are not to be said and expressed is because they are believed to be harmful to its members. It is believed that they

would cause anxiety, embarrassment, or shame in people. Taboo determines that certain objects can only be referred to in certain circumstances, for example, only by certain people, or through deliberate euphemism.

Wardhaugh (1992:240) mentions that taboo and euphemism affect all people. People may not be as deeply conscious of the effects, but the truth is, people refuse to talk directly about certain things. There are some words that are known but are never or hardly ever used because they are too emotional for either the speakers or others. Care is therefore taken so that such words are not expressed at all even though the words are known. Sometimes such words are used very indirectly.

Mabule (2009:45) refers to taboo as terms that are culturally sensitive. The terms are regarded as 'culturally sensitive' because it is not easy to mention them. Such terms are perceived as being offensive or taboo. In most cases, such terms are those that are associated with reproduction. The naming and the functions of organs of human reproduction are perceived by amaNdebele as taboo. The use of euphemisms functions prominently so that the translated text can be acceptable to the targeted readers. Taboo words are mostly replaced by ordinary expressions on euphemism in the translation of certain sensitive documents to avoid unpleasant subjects. In most instances, the targeted reader feels offended when a translator uses direct equivalents of certain words regarded as taboo. The reader will therefore label such translators as vulgar and insensitive to the norms, culture and values of the target readers.

Mabule (2009:51) further mentions that the general public and the translators are divided in the way in which the terms regarded as taboo should be translated. There are two conflicting views. The one group prioritises cultural sensitivity in translation whereas the other group is message oriented. The latter view prioritises the conveying of the message more than the manner in which it is conveyed. This means that translation in this case is done without taking into consideration cultural sensitivity. The main purpose here is to take the message from the source language into the target language. The former view argues that translation needs to be understood within a cultural frame of reference. Translators should therefore serve as cultural mediators. They should preserve the cultural norms and rules in the target language. Preservation of cultural norms suppresses the individual translator's freedom of expression and this may make translation to be inaccurate. This is because the translator has to carefully choose the phrases so as not to offend the target readers in this instance, the isiNdebele speakers. In order to avoid offending the target readers, figurative rather than literal translation

should be implemented. Taboos therefore have a moral or cautionary effect on people's behaviour and speech.

Based on the above, translators should seek to strike a balance between the two approaches. This means that as much as they should be culturally sensitive when translating, they should also strive to convey the message un-altered (from the source language to the target language).

Due to the fact that this project is about the collection of terms with the purpose of compiling a medical glossary (which inevitably includes taboo terms), the researcher also discovered how cultural taboos can hinder effective communication in the medical field. Specific examples are given in Chapter 4.

2.10 COPYRIGHT AND PERMISSIONS

Bowker and Pearson (2002:59) explain the importance of copyright permission by saying that, to ensure safety, it is important to contact the owner of any text that you want to include in your corpus and obtain his or her permission. They also confirm that this process is not straight forward. This is because it is sometimes not clear who holds the copyright, or it may be difficult to track down contact details. They say that sometimes, even if you managed to track the holders, they may not understand what a corpus is or how it is used; the author may think that you will use his/her text for financial gain.

To emphasise Bowker and Pearson's (ibid.) view, Olohan (2004:51) explains that the researcher, when faced with problems when asking for permission to use texts, should stress that they will be used for research purposes only, and to specify what other researchers will have access to.

Regarding the speech corpus from Ikwewezi FM, the researcher went to the Ikwewezi FM station and explained to the programme manager what she is busy with and also that she will be recording the health programme for her research. The programme manager indicated that it was not necessary to ask for permission because the programme is broadcast to the public in any case. She emphasised, however, that it would be proper to acknowledge Ikwewezi FM in the study, which the researcher did.

Regarding the doctors that participated in the study, the researcher contacted them directly. There was no need to ask for any permission from any institution because these doctors have their own surgeries.

2.11 CONCLUSION

This chapter presented an overview of the corpus-driven approach. CDA has proven to be relevant to the corpus of the language under research. This approach ensures that language constructs come from the analyses of the corpus and the hypothesis is based on the evidence found.

The theoretical frameworks that underpin this research, namely Corpus linguistics and Grounded Theory are also in alignment with the CDA explained. Corpus linguistics presents samples of what people have actually said unlike hypothesising about what they should or might have said. It is suitable for the analyses of the corpora of the current study. It is also relevant for the compilation of the first isiNdebele glossary for special language purposes.

The other theory the researcher has implemented in this study is the Grounded Theory. This theory also ensures that assumptions are avoided. It further advocates an understanding of social phenomena that are not pre-formed or pre-theoretically developed with existing theories and paradigms.

When one looks at the corpus-driven approach, corpus linguistics and Grounded Theory, one realises that there are some similarities in that in all three linguistic constructs always come from the analysis of the corpus.

The aim of the present study is to compile an English-isiNdebele glossary of medical terms. Through the works of various scholars, it has been proven that there are no hard and fast rules regarding how large a corpus ought to be. Much can be accomplished with a corpus of a small size.

Various projects, which included the compilation of terminologies for glossary making were also reviewed. The researcher wanted to establish amongst others, the purpose of the project, that is, what the project aimed to produce and how the product would be made available, term extraction methods that were employed as well as the size of the corpus. Not all requirements were met, some projects did not mention the tool they used for term extraction, some did not

mention the size of the corpus and how the product would be made available for access by the public. This research therefore aims to fill those gaps.

What the researcher also discovered is that in some projects, no tool was used to extract terms. Terms were manually extracted. This does not insinuate that manual extraction and analyses are wrong, what is important is that there should be verification. For instance, in the CLTAL and the SANTED projects the researcher reviewed, the extraction, the analyses and the quality assurance of data were done manually. Although the process becomes slower than in the case where the machine is used, it is doable.

In all the projects where terms were manually extracted, subject specialists were brought on board for quality assurance. This is an indication that manual extraction of terms is not a problem as long as there are specialists to verify those manually extracted terms.

Also, some of the projects made use of the WordSmith Tools as is the case in the present study. However, not all projects made use of all the functions in the WordSmith Tools. This is an indication that the choice of using a function is determined by the purpose of the study or project.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 INTRODUCTION

This chapter discusses the approach the researcher followed in collecting and analysing data for this study. The approach involves the research design (which are methods and instruments for data collecting) and the compilation of both the written and spoken corpus. Theories and tools for term extraction and analysis are also described in this chapter. All these aspects will be explained with reference to the aim of this study, which is the compilation of an English-isiNdebele glossary of medical terms.

In the previous chapter, different methodologies used by different terminologists in extracting and compiling terminology were reviewed. The researcher wanted to establish amongst others, how term candidates were identified, and whether they were manually, automatically or semi-automatically identified. The researcher also wanted to know the product (glossary, term list or dictionary) the project wanted to produce, how the product would be made available, etcetera. Based on the works by different scholars, the researcher was able to identify the type of methodologies that would contribute towards the fulfilment of the present study's aim.

3.2 RESEARCH DESIGN

This section discusses the research methods, instruments used for collecting data and the collection of both the written and spoken corpus for this study.

3.2.1 Research methods

This section is about the qualitative and the quantitative research methods used in this study. The quantitative and the qualitative methods were discussed in general, in Chapter 2. In this chapter they will be discussed with regard to the aim of this study.

3.2.1.1 The qualitative research method

The qualitative research method is used in this study for the purpose of understanding why a particular isiNdebele term is preferred and regarded as an appropriate equivalent in comparison to another term. The qualitative research method helped the researcher to

understand people concerning their own definitions of their world. Personal views regarding a particular term were also established, through this method. Participants were expected to provide in-depth, detailed information about an isiNdebele medical term they preferred. That enabled the researcher to gain an in-depth understanding (which would not have been possible through quantitative investigations) of issues surrounding the choice of a particular term, concept, process and experience. This research method ensured that the usage of a medical term is understood in its context.

3.2.1.2 *The quantitative research method*

The study utilises the quantitative method as well. This is the research method that uses numbers and statistics. This method assisted the researcher to be as objective as possible regarding the participants' choice of medical terms. The researcher was also able to decide about the medical terms that were to be prepared for verification by medical doctors.

The computer analysis tool was used in analysing corpus results in this study. The tool provides statistical information which includes word counts, frequencies and other forms of statistical information. The implementation of the tool makes the study quantitative. Terms identified by the tool were considered for the compilation of the glossary.

From what has been presented, it is evident that this research makes use of both the quantitative and the qualitative method of research. This kind of approach, where there is a mixture of both the quantitative and qualitative method, is called the triangulation approach. This approach involves the use of multiple theories, data sources and methods within the study for the investigation or researching of a single phenomenon, situation or question. This means that data is drawn from a much more diverse set of sources and this diversity ensures a more expansive look at the situation.

The triangulation approach has further helped with the integration of both questionnaires and interviews as well as the corpus linguistics and the Grounded Theory. The integration greatly assisted with the provision of a clearer understanding of the problem of this study, which is the shortage of resources and also the main aim of this study which is the compilation of the glossary.

As mentioned in the previous chapter, the present study also uses the corpus-driven approach (CDA). This approach is suitable for this study because it ensured that the researcher would be as objective as possible. The CDA prevented the researcher from approaching corpora and

the WordSmith Tools with pre-conceived ideas. The researcher wanted an approach where every conclusion would be based on the evidence which comes from the corpus at hand, and this approach was relevant.

3.2.2 Data collection instruments

To collect data for this study, open-ended questionnaires and semi-structured interviews were used.

3.2.2.1 Questionnaires

The acceptability of medical terms by isiNdebele speakers has an impact on the isiNdebele language and therefore had to be taken into consideration. Such acceptability was tested by means of English-isiNdebele medical questionnaires distributed to health workers in Mpumalanga clinics.

The researcher drafted open-ended questionnaires. The terms put into the questionnaires were sourced from the pamphlets the researcher obtained from the Mpumalanga Department of Health. The open-ended questionnaire types allowed the respondents to show some creativity. These questionnaires gave participants opportunities to develop ideas. Participants had to choose between the given responses and they also had to justify their choices.

Questionnaires were drafted and sent to Mpumalanga Province, the Nkangala District, to be precise. Nkangala district has six municipalities. The study was, however, conducted in three municipalities, namely Thembisile Hani, Dr JS Moroka and Emalahleni municipality. Out of the 18 clinics that Thembisile Hani municipality has, only ten were sampled for the study. Dr JS Moroka municipality has 28 clinics of which only ten clinics were sampled for the study and out of the ten clinics in Emalahleni, five clinics were sampled for the study.

Nkangala district has three hospitals and questionnaires were sent to all three hospitals. In this study, sampling is purposive. It is purposive because the researcher deliberately chose the isiNdebele speaking health workers to participate in the study. The health workers that were non-isiNdebele speakers were deliberately excluded from the research. The sampling method applied is purposive because the researcher approached the sampling problem with a specific plan in mind. The researcher knew the group she was seeking. She wanted only the health workers who were isiNdebele speaking. Proportionality was therefore not the researcher's primary concern. It is the experts in the field (medical) and language

(isiNdebele) which the researcher was interested in. The purposive sampling enabled the researcher to get the opinions of her target population about some terms. The challenge with this sampling method was that the isiNdebele speaking nurses, Curriculum Implementers, translators and doctors were always weighed down by their work load.

The researcher was aware that health facilities selected for this study are in different geographical locations. Some are in semi-urban and some in rural areas. She therefore had to find out if factors such as the geographical location of health facilities, qualifications of the nurses in isiNdebele, and the age of nurses would influence their choice of particular isiNdebele equivalents (terms) or not.

Questionnaires were not only sent to nurses, they were also sent to doctors, isiNdebele Curriculum Implementers (CIs) for Further Education and Training (FET) as well as the isiNdebele language practitioners who work in the parliament of South Africa.

The questionnaire, containing conditions, instructions and a list of terms was available in isiNdebele only, since the informants were all isiNdebele speakers. However, the English translations are also included in the questionnaire, for the sake of clarity.

The following is just a precept: See a full questionnaire in **Appendix E**.

ISIGABA A

IRHELO LEMIBUZO NGAMATHEMU APHATHELENE NEZAMAPHILO

Yeleda bona uhlanganyela kilelirhubhululo ngokuthanda kwakho begodu nangesikhathi sakho. Ibizo lakho alikazokufakwa kilelirhelo lemibuzo ukuqinisekisa bonyana imininingwana etholakele kilelirhubhululo ibulungeka iyifihlo. Isizo lakho njengomuntu okhuluma isiNdebele begodu nosebenza ngezamaphilo lizokuthatjelwa khulu ngombana imininingwana ozosinikela yona izokusiza ekuthuthukiseni amathemu wesiNdebele aphaathelene nezamaphilo. Nawuthanda ukuhlanganyela kilelirhubhululo, landela imilayo elandelako.

Please note that you participate in this research at your own will and in your own time. To protect your identity, you are requested to provide a pseudonym. Your participation as a mother-tongue speaker of isiNdebele and also as someone who works in Health will be highly appreciated. The information you are going to give us, will help in developing isiNdebele health terminology. If you want to participate in this research, follow the following instructions.

ISIGABA B

Kilesisigaba kunethemu le-English nethemu lesiNdebele. Faka itshwayo [√] nawuvumako, begodu ufake itshwayo [×] nawungavumiko. Lokha nawungavumelani nethemu lesiNdebele, elinikelweko, tlola elifaneleko.

In this section, there is an English term and its isiNdebele equivalent. Put a [√] where you agree and put an [×] where you disagree. Where you do not agree with the isiNdebele term provided, write the appropriate one.

English term	isiNdebele equivalent	Faka itshwayo [√] nawuvumako, begodu ufake itshwayo [×] nawungavumiko. Put a [√] where you agree and put an [×] where you disagree	Lokha nawungavumelani nethemu lesiNdebele, elinikelweko, tlola elifaneleko. Where you do not agree with the isiNdebele term provided, write the appropriate one.
--------------	-----------------------	--	--

- | | | | |
|---------------------|----------------------------------|--|--|
| 1. abdominal pain | ubuhlungu
bamathumbu | | |
| 2. accouchement | ukubeletha | | |
| 3. antibody | isivikelasifo | | |
| 4. antenatal care | ipheko lokuba
sebantwini | | |
| 5. adverse reaction | ukwaliwa sihlahla | | |

English term	isiNdebele equivalent	Faka itshwayo [√] nawuvumako, begodu ufake itshwayo [×] nawungavumiko. Put a [√] where you agree and put an [×] where you disagree	Lokha nawungavumelani nethemu lesiNdebele, elinikelweko, tlola elifaneleko. Where you do not agree with the isiNdebele term provided, write the appropriate one.
6. antiseptic (n)	isivimbela kubhibhidlha		
7. arterial blood	iingazi ezithunyelwako		

ISIGABA C

Kilesi sigaba kunamathemu wesiNdebele angaphezu kwelilodwa. Tlola itshwayo [√] lokha nawuvumelana nethemu elinikelweko, bese ufaka itshwayo [×] la ungavumelani khona nethemu elinikelweko. Ekholomini lokugcina, tlola ithemu olaziko.

In this section there is more than one isiNdebele equivalent. Put a [√] where you agree and put an [×] where you disagree. Write another isiNdebele term that you know, in the last column.

English term	isiNdebele equivalent 1	√ or ×	isiNdebele equivalent 2	√ or ×	Tlola elinye ithemu lesiNdebele olaziko. Write another isiNdebele term you know.
1. AIDS <Acronym>	intumbantonga		i-AIDS		
2. antiretroviral drug	isigogobalisi sentumbantonga		isirhobhisi sentombantonga		
3. bacterial infection	ithelelwano yomulwana		ithelelwano yebhaktheriya		
4. cholera	ikholera		ubulwele bokurhuda		
5. cervix	umlomo wesibe letho		iseviksi		
6. chronic disease	ubulwele oburwahla phazako		ubulwele obungalaphekiko		

ISIGABA D

Lesi sigaba simayelana neminingwana yomuntu ozibandakanya kilelirhubhululo. (Ubulili, iminyaka njalonjalo). Indabuko yomuntu ohlanganyela kilelirhubhululo nokobana usebenze kimiphi imikhakha.

This section consists of Demographic issues (gender, age etc.). Questions are asked in order to establish the region of origin as well as the section of the hospital or clinic where they worked.

a) Uhlala kuphi?

b) Uthome nini ukusebenza kilesisibhedlela namkha itlinigi?

Owing to the fact that the purpose of this study was to compile a bilingual glossary, all questionnaires were designed in such a way that there was an English term and its isiNdebele equivalent, that had to be verified. Questionnaires were formulated in such a way that more

terms would be elicited from the respondents. The respondents also had to explain why they did not agree with a particular isiNdebele equivalent(s). Through open-ended questionnaires, more respondents were reached.

3.2.2.2 *Interviews*

Semi-structured interviews were used in this study. The purpose of using interviews was to verify terms collected through questionnaires. In this way, the researcher would get an indication of which terms to regard as standard, that is, terms that are more commonly used, which would be accepted by isiNdebele translators and language users more easily. Through the semi-structured interviews, the inconsistent usage of medical terms was eliminated. The researcher also managed to get the views of medical doctors as to how the lack of standardised isiNdebele medical terminology has affected the development of the language. The perception and the beliefs of participants about a preferred term were also made clear, through these types of interviews.

These interview types enabled the researcher to gather as much data as possible because they allowed her to change the order of questions easily, depending on the direction of the interview. The semi-structured interviews ensured that additional questions that were not anticipated in the beginning of the interview, could still be asked. This also contributed towards the filling of the gap caused by the shortage of terminology in isiNdebele. The researcher used a tape recorder during the interview. This enabled the researcher to give a more accurate report than relying only on the fact that she was taking notes.

3.2.3 **Collection of written and spoken corpora**

Both written and spoken corpora were collected in this study.

3.2.3.1 *The written corpus*

The written medical corpora of this study were sourced from medical texts mostly downloaded from the internet and also from the SADiLaR website (2021). They also came from matric/Grade 12 question papers, the Constitution of the Republic of South Africa and other documents such as medical pamphlets and brochures. It is also important to note that the written corpora of this study are bilingual. Re-keyboarding was also used to type one medical text that was found in the Mpumalanga clinic, into computer files.

3.2.3.2 *The spoken corpus*

Spoken corpora of this study are derived from the Ikwewezi FM health program and from interviews with the isiNdebele speaking doctors. Unlike written corpora of this study which are bilingual (English and isiNdebele), spoken corpora are monolingual (isiNdebele only). The spoken corpora of this study are literacised. Refer to Chapter 2 for more details.

3.3 THEORIES FOR DATA ANALYSIS

The corpus-driven approach, the Grounded Theory and corpus linguistics theory are applied in this study. Both the written and spoken corpora were analysed through the corpus query software, the WordSmith Tools. Its licence number is: 424647885.

3.3.1 The corpus-driven approach (CDA)

The researcher chose to apply the CDA because through this approach, the English-isiNdebele corpus the researcher collected is used as evidence from which terms to be used in compiling the English-isiNdebele glossary, will be extracted.

The CDA is applied in this study because through it, the corpus of this study is used as evidence on which statements are based. Evidence plays a significant role in this study, hence the application of CDA.

3.3.2 The Grounded Theory

The Grounded Theory contains methods of comparison and difference. Through these methods of comparison and difference, the researcher managed to establish the extent to which an isiNdebele equivalent is acceptable to the end-users, in the medical domain. The Grounded Theory enabled the researcher to establish the medical terms that are preferred. The Grounded Theory was used to analyse the open-ended questionnaires of this study. Through the Grounded Theory the researcher compiled a glossary consisting of terms that are used on an everyday basis.

Amongst all the linguists and projects that the researcher reviewed in Chapter 2, none have applied the Grounded Theory.

3.3.3 The corpus linguistics theory

The present study applies the corpus linguistics theory to analyse both the written and the spoken corpus of this study. After the collection of both corpora, the next step was to identify term candidates. The study employed a semi-automatic term extraction method. The concept term extraction refers to the identification of term candidates in a text containing language for specific purposes (LSP), here a medical corpus.

The WordSmith Tools was also used for identifying term candidates. The WordSmith Tools has WordList, KeyWords and Concordance functions. The researcher experimented with all the functions, however, the KeyWords function is the most relevant function for this study. The Concordance function is also important in this study because this function ensured that the frequencies (occurrences) of a particular medical term as well as its possible, immediate environment were seen at first glance. It also supported the researcher in identifying multi-word terms. Therefore, the KeyWords and Concordance function were most relevant for the realisation of the aim of this study.

Because this is the first study in isiNdebele to be making use of the WordSmith Tools, the researcher had to manually verify the terms identified by the KeyWords function in order to establish to what extent the tool successfully identified term candidates. It has been discovered by many linguists that whether an endeavour is manual or computational, human beings will always remain the final judges.

The researcher manually verified terms identified by the KeyWords function. In the case where the tool identified fewer terms than those identified manually, the researcher had to decide whether to use the terms identified manually, through the tool or through both. Both the single and multi-terms were also identified.

Owing to the fact that the KeyWords function does not only need language for special purposes, but the language for general purposes too, the researcher also had to collect the LGP for the written corpus.

Comparison between the written and spoken corpus also took place. Terms found in the written corpus had to be compared with those in the spoken corpus. Terms collected through questionnaires had to be verified through interviews in order to improve the acceptability of terms, i.e. giving an indication of which terms are actually used and accepted in a specific technical (medical) field.

The purpose was also to establish how factors such as geographical location, age and qualification in isiNdebele influences term-formation processes. Term-formation processes as applied in isiNdebele are explained in the next chapter.

3.4 CONCLUSION

This chapter discussed the approach the researcher followed in collecting and analysing data for this study. The approach involves the research design (which are methods and instruments for data collecting) and the compilation of both the written and spoken corpus. Data analysis involves theories and tools for term extraction and analysis. All these were explained with reference to the aim of this study which is the compilation of an English-isiNdebele glossary of medical terms.

In the previous chapter, different methodologies used by different terminologists in extracting and compiling terminology were reviewed. The researcher wanted to establish amongst others, how term candidates were identified, whether they were manually, automatically or semi-automatically identified, what the product was, whether the product was a glossary, term list or dictionary and finally how the product was made available.

Based on the works by different scholars, the researcher is able to see the type of methodologies that will contribute towards the fulfilment of the present study's aim. In this chapter a description was given about how the researcher plans to apply the theories. The reason for the choice of a methodology and how it would be applied in this study has been discussed in this chapter.

Both the qualitative and the quantitative research methods are applied in this study. Through the qualitative method, the researcher came to an understanding of why a respondent would prefer a particular isiNdebele term as compared to another one. The personal views regarding a particular term were established. Through this method, the researcher managed to establish which terms different respondents would prefer to a particular term because of its insensitivity or vulgarity.

Through the quantitative research method, the researcher decided on the medical terms to prepare for verification by medical staff, Curriculum Implementers, doctors and translators. The quantitative research methodology uses numbers and statistics. The implementation of the tools makes the study quantitative. The tool provides statistical information which includes word counts, frequencies and other forms of statistical information. Terms identified by the tool are considered for the compilation of the glossary.

The study also uses the corpus-driven approach (CDA). Through this approach, every conclusion was based on the evidence which came from the corpus at hand. The open-ended questionnaires and interviews used in this study both fitted well in the aim of the study. The open-ended questionnaires allow for an infinite number of possible answers. Through them, the researcher will discover more details regarding a particular term. The questionnaires will assist in making the researcher to understand the feelings of a respondent about a particular term.

Through the semi-structured interviews, the researcher will gain understanding about a particular term. This kind of interview will enable the participants to freely express their views on a particular term. This interview type is reliable and natural. It also provides an opportunity for learning.

The researcher chose to apply the corpus-driven approach because through this approach, the English-isiNdebele corpus the researcher collected is used as evidence from which terms to be used in compiling the English-isiNdebele glossary, will be extracted. Evidence plays a significant role in this study, hence the application of the corpus-driven approach.

Through the Grounded Theory, the researcher established to what extent an isiNdebele equivalent is acceptable to the everyday users in the medical domain. Through it, the researcher established the medical terms that are preferred.

The WordSmith Tools was used for identifying term candidates of this study. The WordSmith Tools has WordList, KeyWords and Concordance functions. The researcher experimented with all the functions, however, the KeyWords function is the most relevant function for this study. The Concordance function is also important in this study because this function ensures that the frequencies (occurrences) of a particular medical term as well as its possible, immediate environment, are seen at first glance. It also supports the researcher in identifying multi-word terms. Therefore, the KeyWords and Concordance function are most relevant for the realisation of the aim of this study. Because this is the first study in isiNdebele to be making

use of the WordSmith Tools, the researcher had to manually verify the terms identified by the KeyWords function in order to establish to what extent the tool successfully identified term candidates. It is proven in this study that human beings will always remain the final judges in any terminological activity, whether that endeavour is manual or computational.

CHAPTER 4

PRESENTATION AND ANALYSIS OF DATA COLLECTED THROUGH QUESTIONNAIRES

4.1 INTRODUCTION

This chapter presents the interpretation and the discussion of data collected through questionnaires. It also presents term-formation processes applicable in the formation of isiNdebele medical terms of this study. The researcher also compares the term-formation processes of this study and those of the projects reviewed and discussed in Chapter 2. The development of isiNdebele terminology and the challenges experienced are discussed in this chapter as well.

The researcher collected data through both questionnaires and interviews. Only the data collected through questionnaires will be presented, interpreted and discussed in this chapter. Data collected through interviews as well as recordings from Ikwekwezi FM are analysed in the next chapter.

4.2 QUESTIONNAIRES

Questionnaires in this study were prepared for four different categories of respondents, namely the isiNdebele speaking nurses, isiNdebele Curriculum Implementers, English-isiNdebele translators who work in the Parliament of South Africa as well as isiNdebele speaking doctors who were not going to be available for interviews (the interviews will be discussed in the next chapter). All questionnaires were administered between July and December 2017, in the Nkangala District of Mpumalanga Province. Nkangala District has six municipalities, namely Emalahleni, Thembisile Hani, Dr JS Moroka, Victor Khanye, Emakhazeni and Steve Tshwete. The research was, however, conducted in three municipalities, namely Emalahleni, Thembisile Hani as well as Dr JS Moroka. This is because the isiNdebele speaking health workers are mostly found in these three municipalities.

Questionnaires for nurses were personally delivered to their health facilities by the researcher herself. The researcher met with the clinics' Operational Managers (OPMs) first. This was done in order to explain the purpose of this research to them and also to ensure the safe handling of questionnaires. Filled-in questionnaires were later collected from the OPMs' offices by the researcher.

The doctors were the first to be sent questionnaires; they were sent through emails. Nurses in different municipalities were the next. The questionnaires to nurses were personally delivered by the researcher. The first nurses to receive questionnaires were the ones in Emalahleni municipality, followed by the ones in Thembisile Hani and the last ones were those in Dr JS Moroka. Sixty percent of the terms were common in the three questionnaires sent to nurses, in the three municipalities. The purpose was to cover and also verify as many terms as possible.

The researcher ensured that questionnaires are not long. She had to be mindful of the fact that nurses were at work (even if they were using their lunch times to fill in the questionnaires). Very long questionnaires might have resulted in nurses just ticking terms without thorough understanding.

Open-ended questionnaires were used by the researcher to collect data. The questions were designed in such a way that the respondents would be able to agree or disagree with the isiNdebele equivalent(s) provided. Where the respondent did not agree, he/she had to provide the term he/she deemed more appropriate. Questionnaires assisted the researcher in deciding on the medical terms that had to be included in the questionnaire for interviews. Table 4.1 displays the medical pamphlets from which terms that were included in the questionnaires were sourced.

Table 4.1: *Metadata of medical pamphlets from which questionnaire terms were sourced*

Name	Publisher	Country
Sexually Transmitted Infections (STIs) English-isiNdebele	Soul City Khomanani	Republic of South Africa
Anxiety Disorders English-isiNdebele	Department of Health	Republic of South Africa
Healthy Life-style booklet: English-isiNdebele	Department of Health	Republic of South Africa
Patients' Rights Charter English-isiNdebele	Department of Health	Republic of South Africa
HIV/AIDS and Treatment English-isiNdebele	Soul City Institute Health and Development Communication	Republic of South Africa

Name	Publisher	Country
High blood pressure: English-isiNdebele	Soul City Institute Health and Development Communication	Republic of South Africa
Living positively with HIV & AIDS: English-isiNdebele	Soul City Institute Health and Development Communication	Republic of South Africa

After all questionnaires were received back, a questionnaire for the interviews was then prepared. This will be discussed in the next chapter.

In total, one hundred (100) questionnaires were distributed. Out of hundred (100) questionnaires, seventy (70) were completed. Out of the seventy (70), ten (10) were spoilt, according to the researcher since the informants agreed with all terms and did not make an effort to provide any new term they knew. This shows that they just ticked and did not give the questions any thought. Therefore, the final number of valid filled in questionnaires used in this study, are sixty (60). It needs to be emphasised that in this present study, it is not the number of questionnaires but the number of terms (that were verified) that is important.

The tables below provide full details of categories of respondents, number of questionnaires as well as the number of terms presented.

Table 4.2: *Details of doctors who participated in the study*

Doctors	Number of terms
Dr A	145
Dr B	145
Total	145

Table 4.2 shows that only 2 doctors participated in the completion of the questionnaires. The doctors were given questionnaires with 145 similar terms. These are the doctors who work in hospitals. Initially the researcher planned to interview all the doctors she contacted. These two doctors, however, indicated that they would not be available for interviews, later in the study. The researcher had to send the questionnaires through an email. Both doctors did fill in the questionnaires and emailed them back.

Table 4.3: *Details of hospitals and clinics where nurses participated*

Municipalities	Number of clinics & hospitals	Number that participated	Number of questionnaires filled in	Number of terms
Emalaheni	10 clinics	1	2	100
	2 hospitals	1		
Thembisile Hani	18 clinics	10	23	50
	1 hospital	0		
DR JS Moroka	28 clinics	10	31	50

As reflected in Table 4.3, the number of terms that were given to nurses in Emalaheni municipality were 100. As mentioned earlier, besides the two doctors, nurses in Emalaheni municipality were the first (amongst the nurses) to be given questionnaires. This municipality has ten clinics but only one clinic participated in the study. There are two hospitals in Emalaheni but only one hospital participated. Nurses in both the clinic and hospital that participated, decided to complete questionnaires in groups. This means that only two questionnaires were filled in. Thembisile-Hani municipality has 18 clinics but only ten clinics participated in the study. This municipality has one hospital, unfortunately the hospital did not participate. Dr JS Moroka municipality has 28 clinics, of which ten participated.

Table 4.4: *Details of Curriculum Implementers who participated in the study*

Curriculum Implementers	Number of terms
CI 1	100
CI 2	100
Total	100

As reflected in Table 4.4, two isiNdebele Curriculum Implementers (better known as CIs) participated in the study. These were the Curriculum Implementers in the Mpumalanga Department of Basic Education who were implementing the curricula for Further Education and Training phase (Grade 10-12). They were each given the same questionnaire containing 100 different medical terms.

Table 4.5: *Details of isiNdebele translators who participated in the study*

Translators	Number of terms
Translator 1	50
Translator 2	50
Total	50

As reflected in Table 4.5, two English-isiNdebele senior language practitioners who are in the South African Parliament (in Cape Town), participated in the study. They were each given the same questionnaire with 50 similar terms.

4.2.1 Summary of tables of respondents

As mentioned earlier, 60% of terms were similar in all questionnaires. Questionnaires were sent to doctors, nurses, Curriculum Implementers and translators. These are the terms in which there were differences or disagreements as far as their isiNdebele equivalents were concerned, for example, virus, germs, asthma, indigestion, constipation, etcetera. The total number of terms verified through questionnaires was 150.

4.2.2 Analysis of questionnaires

After data had been collected, it was analysed with as much detail as possible in order to reach certain conclusions. Questionnaire data had to be analysed in this study in order to determine the acceptability of isiNdebele medical terms by amongst others, the isiNdebele health workers, especially because they are using those terms on a daily basis.

Data analysis had to take place so as to bring order to the data and also to ensure that data make sense. The researcher had to analyse the questionnaires with the purpose of trying to give meaning to the bulk of collected data. This helped to illustrate evidence and also to expand explanations. Unrefined data had to be scrutinised and analysed in a creative and meaningful manner.

4.2.3 Theory used in analysing questionnaires

Data collected through questionnaires was analysed through the Grounded Theory. Since the theory uses the strategies of constant comparative analysis, similarities and differences were identified in the data that the researcher collected. As mentioned in the literature review, under 2.2.2.1, the Grounded Theory has three coding techniques, namely: the open, the axial and

the selective coding. These coding techniques were applied in this chapter and the techniques relevant for this study were also established.

4.2.3.1 *Open coding*

Only nurses who were in the three municipalities were given questionnaires (Emalahleni, Thembisile Hani and Dr JS Moroka). For those municipalities, codes were created as indicated below.

The letter 'N' is used throughout and it represents 'nurse'. For instance, for Thembisile Hani nurses, the code 'NTH' is used. This code stands for a nurse in Thembisile Hani. The number of nurses who filled in questionnaires in this municipality is 23. For the nurses in Dr JS Moroka the code 'NJSM' is used which stands for a nurse in JS Moroka. The number of nurses in this municipality who filled in questionnaires is 31. For Emalahleni municipality, the code is 'NE' which stands for a nurse in Emalahleni. In Emalahleni, as mentioned before, only two questionnaires were filled in by two groups of nurses. Therefore, the code used is NE1 and NE2.

As mentioned earlier, the second group that participated in the completion of questionnaires was the isiNdebele Curriculum Implementers for Further Education and Training. Two Curriculum Implementers participated; the code is therefore CI1 and CI2. The third group that participated in the filling in of questionnaires was translators. They were coded as T1 and T2. This means that there were two isiNdebele translators that participated in the filling in of questionnaires. The fourth group was the two isiNdebele speaking doctors. They were coded as D1 and D2.

Analysis of responses to questionnaires

Below follows an interpretation of responses of all the respondents who participated in the research study. The following terms and their responses are selected as being representative of the terms that the researcher received.

The first two research questions formulated in 1.2 are considered here, namely:

- How does the paucity of standardised terminology in specialised subject fields negatively impact the development of isiNdebele?
- What role can corpus-driven lexicography play in filling the gap caused by the shortage of terminology?

The questionnaire was made up of two sections, namely Section A and Section B which are described in detail below.

SECTION A

In this section, there is an English term and its isiNdebele equivalent. Respondents were instructed to put a tick [✓] where they agreed with the equivalent provided and a cross [×] where they disagreed with the equivalent provided. Where they did not agree with the isiNdebele term provided, they were expected to provide the appropriate one. The focus here was on the respondents who disagreed with the term provided and also those who added new terms. The analysis is as indicated below.

- Accouchement > *ukubeletha*

NTH13, NTH7, D1, D2 and NTH23 added the term *ukubhebhula*.

Analysis: The term *ukubeletha* means 'to give birth' in isiNdebele. The term *ukubhebhula* is preferred to the term *ukubeletha* because it is euphemistic. The term *ukubhebhula* loosely translates as 'to carry on the back'. D1 agrees with the term *ukubeletha* and also *ukubhebhula* but also added the term *ukuzala*.

- Antibody > *isivikelasifo*

NTH15 added the term *isilwisasifo* loosely translated as 'disease fighter'. NTH9 and NTH8 are in agreement with the term *isilwisi-anthijeni* loosely translated as 'antigen fighter'. NTH10 refers to antibody as *isilwi ne-anthijeni* that loosely translates as 'that which fights with antigen'.

Analysis: From the above responses, it is clear that most respondents preferred the terms that translate an 'antibody' as that which 'fights the disease' rather than that which prevents the disease.

- Antiseptic > *isivimbelakubhibhidlha*

NJSM23 and NJSM12 do not agree with the provided equivalent, *isivimbelakubhibhidlha* but prefer the term *isivimbelamagciwana*, which is loosely translated as ‘germs preventer’ as well as *isivimbelamalwele* loosely translated as ‘diseases-preventer’. NJSM12 prefers the term *isivimbelangogwana* which loosely translates as ‘virus preventer’. D1 added the term, *i-anthisepthiki*.

Analysis: It is clear that as much as most participants agree with *isivimbelakubhibhidlha* loosely translated as ‘that which prevents sepsis’, the term *isivimbelamagciwana* is also added. *Isivimbelamagciwana* also makes sense because the reason why there is ‘sepsis’ is because of the presence of ‘germs’.

- Asthma > *isifubasomoya*

NTH11 added the term *ukuvaleka* ‘to be closed’, NE1 disagrees with the term *isifubasomoya* which is translated as ‘chest air’ but instead she prefers the term, *isifosesifuba* loosely translated as ‘chest disease’, NE2 agrees with NE1 on *isifosesifuba*, NTH9 and D1 disagree with the term *isifubasomoya* and refer to asthma as *i-asma*.

Analysis: When one looks at the provided equivalent, *isifubasomoya*, one realises that the term is semantically incorrect. *Isifuba* is a ‘chest’ and *ummoya* is ‘air’. This term therefore means ‘asthma’ is the ‘chest of the air’. The air does not have a chest. The appropriate equivalent for asthma is therefore, *i-asma* and *isifosesifuba*. To avoid confusion, these two terms should be used.

- Bacteria > *igciwana / ingogwana*

T1 and NTH11 prefers the term *umulwana*, NJSM21 refers to it as *ibhakteriya*.

Analysis: There is confusion as far as the isiNdebele equivalent for ‘virus’, ‘germ’ and ‘bacteria’ is concerned. The isiNdebele equivalents are used interchangeably. For the purpose of this study the term ‘bacteria’ will be referred to as *igciwana*, ‘virus’ will be referred to as *ingogwana* and the term ‘germ’ will be referred to as *umulwana*.

- Burn out > *Ukuphelelwa mamandla*

D2 and D1 do not agree with the term *ukuphelelwa mamandla* instead they prefer the term *ukuphelelwa likareko* loosely translated as 'losing the zeal'.

Analysis: The two isiNdebele terms *ukuphelelwa mamandla* 'to lose power' and *ukuphelelwa likareko* actually mean one and the same thing. They both refer to the state of emotional, mental and physical exhaustion brought on by prolonged or repeated stress.

- Cancer > *ikankere*

NJSM22 disagrees with the given term *ikankere*, borrowed from Afrikaans 'kanker' but prefers the term *umdlavuza*. NJSM23, NJSM22 and NTH1 agree with the equivalent provided but also added the term *umdlavuza*. D1 and CI2 prefer the term *ikhenza* which is borrowed from the English 'cancer'.

Analysis: Most participants agreed with the equivalent provided, *ikankere*. Some preferred the term *umdlavuza*, which is a term also used in isiZulu. Whilst some prefer the Afrikaans loan word *ikankere*, others prefer *ikhenza* which is loaned from English. The latter term is also spelled as *ikhensa* by some. Both *ikankere* and *ikhenza* are therefore accepted.

- Chicken pox > *iintudla*

D2, NJSM12 and D1 disagree with the term *iintudla*, instead they prefer the term *ipoksi* for 'chicken pox'. NJSM3 refers to it as *iwaterpoksi* loaned from Afrikaans 'waterpokkies'.

Analysis: The term *iintudla* is a coined term for 'chicken pox'. The use of the term *iwaterpoksi* and *ipoksi* confirms that Afrikaans has influenced isiNdebele much. Both terms are loaned from the Afrikaans term 'waterpokkies'.

- Constipation > *ukubhinjidelwa*

CI1, NJSM16, NJSM7, NTH11, NJSM11 and NTH6 disagree with the term provided, *ukubhinjidelwa* but they prefer the term *ukuparelwa*. NTH15 prefers the term *ukuqurhelwa* and *ukubophana kwesisu* to *ukubhinjidelwa*.

Analysis: There are three isiNdebele equivalents used for the term 'constipation', as mentioned above. It needs to be mentioned that the term 'constipation' and 'indigestion' are translated as one term whereas they are actually two terms or rather two medical conditions. In isiNdebele, both 'constipation' and 'indigestion' are referred to as *ukubhinjidelwa* and *ukuparelwa*. For the purpose of this study, the term *ukubhinjidelwa* will be used to refer to 'constipation' and for 'indigestion', the isiNdebele terms *ukuqurhelwa* and *ukuparelwa* will be used.

- De-oxygenated blood > *iingazi ezingana-oksijini*

NJSM7 disagrees with the equivalent provided but added the term *iingazi ezinganamoya ohlwengekileko* loosely translated as 'the blood that does not have clean air'. NJSM13 added the terms *iingazi ezingakahlanzeki* and also *iingazi ezifeyila*. Both terms loosely translate as 'blood which is not clean'.

Analysis: Most participants agree with the term *iingazi ezingana-oksijini*. Other terms were added such as *iingazi ezinganamoya ohlwengekileko* which means 'blood that does not have clean air or blood that is not clean'. This means that 'de-oxygenated blood' is blood which is not clean.

- Depression > *ukuba phasi emmoyeni*

D1 and D2 disagree with the provided equivalent *ukuba phasi emmoyeni* 'to be down in the spirit', but instead prefer the term *igandeleleko ngokomkhumbulo* loosely translated as 'having pressure of the mind'.

Analysis: When one looks at the two isiNdebele terms namely, *ukuba phasi emmoyeni* and *igandeleleko ngokomkhumbulo*, one realises that the two terms actually mean one and the same thing. To say one is down in the spirit, does not differ from the fact that one is experiencing some mental pressures.

- Diabetes > *ubulwele beswigiri/ubulwele betjhukela*

NJSM23 added the term *itjhukela* which means 'sugar'. NJSM17 agrees with the two equivalents provided but also added the terms *itjhukela* (sugar) and *iswigiri* (sugar).

Analysis: Whilst most participants use the term *ubulwele beswigiri/ubulwele betjhukela* 'the disease of sugar', it is interesting to come across the term *itjhugela/iswigiri* which is 'sugar'. In other words when a person is diabetic, it is just said *unetjhukela/uneswigiri* 'he/she has sugar'.

- Diarrhoea > *irhudo*

NTH13, D1, D2 and NTH10 added the term *ukuthulula*, loosely translated as 'to spill'. NE2 refers to 'diarrhoea' as *umrhudo*. T1 prefers the term *ukurhuda* for 'diarrhoea'. NTH11 refers to it as *ukujejeza*. D2 also added *ubulwele bokuthulula* which is euphemistic. D1 also added the term *ukurhuda*.

Analysis: Both terms, *ukuthulula* and *ukurhuda* are actually preferred. The term *ukuthulula* is mostly preferred because it is a euphemism. *Ukuthulula* is perceived to be a milder term than *ukurhuda* or *irhudo* which are perceived to be harsher. The term *irhudo* or *ukurhuda* are therefore regarded as vulgar.

- Eczema > *i-ekzima*

NTH10 and T1 added the term *ikhwekhwe*; NJSM23 refers to 'eczema' as *iratjhi esabutjha* loosely translated as 'a rash with burning sensation'. NTH15 refers to the term 'eczema' as *isibabo*, NTH14 added the term *ubulwele besikhumba*. NJSM6 refers to it as *iratjhi*.

Analysis: The coined isiNdebele equivalent for 'eczema' is *ikhwekhwe*. The term *isibabo* is a term loaned from Sesotho sa Leboa *sebabo*. The *ubulwele besikhumba* 'the disease of the skin' is a bit generic. The term *iratjhi* 'rash' does not refer to 'eczema'. There is a difference between 'eczema' and a 'rash'. They are both skin conditions but different.

- Epilepsy > *isithunthwana*

NTH4 disagrees with the equivalent term provided but instead prefers the term *isifo sokuwa* loosely translated as 'the disease of falling'. D2, NJSM23, NJSM17, T1, NJSM23 and C11 added the term *ubulwele bokuwa* which still means 'the disease of falling'. CI1 and D1 added the term *i-ephilepsi*, the loan term.

Analysis: Most participants agreed with the term *isithunthwana* but also added *ubulwele bokuwa* and *isifo sokuwa*. There are also those that prefer to use the term *i-ephilepsi*. It was noticed that the term 'epilepsy' was spelled differently. Some respondents spelled it with an -h- and others without, that is, *i-ephilepsi* and *i-ephilepsi*.

- Germ > *umulwana*

T1 prefers the term *ingogwana yobulwele*, NJSM21 refers to it as *imbewana yobulwelwe*. NTH11 disagrees with the term *umulwana* but instead prefers the term *igciwana*.

Analysis: The terms 'virus' and 'germ' are analysed simultaneously. This is because the two terms 'germ' and 'virus' are mostly translated as one term in isiNdebele. Some Ndebele speakers use the term *ingogwana* when referring to a 'germ' and *umulwana* for a 'virus'. Others are unaware which term denotes a 'germ' and which one a 'virus'.

There are three equivalents of 'virus', namely *umulwana*, *ingogwana* and *ivayirasi*. There is confusion as far as the isiNdebele equivalent for 'virus' and 'germ' is concerned. The isiNdebele equivalents are used interchangeably. For the purpose of this study the term 'virus' will be referred to as *ingogwana* and the term 'germ' will be referred to as *umulwana*.

- Gonorrhoea > *igonoriya*

NE2 agrees with the provided equivalents but also adds the term *isifo esithathelana ngokomseme* which loosely translates as 'the disease transmissible through the reed mat'. NTH10 disagrees with the provided equivalent, *igonoriya*. Instead she prefers the term *ubulwele obuthathelana emabhayini* which loosely translates as 'the disease which is transmissible through the blankets'. NJSM3 added the term *idrophu*.

Analysis: The two terms, *isifo esithathelana ngokomseme* and *ubulwele obuthathelana emabhayini* actually refer to one thing. Both refer to sexually transmitted diseases. The two paraphrased terms are a bit generic. They do not necessarily talk to the term 'gonorrhoea'. The loaned term *igonoriya* is a better equivalent.

- Groin > *imbilapho*

NTH15, NJSM17 and D2 disagree with the equivalent *imbilapho* for 'groin'. They maintain that *imbilapho* is 'lymph node' in English and not a 'groin'.

Analysis: A very interesting finding here is that the isiNdebele equivalent *imbilapho* is actually a 'lymph node' and not a 'groin'. A 'groin' is a body part and not a medical condition. It is that part between the abdomen and the upper thigh on either side of the body. It was therefore incorrect to refer to 'groin' as *imbilapho*.

- High blood pressure > *isigandelelo seengazi esiphakamileko*

T1 prefers the term *umfutho weengazi ophakamileko* which loosely translates as 'the pressure of the blood which is high'. NTH6, D1, NJSM7, NTH1 and D2 added the loan term *ihayibhladi*. NJSM7 and NTH1 added the term *ihayi-hayi*.

Analysis: The provided term *isigandelelo seengazi esiphakamileko* which translates also as 'the pressure of the blood which is high' is appropriate. It is a coined term, however, the terms *ihayibhladi* 'hypertension' and *ihayihayi* 'high high' are used on a day to day basis in health institutions. More participants preferred them to the coined terms, *isigandelelo seengazi esiphakamileko* and *umfutho weengazi ophakamileko* which also translate as 'the pressure of the blood which is high'.

- Incubator > *ibhodlelo*

NTH10, T1, NTH6 and NJSM23 added the term *isifukamiseli*.

Analysis: The term 'bottle' *ibhodlelo* for 'incubator' is the term used more than any other term for an 'incubator'. The term *isifukamiseli* does explain what an 'incubator' is. To 'incubate' is *ukufukamela* in isiNdebele. This makes the term *isifukameli* to be more accurate than *ibhodlelo* 'bottle' which might be misleading.

- Medicine > *isihlahla*

NTH10 added the term *imitjhoga* for 'medicines'. NTH3, C11 and NTH7 agree on the term *intatha*. C11 and D2 agree on the term *umtjhoga* which is a singular form of the term *imitjhoga*.

Analysis: Although the terms *umtjhoga*, *intatha*, *ipengu* and *isihlahla* refer to 'medicine', the two terms *intatha* and *ipengu* are used mostly to refer to 'traditional medicine'.

- Pap smear > *ipapsmiye*

CI1, CI2, D1 and D2 added the term *ukuhlolwa komlomo wesibeletho*.

Analysis: In real, daily conversations, the term *ipapsmiye* 'pap smear' is mostly used, as compared to the paraphrased *ukuhlolwa komlomo wesibeletho*, which loosely translates as 'the screening or testing of the cervix'.

The terms 'miscarriage', 'menstruation', 'pap smear' and 'accouchement' belong to the 'gynecology' or 'female ailments' category.

- Pulse > *ibetho*

T2 agrees with the two equivalents provided but adds the noun, *ihliziyo*. She prefers the term *ibetho lehliziyo* as well as *ukupompa kwehliziyo* for 'pulse' while NTH4, NE2 and NTH9 concur with T2 only on the term *ibetho lehliziyo*. NE2 added the term *iphalsi* loaned from English.

Analysis: It is interesting that most participants qualified the term *ibetho* by adding the noun, *ihliziyo* 'heart'. This is actually proper because to say *ibetho* 'beat or pulsation' only, is vague.

- Vaccine > *umu-endo*

D2 added the term *ukuhlatjelwa ubulwele obuthileko* loosely translated as 'to be injected for a particular disease'. D1 added the term *ivaksini*, which is a loan term. NTH22 added the term *ihlabelo* 'vaccination'.

Analysis: All isiNdebele alternative equivalents of the term 'vaccine' are relevant. The paraphrased isiNdebele equivalent, *ukuhlatjelwa ubulwele obuthileko*, *ivaksini* and *ihlabelo* are acceptable. They are all talking to the English term, 'vaccine'.

- Viral load > *ubungako be-HIV*

D2 prefers the term *ubungako bevayirasi* 'viral load'. D1 on the other hand prefers the term *ivayiralilowudu*, a transliterated term.

Analysis: The term *ubungako bevayirasi* better explains the term 'viral load'. The term gives a clearer understanding of what 'viral load' stands for. It is actually more informative as compared to the transliterated *ivayiralilowudu*. This is because a person who does not know what a 'viral load' is, may not be assisted by the term *ivayiralilowudu* however *ubungako be-HIV* that loosely translates as 'the amount of HIV' and *ubungako bevayirasi* 'the amount of virus' give clarity.

- Virus > *ivayirasi*

NTH14 and T1 say that the equivalent for 'virus' is *umulwana* whilst NTH11, NTH3, NTH7 and NTH9 say it is *ingogwana*. T2 refers to it as *umulwana othathelwanako* which loosely translates as 'an infectious virus'.

Analysis: As mentioned above, when the term 'germ' was analysed, the term *ingogwana* will be used in this study to refer to 'virus'.

SECTION B

In this section there is more than one isiNdebele equivalent. Respondents were instructed to put a tick [✓] where they agree and a cross [×] where they disagree. They were requested to write another isiNdebele term that they know in the last column. The following responses were received regarding this section:

- Anemia > *i-anemiya/ukuthayela kweengazi*

NTH9 added the term *emzimbeni* which then becomes *ukuthayela kweengazi emzimbeni*. T1 prefers the term *ithayelongazi emzimbeni*, both translations mean 'the shortage of blood in the body'.

Analysis: It means there is an agreement between participants regarding the term *ukuthayela kweengazi emzimbeni* which translates as 'the shortage of blood in the body'.

- Antiretroviral drug > *Isigogobalisi sentumbantonga/isirhobhisi sentumbantonga*

D1 and D2 prefer the term *ama-ARV* 'The ARVs'.

Analysis: Although the two provided equivalents explain the term ‘antiretroviral drug’ well, the term *ama-ARV* is actually mostly used on a daily basis. The coined terms *isigogobalisi sentumbantonga* as well as *isirhobhisi sentumbantonga* may not be in use as compared to the abbreviated, *ama-ARV*.

- Aorta > *i-ayotha/umthambo omkhulu othumelako*

NTH10 and T1 added the term, *umthambo omkhulu wehliziyo* which loosely translates as ‘the big artery or vein of the heart’.

Analysis: The term *umthambo omkhulu wehliziyo* is a better equivalent as compared to *umthambo omkhulu othumelako* which loosely translates as the ‘big artery that sends’.

- Cholera > *ubulwele bokurhuda/ikholera*

NE4 added the term *indeni ebovu*. NTH15 refers to ‘cholera’ as *ubulwele bokuthulula* which loosely translates as ‘the disease of the diarrhoea’.

Analysis: The term ‘cholera’ and ‘diarrhoea’ are translated as one term in isiNdebele equivalents. The euphemism *ukuthulula* is preferred and regarded as less harsh as compared to *ukurhuda* which is regarded as vulgar and harsh.

- Constipation > *ukubhinjidelwa/ukuqurhelwa*

NTH9 disagreed with the term *ukubhinjidelwa* but agrees with the term *ukuqurhelwa*. She argues that *ukubhinjidelwa* means to be ‘bloated’. NTH7 disagreed with both equivalents and added *ukufasa/ukubopha kwendeni* loosely translated as ‘tightening of the tummy’. NJSM8, D1, D2, C1 and C2 added the term *ukuparelwa*.

Analysis: Considering the above, it is clear that the terms ‘constipation’ and ‘indigestion’ are translated as one term in isiNdebele. In this study the term ‘constipation’ will be referred to as *ukubhinjidelwa* and ‘indigestion’ as *ukuqurhelwa* and *ukuparelwa*.

- Germ-cell > *iseli lomulwana/iseli lembewana*

NTH9 added the term *iseli legciwana*.

Analysis: This shows that according to this respondent, an equivalent for 'germ' is *igciwana*. T1 prefers *iseli lengogwana lobulwele* which loosely translates as 'the cell of the virus of the disease'.

- Haemorrhage > *ukopha/ihomoreji*

T1 added the term *ukubhluya*; this is a term borrowed from Afrikaans 'om te bloei' (to bleed). NTH10 added the term *ukukghuthuka kweengazi* loosely translated as 'the pouring out of blood'; this emphasises the degree to which a person loses blood. NTH9 added the term *ukopha kweengazi zomthambo* loosely translated as 'bleeding of the veins'.

Analysis: Equivalent terms added indicated the quantity of blood loss. This is emphasised by NTH10 who referred to it as *ukukghuthuka kweengazi* which loosely translates as 'the pouring of blood'.

- Hypodermic syringe > *isirinji/isipeyidi*

D2 added the term *isirinji esihayiphodemikhi* 'hypodermic syringe'. D1 disagrees with the provided term, *isipeyidi* (which is a term borrowed from Afrikaans 'spuit') simply because it is a generic term.

Analysis: The participants provided the transliterated terms. There is no coined isiNdebele term provided for the hypodermic syringe. Even the two terms provided *isirinji* 'syringe' and *isipeyidi* 'spuit' are also not the coined terms.

- Indigestion > *ukubhinjidelwa/ukuqurhelwa*

NJSM24 added the term *ukuqunjelwa*. NJSM3 added the term *ukuparelwa*.

Analysis: As mentioned above, the term 'constipation' will be referred to as *ukubhinjidelwa* and 'indigestion' as *ukuqurhelwa* and *ukuparelwa*.

- Menstruation > *ukuba senyangeni/ukuba sesikhathini*

D2, D1, C1, C2, T1, T2, TH1-TH10 added the term *ukuba semalangen*, loosely translated as 'to be in the days'. D1 also added the term *ukumensa*, which is a loan term.

Analysis: All the terms *ukuba senyangeni* 'to be in the month', *ukuba sesikhathini* 'to be in the time', *ukuba semalangeneni* 'to be in the days' and *ukumensa* 'to menstruate' are all used, and equally so.

- Miscarriage > *ukubuya endleleni/ukonakalelwa*

D2 added the term *ukuphunyelwasidisi* 'the coming out of pregnancy' and *ukuphunyelwambungu* 'the coming out of a fetus'.

Analysis: The two terms *ukuphunyelwasidisi* and *ukuphunyelwambungu* are appropriate. They both refer to miscarriage. The provided equivalents *ukubuya endleleni* 'to come back from the road' and *ukonakalelwa* 'to have your things spoilt or destroyed' are actually euphemisms as compared to *ukuphunyelwasidisi* and *ukuphunyelwambungu*.

- Ulcer > *i-alsa/isilonda sangendeni*

NTH9 disagrees with the second equivalent *isilonda sangendeni*. She prefers the term *isilonda esivulekileko* which means 'an open sore'.

Analysis: In fact, NTH9's argument is correct. To say *i-alsa* is *isilonda sangendeni* is misleading because the term *isilonda sangendeni* refers to 'stomach ulcer'. The term 'ulcer' refers to any 'open sore'. There is also a 'mouth ulcer' which is *isilonda esivulekileko sangemlonyeni*. Therefore, the term 'ulcer' should be translated as *isilonda esivulekileko* and *i-alsa* unless it is specified as to what kind of an 'ulcer' it is.

- Virus > *ivayirasi/umulwana othathelanako*

NTH9, NJSM6 and NTH7 disagree with the provided equivalent, they instead refer to the term 'virus' as *ingogwana*.

Analysis: As mentioned before, the term 'virus' and 'germ' are used interchangeably. For the purpose of this study, the term *ingogwana* will be used for the term 'virus', and the term *umulwana* will be used for the term 'germ'. See also the earlier explanation on 'virus' and 'germ' under Section A.

Interpretation of collected terms (Open Coding)

When one looks at the above, one realises that the method of difference and the method of agreement are applicable in the terms as filled in by various respondents of the questionnaires. A variety of term-formation strategies have been used to form isiNdebele terms. Transliterated terms are preferred to coined terms. Term-formation processes are discussed in detail in 4.3.

4.2.3.2 Axial coding

In open coding, medical terms were merely identified but it is in axial coding that the researcher placed each term in a particular category (theme or sub-theme). Axial coding is about links and relationships. Through the axial coding, data that was fractured during open coding is re-assembled. Categories are related to their sub-categories to form more precise and complete explanations about the phenomena. No analysis will be done here, otherwise there will be repetitions. Terms are organised in terms of a sense relation of superordinate terms versus hyponyms. The following five categories are discussed below.

(a) Primary Health Care

anemia > *ukutlhayela kweengazi*
antibody > *isivikelasifo*
antiseptic > *isivimbela-kubhibhidlha*
asthma > *isifuba somoya*
aorta > *umthambo omkhulu*
ulcer > *isilonda esivulekileko*
hypodermic syringe > *isirinji/isipeyidi*
haemorrhage > *ukopha/ihomoreji*
cholera > *ikholera*
germ-cell > *iseli lomulwana*

(b) Gynecology

miscarriage > *ukubuya endleleni/ukonakalelwa*
menstruation > *ukuba senyangeni*
accouchement > *ukubeletha*
pap smear > *ukuhlolwa komlomo wesibeletho*

(c) Mental Health

epilepsy > *ubulwele bokuwa*

burn out > *ukuphelelwa likareko*

depression > *ukugandeleleka ngokomkhumbulo*

(d) HIV/AIDS

virus > *ingogwana*

antiretroviral drug > *isigogobalisi sentumbantonga/isirhobhisi sentombantonga*

viral load > *ubungako be-HIV*

PCR test > *ukuhlolwa kwe-PCR*

(e) Sexually Transmitted Disease

gonorrhoea > *igonoriya*

genital herpes > *iinlonda zezitho zobulili/amatjhatjhazi wezitho zobulili*

genital warts > *isupa yezitho zobulili*

In axial coding, terms are now grouped and put under a particular category. For instance, in the case of the term 'miscarriage', one knows that this term is applicable only to women, hence it is grouped together with terms such as 'accouchement', 'menstruation', etcetera. In terms of sense relation, the terms, 'accouchement', 'menstruation', 'miscarriage' and 'pap smear' are hyponyms to the superordinate gynecology. Terms such as 'depression', 'epilepsy' and 'burn out' are hyponyms of the superordinate 'mental health'. Also, terms such as 'viral load' and 'antiretroviral drug' are hyponyms of the superordinate 'HIV/ AIDS'.

Table 4.6: Order of preference of the use of isiNdebele medical terms (Open & Axial Coding)

Term	English	isiNdebele			
		First	Second	Third	Fourth
1	virus	ivayirasi (53)	ingogwana (51)	umulwana (2)	umulwana othathelanwako (1)
2	diabetes	ubulwele beswigiri (58)	ubulwele betjhukela (57)	itjhukela/iswigiri (2)	
3	eczema	i-ekzema (58)	ikhwekhwe (2)	iratjhi (1)	iratjhi esabutjha (1)
4	asthma	i-asma (58)	isifubasomoya (50)	isifosesifuba (2)	
5	epilepsy	isithuthwana (52)	isifo sokuwa/ubulwele bokuwa (5)	i-ephilepsi (3)	ubulwele bokutsirimezeka (1)
6	cancer	umdlavuzi (59)	ikankere (56)	ikhenza (54)	
7	antiseptic	isivimbela kubhibhidlha (57)	isivimbela-magciwana (1)	isivimbela-malwele (1)	isivimbela- ngogwana (1)
8	constipation	ukubhinjidelwa (56)	ukuparelwa (46)	ukuqurhelwa (1)	ukubophana kwesisu (1)
9	germ	umulwana (50)	imbewu yobulwele (40)	umulwana obanga ubulwele (1)	
10	high blood pressure	isigandelelo seengazi esiphakamileko (57)	ihayibhladi (54)	ihayi hayi (3)	
11	de-oxygenated blood	iingazi ezingenaoksijini (40)	iingazi ezingenamoya ohlwengekileko (1)	iingazi ezingakahlanzeki (1)	iingazi ezifeyila (1)
12	cholera	ikholera (59)	ubulwele bokurhuda (56)	ubulwele bokuthulula (1)	
13	medicine	isihlahla (48)	umtjhoga (3)	intatha (2)	ipengu (1)
14	indigestion	ukubhinjidelwa (50)	ukuqurhelwa (48)	ukuparelwa (5)	ukuqunjelwa (1)

Term	English	IsiNdebele			
		First	Second	Third	Fourth
15	anemia	ukuthayela kweengazi (48)	itlhayelongazi (38)	i-anemiya (30)	
16	antibody	isivikelasifo (53)	isivikelamalwele (39)	isilwisamalwele (35)	i-anthibhodi (30)
17	ulcer	i-alsa (56)	isilonda esivulekileko (50)	iisilonda sangendeni (46)	
18	hypodermic syringe	isirinji (58)	isipeyidi (38)		
19	haemorrhage	ukopha (50)	ihomoreji (45)		
20	miscarriage	ukubuya endleleni (54)	ukonakalelwa (50)	ukuphunyelwa sidisi (48)	
21	menstruation	ukuba sesikhathini (58)	ukuba semalangenani (56)	ukumenza (46)	
22	pap smear	ipapsmiye (56)	ukuhlolwa komlomo wesibeletso (54)		
23	accouchement	ukubeletha (53)	ukubhebhula (50)		
24	depression	ukugandeleleka ngokomkhumbulo (56)	idiphreshini (46)	ukuba phasi emmoyeni (4)	
25	genital herpes	iinlonda zezitho zobulili (58)	amatjhatjhazi wezitho zobulili (30)		

It should first be noted that not all terms are presented in Table 4.6. The purpose of the table is to illustrate the order in which terms are being ranked as informed by the data the researcher got from the questionnaires. The researcher manually counted the number of occurrences of the various terms in the questionnaires in order to determine the order of preference. The table also reflects the various term-formation processes applicable in the formation of the above terms, the isiNdebele equivalents. Most of the terms in Table 4.6 are formed using the paraphrasing technique. For example, the term 'miscarriage' is rendered as *ukubuya endleleni* which loosely translates as 'to come back from the road' and *ukonakalelwa mbungu* which loosely translates as 'to have your fetus spoil' and *ukuphunyelwa sidisi* which loosely translates as 'the coming out of pregnancy'. A further example is the term 'menstruation' which is rendered as *ukuba sesikhathini* which translates as 'to be in the time' and *ukuba senyangeni* which translates as 'to be in the month'. One realises that paraphrasing has been extensively used to create these terms. Term-formation processes will be discussed in detail in 4.3.

4.2.3.3 *Selective coding*

For this study, it was not necessary to implement selective coding. The aim of this study is to compile a bilingual glossary of medical terms. The open coding was sufficient to establish terminology acceptable to health workers in the health fraternities. It was also enough to categorise terms identified through the opening coding, into their respective categories (through axial coding). Therefore, selective coding does not contribute in any way to the aim of this study. The following are term-formation processes applicable to isiNdebele medical terms. The theories regarding term-formation processes were discussed in Chapter 2.

4.3 TERM-FORMATION PROCESSES

Primary and secondary term-formation processes as applicable in isiNdebele, are discussed in this section.

4.3.1 Primary term-formation strategies (foreign) in isiNdebele

Primary term-formation strategies include paraphrasing, borrowing, derivation, deideophonisation and abbreviation.

4.3.1.1 Paraphrasing

Table 4.7: Examples of paraphrasing

English	isiNdebele
dentist	<i>udorhodere wamazinyo</i>
gynecologist	<i>udorhodere wabafazi</i>
cholera	<i>ubulwele bokurhuda</i>
depression	<i>ukugandeleleka ngokomkhumbulo</i>
antibody	<i>isivikela sifo</i>

The term 'gynecologist' will be easily understood by an isiNdebele speaker when it is referred to as *udorhodere wabafazi* rather than *igayinakholoji*. The same applies to the concept 'antibody'. IsiNdebele speakers will understand the term *isivikela sifo* better than *i-anthibhodi*.

4.3.1.2 Borrowing

Borrowing is classified into:

(a) Direct loan

Table 4.8: Example of direct loan

English term	isiNdebele term
HIV	<i>i-HIV</i>

When one looks at the above example, one realises that the English term is taken as it is (with its English spelling). Only the isiNdebele class 9 noun class prefix has been added.

(b) Coinage

Table 4.9: Examples of coinage

English term	isiNdebele term
AIDS	<i>umbulalasihlangu/intumbantonga</i>
high blood pressure	<i>umtjhisu weengazi</i>

(c) Transliteration

Table 4.10: *Examples of transliteration*

English term	Transliteration	Coinage
cancer	<i>ikhensa/ikhenza</i>	<i>umdlavuza</i>
AIDS	<i>i-eyidzi</i>	<i>intumbantonga/umbulalasihlangu</i>
virus	<i>ivayirasi</i>	<i>ingogwana</i>
high blood (pressure)	<i>ihayibhladi</i>	<i>ukugandeleleka kweengazi/ umtjhiswo weengazi</i>

Table 4.10 reflects that isiNdebele speakers prefer the transliterated terms to their coined counterparts. For example, the term *ihayibhladi* is preferred to its coined equivalent *ukugandeleleka kweengazi* or *umtjhiswo weengazi*. This is because the transliterated lexical items are short, straight to the point and easy to comprehend as opposed to the coined terms which are very lengthy and difficult to understand because the manner in which they are coined is logical but complex.

As mentioned in Chapter 2, transliteration is the most productive method of developing terms, but for isiNdebele to fully develop, it must strive for the coinage of terms. Relying on transliteration seriously impacts the development of isiNdebele especially because most transliterated terms are inconsistently spelled.

4.3.1.3 Derivation

Table 4.11: *Examples of derivation*

English term	isiNdebele term	Derivation
patient	<i>isiguli</i>	isi- class 7 noun prefix -gul- verbal root -i personal deverbative suffix
childbirth	<i>ukubeletha/ ukubhebhula</i>	uku- class 15 prefix -beleth- verbal root -a verbal suffix

Isiguli is a deverbative formed from the verb *ukugula*. A class 7 noun prefix was prefixed to the verbal root *-gul-*, then the suffix *-a* of the verb stem was replaced by the personal deverbative suffix morpheme *-i*.

4.3.1.4 Deideophonisation

Table 4.12: *Example of deideophonisation*

English term	isiNdebele term	isiNdebele ideophone
phlegm/sputum	<i>isikhohlela</i>	<i>khohlo</i> '(to) cough'

In the case of *isikhohlela*, the noun is derived from the ideophone *khohlo* which indicates the sound of coughing. Thus, *isikhohlela* refers to 'a substance or product resulting from the sound of a cough'. In this process an object is given a name resembling its sound.

4.3.1.5 Abbreviation

Abbreviation has sub-categories of acronyms, clipping and blending.

(a) Acronyms

Table 4.13: *Examples of acronyms*

English	isiNdebele	Term in full
ELISA	<i>i-ELISA</i>	Enzyme Linked Immuno Sorbent Assay
TB	<i>iTB</i>	Tuberculosis

In both the examples, class 9 prefixes have been added, for example TB = *iTB*. The first example confirms that the hyphen in isiNdebele is also used for practical purposes, that is, to separate two adjoining vowels.

(b) Clipping/Shortening

Table 4.14: *Examples of clipping*

English	isiNdebele
gynaecologist	<i>i-gayini</i>
laboratory	<i>i-lebhu</i>
influenza	<i>i-flu</i>

In the above examples, it is clear that the isiNdebele coined equivalents are shortened forms of the English terms.

(c) Blending/Compounding

In compounding, terms are created using other words or parts of speech e.g. *isidaka* + *imizwa*. This is the combination of *isidaka* (intoxicator) + *imizwa* (senses) and the noun formed from these two terms is *isidakamizwa*. When looking at this term, one can still determine the two components which are *isidaka* (intoxicator) + *imizwa* (senses) and therefore, this term-formation process is compounding and not blending as most African languages scholars refer to it. For instance, Van Huyssteen (2003:124) says that blending has taken place in examples such as *izakha* 'builders' + *umzimba* 'body' = *izakhamzimba*. Buthelezi (2008:186) also refers to terms such as *ibala* 'spot' + *izwe* 'land' = *ibalazwe* 'map' as blends whilst it is actually compounding. This is an indication that there is confusion in African languages in as far as the two term-formation processes are concerned, namely blending and compounding.

4.3.2 Secondary term-formation strategy (internal resources) in isiNdebele

Semantic transfer is a secondary term formation strategy whereby already existing words in isiNdebele have their meanings broadened and expanded in order to express a new concept.

Table 4.15: *Examples of semantic transfer*

English term	isiNdebele term	Meaning of isiNdebele term
incubator	<i>ibhodlelo</i>	'bottle'
department (of Health)	<i>umnyango</i> (<i>wezamaPhilo</i>)	'door'

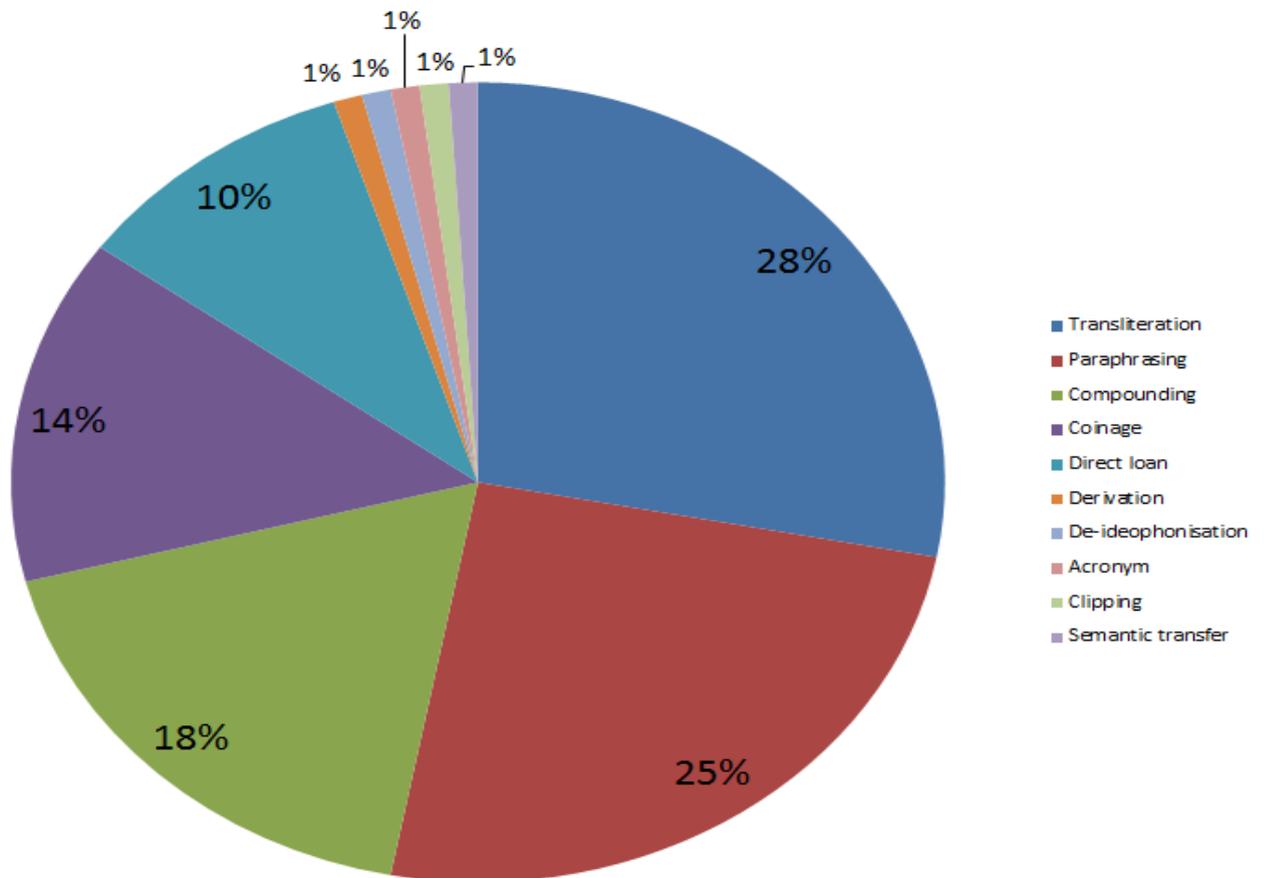
The common word 'bottle' extends its general meaning and embraces a new concept of an 'incubator', which was not imagined when the word was first created. It is the same with the word 'door'; it extends its general meaning and embraces a new concept of a 'department'.

4.3.3 Overview of term-formation strategies

Looking at the data, the two term-formation processes that clearly stood out from the rest of the data were transliteration and paraphrasing. These two made up 53% of all the words included in the data. The remaining 47% was divided between compounding, coinage, direct loan, derivation, acronym, clipping and semantic transfer.

The following are the findings of term-formation strategies mostly used by amaNdebele. See Table 4.16 below.

Table 4.16: Summary of term-formation strategies



The analyses of questionnaires indicate that amaNdebele prefer transliterated terms to the coined ones. This is because coined terms are ambiguous, they may not have one obvious meaning or interpretation. The speakers have to always explain coined terms, unlike the transliterated ones. Transliterated terms are closer to the source language. For instance, the term *i-asma* for 'asthma' is preferred to the coined *isifubasomoya*. The term *eyidzi* for 'AIDS' is preferred to the coined *umbulalasihlangu* or *intumbantonga*. The term *ikhenza* is preferred to the coined *umdlavuzza*. The transliterated term *ihayibhladi* for 'high blood pressure' is used more than its coined counterpart *isigandelelo seengazi*. It may not be easy for one to understand the meaning of *isigandelelo seengazi* compared to *ihayibhladi*. One may not even know that the term *isigandelelo seengazi* actually refers to 'high blood pressure'.

Paraphrasing is the second most preferred method of term formation used by amaNdebele when creating terms. For example; 'optometrist' is referred to as *udorhodere wamehlo* and 'diabetes' is referred to as *ubulwelele beswigiri*.

Compounding is another method used in isiNdebele. The challenge with compounding has been the issue of hyphenation. There is a lot of inconsistency as far as compounds are concerned. Mahlangu (2013:101) says that some compounds are hyphenated, and some are unhyphenated for no apparent reason.

What Mahlangu (ibid.) says, is proof enough that most people just write without verification with the orthography and spelling rules booklet (PanSALB, 2018). It shows that some are not even aware of such a resource. This causes confusion and inconsistencies. This study intends to address inconsistencies especially in the writing of medical compound nouns, for instance the isiNdebele term *intumbantonga* (AIDS). Sometimes it is written as *intumba-ntonga* (with a hyphen) and in other instances as *intumbantonga* (without a hyphen) (Mahlangu, 2013:101 and 108).

Ndimande-Hlongwa (2010:213) says that the purpose of an orthography of a language is to specify the correct way of using a specific writing system for writing the language. Van Huyssteen (2003:59) maintains that orthography in African languages does not only include spelling but it includes terminology too. This is an indication that it is important to use the orthography and spelling rules correctly.

According to Skhosana (1998:31 and 42), the Southern Ndebele Language Board played a role in the publication of isiNdebele orthography and spelling rules. Amongst them were the ones published by the government printers in 2001 and it was called isiNdebele Terminology and Orthography No 1. The other one was published by the Pan South African Language Board in 2008 and it is called, *Imithetho yokutlola nokupeleda isiNdebele*.

In the 1995 orthography, the use of the hyphen in isiNdebele is discussed. The hyphen is used to separate two vowels coming together, to separate longer words as well as to join concords to numerals, e.g. *abali-9*, to separate a class prefix from foreign words beginning with vowels, e.g. *i-AIDS*.

Other strategies such as semantic transfer, blending, clipping, the formation of acronyms and deideophonisation are minimally used in term formation by amaNdebele. What the researcher discovered through the questionnaires is that mechanisms such as clipping, the use of acronyms and deideophonisation are mostly used by nurses who are below 40 years of age. The nurses above 40 years of age mostly use paraphrasing. For instance, the term 'gynecologist' is referred to as *igayini* particularly by younger nurses. However, the older nurses refer to this as *udorhodere wabafazi* 'the doctor of women', which is paraphrasing. The

researcher believes that the reason why this is so, is because younger nurses come into the nursing field while still under the influence of the source language (English), unlike their older counterparts who have been in the field for some time and have already spent some time with patients who would understand paraphrased terms, quicker. A nurse who has been in the field for some time knows that older patients may not know what *igayini* is for instance but will easily understand the term *udorhodere wabafazi*.

The following is a discussion on taboo as applied in isiNdebele.

4.4 TABOO AND TERM-FORMATION PROCESSES IN ISINDEBELE

The subject of taboo cannot be excluded when one deals with term-formation strategies in African languages. AmaNdebele avoid terms with sexual connotations. They do not mention them in a direct manner. For instance, the term 'sexually transmitted infection' is translated as *ubulwele obuthathelwana ngokomseme* which means the 'disease that one gets from the reed mat'. 'Sexual intercourse' is translated as *ukuya emsemeni* which means 'going to the reed mat'. In this example, the term *ukulalana* 'sexual intercourse' has been avoided, instead the term *umseme* 'reed' is used. It is also important to note that cultural taboos hinder effective communication in the medical field. In the spoken corpus the term *ukukha umrorho* 'to pick up the vegetables' is also used for sexual intercourse. When a Ndebele patient tries to explain the challenges of impotence during sexual intercourse using the term *ukukha umrorho*, the doctor may be lost. The doctor may assume that the patient has a challenge of bending when picking up vegetables, whereas the expression actually relates to challenges experienced during sexual intercourse. The ineffective communication may unfortunately lead to a misdiagnoses and wrong prescription.

The names of those parts of the body that perform 'private' functions such as 'urinating' and 'passing stool', and private reproductive functions, are also avoided. For example the term 'penis' is referred to as *intonga* 'stick' in isiNdebele instead of *ipipi*. 'Urine' is referred to as *umhlambuluko* instead of *umthundo* which is regarded as vulgar. 'Passing of stool' is referred to as *ukuzithuma* 'to send oneself' instead of *ukubhabha* or *ukukaka*, which is considered as vulgar and harsh.

Euphemism is used in this study as a taboo avoidance strategy. Through euphemism certain matters are not directly mentioned or expressed. For instance, one may be talking about 'condoms' and instead of using the transliterated term *ikhondomu* which everyone knows, one

may prefer the term *ijasi yomkhwenyana* ‘the bridegroom’s coat’. Not many people would understand that one is actually referring to ‘condoms’. One would think a person is referring to the bridegroom’s jacket. This confirms what Allan and Burridge (2006) state that the things that create euphemism are amongst others, extension of meaning, re-introducing into the vocabulary, rarely used words, borrowing or coining of new words.

The term *ukufa* ‘to die’, for instance, is seldom used when a death case is reported. Instead, the term *ukulala* ‘to sleep’ or *ukukhamba* ‘to go’ is used.

Also see the following examples of taboo terms.

Table 4.17: *Examples of taboo*

English	isiNdebele	Meaning
menstruation	<i>ukuba sesikhathini/ ukuba semalangenini/ ukuba senyangeni</i>	‘be in a time/to be in the days/to be in the moon’
genitals	<i>izitho zangasese</i>	‘body parts that are private’

Table 4.17 reflects that, in order to avoid words that are regarded as taboo, amaNdebele prefer to use paraphrasing.

4.5 CONCLUSION

This chapter presented the interpretation and the discussion of data collected through questionnaires. Terms that were put in the questionnaires were manually sourced from pamphlets. From the projects that the researcher reviewed in Chapter 2, the researcher discovered that even in corpus linguistics, even when one is applying the corpus-driven approach, manual sourcing and analyses can still be implemented, hence, the term semi-automatic term extraction. What must always happen is verification of terms by language experts. After the manual sourcing of terms, terms were put in questionnaires and were verified by doctors, nurses, translators, Curriculum Implementers and finally verified by other medical doctors, through interviews.

The usage of the Grounded Theory in analysing the questionnaires has produced satisfactory results in this study. Through its (Grounded Theory) methods of agreement and disagreements, the researcher managed to locate cases that are similar in many respects, that is, what is common across cases. The researcher managed to identify respondents who agreed and those who were in disagreement as far as a particular term is concerned.

In one hospital and one clinic, instead of nurses filling in the questionnaires individually, they filled in one questionnaire as a group. More terms were harvested in such instances because the nurses had an opportunity of debating the equivalents amongst themselves. It is because of such debates that a term such as 'groin' was clarified. In the questionnaire the term *imbilapho* was put as an equivalent of 'groin' but through the debate amongst nurses, it was discovered that 'groin' is actually a body part and not *imbilapho*. *Imbilapho* is actually a 'lymph node'. This was later confirmed by the doctors.

This chapter also discussed term-formation processes as applicable in isiNdebele. Transliteration was found to be a technique mostly used in isiNdebele. The challenge with transliterated terms is spelling. Unfortunately, the isiNdebele spelling rules do not help to address this issue. In 2008, the spelling rules were revised by the isiNdebele National Language Body and published by PanSALB, but the issue was still not addressed. Presently there is a 2018 spelling rules booklet, but still nothing has been mentioned regarding transliterated terms.

Another term-formation process that poses a challenge in isiNdebele is compounding. There are inconsistencies in the writing of isiNdebele compounds. Some are hyphenated, and some are not, without any valid reason. For instance, *intumbantonga* (AIDS), is written as *intumbantonga* (with a hyphen) by some authors and as *intumbantonga* (without a hyphen), by others. To address these inconsistencies, compound nouns will be unhyphenated in this study.

Regarding term-formation processes that the researcher reviewed (from other languages), the researcher discovered that every language has its own term-formation technique mostly used for creating technical terms. For instance, word borrowing is one of the methods used in the terminology development of ChiShona. In isiXhosa, the effective ways of creating technical terms were through loaning and compounding.

CHAPTER 5

DATA ANALYSIS OF THE WRITTEN AND THE SPOKEN CORPUS

5.1 INTRODUCTION

This chapter discusses the design, the compilation as well as the analysis of both the written and the spoken corpora. The analysis of corpora is done semi-automatically. It is done with the purpose of answering the question on the role corpus-driven lexicography can play in filling the gap caused by the shortage of terminology and also the question on ways in which existing methods of producing terminologies can be transferred to isiNdebele where only sparse resources are available.

As mentioned in the previous chapters, this study uses corpus linguistics as a basis for theoretical analysis. The advantages of corpus linguistics have been discussed in detail in Chapter 3, one of which is that the corpus is stored electronically. This makes it easy to find, sort and to count items. Corpora collected for this study are used as a basis for investigating language use and also for developing an English-isiNdebele glossary of medical terms. The discussion will first be on the spoken corpus, then on the written one.

5.1.1 The spoken corpus

As mentioned earlier, spoken corpora in this study refer to corpora from spoken material which has been literacised. The spoken material of this study is from Ikwewezi FM health program as well as from the interviews with isiNdebele speaking doctors. The interviews from Ikwewezi FM consist of spontaneous speech whilst the ones from the interviews with doctors consist of planned speech.

Spoken language corpora are explored in this study because in isiNdebele, just like in many other languages, such corpora are a dominant modality. This means that they are the dominant methods of communication, compared to written corpora. IsiNdebele has also been a spoken language for some time. Spoken corpora play a role by revealing how terms that already exist in societies are used. They promote the natural development of terms and also help to determine technical terms that have been incorporated into the Ndebele lexicon.

Spoken corpora further present varieties of languages regarding the vocabulary and the pronunciation of words. Although the advantages of spoken corpora have been outlined, it needs to be emphasised that such corpora have been less explored, whilst they are complementary to written language corpora in several senses.

The challenges of collecting spoken corpora are highlighted by Ngcobo and Nomdebevana (2010:190 and 192) and referred to in the literature review chapter, under 2.2.1, paragraph 15. One of the challenges of spoken corpora is the issue of word-hood in agglutinating languages. It is easier to identify words in isolating languages than is in the case in agglutinating languages, to which isiNdebele belongs. This is illustrated by means of the following two examples:

Uzombelethela esibhedlela sangeqadi 'She will give birth to him/her in a private hospital'

U- subject concord (class 1)

-zo- future tense morpheme

-m- object concord (class 1)

-beleth- verbal root

-el- applied extension morpheme

-a terminative vowel

Bamhlabisele inyumoniya 'They caused him/her to be injected for pneumonia'

Ba- subject concord (class 2)

-m- object concord (class 1)

-hlab- verbal root

-is- causative extension morpheme

-el- applied extension morpheme

-e terminative vowel/past tense

As mentioned under 5.1.1, paragraph 3, the written and spoken corpora complement each other. Table 5.1 shows a few examples of how the written and the spoken corpus have complemented each other.

Table 5.1: Comparison between the written and spoken corpus

Written corpus	Spoken corpus
<i>umtholapilo/itlinigi/itliniga</i>	<i>itlinigi</i> ‘clinic’
<i>umhlengikazi</i>	<i>unese</i> ‘nurse’
<i>ukugandeleleka ngokomkhumbulo</i>	<i>ukugandeleleka ngokomkhumbulo/ idiphreshini</i> ‘depression’
<i>itjhukela/ubulwele betjhukela/ubulwele beswigiri</i>	<i>itjhukela/iswigiri/ idayibhethisi</i> ‘diabetes’
<i>ukuzithwala/ukuba sidisi</i>	<i>ukuzithwala/ukuba sidisi/ukuba sebantwini</i> ‘to be pregnant’
<i>ibhodlelo/isifukamisi/isifukamiseli</i>	<i>ibhodlelo/i-inkhyubheyitha</i> ‘incubator’
<i>ikholera, ubulwele bokurhuda/indenebovu</i>	<i>ikholera</i> ‘cholera’

When one looks at two of the terms, namely ‘clinic’ and ‘nurse’, one realises that in spoken language the term *itlinigi* is used more frequently than its coined counterpart, *umtholapilo*. The same applies to the term ‘nurse’. In spoken language, the transliterated term *unese* is used more frequently in comparison to its coined counterpart, *umhlengikazi*. This makes it clear that in spoken language, coined words are less used compared to the transliterated ones.

(a) Recording

Spoken corpora of the present study comprise of recordings from the Ikwewezi FM health program as well as interviews with isiNdebele speaking doctors. Interviews were conducted by the researcher. Spoken corpora of this study were compiled from January 2018 to June 2019.

Regarding the radio station Ikwewezi FM, it was previously known as Radio Ndebele. It is one of the radio stations that are owned by the South African Broadcasting Corporation (SABC). It is an isiNdebele broadcasting radio station, serving the isiNdebele speaking community. It was founded on 1 January 1983. The channel is found on Frequency: 90.6 - 107.7. For more information on Ikwewezi FM, one can visit: https://en.wikipedia.org/wiki/Ikwewezi_FM. The health programs which the researcher recorded were aired on [Wednesdays and Fridays for 30 to 45 minutes](#).

Recording was the first step in the collection of the isiNdebele spoken corpus. The researcher met with the then Ikwewezi FM program manager. The meeting was held in her office. The researcher presented her intention to record the health programs for the purposes of her study.

The program manager verbally granted the researcher permission to record the health program. She maintained that the health programs are meant for the public anyhow. She further requested that the researcher should acknowledge Ikwewezi FM in her study.

The researcher also met with the presenter of the health program, who was keen to assist the researcher, even to the extent that she would send the researcher her schedules from time to time which consisted of the health topics to be presented as well as the times and the name of the doctor who would be on radio. This she did so that the researcher would not miss the coming health topics.

The doctor who was interviewed on most of the ailments was also made aware of the researcher's study. He also updated the researcher about some programs that the researcher might have missed. He sent some recordings and would inform the researcher about the coming presentations. The researcher used a recorder for recording this health program from Ikwewezi FM. In the health program there were also call-ins as well as questions asked through social platforms like *Face Book* and *Twitter*. The presenter would read the questions asked through the social platforms to which the guest doctor or sister would respond. The callers' names are not included in the corpus. Most callers opted to remain anonymous but even those who introduced themselves with their real names, were not identified by their names, in the study. They are referred to as *umlaleli* 'listener' 1, *umlaleli* 'listener' 2, etcetera. The Ikwewezi FM broadcasts that resulted in the spoken corpus are represented in Table 5.2.

Table 5.2: *Summary of the spoken corpus from Ikwewezi FM*

Topic	Date	Time	Size of the corpus (Tokens)
1.1 Mental health			
1.1.1 <i>ihloko</i> 'headache'	15-01-2018	11H15	647
1.1.2 <i>idiphreshini</i> 'depression'	18-01-2018	10H30	442
1.1.3 <i>ukwethuka</i> 'anxiety'	16-07-2018	11H15	408
1.1.4 <i>i-OCD</i> 'Obsessive Compulsive Disorder'	30-07-2018	10H15	458

Topic	Date	Time	Size of the corpus (Tokens)
1.1.5 <i>iskizofreniya</i> 'schizophrenia'	23-07-2018	11H15	452
1.1.6 <i>istrowuku</i> 'stroke'	03-05-2018	10H30	380
1.1.7 <i>ibhayiphola</i> 'bipolar 1'	28-03-2018	10H30	243
1.1.8 <i>ibhayiphola</i> 'bipolar 2'	13-06- 2018	11H15	82
1.2 Gynecological ailments			
1.2.1 <i>ihisterekhomi nesterilizeyitjhini</i> 'hysterectomy & sterilisation'	14-05-2018	10H30	549
1.2.2 <i>amafayibhrotsi</i> 'fibroids'	06-08-2018	10H30	658
1.2.3 <i>idawuni sindromu</i> 'down syndrome'	14-03-2018	11H15	196
1.2.4 <i>imolaphregnensi</i> 'molar pregnancy'	09-03-2018	10H30	630
1.2.5 <i>ipapsmiye</i> 'pap smear'	07-11-2018	10H30	290
1.2.6 <i>utjwala nesidisi</i> 'alcohol and pregnancy'	06-08-2018	15H00	272
1.3 Primary health care			
1.3.1 <i>abantwana abanokukhubazeka</i> 'children with disability'	25-05-2018	10H30	573
1.3.2 <i>imalnyuthrishini</i> 'malnutrition'	25-05-2018	10H30	367
1.3.3 <i>ukuhlatjelwa umgomani</i> 'flu vaccine'	16-04-2018	11H15	302
1.3.4 <i>ihayiphehayidrosis</i> 'hyperhydrosis'	22-01-2018	10H30	767
1.3.5 <i>amalwele angathathelaniko</i> 'non-communicable diseases'	23-08-2018	10H30	573
1.3.6 <i>ikholikhi</i> 'colic'	22-08-2018	10H30	199
1.3.7 <i>ama-anthibhayothiksi</i> 'antibiotics'	07-11-2018	10H30	255
1.3.8 <i>i-TB</i> 'TB'	15-11-2018	10H30	262
1.3.9 <i>ubulwele beswigiri</i> 'diabetes'	02-07-2018	10H30	307
1.3.10 <i>ipholiyo</i> 'polio'	28-11-2018	10H30	300
1.3.11 <i>imalariya</i> 'malaria'	03-10-2018	11H15	258
1.3.12 <i>ukumilisa</i> 'teething'	31-10-2018	11H15	252
1.3.13 <i>i-CFS</i> 'Chronic Fatigue Syndrome'	02-09-2018	10H30	328
1.3.14 <i>ivayithaligo</i> 'vitiligo'	05-09-2018	10H30	613

Topic	Date	Time	Size of the corpus (Tokens)
1.4 Cancer			
1.4.1 <i>ikhenza yesikhumba</i> 'skin cancer'	08-01-2019	10H30	582
1.4.2 <i>ikhenza yephrosteyiti</i> 'prostate cancer 1'	29-01-2019	10H30	589
1.4.3 <i>ikhenza yephrosteyiti</i> 'prostate cancer 2'	02-07-2019	10H30	179
1.4.4 <i>i-HPV</i> 'Human Papilloma Virus'	12-03-2019	10H30	403
1.4.5 <i>ikhenza yesibe letho neyomlomo wesibe letho</i> 'cancer of the cervix and cancer of the uterus'	09-07-2019	10H30	549
1.4.6 <i>umbala odalwa mahomowuni</i> 'hormonal pigmentation'	06-08-2019	10H30	326
Total			13,691

Table 5.2 reflects the spoken corpus composition as recorded from Ikwekwezi FM. It indicates the title of the ailment, the date and the time at which it was presented. In the last column, the size of the corpus is indicated. Spoken corpora collected are classified into four categories, namely mental health, gynecological ailments, primary health care and cancer.

(i) ***Ipilo yomkhumbulo* 'mental health'**

The corpus on mental health collected in this study includes *ihloko* 'headache', *idiphreshini* 'depression', *ukwethuka* 'anxiety', *i-OCD* 'Obsessive Compulsive Disorder', *iskizofreniya* 'schizophrenia', *istrowuku* 'stroke', *ibhayiphola 1* 'bipolar 1' and *ibhayiphola 2* 'bipolar 2'. In the case of 'bipolar' 1 and 2, it does not mean that there are two types of 'bipolar', but it simply means that this topic was presented on two occasions by different doctors and on different dates. The first one was broadcast on the 28th of March 2018 and the second one on the 13th of June 2018.

(ii) ***Amalwele apha thelene nesifazi* 'gynecological ailments'**

Miller-Keane Encyclopaedia and Dictionary of Medicine (2003) define gynecology as the branch of medicine that is particularly concerned with the health of the female organs of

reproduction. <https://www.elsevier.com/books/miller-keane-encyclopedia-and-dictionary-of-medicine-and-nursing-and-allied-health/miller-keane/978-0-7216-9791-8>.

The corpus on 'gynecological ailments' includes *ihisterekhthomi nesterilizeyitjhini* 'hysterectomy and sterilisation', *amafayibhrotsi* 'fibroids', *idawunisindromu* 'down syndrome', *imola pregnensi namkha isidisi semola* 'molar pregnancy', *ipapsmiye* 'pap smear' and *utjwala nesidisi* 'alcohol and pregnancy'.

(iii) **Ukuthogonyelwa kwepilo okusisekelo 'primary health care'**

Primary health care (PHC) forms an integral part of the country's health system. It is the first level of contact of individuals, the family and the community with the national health system. It brings health care as close as possible to where people live and work. It constitutes the first element of a continuing health care service. It is based on practical, scientifically sound and socially acceptable methods and technology, made universally accessible to individuals and families in the community through their full participation, and at a cost that the community and country can afford to maintain at every stage of their development in the spirit of self-reliance and self-determination (WHO, 1978).

Spoken material on primary health care includes the following:

Abantwana abanokukhubazeka 'children with disability', *imalnyuthrithini* 'malnutrition', *ukuhlatjelwa umgomani* 'flu-vaccine', *ihayiphehayidrosisi* 'hyper-hydrosis', *amalwele angathathelaniko* 'non-communicable diseases', *ikholikhi* 'colic', *ama-anthibhayothiksi* 'antibiotics', *i-TB* 'TB', *ubulwele beswigiri* 'diabetes', *ipholiyo* 'polio', *imalariya* 'malaria', *ukumilisa* 'teething' and *i-CFS* 'Chronic Fatigue Syndrome'.

(iv) **Ikhenza 'cancer'**

The corpus on cancer includes *ikhenza yesikhumba* 'skin cancer', *ikhenza yephrosteyiti 1* 'prostate cancer 1', *ikhenza yephrosteyiti 2* 'prostate cancer 2' and *i-HPV* 'Human Papilloma Virus'. Also, here, the fact that there is prostate cancer 1 and 2 simply means that this topic was presented by two different doctors on different dates. Firstly, it was presented on 29 January 2018 followed by a second presentation on 2 July 2018.

Regarding *ikhenza yesibeletho* ‘cancer of the uterus’, *nekhenza yomlomo wesibeletho* ‘and cancer of the cervix’, the researcher decided to include them under the category of ‘cancer’ instead of discussing them under gynecological ailments, although these two cancer types only affect women.

Table 5.3 is a summary of doctors who were interviewed:

Table 5.3: *Summary of doctors who participated in interviews*

Date	Participant	Venue	Duration	Size of the corpus
19 May 2018	DOCTOR 1	Doctor’s Surgery	1 hour	5,977
12 April 2018	DOCTOR 2	Doctor’s Surgery	1 hour	6,213
Total				12,190

It should be noted that the spoken corpus of this study is monolingual, it does not have a corresponding English translation. Table 5.2 reflects the recordings from Ikwewezi FM. The recordings were presented in isiNdebele, to isiNdebele speaking listeners. After the recording of the spoken corpus, transcription and term extraction followed.

(b) Transcription

For this study, the denaturalised-literacised transcription was implemented. It is literacised because the spoken corpus was converted to a written corpus. It is denaturalised because aspects such as fillers and purposeless repetitions were excluded. They were excluded because they may distract one from focusing on the content of the speech. It needs to be emphasised that no sentences are altered or edited during this type of transcription. See Table 5.4.

Table 5.4: An extract from the spoken corpus showing sentences with and without fillers

Topic	Sentences with fillers	Sentences without fillers
Skizofreniya 'Schizophrenia' Presented on 23 July 2018	<p>1. Mrhatjhi: Namhlanje sizokukhuluma ngamalwele naka athinta umkhumbulo ke... esiliphetheko namhlanjesi. <i>e.e.e.e.e</i> <i>i...</i>isihloko ngeseskizofreniya... Sikhamba nesithekeli sethu uSimphiwe Jiyana ongunohlalakuhle... Ses Simphiwe lotjha...</p> <p>2. UNohlalakuhle: Lotjha Busayi... ngiyaphila ninjani...ngibingelela nabalaleli boke beKwekwezi FM namhlanje ekuseni...</p> <p>3. Mrhatjhi: Siyathokoza... libudisana igama leskizofreniya leli <i>e e e e</i> <i>e...</i>nangilibizako (hi hihi (laughing) akhe ungibizele lona ngilizwe kuhle bona libizwa bunjani...</p> <p>4. UNohlalakuhle: <i>hi hi hi</i> (Laughing) iskizofreniya....</p> <p>5. <i>Tring tring...tring tring....</i> (ringing of the telephone) <i>Tring</i></p>	<p>1. Mrhatjhi: Namhlanje sizokukhuluma ngamalwele naka athinta umkhumbulo ke. Esiliphetheko namhlanjesi isihloko ngeseskizofreniya. Sikhamba nesithekeli sethu uSimphiwe Jiyana ongunohlalakuhle. Sesi Simphiwe lotjha.</p> <p>2. UNohlalakuhle: Lotjha Busayi. Ngiyaphila ninjani? Ngibingelela nabalaleli boke beKwekwezi FM namhlanje ekuseni.</p> <p>3. Mrhatjhi: Siyathokoza, libudisana igama leskizofreniya leli nangilibizako. Akhe ungibizele lona ngilizwe kuhle bona libizwa bunjani.</p> <p>4. UNohlalakuhle: Skizofreniya.</p>

In Table 5.4 the researcher numbered the turn-takings in the dialogue between the Ikwekwezi FM presenter and the social worker just for the sake of easy reference. The researcher further italicised the fillers. Fillers include the laughter, the stuttering and the ringing of the telephone by the listeners calling in. For instance, in sentence number 1, under the sentences with fillers, the unnecessary *e e e e e e* have been removed in the righthand column, under sentences

without fillers. In sentence 3, the fillers e e e e e were also removed, these are just mannerisms by the presenter and as one can see in the righthand columns, even when they are removed, they do not alter the meaning of the sentence at all.

The laughter by *uNohlalakhle* 'social worker' has also been removed in the righthand column. It does not impact the meaning and the sense of the sentence. The ringing of the telephone by listeners calling in, was excluded as well.

5.1.2 The written corpus

The design of the written corpus in this study is directed towards a written parallel corpus which comprises of 41 English-isiNdebele medical texts. These texts represent a medical isiNdebele discourse which reveals the use of medical terms in context.

In this study, re-keyboarding and electronic transfer have been used to enter medical texts into computer files. Through electronic transfer, texts were retrieved or downloaded from the internet into the computer files. One printed brochure on 'diabetes' was typed into computer files through re-keyboarding.

The written, parallel corpora collected in this study were adequate for solving the research problem of the present study which is the shortage of linguistic resources in isiNdebele. The four dictionaries that are presently available in isiNdebele are dictionaries for general purposes. This means that even if their development was corpus-driven, the corpora utilised were collected for general purposes. There is a shortage of resources in isiNdebele and this implies a shortage of terminology in the language, in general. This study has therefore been designed to investigate the role of corpora in the development of isiNdebele terminology and the compilation of a specialised bilingual glossary of medical terms.

Text availability has played a significant role in the collection of texts for this study. The compilation of the parallel corpora is determined by the availability of texts and their translations. It is the availability of texts that has determined the size of the corpus collected for this study. A summary of the components of the written isiNdebele corpus is given in Figure 5.1.

N	text file	file size	tokens (running words) in	tokens used for word list	sum of	types (distinct)	type/token ratio	standard	STTR	STTR	mean word length	word length	mean (in std.dev.)	mean (in std.dev.)	
1	Overall	38,588,82	2,077,086	1,993,776		106,366	5.34	59.27	40.06	1,000	8.00	3.41	132,115	56.96 4,028.4	41 48,628 217,024
2	(1124200880058 AM) GEMS_APP(NDEBELE)S/	20,458	2,209	2,167		1,032	47.62	55.00	31.82	1,000	8.04	3.17	117	18.52 14.84	1 2,167.0
3	(128200995441 AM) ndebeleSADiLaR.txt	20,739	2,249	2,214		1,337	60.39	68.95	21.96	1,000	8.00	3.10	148	14.96 9.76	1 2,214.0
4	(222201095205 AM) NdebeleSADiLaR.txt	15,352	1,727	1,649		999	60.58	59.20		1,000	7.68	3.23	106	15.56 9.72	1 1,649.0
5	(310201010444 PM) FINAL SELECTION PACK N	16,185	1,792	1,640		590	35.98	45.80		1,000	7.81	3.48	44	37.27 142.76	1 1,640.0
6	(310201012440 PM) FINAL MB NDEBELESADi	16,633	1,819	1,680		669	39.82	53.40		1,000	7.91	3.38	43	39.07 144.18	1 1,680.0
7	(3152010104321 AM) NdebelenewformSADiLaR.t	21,649	2,332	2,290		1,079	47.12	55.15	31.71	1,000	8.06	3.19	132	17.35 13.20	1 2,290.0
8	(3172010120110 PM) FINAL MB NDEBELESADi	16,633	1,819	1,680		669	39.82	53.40		1,000	7.91	3.38	43	39.07 144.18	1 1,680.0
9	(382010120836 PM) marketing ndebeleSADiLaR.	5,318	595	570		408	71.58			1,000	7.70	3.09	35	16.31 12.84	1 570.00
10	(43200923541 PM) GEMS_APP(NDEBELE)SAD	21,760	2,342	2,302		1,082	47.00	55.05	31.78	1,000	8.07	3.18	126	18.27 14.06	1 2,302.0
11	(49201013614 PM) PHR brochure_NdebeleSADiL	1,352	149	140		109	77.86			1,000	7.79	3.13	13	10.77 6.47	1 140.00
12	(5102010114932 AM) Q1 newsletter_Ndebele_Re	14,913	1,609	1,577		982	62.27	67.50		1,000	8.06	3.23	103	15.31 8.30	1 1,577.0
13	(527201092947 AM) Ndebele Maternity Programr	5,218	547	535		351	65.61			1,000	8.27	3.02	30	17.83 15.87	1 535.00
14	(712201030934 PM) GEMS AFFIDAVIT A - NDEI	1,197	151	151		102	67.55			1,000	6.66	3.95	9	16.78 19.29	1 151.00
15	(712201031915 PM) GEMS AFFIDAVIT B - NDEI	1,204	154	154		81	52.60			1,000	6.40	4.01	8	19.25 18.11	1 154.00
16	(712201041954 PM) Member newsletter Ndebele:	11,021	1,189	1,174		727	61.93	64.50		1,000	7.94	3.30	78	15.05 8.31	1 1,174.0
17	(7302010105055 AM) GEMS AR09-10_NdebeleS	60,924	6,873	6,074		2,590	42.64	55.10	36.11	1,000	7.72	3.40	269	22.58 14.74	1 6,074.0
18	(762010100643 AM) 1495 Final Gems Birthing C	3,722	384	370		269	72.70			1,000	8.44	3.12	28	13.21 7.16	1 370.00
19	(87200915043 PM) isiNdebeleSADiLaR.txt	15,317	1,673	1,635		1,038	63.49	67.00		1,000	7.93	3.16	131	12.48 6.56	1 1,635.0
20	2000-04-ndebeleSADiLaR.txt	53,099	5,750	5,561		2,007	36.09	52.28	39.39	1,000	8.02	3.39	90	61.79 77.96	1 5,561.0
21	CORP.NCHLT.nr.CLEAN.2.0SADiLaR.txt	14,829.57	797,323	774,189		87,420	11.29	62.37	38.80	1,000	8.07	3.35	58,991	42.77 2,988.7	1 774,189
22	CORP.NCHLT.nr.RAW.2.0SADiLaR.txt	23,280.47	1,233,432	1,175,356		105,082	8.94	57.32	44.71	1,000	7.95	3.44	70,813	70.02 4,778.7	1 1,175,3
23	DAC-NLS.IhlahluboYokobanaUsidisi.2007-05-18.r	5,264	243	238		172	72.27			1,000	8.08	3.29	12	19.83 25.05	1 238.00
24	DAC-NLS.IsibawoSasabeloSokurholophala.2007-	25,300	678	649		419	64.56			1,000	7.87	3.31	27	24.04 38.32	1 649.00
25	DAC-NLS.UkwelatjhelwaUkusebenzisaK(1).2007-	12,090	287	272		186	68.38			1,000	8.28	3.71	12	22.67 27.93	1 272.00
26	DAC-NLS.UkwelatjhelwaUkusebenzisaKumbi.200	16,222	384	379		265	69.92			1,000	8.65	3.38	16	23.69 45.76	1 379.00
27	GCIS.10ReasonsToTestForHivAids.2010-10-19.nr	1,205	130	126		104	82.54			1,000	7.99	3.42	11	11.45 8.71	1 126.00
28	GCIS.HealthMinisterLetter.2010-10-19.nrSADiLaF	4,206	408	406		307	75.62			1,000	9.09	3.04	18	22.56 17.91	1 406.00
29	GCIS.Healthposter.2010-10-19.nrSADiLaR.txt	4,010	400	396		299	75.51			1,000	8.81	3.03	22	18.00 12.99	1 396.00
30	16 Days of Activism Leaflet Ndebele.txt	10,344	1,016	896		613	68.42	60.30		1,000	7.97	3.47	70	12.80 16.21	1 896.00
31	Abantu abaphila nokukhubazeka HL 2015.txt	5,693	574	571		399	69.88			1,000	8.76	3.21	43	13.28 4.73	1 571.00
32	Amalwele ayingozi ebantwini abatjha HL 2010.txt	1,778	188	188		149	79.26			1,000	8.13	3.03	14	13.43 4.69	1 188.00
33	Ikholera from 2010 FAL P1.txt	2,753	294	290		240	82.76			1,000	7.96	3.04	29	10.00 4.36	1 290.00
34	Isigaba 2 somthosisekelo.txt	2,400	257	245		196	80.00			1,000	7.74	3.62	7	35.00 55.11	1 245.00
35	IsiNdebele ibandulo labelaphi bendabuko.txt	27,260	2,596	2,570		1,404	54.63	61.70	27.08	1,000	8.35	2.91	182	14.12 13.89	1 2,570.0
36	I-TB Netjhukela.txt	1,148	130	130		99	76.15			1,000	6.53	4.01	15	8.67 4.78	1 130.00
37	nde mnt-booklet.txt	11,431	929	912		525	57.57			1,000	8.43	5.31	92	9.91 8.10	1 912.00
38	Nde Treatment Adherence Flyer.txt	84	12	12		2	16.67			1,000	5.00	1.04	1	12.00	1 12.00
39	Ukuhlololwa i-HIV.txt	10,994	971	952		579	60.82			1,000	8.40	3.59	79	12.05 8.86	1 952.00
40	Ukuhlongakala kwamasana HL Gauteng 2014.txt	3,338	344	343		274	79.88			1,000	8.49	3.09	33	10.39 4.31	1 343.00
41	Ukunikela ngeengazi FAL G.12 2010.txt	5,735	601	567		369	65.08			1,000	7.62	3.23	37	15.32 6.25	1 567.00
42	Umzimba omkhulu Exempler HL G.12 2014.txt	4,837	526	526		394	74.90			1,000	8.02	2.92	42	12.52 5.60	1 526.00

Figure 5.1: Summary of the written isiNdebele corpus

It should be noted that the English versions of fliers are also available. They are being saved for future research. It was easy to access the English versions of fliers.

5.2 SUMMARY OF ALL THE DATA COLLECTED (SPOKEN AND WRITTEN ISINDEBELE)

The isiNdebele corpus compiled from written text has a size of 2,077,086 tokens. The spoken literacised corpus numbers 25,881 tokens. In total, the isiNdebele corpora consist of 2,102,967 tokens.

5.3 ALIGNMENT OF TEXTS

An alignment of texts had to be done before the process of term extraction. After electronic texts had been downloaded and stored locally, they were proof-read by the researcher. There are 29 texts that were taken from the SADiLaR website and six from grade 12 question papers of the Department of Basic Education. These texts are on a variety of topics, namely cholera, obesity, infant mortality, blood transfusion, diseases affecting the youth of South Africa as well as disability in South Africa. The isiNdebele question papers set by the Department of Basic Education acknowledged the English sources. All the English texts had been taken from the internet by the isiNdebele examiners and translated into isiNdebele. It was thus easy for the researcher to find those English sources.

Table 5.5 indicates the names and publishers (if known) of the fliers and posters used in this study. It needs to be mentioned that some fliers were downloaded from the internet. *Ibandulo labelaphi* 'Palliative care' was translated by the Department of African Languages, Unisa. Most of the texts were accessed from the SADiLaR web-page (2021), where they are freely made available. All the medical texts taken from the SADiLaR web-page, are from the Government Employees Medical Scheme (GEMS). They were published by GEMS, SADiLaR translated the texts only.

Table 5.5: *Metadata of written corpus*

Name	Publisher	Country of publication
Ibandulo labelaphi/Palliative Care	Unpublished Translated in the Dept. of African Languages, Unisa	Republic of South Africa
HIV/AIDS treatment sheet	Soul City Khomanani	Republic of South Africa
Itjhidi lelwazi lokuhlahubela i-HIV/HIV testing information sheet	Dilicom language and communication	Republic of South Africa
16 Days of Activism Isibawo sobulunga	Unknown	Republic of South Africa
MNT booklet	Department of Justice	Republic of South Africa
The Approved TB & Diabetes poster	Department of Health in partnership with URC and SANOFI	Republic of South Africa
Isibawo sobulunga	GEMS	Republic of South Africa
Umhlahlandlela wobulunga	GEMS	Republic of South Africa
Umlayezo ovela kusiphathiswa esikhulu	GEMS	Republic of South Africa
Umlayezo ovela kusiphathiswa esiyihloko	GEMS	Republic of South Africa
Imali ebhadelwa esikimini	GEMS	Republic of South Africa
Ukukhetha ikhetho elifanele umndenakho	GEMS	Republic of South Africa
Ukuzibophelela kweGems	GEMS	Republic of South Africa
Ihlelo lokubelethisa	GEMS	Republic of South Africa
Gems Afidavita-A	GEMS	Republic of South Africa
Gems Afidavita-B	GEMS	Republic of South Africa
Ubujamo oburhabako	GEMS	Republic of South Africa
Ukubelethwa komntwanakho	GEMS	Republic of South Africa
Ukuzaliswa kweforomo	GEMS	Republic of South Africa
Iimbhedlela zangeqadi	GEMS	Republic of South Africa
Foromo elitjha	GEMS	Republic of South Africa
Ihayiphathenshini	GEMS	Republic of South Africa
Umbiko womnyaka ka 2009	GEMS	Republic of South Africa
Umthetho wokuthuthukisa ukulingana	GEMS	Republic of South Africa
Abadlalifa	GEMS	Republic of South Africa
Umhlahlandlela wokuba lilunga	GEMS	Republic of South Africa
Ihlahlubo yokobana usidisi	GEMS	Republic of South Africa
Isibawo sesabelo sokurholophala	GEMS	Republic of South Africa

Name	Publisher	Country of publication
Ukwelatjhelwa ukusebenzisa iindaki	GEMS	Republic of South Africa
Ukubuyiselwa	GEMS	Republic of South Africa
Ebujameni bangaphambili	GEMS	Republic of South Africa
Ukuzihlolisela ingogwana yeNtumbantonga	GEMS	Republic of South Africa
Ungqongqotjhe wezamaphilo	GEMS	Republic of South Africa
Ilwazi lomphakathi mayelana ne-N1H1	GEMS	Republic of South Africa
Iinzathu ezili-10 zokuhlalubelwa i- HIV/AIDS	GEMS	Republic of South Africa
Isibawo sesabelo sokurholophala kwesikhatjhana	GEMS	Republic of South Africa

Bowker and Pearson (2002:96-97) as well as Taljard and De Schryver (2002:58) mention that alignment software makes several assumptions about texts and their translations. One of the assumptions made by alignment programs is that source texts and their translations have the same number of paragraphs and the same number of sentences. The researcher therefore, had to manually align the corpus by ensuring that the English paragraphs correspond with their isiNdebele counterparts. The researcher had to ensure that superfluous hard returns (that is, line breaks) are removed. This was done to ensure that the source text is parallel to the target text. This proves that the human factor cannot be discarded in terminological practice. See the following un-aligned parallel English-isiNdebele texts in Figure 5.2.

<p>Information sheet for testing HIV If you have any difficulties in understanding this document, you may ask the nurse, laboratory assistant or doctor to explain it to you.</p> <p>What are my rights? You have the following rights: 1. Not to be tested for the AIDS virus without your free and informed consent. 2. To be given all relevant information regarding the harms, risks and benefits of taking, or not taking, the HIV test. 3. To refuse to take the test. If you do this, your application for insurance may be denied. You may, however, wish to consider alternatives such as specialist life products offered by some companies, endowment or other 'pure' financial products. Consult your financial adviser for further information on the options available to you. 4. To receive pre-test counselling which is private and confidential, and which will inform you more fully about the test and its implications before you give consent. If you are in any way unfamiliar with the issues involved, you are strongly advised to seek pre-test counselling. You are also within your rights to waive any personal pre-test counselling. 5. To have your test result treated confidentially. An abnormal test result will be made available to your doctor and this test result will also be stored on the database operated for and on behalf of the life offices by an applicable life assurance association in an encoded form. This information can only be accessed by other insurance companies with your consent. You also have the right to access this information to check that it is correct. 6. To post-test counselling if the test is positive, at the expense of Unilife.</p>	<p style="text-align: center;">ITJHIDI LELWAZI LOKUHLAHLUBELWA I-HIV</p> <p style="text-align: center;">NANGABE UNOMRARO WOKUZWISISA LOMTLOLO, BAWA UNESI NAMKHA UMSIZI WELABHORATHRI NAMKHA UDORHODERE BONA AKUHLATHULULELE WONA</p> <p>NGABE NGIMAPHI AMALUNGELWAMI? Unamalungelo alandelako:</p> <ol style="list-style-type: none"> 1. <i>Ukungahlahlubelwa</i> umulwani obanga intumbantonga (HIV) ngaphandle kwemvumo yakho edzimelele ekuzithandeleni kwakho kanye nekuzwisiseni ngokuzeleko. 2. <i>Ukunikelwa yoke imininingwana ngobungozi, irishki kanye nenzuzo, yokuhlahlubelwa</i> namkha ukungahlahlubelwa umulwani we-HIV. 3. <i>Ukwala ukuhlahlubelwa umulwani.</i> Nangabe ukhetha ukwenza njalo, isibawo sakho sokuthatha ijthorensingarahwa, nangabe bejthorensi leyo bafuna wenze iinhlalubo ze-HIV njengengonye yokulinganisa irishki ekunikeleni ijthorensi. Kodwana ungakhetha ezinye iindlela zokusisa ezingafuni ijhejo lerishki, ezifana netjthorensi ye-endowment namkha eminye imikhqizo esebenza ngeemali kwaphela. Thintana nomyelelisakho wezemali, ukuthola isizo. 4. <i>Ukufumana ukuyeleliswa ngaphambi kokuhlahlubelwa umulwani</i> okungokufihlakeleko. Ikhoh okuzakuyelelisa ngokunabileko ngehlahlubwakho kanye nemithelela yakhona, ngaphambi kobana uvume ukuhlahlubelwa umulwani we-HIV. Nangabe awunalwazi ngezinto ezimayelana nokuhlahlubwakho, sikukhuthaza bona ufune iinyeleliso zangaphambi kokuhlahlubwa. Ungakhetha indlela yinye kwezintathu zokuyeleliswa kwangaphambi kokuhlahlubwa; <ol style="list-style-type: none"> a) Ukufunda lomtolo welwazi. b) Ukuyeleliswa okuyifihlo ngelimi lekhenu kuyafumaneka simahla ukusukela nge-ri le-7 ekuseni kufikela nge-iri le 7 ntambama phakathi kweveke, enomborweni yasimahla ethi-0800 562 562. Godu unelungelo lokwala ukuyeleliswa okwenzeka ngesikhathi sangaphambi kokuhlahlubwa. c) Ukuyeleliswa ngaphambi kokuhlahlubelwa umulwani emalaborathri akhethiweko eendaweni zemadorobheni. Sibawa utshintane nebhrowukha yakho/umhlanganisi mayelana nalokhu. 5. <i>Ukufumana udorhodere lapho ozakufumana khona imiphumelakho</i>, nangabe awunayе udorhodere ongamketha lowo ozakunikelwa imiphumela yeenhlalubo, ungakhetha isentha yezokuthintana ngomtato ebalwe ngehla, mayelana nomnqopho wokufumana imiphumela. 6. <i>Ukufuna imiphumelakho ibe yifihlo.</i> Imiphumela yehlalubo engakajayeleki izakunikelwa udorhoderakho begodu lemiphumela izakubulungwa erhelweni elikhulu ledatha lakwa-ASISA ngendlela yokungavezi igama lakho, ngombana kuzakusetjenziswa ikhowudu. Lemininingwana ingafunywana ngezinye iinkampani zetjthorensi kwaphela, nakhona ngemvumo yakho. Godu unelungelo lokufumana lemininingwana ukuhlola kobana imumethi okungikho na. 7. <i>Ukufumana umzombe munye wokuyeleliswa ngemva kokuhlahlubelwa umulwani nangabe iinhlalubo zikhombisa ukutshwayeleka</i>, ngokusebenzisa iindleko zekampani yetjthorensi yepilo.
---	---

Figure 5.2: Un-aligned parallel English-isiNdebele texts

Figure 5.2 reflects the original English-isiNdebele texts before their alignment. When comparing number 4 of the English version to that of isiNdebele, one realises that number 4 of the isiNdebele version is sub-divided into a, b and c, whereas it is just one paragraph in the case of English. The content of English number 5, is covered in number 5 and 6 in the case of isiNdebele. Number 6 in English is number 7 in isiNdebele. Based on this, it was necessary to align the two languages before using the WordSmith Tools. The aligned English-isiNdebele text follows below in Figure 5.3.

<p>HIV Testing Information Sheet</p> <p>If you have any difficulties in understanding this document, you may ask the nurse, laboratory assistant or doctor to explain it to you.</p> <p>What are my rights?</p> <p>You have the following rights:</p> <ol style="list-style-type: none"> 1. Not to be tested for the AIDS virus without your free and informed consent. 2. To be given all relevant information regarding the harms, risks and benefits of taking, or not taking, the HIV test. 3. To refuse to take the test. If you do this, your application for insurance may be denied. You may, however, wish to consider alternatives such as specialist life products offered by some companies, endowment or other 'pure' financial products. Consult your financial adviser for further information on the options available to you. 4. To receive pre-test counselling which is private and confidential, and which will inform you more fully about the test and its implications before you give consent. If you are in any way unfamiliar with the issues involved, you are strongly advised to seek pre-test counselling. You are also within your rights to waive any personal pre-test counselling. 5. To have your test result treated confidentially. An abnormal test result will be made available to your doctor and this test result will also be stored on the database operated for and on behalf of the life offices by an applicable life assurance association in an encoded form. This information can only be accessed by other insurance companies with your consent. You also have the right to access this information to check that it is correct. 6. To post-test counselling if the test is positive, at the expense of Unilife. 	<p>ITJHIDI LELWAZI LE-ASISA LOKUHLAHLUBELWA I-HIV</p> <p>Nangabe unomraro wokuzwisa lomtlo, baw a unesi namkha umsizi welaborathri namkha udorhodere bona akuhlathululele wona</p> <p>Ngabe ngimaphi amalungelwami?</p> <p>Unamalungelo alandelako:</p> <ol style="list-style-type: none"> 1. Ukungahlalubelwa umulwani obanga intumbantonga (HIV) ngaphandle kwemvumo yakho edzimelele ekuzithandeleni kwakho kanye nekuzwisiseni ngokuzeleko. 2. Ukunikelwa yoke imininingwana ngobungozi, iriskhi kanye nenzuzo, yokuhlalubelwa namkha ukungahlalubelwa umulwani we-HIV. 3. Ukwala ukuhlalubelwa umulwani. Nangabe ukhetha ukwenza njalo, isibawo sakho sokuthatha itjhorensa singlararhwa, nangabe betjhorensa leyo bafuna wenze iinhlalubo ze-HIV njengengcenywe yokulinganisa iriskhi ekunikeleni itjhorensa. Kodwana ungakhetha ezinye iindlela zokusisa ezingafuni itjhejo leriskhi, ezifana netjhorensa ye-endowment namkha eminye imikhizozo esebenza ngeemali kwaphela. Thintana nomyeelisakho wezeemali, ukuthola isizo. 4. Ukufumana ukuyelelisa ngaphambi kokuhlalubelwa umulwani okungokufihlakeleko, lokho okuzakuyelelisa ngokunabileko ngehlalubwako kanye nemithelela yakhona, ngaphambi kobana uvume ukuhlalubelwa umulwani we-HIV. Nangabe awunalwazi ngezinto ezimayelana nokuhlalubwako, sikukhuthaza bona ufune iinye eliso zangaphambi kokuhlalubwa. Ungakhetha indlela yinye kwezintathu zokuyelelisa kwangaphambi kokuhlalubwa; 5. Ukufuna imiphumelakho ibe yifihlo. Imiphumela yehlahlubo engakajayeleki izakunikelwa udorhoderakho begodu lemiphumela izakubulungwa erhelweni elikhulu ledatha ngendlela yokungavezi igama lakho, ngombana kuzakusetjenziswa ikhowudu. Leminingwana ingafunywana ngezinye iinkampani zetjhorensa kwaphela, nakhona ngemvumo yakho. Godu unelungelo lokufumana lemininingwana ukuhlola kobana imumethe okungikho na. <p>6 Ukufumana umzombe munye wokuyelelisa ngemva kokuhlalubelwa umulwani nangabe iinhlalubo zikhombisa ukutshwayeleka, ngokusebenzisa iindleko zekampani yetjhorensa yepilo.</p>
---	---

Figure 5.3: Aligned parallel English-isiNdebele text

Figure 5.3 shows the alignment results. The isiNdebele sub-divisions were removed so that the remaining paragraphs correlate with their English counterparts. It needs to be emphasised that no editing took place except that the sub-divisions were deleted to form one paragraph. The isiNdebele number 5 and 6 were merged and aligned with the English number 5. The original number 7 in isiNdebele, was aligned with the English number 6. After the manual text alignment, text conversion took place. This was done before the Wordsmith Tools could be used. The corpus files were converted to *text* (.txt) format. Thereafter, term extraction took place.

5.4 TERM EXTRACTION

The purpose of this chapter is to identify term candidates for extraction. Term extraction is the recognition of a term candidate in a specific text, here a medical corpus. Term extraction is the mining or retrieval of relevant terms from a given corpus. Term candidates on the other hand, are words or phrases that appear to be terms. Term extraction in this study is done semi-automatically. This means that it is done both through the corpus processing tool (which will be discussed later) as well as manually.

The identification of term candidates is done methodologically by comparing the corpus of a language for special purposes (LSP) to its general purpose (LGP) corpus. The general corpus is also called the reference corpus. The general or reference corpus is important because it represents a language across a wide range of speakers/writers, contexts, and registers. The general or reference corpus is non-technical and can be used together with the special purpose corpus for the purposes of comparison. This is done in order to identify features of a specialised language that differ from general language. The general corpus used in this study is summarised in Table 5.6.

Table 5.6: *Summary of isiNdebele general corpus*

N	Name	Publisher	Size of the file	Number of tokens
1	Ukuthwala 'abduction'	Department of Justice	15,863	2,439
2	Umthethokambiso we-Sewula Afrika 'Constitution of the Republic of SA'	Republic of South Africa	303,361	32,107
3	Isitatimende somHlangano weKhabinethe	Parliament of SA	83,456	10,035
4	IsiNdebele iLimi leKhaya, iphepha lokuThoma (iGreyidi 12)	Department of Basic Education	299,914	28,286
5	IsiTatimende somThethokambiso weKharikhyulamu nokuHlola (iLimi leKhaya: AmaGreyidi 4-6)	Department of Basic Education	303,361	32,107
6	UkuHlolwa kweliZwelo ke lomNyaka (2013)	Department of Basic Education	23,623	2,138
7	IsiNdebele iLimi leKhaya, P1 Feb Matjhi 2018	Department of Basic Education	23,623	2,138
8	IsiNdebele HL P2 Nov 2017	Department of Basic Education	63,823	6,064
9	IsiNdebele FAL P1 G 12 2017 Feb Matjhi 2017	Department of Basic Education	20,706	1,927
10	IsiNdebele FAL P2 G12 Feb-Matjhi 2017	Department of Basic Education	14,799	1,373
11	IsiNdebele iLimi leKhaya P1 Feb Matjhi 2018 (Memo)	Department of Basic Education	299,914	28,286
12	IsiNdebele FAL	Department of Basic Education	21,109	2,007

N	Name	Publisher	Size of the file	Number of tokens
13	isiNdebele iLimi lesiBili lokuNgezelela (P1) 2008 Nov	Department of Basic Education	26,113	3,115
14	Memo p1	Department of Basic Education	16,734	1,543
15	Nde New 2013 Grade 2 HL	Department of Basic Education	10,835	742
16	SAL Paper 2	Department of Basic Education	26,113	2,591
17	SADiLaR (web-site)	SADiLaR	18,951,15	1,186,015
	Total			1,342,913

Table 5.6 reflects the diversity of texts collected for this corpus. There are isiNdebele question papers from the Department of Basic Education's official website. [www.education.gov.za>Curriculum>Examinations](http://www.education.gov.za/Curriculum/Examinations)> NSC Past Examination papers, the Constitution of the Republic of South Africa in isiNdebele, the cabinet meeting statement and the document on *ukuthwala* 'abduction'. Most of the documents in these general corpora are from the Department of Education. Since it is mandatory for the Department of Education to provide documents for learners in all official languages, it also has to have its examination papers set in isiNdebele, hence it was easy to acquire the isiNdebele question papers for the corpus. The biggest size of the general corpora is from the SADiLaR web-site (2021). Corpora here are made up of 1,010 texts which include amongst others, the State of the Nation Address (SONA) by various presidents of the Republic of South Africa, articles from National Language Services, speeches by various political ministers, various announcements from government departments such as the Department of Arts and Culture and National Language Services as well as minutes by various parliamentary committees.

The general corpora of this study contain 1,342,913 tokens and the specialised corpora (both the written and the literacised one) contain 2,102,967 tokens. This is significant because the size of the general corpora need to be bigger than specialised corpora especially because the KeyWords function was also used.

Frequency and key words analysis were explored for generating domain specific candidates for headword selection. Using such statistical approach has proven in this study to be faster, reliable and free from human error or bias. A discussion on the WordSmith Tools follows.

5.4.1 Term extraction through the WordSmith Tools

The WordSmith Tools has been used in this study to identify term candidates from both the written and the spoken corpus (literacised). The WordSmith Tools has the following functions, namely: The Concord, KeyWords and WordList functions. All these functions were used in analysing the written corpora of this study. The corpus had to be converted into plain text (.txt) format first, before the WordSmith Tools, version 6 could be used. See Figure 5.4.



Figure 5.4: WordSmith Tools functions

(i) The WordList/The Frequency-ranked WordList

The Frequency-ranked WordList is the function which ranks words in a corpus in a descending order of frequency. The frequency lists give an indication of the topic of a text. The conducting of a frequency count of a word or term can be regarded as an important and valuable tool for terminological and lexicographical purposes.

The WordList comprises a list of all the words in the corpus. This is the first function the researcher used. After selecting a WordList function, the output came in three different formats, namely the frequency-ranked word list, the statistical analysis and an alphabetically ordered word list. It needs to be highlighted that for this research, because of its size of corpora, the statistical analysis and an alphabetically ordered word list are not that relevant.

It is also problematic to apply the frequency of occurrence WordList to a language like isiNdebele because of its class system. The singular and plural forms are shown separately on the frequency word list. The other reason is that the researcher does not have a lemmatised form of texts that could be offered to the tool which makes it difficult to get the expected results through the frequency WordList. The Frequency count or WordList tool on its own cannot always be enough and should always in languages such as isiNdebele, be combined with Concordance to grasp a term or phrase in its proper context.

Having mentioned the above about the disadvantages of the frequency WordList function in isiNdebele, it needs to be mentioned that there are, however, also advantages of using the WordList function. The advantages are that through the frequency WordList, synonyms and terms with variant forms are extracted. Consider the following terms with variant forms:

udorhodere vs udorhodera 'a doctor'

umulwani vs umulwana 'a germ'

ingogwana vs ingogwani 'a virus'.

Another problem is that of the function words: with no part-of-speech tagging available, the tool cannot distinguish between important (=content) words and non-important (=function) words. Before one can understand how functions such as KeyWords and Concordance function, one needs to understand how the word list functions. The researcher illustrates the word list with function/grammatical words (Table 5.7) and the one without function words (Table 5.8).

Table 5.7: *isiNdebele most frequent first 100 types*

N	Type	Frequency	N	Type	Frequency
1	#	236	51	la	16
2	bona	120	52	TB	16
3	umuntu	70	53	abaphila	15
4	namkha	67	54	ilanga	15
5	abantu	56	55	isibawo	14
6	iinhlungu	53	56	izinto	14
7	nangabe	50	57	ku	14
8	wokugula	47	58	ngaphandle	14
9	nanyana	46	59	ngemva	14
10	kudambise	41	60	okumele	14
11	namatshwayo	41	61	abanye	13
12	begodu	40	62	ebantwini	13
13	okusingatha	40	63	imiphumela	13
14	kanye	35	64	kodwana	13
15	bese	34	65	kufanele	13
16	i	34	66	umulwani	13
17	kobana	33	67	yesondlo	13
18	ukutjheja	33	68	amalwele	12
19	inomboro	30	69	isizo	12
20	khulu	30	70	izinga	12

N	Type	Frequency	N	Type	Frequency
21	kumele	30	71	nofana	12
22	iingazi	29	72	ubujamo	12
23	isiguli	29	73	umlandu	12
24	nokukhubazeka	29	74	iinguli	11
25	ngeengazi	28	75	kungaba	11
26	HIV	27	76	leyo	11
27	khona	27	77	umtato	11
28	ngaphambi	25	78	e	10
29	imali	23	79	ilwazi	10
30	isikhathi	23	80	kube	10
31	lokhu	23	81	kuqakathekile	10
32	na	23	82	njani	10
33	le	22	83	yebhanga	10
34	lokha	22	84	zoke	10
35	lokho	22	85	amathuba	9
36	bonyana	21	86	eendaweni	9
37	umntwana	21	87	hlangana	9
38	africa	20	88	imininingwana	9
39	ama	20	89	kanti	9
40	kuhle	20	90	ngayo	9
41	ikhotho	19	91	ngombana	9
42	ngendlela	19	92	ukwenza	9
43	ukuze	19	93	wena	9
44	ekhotho	18	94	yesizo	9
45	iinhlahlubo	18	95	abatjha	9
46	ephasini	17	96	amanzi	8
47	umzimba	17	97	ihlangano	8
48	we	17	98	iindleko	8
49	indlela	16	99	iindlela	8
50	ingabe	16	100	ipilo	8

Table 5.7 reflects that grammatical words or function words dominate all frequency word lists. These include nouns (*umntwana*, *iindlela*, *amathuba*, *imininingwana*, etcetera), adverbs (*ekhotho*, *eendaweni* and *ephasini*), pronouns (*lokhu* and *wena*), conjunctives (*ukuze*, *bonyana*, *kanti* and *ngombana*), etcetera. According to this table, medical terms are few, for instance ranked number 6 is the term *iinhlungu* 'pain' (noun) with 53 occurrences. The term *wokugula* 'of disease'

(possessive) is ranked number 8 and it has 47 occurrences. Ranked from number 22 to 24 are medical terms *iingazi* 'blood' (noun), *isiguli* 'a patient' (noun) and *nokukhubazeka* 'with disability' (associative). They all have 29 occurrences. The term *ngeengazi* 'with the blood' is ranked number 25 and it has 28 occurrences. Consider the following word list without function words, from doctors' interviews.

Table 5.8: *Doctors' interviews frequency-ranked word list without function words*

N	Word	Frequency	N	Word	Frequency
10	AIDS	1	111	emalangeneni	1
25	amagciwana	1	112	embriyo	3
26	amakhonthrasepthivi	2	117	endomethriyamu	5
27	amalangakho	1	138	ifantom	1
29	amalwele	2	142	ibhayiphola	1
33	amatshwayo	1	147	igroyini	4
39	angathathelaniko	2	148	ihayiphehayidrosisi	1
46	anjengemasela	1	149	ihephayithisi	1
47	anjengesifo	1	150	i-HIV	2
49	anthidiphresentsi	1	156	iinguli	3
50	anthibhayothiki	1	158	iinkhandelimbeleko	1
51	anthibodi	2	159	ikankere	1
52	anthibhayothiksi	1	160	ikhenza	4
54	asidisi	1	163	imbilapho	2
59	athathelanako	1	164	imelanono	2
71	banetjhukela	1	165	imenophosi	1
80	bezamaphilo	2	171	influwenza	1
81	bodorhodere	2	176	ingogwana	2
87	khenza	2	181	iphrosteyiti	1
88	khomunikhebuli	5	182	iplasentha	2
92	dayibhethisi	1	185	isidisi	1
93	diphreshini	2	186	isifo	2
99	ekthophiki	2	190	isindromu	2
110	emagciwaneni	1	193	istrowuku	1

Table 5.8 reflects more transliterated terms than coined ones. Out of the 48 terms in the table, 30 are transliterated terms, for example term number 181 is *iphrosteyiti* 'prostate' and number 165 is *imenophosi* 'menopause'. This shows that doctors are more exposed to English words than to the coined ones. The number of coined terms is 16, examples of which are *amagciwana* 'bacteria' at number 25 and *ingogwana* 'virus' at number 176. The two remaining terms are abbreviations (the acronyms 'HIV' and 'AIDS').

(ii) The KeyWords

The researcher applied the KeyWords function with the purpose of extracting all term candidates from the corpus. It needs to be emphasised that this is the first time that the tool is used for isiNdebele and for that reason the researcher had to manually validate the resulting list of term candidates.

The KeyWords function was used to calculate words which are key in a text, that is, words used much more frequently or much less frequently in each corpus. Through this function, terms used in the medical domain were identified. This was done through the calculation of keyness, which isolates words which are key to the analysis corpus. The analysis corpus is domain specific and technical. To calculate keyness, the frequency of each word in the word list of the text under investigation is compared with the frequency of the same word in the reference word list. The output is a list of key words, or words whose frequencies are higher in the analysis corpus than in the reference corpus. The reference corpus is made up of the language for general purposes. It is non-domain specific as well as non-technical.

Any word which is found to be outstanding in its frequency in the text, is then considered to be 'key'. Key words are presented in their order of outstandingness. The KeyWords function provided the term candidate list which is represented in Table 5.9.

Table 5.9: *The resulting KeyWords function terms*

Number	Resulting term
1	iinhlungu
2	wokugula
3	kudambise
4	isiguli
5	ngeengazi
6	iingazi
7	nokukhubazeka
8	HIV
9	ukutjheja
10	TB
11	umulwani
12	iinguli
13	amalwele
14	ukutjhejwa
15	ubulwele
16	ukutshayeleka
17	iinhlahlubo
18	imiphumelakho
19	udorhodere
20	isilonda
21	nokuhlanza
22	beswigiri
23	amatshwayo
24	wesibhedlela

Number	Resulting term
25	dorh
26	kwesibhedlela
27	obungalaphekiko
28	lezamaphilo
29	sista
30	sezokwelapha
31	abodorhodere
32	iinhlungu
33	zetjhejo
34	kudambise
35	okusingatha
36	esibhedlela
37	ukubelethela
38	imitjhoga
39	nokukhubazeka
40	sokwelapha
41	lokwelapha
42	wamazinyo
43	amahomowuni
44	ikhenza
45	singabelapha

This table reflects the first version of the semi-automatically found term candidates. According to Table 5.9, multiwords are not identified and extracted by the KeyWords function. For instance, when one looks at number 1 and 3, it is a multi-word *kudambise iinhlungu* 'to relieve pains', but it is identified as two separate words with different numbers and frequencies. The same is evident with number 15 and 22. This is one word, *ubulwele beswigiri* 'the disease of sugar/ diabetes' but the two words are identified by the KeyWords function as two separate words, with different frequencies.

The terms that resulted from the KeyWords function had to be validated manually. This is because the KeyWords function cannot sort the variants of the term candidate list, neither can it identify the multi-word term candidates. The researcher therefore had to sort the variants and also identify multi-word term candidates manually. The manual identification of term candidates will be discussed in detail in 5.4.2.

(iii) The Concordance

Using the Concordance software helped with the creation of an in-depth knowledge of the usage of terms in an acceptable way. The Concordance displays the results of a search so that the linguist can read them in the context in which they occur in the text. The Concordance helps to visualise data. It enabled the researcher of the current study to find patterns of similarity or contrast in the words surrounding the search term.

The search term and its co-text are arranged so that the textual environment can be assessed and patterns surrounding the search term can be identified visually. Furthermore, reading concordances allows the user to examine what occurs in the corpus, to see how meaning is created in texts, how words co-occur and are combined in meaningful patterns, without any fixed pre-conceptions about what those units are. See Figure 5.5.

N	Concordance	Set	Word #	Sen	Sen	Par	Par	Hea	Hea	Sec	Sec	File	Date	%
1	sithokozile. 11. 02 Julayi 2018: <i>ikankere</i> yePrhosteyiti: Busayi and Dr		4,793	551	30%	0	33%			0	33%	Oral corpus tras	2019/Jan/14 00:	33%
2	- umzimba sewuya ku-menopause. B: <i>ikankere</i> yesibelesho ibonakala ngani?		9,680	1,0	50%	0	67%			0	67%	Oral corpus tras	2019/Jan/14 00:	66%
3	uthola kuyi cancer of the <i>cervix</i> ; <i>ikankere</i> yomlomo wesibelesho. Busayi		4,390	501	83%	0	30%			0	30%	Oral corpus tras	2019/Jan/14 00:	30%
4	yesibelesho. Cancer of the <i>cervix-ikankere</i> yomlomo wesibelesho.		9,429	1,0	71%	0	65%			0	65%	Oral corpus tras	2019/Jan/14 00:	64%
5	ikhenza yomphimbo kibomma <i>ibanga</i> <i>ikankere</i> yomlomo. B: Ingabe		8,314	90%	89%	0	57%			0	57%	Oral corpus tras	2019/Jan/14 00:	57%
6	Kuba yini kibomma <i>ibanga</i> <i>ikankere</i> yomlomo wesibelesho. Dr:		8,296	907	75%	0	57%			0	57%	Oral corpus tras	2019/Jan/14 00:	57%
7	isikhathi uthola umma ane <i>khenza</i> , <i>ikankere</i> yesibelesho. Kesinye isikhathi		4,380	50%	89%	0	30%			0	30%	Oral corpus tras	2019/Jan/14 00:	30%
8	njalonjalo nesikhumba siyaba <i>nayo</i> <i>ikankere</i> . Khulukhulu eendaweni		1,068	117	100%	0	7%			0	7%	Oral corpus tras	2019/Jan/14 00:	7%
9	sehlile kesinye isikhathi. Cancer of the <i>uterus-ikankere</i> yesibelesho. Cancer of		9,424	1,0	83%	0	65%			0	65%	Oral corpus tras	2019/Jan/14 00:	64%
10	amihlobo eyahlukeneko <i>yekhenza/</i> <i>ikankere</i> .Iprostate cancer ne		14,212	1,5	100%	0	98%			0	98%	Oral corpus tras	2019/Jan/14 00:	98%

Figure 5.5: Multi-words on cancer

- **The terms *ikhenza* and *ikankere* ‘cancer’**

Because the two terms *ikhenza* (derived from the English ‘cancer’) and *ikankere* (derived from the Afrikaans ‘kanker’) are used interchangeably in isiNdebele, the researcher had to establish which of the two terms (*ikhenza* and *ikankere*) are mostly used. The researcher discovered that they share the same percentage (50-50) as far as their usage is concerned. See the following extract on *Ikhenza yesikhumba* ‘skin cancer’ which was presented on 08 January 2018 as well as ‘Hysterectomy and Sterilisation’ presented on 14 May 2018 and ‘HPV (Human Papilloma Virus)’ presented on 12 March 2018.

A. IKHENZA YESIKHUMBA 08 JANUARY 2018

1. **Mrhatjhi:** Siyakwamukela dorhodere, akhese usihlathululele indaba yekankere yesikhumba.
2. **Dorhodere:** Njengezinye izitho eziba nekhenza njengekhenza yamathumbu nesibeletho njalonjalo, nesikhumba siyaba nayo ikankere. Khulukhulu eendaweni eziqalene khulu nelanga njengobuso. Iimpumulo zona ziqalene nelanga khulu, imilomo iindlebe njalonjalo. Nakezinye iindawo la ilanga lingafiki khona solankana kunesikhumba. Iyaba khona ikhenza yesikhumba njengangaphasi kweenyawo namkha ingaphakathi lesandla. Eminweni nezithweni zangasese. Nokho ilanga linomthelela omkhulu kukankere yesikhumba.
3. **Mlaleli:** Ngiyathokoza mrhatjhi, nawunamabala amhlophe emkhonweni ingabe kungaba ngiyo ikhenza na. Nginobuhlungu ezisweni ingaba ngiyo ikhenza na?

B IHISTEREKTHOMI NE STERELIZEYISHINI 14 MAY 2018

Matron: Kesinye isikhathi kune 'cancer of the uterus' ikankere yesibeletho. I'cancer of the cervix', yikankere yomlomo wesibeletho..i-'endometrium cancer'. Busayi i-endomethriyamu liboda lesibeletho, ibodeli nalinamasele wekhenza.

C. I-HPV (I-HUMAN PAPILLOMA VIRUS) 12 MARCH 2018

1. **Mrhatjhi:** Asingene endabeni Dorh Masombuka sikhulume nge-HPV, phana yini i-HPV?
2. **Dorh:** Busayi nasikhuluma ngeHPV sikhuluma nge'Human Papilloma Virus'. Leli ligciwana elibanga iinsumpa ebantwini, uthole baneensumpa ebusweni, ezandleni iindawo zoke. Kodwana okuspecial ngegciwaneli kukuthi libanga ikhenza yomlomo wesibeletho kibomma, ngikho okwenza ubulwelobu bube serious khulu.
3. **Mrhatjhi:** Ingabe i-HPV iba kibomma kuphela namkha nakibobaba? Kuba yini kibomma ibanga ikankere yomlomo wesibeletho.
4. **Dorh:** Inarha inamavayirasi amanengi amihlobohlobo-kanengi elinye nelinye lihlala endawenalob. Kibobaba ibanga ikhenza yomphimbo kibomma ibanga ikankere yomlomo.
5. **Mrhatjhi:** Ingabe inamtshwayo i-HPV ebonakala ngawo namkha kumele sibone umuntu oneentsumpa bese sithi unayo?

6. **Dorh:** Bussai akutjho bona umuntu oneentsumpa unayo **ikhenza** le. I-Immune system yakho nayinamandla kungenzeka ingakubamba ugcine uneensumpa kwaphela.
7. **Mrhatjhi:** Babantu abanengi abanayo i-HPV?
8. **Dorh:** Abantu abanengi bangaba nayo, sekufana ne-HIV. Abanye baba neensumpa kwaphela kodwana kungabi **nekhenza**.

The above extracts reflect that the two isiNdebele terms (*ikankere* and *ikhenza*) for 'cancer' are used interchangeably, even by the same speaker.

In extract **A**, number 2, the doctor uses both terms interchangeably in the same sentence. The presenter refers to 'cancer' as *ikankere* and as *ikhenza* in number 1 and 3 respectively.

In extract **B**, the matron who was on the show also referred to 'cancer' as both *-kankere* and *-khenza*.

In extract **C**, number 4, both *ikhenza* and *ikankere* are used in the same sentence by the doctor while he is explaining the types of 'cancers'.

Key Word In Context (KWIC) displays of the search pattern *ikankere*

Just as in the case of the term *ikhenza*, the term *ikankere* is preceded by the verb *ibanga* 'it causes'. The search word *ikankere* is followed by the type of 'cancer'. *Ikankere yesibelesho* 'cancer of the womb'.

Before the search word *ikankere* in Figure 5.5

Line 5 and 6 *ibanga ikankere* 'it causes cancer'

After the search word *ikankere* in Figure 5.5

Line 1 *Ikankere yephrosteyiti* 'cancer of the prostate'

Line 2 *Ikankere yesibelesho* 'cancer of the womb'

Line 3 & 4 *Ikankere yomlomo wesibelesho* 'cancer of the cervix'

The discussion on concordance includes collocates. See the following:

(iv) The Collocates

Bowker and Pearson (2002:124) say that collocates are words which typically occur in the vicinity of your search pattern. Collocates play an important role in corpus linguistics, they make it easier for the learner to understand the usage of words. Collocates assist with the understanding of how two words come together meaningfully. Some concordances offer an additional facility which frequently ranks the words that appear in the vicinity of the search pattern. This facility computes and displays the most frequently occurring collocates. See Figure 5.6.

N	Word	With	Relation	Texts	Total	Total	Total	f	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	IKHENZA	ikhenza	0.000	1	25	0	0	0	0	0	0	0	0	25	0	0	0	0	0
2	LE	ikhenza	0.000	1	5	0	5	0	0	0	0	0	0	0	5	0	0	0	0
3	YESIKHUMBA	ikhenza	0.000	1	5	0	5	0	0	0	0	0	0	0	5	0	0	0	0
4	BONA	ikhenza	0.000	1	5	3	2	0	1	2	0	0	0	0	0	0	1	1	0
5	KODWANA	ikhenza	0.000	1	3	2	1	1	1	1	0	0	0	0	0	0	0	1	1
6	NAYO	ikhenza	0.000	1	3	2	1	0	0	0	0	2	0	0	0	0	1	0	0
7	IBIZWA	ikhenza	0.000	1	3	1	2	0	1	0	0	0	0	0	0	2	0	0	0
8	NGIYO	ikhenza	0.000	1	2	1	1	0	0	0	0	1	0	0	0	0	0	0	1
9	NAMKHA	ikhenza	0.000	1	2	1	1	0	0	0	0	1	0	0	0	0	0	1	0
10	YI	ikhenza	0.000	1	2	1	1	0	0	0	1	0	0	0	0	0	0	1	0
11	YOMLOMO	ikhenza	0.000	1	2	0	2	0	0	0	0	0	0	0	1	0	0	0	1
12	UKUHLATHULL	ikhenza	0.000	1	2	2	0	0	0	0	0	2	0	0	0	0	0	0	0
13	YE	ikhenza	0.000	1	2	1	1	0	0	1	0	0	0	0	1	0	0	0	0
14	NA	ikhenza	0.000	1	2	1	1	0	0	0	0	1	0	0	1	0	0	0	0
15	IYINI	ikhenza	0.000	1	2	1	1	0	0	0	0	1	0	0	0	0	0	0	1
16	KHONA	ikhenza	0.000	1	2	2	0	1	0	0	0	1	0	0	0	0	0	0	0
17	INGABE	ikhenza	0.000	1	2	2	0	0	0	2	0	0	0	0	0	0	0	0	0
18	IYATHELELANI	ikhenza	0.000	1	2	1	1	0	0	0	1	0	0	0	0	0	0	0	1
19	LEYO	ikhenza	0.000	1	2	1	1	1	1	0	0	0	0	0	1	0	0	0	0
20	MELANONO	ikhenza	0.000	1	2	1	1	0	0	0	0	0	1	0	0	0	0	0	1
21	KIBOMMA	ikhenza	0.000	1	2	0	2	0	0	0	0	0	0	0	0	1	1	0	0
22	IBANGA	ikhenza	0.000	1	2	1	1	0	0	0	0	0	1	0	0	0	1	0	0

Figure 5.6: Display of collocates for the search pattern *ikhenza*

Figure 5.6 displays the search pattern *ikhenza* in the middle column and then shows, for each position to the left and right of the search pattern, which word occurs most frequently. Thus, *ikhenza* appears in line 1 a total of 25 times. Of these, it occurs 25 times in the center, 0 (zero) times to the left and 0 (zero) times to the right.

Bowker and Pearson (2002:126) say that the advantages of using the collocate viewer is that it does a lot of the hard work for you and saves you having to sort and resort your concordances to bring typical patterns into relief. If you take the term *ikhenza*, you will notice that it is more usual to use it with *ebanga* ‘causes’ of cancer than *eyenza* ‘makes’.

Figure 5.7 reflects the search pattern which is *iingazi* in the middle column and then shows, for each position to the left and to the right of the search pattern which word occurs most frequently. In this figure, *iingazi* appears in line 3 a total of 4 times. Of these, it occurs 0 times in the centre, 6 times to the left of the search pattern, and 2 times to the right of the search pattern. Therefore, *bona* is the next most frequently occurring collocate. See the following collocates of the term *iingazi* in Figure 5.7.

N	Word	With	Relation	Texts	Total	Total	Total	f	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	IINGAZI	iingazi	0.000	1	16	0	0	0	0	0	0	0	0	16	0	0	0	0	0
2	BONA	iingazi	0.000	1	7	6	1	0	1	1	2	2	0	0	0	0	0	1	0
3	EZINTATHU	iingazi	0.000	1	4	2	2	1	0	1	0	0	0	0	0	0	0	2	0
4	ABANTU	iingazi	0.000	1	3	2	1	0	1	1	0	0	0	0	0	0	0	1	0
5	UMUNTU	iingazi	0.000	1	3	2	1	1	1	0	0	0	0	0	0	1	0	0	0
6	AMA	iingazi	0.000	1	3	0	3	0	0	0	0	0	0	0	0	2	1	0	0
7	NANGABE	iingazi	0.000	1	2	1	1	0	0	0	1	0	0	0	0	0	0	0	1
8	NAMA	iingazi	0.000	1	2	0	2	0	0	0	0	0	0	0	0	0	0	0	2
9	ZABO	iingazi	0.000	1	2	0	2	0	0	0	0	0	0	0	1	0	0	1	0
10	PLASMA	iingazi	0.000	1	2	0	2	0	0	0	0	0	0	0	0	0	1	1	0
11	NGAPHAMBI	iingazi	0.000	1	2	2	0	0	1	1	0	0	0	0	0	0	0	0	0
12	OKUMASOTJA	iingazi	0.000	1	2	2	0	0	0	0	2	0	0	0	0	0	0	0	0
13	BAKHETHA	iingazi	0.000	1	2	1	1	0	0	1	0	0	0	0	0	0	0	0	1
14	AKHIQHIZA	iingazi	0.000	1	2	2	0	0	0	0	0	0	2	0	0	0	0	0	0
15	ABANYE	iingazi	0.000	1	2	1	1	1	1	0	0	0	0	0	0	0	1	0	0
16	EZIBALWE	iingazi	0.000	1	2	1	1	0	1	0	0	0	0	0	0	0	0	0	1
17	KOBANA	iingazi	0.000	1	2	2	0	0	0	1	1	0	0	0	0	0	0	0	0
18	IHLANGANO	iingazi	0.000	1	2	1	1	0	0	1	0	0	0	0	0	1	0	0	0
19	EZIBOVU	iingazi	0.000	1	2	0	2	0	0	0	0	0	0	0	2	0	0	0	0

Figure 5.7: Collocates of the term *iingazi*

As mentioned initially, extraction of terms happened semi-automatically in this study. A discussion on manual extraction follows.

5.4.2 The manual extraction of terms

Table 5.9 presented the resulting KeyWords function terms. The output of the tool is far less than that of manually identified term candidates. Because the tool found fewer term candidates, the researcher had to manually find candidates and later compare the term candidates found by the tool and those found manually.

5.4.2.1 *The manual extraction of terms from the corpora*

The manual extraction was first done with the written corpora. Table 5.10 displays the manually identified term candidates from the corpora (both the written and the spoken, literacised corpus) of this study.

Table 5.10: *Manually identified terms from the corpora of this study*

ngokuhlukumeza umkhumbulo	ukuya emsemeni	itjhukela	umhlengikazi
ukwabelana ngokomseme	amatshwayo wokugula	nawunetjhukela	udorhoderakho
ukuzwisa ubuhlungu	ukudambisa iinhlungu	iinguli	ukugula
izinga eliphezulu lokuzithwala	ukutjhejwa kwesilonda	iinhlungu	uyarholophala
iimfundo eziphathelene	ukubeletha ngokusikwa	amaphanado	HIV
nezomseme	abasebenzi bezamaphilo	ibhrufeni/imofimi	iinhlahlubo
ukusuleleka kwamalwele	iingazi ezihlwengileko	ekhubazako	namalaborithri
ayingozi	ukuhlongakala kwamasana	iindakamizwa	ukungatshwayeleki
ukubogaboga kwamasotja	ukunikela ngeengazi	emtholapilo	isifilisi
womzimba	ukubuya endleleni	ibhanditjhi	igonoriya
ukuhlolisa isifo setjhukela	imiphumela enembako	ipulasta	ikholera
ukugandeleleka	okubhebhedhlisa amalwele	amagciwana	mumulwana
ngokomkhumbulo	ukutshwenywa mgogodlha	iphika	ukuhlongakala
amanceba neembazi	ukutshwayeleka	abanephika	ukwelatjhwa
amathambo aphukileko	ngomulwani	ubuhlungu	amakhondomu
ukuzikhupha ebanganini	ukuhlolilewa iHIV	ubulwelobu	esibhedlela
ukuhlolilewa ingogwana	umulwani weHIV	isilonda	ukunona
yeNtumbantonga	iTB netjhukela	isiguli	bodorhodere
imitjhoga egogobalisa	umulwana weTB	imitjhoga	yikholesteroli
intumbantonga	iTB iyalapheka	istresi/sitresi	iTB
amalwele ayingozi ebantwini	imitjhoga yeTB	imigreyini	iHIV nentumbantonga
abatjha	mumulwani we-AIDS	abane albhinzimu	sekunemitjhoga
amalwele athathelana	ukugonyuluka nokuhlanza	imelanini	ukwelapha
ngokomseme	amathumbu asebenzako	imelanono	dorhodere
ipilo yomzimba nengqondo	ukugijinyiswa yindeneni	iingazi	i-SANBS
abantu abaphila	ukurhuda nokuhlanza	ikhanselingi	mavayirasi
ngokukhubazeka	ubulwele obungalaphekiko	sayikhothraphi	iingulani

amatshwayo wekhenza	ukukhandela ikholera	idiphreshini	ubulwelobu
yesikhumba	amalwele athathelanako	ihayiphehayidrosisi	iintjhubaba
ikonzo yabantu abadala	ukutjhejwa komlomo	emtholapilo	yimelanini
ukwelatjhwa-ngokwengqondo	ukuhlanzwa kwesiguli	i-aleji	mahomowuni
iinhlahla zobulwele beswigiri	ukudluliswa kwesiguli	ivasekthomi	iinkhandelimbeleko
ukurhatjheka kobulwele	ukutjhejwa kwesiguli	ilaborathri	iHPV
ukuhlololwa umulwana	amatshwayo wokuthulula	ibhayopsi	igciwana
ukusikinywa kwesiguli	ibandulo labelaphi	nesilonda	iinsumpa
nokusehlisa embhedeni	eminyangweni	ukwethuka	nesilonda
ukuphela kwamanzi	yezamaphilo	iphenkhrisi	ukwethuka
emzimbeni	amatshwayo wekholera	i-OCD	amafobiya
ilungelo lezamaphilo	umzimba omkhulu	iskizofreniya	nginetjhukela
nokubeletha	amasotja womzimba	ukuhalusineyitha	skizofreniya
iinhlahlubo ezikhombisa	ukutjheja iinguli	amafayibhrotsi	ngemitholapilo
ukutshayeleka	ikhenza yamathumbu	amaphiriyodipheyi	ozithweleko
umulwana obanga	ikhenza yesibelesho	ni	imbeleko
intumbantonga	ukuhlangana kwehloko	i-Op	ivayithiligo
abanikeli besizo lezamaphilo	ukuvimbela ukopha	papsmiye	amaGP
ukuba semalangeneni	igandeleleko	ukumunyisa	i-wheel-chair
ukuba semzombelezweni	ngokomkhumbulo	iphregnensi	irhudo
ukubuya endleleni	bafakelwe iingazi	homowuni	nesilonda
ubulwele beswigiri	amasotja womzimba	iTOP	ukwethuka
obulawulekako	ukukhupha isibelesho	namaringiwemu	amafobiya
ukwelatjhwa kwamazinyo	ukubopha isibelesho	amasayikholoji	iHPV
ukuphila nomulwana weHIV	isidisi samatjhubhu	umgomani	igciwana
AIDS	isidisi somkhumbulo	umkhuhlani	iinsumpa
ikankere yomlomo wesibelesho	isidisi seenyama/	nginehayiphathe	ibhayiphola
ukusilingeka nokuba nehliziyo	samathitjhu	nshini	ijemu
encani	ukungaboni isikhathi	istrowuku	imbilapho
ihlahlubo yokobana usidisi	ukukha umrorho	iFAST	esibhedlela
ukurhululwa kombungu	amapilisi wehloko	imalariya	wokuhlinzwa
ukutlhogonyelwa kwangemva	ihloko ekhronikhi	ukumilisa	ama-aleji
kokubeletha	ihisterkthomu nesterilizeyi	ukuthulula	ukuhlahluba
	shini	ifiva	kwezokwelatjhwa

abantu bobulwele	ukuya emsemeni	ekhemisi	ukuzithwala
oburhwahlaphazako behliziyo	udorhodere womkhumbulo	amatshwayo	udorhodere
abantu abaphila nomulwana	isifo setjhukelai	wokugula	abodorhodere
wentumbantonga	ukuhlatjwa emnweni	ivayirasi	iflu
ubulwele bamaphaphu	phrosteyiti khenza	iingogobalisi	umdlavuzwa
abasebenzi bamalaborathri	ukuhlolwa kwesibeletso	ingogwana	umulwani
umgamani	amatshwayo wephrosteyiti	lezamaphilo	ustrejini
othuwelelako(N1H1)	amalwele angathathelaniko	wesiguli	sista
iimpumulo eziphuma	izifo ezingathathelaniko	wezokwelatjhwa	ihomonal-imbhalansi
amathimila	abonese nabodorhodere	kumagayini	imalnyuthrishi
izinga eliphezulu lokutjhisa	ukwelulekwa	ikholikhi	idayiriya
amatshwayo weTB	ngokomkhumbulo	ama-	amabhaktheriya
amazinga wedlhukhosi	ubuthi bembuzana	anthibhayothiksi	
ukuphelelwa mumoya	ubulwele besikhumba	amaphenisilini	
amatshwayo wobulwele	ubuhlungu bokumilisa	iingogwana	
ukuphelelwa mamanzi	kudambise iinhlungu	ihayiphathentjhini	
emzimbeni	irekhotso lezamaphilo	ama -	
ukuphathwa siyezi	ilunga eliyihloko	anthibhayothiksi	
ihlelo lokubelethisa	itjhejo lezamaphilo		
umma osidisi	udorhodere wengqondo		
ukuba sidisi	ukwelulekwa		
amalwelele athathelanako	ngokomkhumbulo		
amalwele angathathelaniko	udorhodere womkhumbulo		
amalwele wehliziyo	ukubeletha okujayelekileko		
ingubo yesibeletso	ukubelethela ekhaya		
isifo setjhukela	idawuni sindromu		
idawuni-sindromu	iimbhedlela zangeqadi		
ukuphungula ubuhlungu	ubulwele besifuba		
ukubelethiswa ngamanzi	udorhodere wamehlo		
ukubeletha ngokuhlinzwa	udorhodere wamazinyo		
ukuba sidisi ngamawele	ithratjhi esemlonyeni		
ukulimala esinyeni			
ungqongqotjhe wezamaphilo			

Table 5.9 confirms that the KeyWords function identified very few terms in both written and spoken corpora. The number of medical terms identified by the KeyWords function is 48 in total. The researcher then manually identified the terms from both the written and the spoken corpus file (as shown in Table 5.10). The number of terms manually identified from the present study's corpora is 289. It is evident that the manually extracted term candidates are far more than those identified through the KeyWords function. It is for this reason that the researcher opted to use the term candidates identified manually especially because all terms identified by the tool are also found in the manually extracted ones.

What is evident is that there are many more term candidates that have been manually identified as compared to the ones identified by the KeyWords function. Multi-terms have also been identified manually. Table 5.9 shows that there are no multi-terms identified by the KeyWords function. A multi-term such as *kudambise iinhlungu* 'to relieve pains' is identified as separate terms: *iinhlungu* (1) and *kudambise* (3).

5.4.2.2 *Comparison between terms manually extracted from the written and the spoken (literacised) corpus*

Although the spoken corpus of this study became literacised, it became like the written corpus after it was transcribed. The researcher was also interested in comparing term candidates from the spoken material with the ones from the written corpus. There are common terms found in both the written and spoken corpus. What is also observed is that the spoken corpus, as compared to its written counterpart, has more transliterated terms. See the following examples.

The isiNdebele term for a 'nurse', transliterated as *unese* and *usista* in the spoken corpus, is *umhlengikazi* (a coined term) in the written corpus. The term *unese* is borrowed from the English 'nurse' and the term *usista* is borrowed from an English term 'sister'. The term 'clinic' is *umtholapilo* (a coined term) in the written corpus, whereas it is *itlinigi* (a transliterated term) in the spoken corpus. The term *itlinigi* is borrowed from the English 'clinic'.

Choice of words is evident in the spoken corpus regarding terms that are regarded as vulgar. For instance, sexual intercourse is referred to as *ukuya emsemeni* in the written corpus. In the spoken corpus it is referred to as *ukuya emsemeni*, *ukukha umrorho* and *ikonzo yabantu abadala*. The term for 'sexual intercourse' is *ukuya emsemeni* in the written corpus; the very act is called *ukuya*

emsemeni which loosely translates as 'to go to the reed-mat' or *ukukha umrorho* which loosely translates as 'to pick up the vegetables' or *ikonzo yabantu abadala* which loosely translates as 'old people's service or fellowship'. This is euphemistic. Spoken language is characterised by more euphemism as compared to its written counterpart. More examples are detailed in **Appendix F**.

The following is an extract highlighting the difference between candidates identified manually and those identified automatically:

An extract on manually versus automatically recognised term candidates

Ibandulo labelaphi bendabuko....

Ihloso yeemfundwezi kukuthi **abelaphi** bendabuko bafunde bona: Kuyini **ukutjheja** okusingatha ***kudambise iinhlungu namatshwayo wokugula?*** Ngiyiphi indlela engcono ***yokutjheja iinguli*** nasele ***zikhamba ephasini?*** Ukutjheja okusingatha ***kudambise iinhlungu namatshwayo wokugula*** kutjhaphulula ***isiguli eenhlangwini*** nakwamanye ***amatshwayo wokugula akhathaza umphefumulo*** ngaphandle kokuhlosa ukuqeda nya ***ubulwele*** obuphethe umuntu. ***Ukutjheja*** okusingatha ***kudambise iinhlungu*** namatshwayo wokugula kukhuthaza ***ipilo*** bese kuthatha ***ukufa*** njengendlela ejayelekile. Akuhlosileko kukobana abantu ***bakhambe msinyana ephasini*** namkha ***ukubambezela ukufa***.

This is the extract from the isiNdebele Palliative care flier. Terms italicised and bolded are term candidates that were detected automatically. Term candidates that are bolded and underlined are those the tool could not identify, they were manually extracted. This is proof that manual verification is always necessary to ensure that those term candidates that were not identified by the tool, are identified manually.

The KeyWords function is unable to identify multi-term words. It always extracts them separately for example in the case of a multi-term such as *ikhensa yomlomo wesibelesho* 'cancer of the cervix' the tool identifies these words as 3 different terms (namely as *ikhenza*, *yomlomo*, *wesibelesho*). This is also the case with the term *ukubelesho ngokusikwa* 'c-section'. This term is also extracted separately (as *ukubelesho* and *ukusikwa*). As mentioned before, to address this

challenge (the one of being unable to recognise multi-words), the Concordance function was used.

When extracting terms from a technical corpus, there are linguistic and technical factors to consider. The linguistic factors include the nominal, verbal and adjectival word categories.

(i) Nominal categories

When dealing with term extraction in the nominal category, uninflected nominal terms, inflected nominal terms and multi-terms should be considered.

- **Uninflected nominal terms:** The researcher started by extracting single nouns in their uninflected forms. A single noun which was extracted for instance was *amatshwayo* 'symptoms'. The term *amatshwayo* is ranked number 40 in a wordlist with 9,124 entries and it has a frequency of 52. Another term is *iinhlungu* 'pains' which is ranked number 35 and has a frequency of 55. Both these terms are never used in their singular forms.
- **Inflected nominal terms:** To extract inflected nominal terms is much more complex especially in an agglutinating language such as isiNdebele. The term *itjhukela* (noun) is ranked number 93 and it has a frequency of 26. It appears also as *setjhukela* (possessive) (N105 with 25 frequencies), *wetjhukela* (possessive) (N372 with 9 frequencies), *yetjhukela* (possessive) (N808 with 5 frequencies), *yitjhukela* (copulative) (N811 with 5 frequencies), *letjhukela* (possessive) (N349 with 3 frequencies) and *netjhukela* (associative) (N349 with 3 frequencies).

The term *ikhenza* (noun) is for instance ranked number 71 with 32 frequencies, it also appears as *khenza* (noun) (N127 with 22 frequencies), *nekhenza* (associative) (N241 with 13 frequencies), *yikhenza* (copulative) (N336 with 10 frequencies), *ngekhenza* (instrumental) (N999 with 4 frequencies), *wekhenza* (possessive) (N1,085 with 4 frequencies).

Nominal multi-terms: A corpus of lemmatised medical terms does not only contain single terms such as *amatshwayo* 'symptoms' but also multi-terms such as *amatshwayo wokugula* 'symptoms of the disease'. In the word list such terms are listed as separate words with different frequencies. For instance, the term *amatshwayo* 'symptoms' is ranked number 40 and *wokugula* is ranked

number 45, with 47 frequencies. The multi-term *ukutjheja isiguli* 'to take care of the patient' is also ranked separately on the word list. *Ukutjheja* 'to take care' is ranked number 69, with 33 frequencies and *isiguli* 'a patient' is ranked number 72, with 31 frequencies. It is the Concordance tool that was able to identify multi-terms, otherwise other functions identify multi-terms as single terms. See (iii) Concordance (Par. 5.4.1), above.

- **Variant forms:** Terms such *ikankere* and *ikhenza* are used interchangeably and are equally representative since they have the same frequencies in the corpus on cancer. The solution is to list both terms as variants in the glossary.
- **Loan words:** *Ikankere* has a higher frequency (N232 with 14 frequencies) compared to *umdlavuzza*, on the other hand, which is the indigenous coined word and which is ranked number 1,361 with 3 frequencies. Also compare *diphreshini* (N261 with 12 frequencies), *idiphreshini* (N237 with 3 frequencies) and the isiNdebele coined *ukugandeleleka* (217 with 14 frequencies); *ivayirasi* (N561 with 6 frequencies) and *umulwana* (N791 with 5 frequencies) and *ingogwana* (N2067 with 2 frequencies).
- **Acronyms:** HIV (N64 with 34 frequencies), TB (N67 with 34 frequencies), HPV (N263 with 12 frequencies), OCD (N325 with 10 frequencies) and AIDS (N630 with 5 frequencies).

(ii) Verbal categories

It is less complex to extract and lemmatise nouns than it is to extract and lemmatise verbs. For instance, it may be difficult to extract the verb *ukuhlaluba* 'to test' when it appears as *ukuhlalutjwa* (where palatalisation has taken place), or the verb *ukubeletha* 'to give birth', when a causative extension *-is-*, has been added to it and it appears as *ukubelethisa*.

The concordance is not only very handy in establishing the actual frequencies of a term but also in indicating how a certain word is used in context in relation to other words in its immediate environment.

Let us consider the occurrences of the verbal root *-beleth-*. The term *lokubelethisa* is ranked number 1 (N1) in the concordance, it has 8 frequencies in the word list. The term *ekubeletheni* is ranked number 7 (N7), with 1 frequency (F1) in the word list. The term *kokubelethwa* is ranked number 16 (N16) in the concordance and has 3 frequencies (F3) in the word list. According to Figure 5.8 with 57 entries of *-beleth-* considered, it becomes obvious that the term *ukubelethisa*

which has 8 frequencies, is quite popular. The causative extension *-is-*, plays a very significant role here. One cannot give birth on one's own, a midwife is always necessary.

N	Concordance	Set	Word #	Sen	Sen	Para	Para	Hea	Hea	Sec	Sec	File
1	Ihlelo lokuBelethisa Njengelunga	IHle		0	0	0%	0	0%		0	0%	(527201092947 2020/Sep/1
2	yesibhedlela: UMTATO 0860 00 4367 IFEKSI 0861 00 4367	IFE		527	29	88%	0	99%		0	99%	(527201092947 2020/Sep/1
3	kobana amukelwe esibhedlela/alatjwe. TJHEJA: Ukuthabela iinzuzo	TJH		347	19	15%	0	65%		0	65%	(527201092947 2020/Sep/1
4	isiyeleliso esiragela phambili begodu IHlelo lokuBelethisa lidoswa phambili	IHle		75	5	47%	0	14%		0	14%	(527201092947 2020/Sep/1
5	isiyeleliso esiragela phambili begodu IHlelo lokuBelethisa lidoswa phambili	IHle		75	5	47%	0	14%		0	14%	(527201092947 2020/Sep/1
6	kobana ukuzitlolisa ehlelweni lbizo nesibongo seLunga eliyiHloko.	Ibiz		355	19	77%	0	67%		0	67%	(527201092947 2020/Sep/1
7	wakho nokukhetha ekubeletheni. elizakukghonakalisa kobana uzwisise	eliz		140	6	72%	0	26%		0	26%	(527201092947 2020/Sep/1
8	lomphathi weendaba: Ilanga ekungilo: Inomboro yomtato yesibhedlela:	Inoi		520	29	68%	0	98%		0	98%	(527201092947 2020/Sep/1
9	Ibizo nesibongo seLunga eliyiHloko. Amabizo azeleko nesibongo sakamma	Am		359	20	33%	0	68%		0	68%	(527201092947 2020/Sep/1
10	iforomo lokuzitlolisa elizalisiweko. Ilunga lihlole mphathi weendaba	Ilun		217	11	25%	0	41%		0	41%	(527201092947 2020/Sep/1
11	kobana amukelwe esibhedlela/alatjwe. TJHEJA: Ukuthabela iinzuzo	alat		346	18	100%	0	65%		0	65%	(527201092947 2020/Sep/1
12	kokwamukelwa kwakho esibhedlela/ Ilwazi ezinikelwa yi-GEMS ngesikhathi	Ilwz		148	6	82%	0	28%		0	28%	(527201092947 2020/Sep/1
13	ahleliweko ozowahlala esibhedlela. Akunanzuzo ezizokubadelwa nakhibe:	Aki		316	17	25%	0	60%		0	60%	(527201092947 2020/Sep/1
14	zibandakanya ekuthatheni iinqunto; funyana itjhejo lezemisebenzi	funj		408	24	53%	0	77%		0	77%	(527201092947 2020/Sep/1
15	kobana uzwisise iinzuzo Ukuvunyelwa kokwamukelwa kwakho	Uku		144	6	77%	0	27%		0	27%	(527201092947 2020/Sep/1
16	atjhegwa khona ngenwa kokubelethwa. EZINYE iinomboro zokuthintana	EZI		499	29	6%	0	94%		0	94%	(527201092947 2020/Sep/1
17	kokubelethwa komntwanakho. Umnoooho kukunikela isiveleliso	Um		69	5	12%	0	13%		0	13%	(527201092947 2020/Sep/1

Figure 5.8: Verbal root *-beleth-*

The verbal root *-beleth-* appears in number 1, 4 and 5 as *lokubelethisa*; in number 7 as *ekubeletheni* and in number 16 as *kokubelethwa*.

Even though the spoken corpus of this study is literacised, the researcher had to compare a corpus from written corpora and from spoken material. A comparison follows.

5.5 COMPARISON BETWEEN THE SPOKEN (LITERACISED) AND WRITTEN CORPUS OF THIS STUDY

Although the spoken corpus of this study is literacised and added to the written one, it is important to indicate that there is a difference between corpora from written and spoken material.

In this section a discussion will be presented with regard to term-formation processes. Aaron and Joshi (2006:272) maintain that the written language is not as natural as its spoken counterpart. Written language is also more formal than the spoken language. In spoken language, speakers have a tendency of using elliptical forms. This means that there is a tendency of omitting certain

words, although the meaning gets retained. Spoken language also uses abbreviated forms. In written language, on the other hand, complete sentences are used.

In support of Aaron and Joshi (ibid.), Thanh (2015:144) says that the spoken language as compared to its written counterpart, often expresses politeness and attitude. Written language is formal and academic as compared to the spoken one. The written language uses strict and appropriate words. Euphemism is applied more in spoken than in written corpora.

5.5.1 Term-formation processes

The researcher needed to compare term-formation processes in the written and spoken corpora collected. The researcher's findings are that although transliteration has been used in the written corpus, it is a dominant term-formation strategy in the spoken corpus collected from Ikwewezi FM. Transliteration is the term-formation strategy in which the stems of foreign language terms are used to coin the equivalents of the target language. The stem of a foreign language is retained and used in the formation of a target equivalent, isiNdebele equivalent in this case.

The use of transliteration in the Ikwewezi health program is discussed below.

N2 is *istrowuku* for 'stroke', N4 is *ibhayiphola* for 'bi-polar', N9 is *iskizofreniya* for 'schizophrenia'; N2 is *idawuni sindromu* for 'down syndrome', N3 is *amafayibhrotsi* for 'fibroids', N5 is *imola pregnensi* for 'molar pregnancy', N6 is *ipapsmiye* for 'pap smear'. The latter is *ukuhlolwa komlomo wesibelesho* in Multilingual HIV/AIDS Terminology (2013a:86).

The duration of time during which the health program is aired is 30-45 minutes. This is a short time and is the reason why transliteration is excessively used. It is for instance quicker and easier to use the term *istrowuku* rather than *ubulwele bokufa ihlangothi* which is paraphrasing. It may be more time consuming to use the term *ukuhlolwa ikhenza yesibelesho* than the term *ipapsmiye*. In cases such as 'bi-polar', the term *ibhayiphola* may be preferred because of the lack of an isiNdebele term.

The other reason for resorting to transliteration is of course the shortage of isiNdebele terms. One of the research questions this study seeks to answer is **how the paucity of standardised terminology in specialised subject fields negatively impacts the development of**

isiNdebele. As mentioned in 2.3.2.2, Alberts (2017:148) maintains that the lack of standardised terminology leads to the random use of terminology. Terminology plays a pivotal role in language development. The development of a language is impacted negatively if there is little or no standardised terminology.

What Alberts (ibid.) says, is true in the case of the language under research. Due to the lack of standardised isiNdebele medical terminology, terminology is therefore used randomly. Consider the following concepts.

- ‘Constipation’ and ‘indigestion’ are the terms that are often used in isiNdebele as the same concept. These are actually two different concepts. ‘Constipation’ refers to an abnormally delayed or infrequent defecation (eliminating solid, semi-solid or liquid waste material from the digestive tract via the anus). This is caused by a reduction in intestinal tract motility (movement of food from the mouth until it is out of the body) leading to dry, potentially painful faecal evacuation. This may result from stress, insufficient fibre in the diet or lack of exercise (Merriam-Webster’s online dictionary, 2021). ‘Indigestion’ on the other hand is the condition of impaired or difficult digestion. It is marked by a burning sensation or discomfort in the upper abdomen.
- ‘Virus’, ‘bacterium’ and ‘germ’ are mostly translated as one term in isiNdebele. Some isiNdebele speakers use the term *ingogwana* when referring to a ‘germ’ and *umulwana* for a ‘virus’. Others use the term *umulwana* when referring to a ‘germ’ and *ingogwana* for a ‘virus’. Others are unaware which term denotes a ‘germ’ and which one a ‘virus’. Some regard the two terms *umulwana* and *ingogwana* as synonyms. This error is not committed by lay persons only but even translators commit this error. When one listens to Ikwekwezi FM, the two terms are used interchangeably. Until there is a standardisation, the translation of the two terms will continue this way and this has an impact on the development of this language because two different terms are regarded as one. There are three equivalents of ‘virus’, namely *umulwana*, *ingogwana* and *ivayirasi*. There is confusion as far as the isiNdebele equivalent for ‘virus’ and ‘germ’ is concerned. For the purpose of this study the term ‘virus’ will be referred to as *ingogwana* and the term ‘germ’ will be referred to as *umulwana*.

'Bacterium' *igciwana* is a single, but complex, cell. It can survive on its own, inside or outside the body (Merriam-Webster's online dictionary, 2021). 'Virus' *ingogwana*, is smaller. A virus, unlike 'bacterium', needs a host such as a human or animal for it to be able to multiply (Merriam-Webster's online dictionary).

As mentioned in 4.2.3.1, when the term 'germ' was analysed, for the purpose of this study, 'virus' is referred to as *ingogwana*, 'bacterium' is referred to as *igciwana* and a 'germ' as *umulwana*.

- The terms 'cholera' and 'diarrhoea' are translated as one term in isiNdebele equivalents. The euphemism *ukuthulula* is preferred instead of *ukurhuda* which is regarded as vulgar.

'Diarrhoea' is a condition in which faeces is discharged from the bowels frequently and in liquid form. 'Cholera' on the other hand, is an intestinal disorder characterised by abnormal frequency and fluidity of fecal evacuations. 'Cholera' is an infectious and often fatal bacterial disease of the small intestine, typically contracted from infected water supplies and causing severe 'vomiting' and 'diarrhoea'. <https://www.dictionary.com/browse/cholera#>.

Based on the two definitions, it is an error to regard the two concepts as the same. Cholera causes not only 'diarrhoea' but also vomiting. It is surprising that the emphasis has been on 'diarrhoea'. It is actually incorrect to refer to cholera as *ubulwele bokurhuda* 'the disease of diarrhoea', since diarrhoea excludes vomiting. Using the transliterated term *ikholera* for 'cholera' is the best option.

Such confusion impacts negatively on the development of a language. Health workers resort to using transliterated terms instead of the coined ones, in order to avoid this confusion. For example, instead of using *ubukubhinjidelwa*, *ukuqurhelwa* and *ukuparelwa* for 'indigestion' or 'constipation', they end up using *ikhonstipheyishini* for 'constipation' and *indayijesitjhini* for 'indigestion'. This impacts the development of isiNdebele negatively.

If the correct terminology is used, the technical and scientific communication skills of all citizens will be developed. Terminology must have characteristics such as specialisation, technicality, single acceptance and conciseness (Alberts, 2017:64 and 67).

When looking at the above terms, one would realise that the terms do not meet the criterion of single acceptance and conciseness. When one listens to an isiNdebele speaker talking about *igciwana*, one may think he/she is referring to 'bacterium' whereas the person is actually referring to a 'virus' *ingogwana*.

This confirms what Roets (2001:10) observed, that many equivalents are formed through transliteration, particularly in subjects such as economics, medicine and physics. If transliteration is excessively used in terminology, it creates a linguistic gap. Transliteration causes the technological vocabulary of a language not to develop. Transliterated equivalents do not adhere to the orthography of a language because they have not been used in written language before. Since transliterated terms do not adhere to most of the general guidelines regarding terminology, Roets (ibid.) therefore advises that transliteration be used as a last resort when one has run out of options.

There are inconsistencies in the writing of transliterated terms. The official Terminology and Orthography/*Imithetho yokutlola nokupeleda isiNdebele* (2008), does not really assist in spelling options, especially if the enquirer is not a native language speaker of isiNdebele. The purpose and function of formalised spelling rules is to enable language users to spell words correctly which do not necessarily appear in dictionaries and/or term lists. For instance, on page 13 of *Imithetho yokutlola nokupeleda isiNdebele* (2008), the following transliterated terms are discussed. An Afrikaans word like 'draad' (wire) should be written as *idrada* in isiNdebele. However, the term *idarada* or *idradi* is being used by amaNdebele. The Afrikaans term 'blaar' (leaf) is to be written as *ibhlara* according to the isiNdebele Orthography and Spelling Rules. However, most amaNdebele write and pronounce this as *ibhlari*, with an -i instead of an -a as the ending vowel.

There are many inconsistencies as far as transliterated terms are concerned. For instance, the term 'fibroids' is pronounced as *amafayibhrothi*, *amafayibrothi* and *amafayibhrotsi*. The same applies to the term 'hormones' which is pronounced as *amahomoni*, *amahomowuni* and *amahomoyini*.

Owing to the above, it is true that transliteration does not assist in the development of technical language, when it is used excessively as is the case in the spoken corpus collected from the Ikwekwezi FM health program.

5.5.2 Corpus size

Size here refers to the number of tokens, or running words, making up the corpora. The size of the specialised corpora of this study is 2,102,967 tokens. Litosseliti (2010:95) mentions that the corpus size should be just big enough to reveal something about frequencies of certain linguistic phenomena, enabling researchers to examine what is typical and what is rare in a language. There are no hard and fast rules regarding how large a corpus ought to be, instead size is dictated by several criteria. One of these criteria concerns the purpose for which the corpus is intended.

5.5.3 WordSmith Tools and corpora (written and spoken)

In both the written and the literacised spoken corpora of this study, the WordSmith Tools was used for semi-automatic term extraction. All the functions/suites (WordLister, KeyWords and Concordancer) of the WordSmith Tools were applied in both the spoken and the written corpora.

5.6 CONCLUSION

Because both the written and spoken, literacised corpora are discussed in this chapter, a comparison was made between terms just to check as to what kind of terms are preferred in the written as compared to their spoken counterpart. It was discovered that in the spoken corpus more transliterated than coined terms are used. The term 'nurse' for instance is referred to as *unese* in the spoken corpus and *umhlengikazi*, a coined term, in the written texts. The same applies to the term 'depression', in written corpora it is *ukugandeleleka ngokomkhumbulo*, whereas it is referred to as *idiphreshini* in spoken corpora.

The spoken corpus of this study was literacised through transcription and then added to the written corpus. Only two types of transcriptions were applied in this study, namely the denaturalised or smart transcription, whereby fillers were not translated and the literacised transcription, whereby the spoken corpus was converted to a written corpus. Therefore, the denaturalised-literacised transcription was implemented in this study.

The identification of term candidates was done semi-automatically, that is, manually and also through the WordSmith Tools. The researcher had to use the hybrid approach so as to establish which approach would yield better results. All the functions of the WordSmith Tools were used.

The researcher started with the WordList function and discovered first hand, after producing a frequency word list, that the WordList function is unable to distinguish between function and content words. In most cases function or grammatical words are ranked higher than the content words and they have higher frequencies too, than the content words. The KeyWords function was also implemented in order to identify term candidates.

After the KeyWords function could not identify many term candidates, manual identification took place. Indeed, more terms were identified manually than through the KeyWords function. The other challenge with both the WordList and KeyWords function is that, both these tools are unable to identify multi-words. They always identify a multi-word separately. For instance, a multi-word such as *umzombelezo wokuba sesikhathini* 'menstrual cycle' was extracted as *umzombelezo* (with its own number and frequency), *wokuba* (with its own number and frequency) and *sesikhathini* (with its own number and frequency). The tool that yielded the best results is the Concordance. This tool managed to identify multi-words. It proved that the Concordance can identify both single and multi-words.

This chapter has proven that the human factor cannot be discarded in terminological practice. For instance, the alignment of English-isiNdebele texts had to be done manually before the WordSmith Tools could be applied, to ensure that the English sentences and paragraphs correlate with the isiNdebele ones.

It is true that the ease or difficulty of using a parallel corpus depends to a large extent on availability of texts and their translation. The study is about the medical field as a specialised field. The total size of the medical corpora is 2,102,967 tokens. The corpora of this study managed to provide answers to the present study's research questions, namely the role corpus-driven lexicography can play in filling the gap caused by the shortage of terminology and the ways in which existing methods of producing terminologies can be transferred to isiNdebele where only sparse resources are available. These two questions have been answered by this chapter.

Term-formation processes, with particular focus on term-formation processes in questionnaires, were discussed in Chapter 4. It was also necessary to explore term-formation processes in both the written and literacised, spoken corpora. For instance, it is clear from this chapter that borrowing, or loan words are mostly used in the spoken corpus whereas the written corpus uses mostly coined words. As mentioned in 5.5, written language tends to be more formal than its

spoken counterpart. It was observed that a term such as 'gynecologist' is referred to as *igayini* in the spoken corpus, whilst it is referred to as *udorhodere wabafazi* 'a doctor of women' in the written corpus.

The other reason for resorting to transliteration is of course the shortage of isiNdebele terms. The paucity of standardised terminology in specialised subject fields has negatively impacted the development of isiNdebele. It was proven that the lack of standardised terminology leads to random use of terminology.

The objective of this chapter, namely that of identifying term candidates that could be used in the compilation of an English-isiNdebele glossary of medical terms, has been realised (See **Appendix G**).

The resultant glossary has answered the research question on how a bilingual glossary can be designed and represented in a user-friendly way in the case of isiNdebele, and how such a glossary can be made available. The resultant English-isiNdebele glossary was then compared with the Basic Health Terms NTS-ENG-ZUL glossary compiled by the Zulu Language Committee of the Department of Arts, Culture, Science and Technology (1997).

CHAPTER 6

FINDINGS AND INTERPRETATIONS

6.1 INTRODUCTION

The chapter starts by revisiting the aim, objectives and the research questions of this study. This is followed by an overview of chapters. Under the overview of chapters, it is established to what extent the aim, objectives and research questions set out at the beginning of the study, have been achieved. This is followed by the examination of contributions this study has made to the body of knowledge. Recommendations as well as the limitations of this study are also presented and finally, suggestions for future research are proposed.

6.2 AIM, OBJECTIVES AND RESEARCH QUESTIONS OF THE STUDY

The main aim of the current study was to investigate how specialised corpora can be used in the compilation of an English-isiNdebele glossary of medical terms. To achieve this, written medical texts were used to extract the parallel, bilingual corpora. The monolingual corpora were extracted from the spoken language (from Ikwewezi FM health program). Term candidates were therefore extracted from the written medical corpora and also from the spoken corpora which were added to the written corpora after transcription.

Firstly, term candidates were manually identified and extracted from the pamphlets which the researcher found from the Mpumalanga Department of Health. Such terms were put in questionnaires for verification by nurses in Mpumalanga health facilities, doctors, Curriculum Implementers and isiNdebele translators.

The bilingual, English-isiNdebele terminology was sourced from the medical pamphlets, grade 12 question papers, the SADIaR website, the Constitution of the Republic of South Africa (1996), etcetera. The monolingual (isiNdebele only) corpus was extracted from the recordings from the Ikwewezi FM health program and also interviews with the medical doctors. The WordSmith Tools through its functions or suites, was then applied to analyse both the monolingual and the bilingual data.

The aim and objectives of this study were addressed by answering the four research questions, which are:

Research question 1: How does the paucity of standardised terminology in specialised subject fields negatively impact the development of isiNdebele?

The lack of standardised isiNdebele medical terminology causes random use of terminology by language speakers. For example, while the terms ‘constipation’ and ‘indigestion’ are actually two different concepts, they are conceived as the same concept in isiNdebele, hence they are translated as one term. The isiNdebele terms *ukubhinjidelwa*, *ukuqurhelwa* and *ukuparelwa* are randomly used for either ‘constipation’ or ‘indigestion’. They are referred to as one medical condition.

Other terms that pose a challenge in isiNdebele for instance, are the terms ‘virus’, ‘germ’ and ‘bacteria’. Standardisation is also needed here. The isiNdebele terms *ingogwana*, *umulwana* and *igciwana* are also randomly used to refer to ‘virus’, ‘germ’ and ‘bacterium’. This random use of terminology negatively impacts the development of isiNdebele.

Research question 2: What role can corpus-driven lexicography play in filling the gap caused by the shortage of terminology in isiNdebele?

A corpus-driven approach has been used in this study. This approach has ensured that all term candidates are extracted from the collected corpora. Every conclusion that was made in this study was based on the corpus or evidence provided.

The WordSmith Tools has been used to extract and analyse corpora. The KeyWords function was used to extract key words from the study’s corpora. The Concordance was also used in extracting multi-words, which could not be extracted through the WordList and KeyWords function. Because the KeyWords function could not identify most of the term candidates, the researcher had to manually extract term candidates. Most term candidates were extracted manually.

The manual extraction of terms does not render the study less corpus driven. It has been proven through other studies reviewed that manual intervention is part of lexicography.

Research question 3: How can existing methods of producing medical terminologies be transferred to isiNdebele where only sparse resources are available?

There are existing methods, namely those that have already been implemented by other scholars, in various projects. The researcher has learnt from projects such as the ALLEX, UKZN, CLTAL, SANTED multilingual and the SPeLCAL projects. She also learnt from the studies conducted, where medical terms were compiled, for example the studies where the English-isiZulu medical terms as well as the English-Shona biomedical terminology were compiled. The researcher also learnt methods of producing medical terminologies through reviewing some projects in which glossaries were compiled.

Research question 4: How can a bilingual glossary be designed and represented in a user-friendly way in the case of isiNdebele, and how can such a glossary be made available to its intended users, inter alia, medical fraternity, translators of medical texts, lexicographers and other end-users?

The English-isiNdebele glossary of medical terms has been successfully compiled in the present study, through the corpus-driven approach. The glossary is presented in a user-friendly way. It has an English medical term and its isiNdebele equivalent(s). There is also a definition of terms in both English and isiNdebele. The glossary will be made available on SADiLaR's website and also on the website of the Mpumalanga Department of Health.

6.3 OVERVIEW OF CHAPTERS

This section presents the summary of the thesis from Chapter 1 to Chapter 5.

Chapter 1 presented amongst others, an overview of the background to the proposed study, the research problem, the rationale and reasons for undertaking the study. It was established that, although isiNdebele is amongst the South African languages that have been granted official status, the language is still experiencing a shortage of specialised terminology. In order to close this gap of a shortage of specialised terminology, the researcher focused on medical terminology. This was done by collecting both monolingual and bilingual corpora which served as useful resources for the compilation of an English-isiNdebele glossary of medical terms.

Chapter 2 presented a literature review, including the two theories used in this study which are the corpus linguistics theory and the Grounded Theory. How the two theories complement each other was also discussed. Whilst corpus linguistics was used in this study for the analyses of the study's corpora, the Grounded Theory was used to analyse questionnaires. Both theories are similar in that they both promote the objective approach of studying a language. They both avoid making assumptions but instead advocate that conclusions be made based on the evidence. The corpus-driven approach which is also objective in its approach, was employed for the purpose of realising the aim of this study. Every piece of information was drawn from the corpus, with no pre-conceived information.

Various projects on glossary and terminology compilation were reviewed. The researcher needed to establish, amongst others, the size of the corpus collected, the method used to extract terms, the purpose of term extraction and the approach (corpus-based or corpus-driven) used. Since the researcher used the CDA, she wanted to learn more from the projects that employed this approach too. The researcher discovered that it is only in the compilation of an English-Portuguese glossary of cooking terms, that CDA was implemented. A WordSmith Tools function was also used in this project. This project is very relevant to the present study because of its usage of CDA and the WordSmith Tools.

Studies conducted on the compilation of medical terms, in various African languages were also reviewed. This included studies on ChiShona, isiZulu, Xitsonga, Zimbabwean Ndebele and Shona. The success of these indigenous languages in compiling medical term lists made the researcher realise that this study is also feasible.

Term-formation processes, as employed in other African languages, were also reviewed in this chapter. Taboo is part of term-formation strategies, particularly when one deals with African languages' terminology. Taboo was therefore also discussed in this chapter.

Chapter 3 contains a description of the researcher's envisaged application of the theories and methodologies. The reason for the choice of every methodology and how it would be applied in this study was discussed in this chapter.

Both the qualitative and the quantitative research methods were applied in this study. Through the qualitative method, the researcher came to an understanding of why a respondent would

prefer a particular term as compared to another one. The personal views regarding a particular term were established.

Through the quantitative research method, the researcher managed to decide on which medical terms to prepare for verification by medical staff, Curriculum Implementers, doctors and translators. This methodology uses numbers and statistics. The implementation of the WordSmith Tools makes the study quantitative. The tool provides statistical information which includes word counts, frequencies and other forms of statistical information.

The researcher also explained the importance of applying the corpus-driven approach (CDA) in the study. This approach was relevant since it enabled the researcher to approach corpora and the WordSmith Tools with no pre-conceived ideas. Through this approach, every conclusion was based on the evidence emerging from the corpus at hand. The English-isiNdebele corpus the researcher collected was used as evidence from which terms used in compiling the English-isiNdebele glossary, were extracted. Evidence plays a significant role in this study, hence the application of the corpus-driven approach.

The open-ended questionnaire as well as the semi-structured interviews were the ideal research tools to achieve the aim of the study. The open-ended questionnaires allowed an infinite number of possible answers. The researcher collected more details regarding a term. The questionnaires assisted in making the researcher understand how respondents think as they reflected their feelings about a particular term.

Through the semi-structured interviews, the researcher got to understand particular terms better. This kind of interview enabled the participants to freely express their views on a particular term. The researcher found this interview type reliable and natural. It also provided an opportunity for learning.

The Grounded Theory has methods of comparison and difference. Through these methods, the researcher established to what extent an isiNdebele equivalent is acceptable to the everyday users in the medical domain. Through them, the researcher established the medical terms that are preferred by the medical staff.

The present study also applied the corpus linguistics theory to analyse corpora of this study. After the collection of corpora, the next step was to identify term candidates. The study employed a semi-automatic term extraction method. The concept term extraction refers to the identification of term candidates in a text containing language for specific purposes (LSP), here a medical corpus.

The WordSmith Tools was used for identifying term candidates of this study. The WordSmith Tools has WordList, KeyWords and Concordance functions. The researcher experimented with all the functions, however, the KeyWords function was the most relevant function for this study. The Concordance function proved to be important in this study because this function ensures that the frequencies (occurrences) of a particular medical term as well as its possible, immediate environment are seen at first glance. It also supports the researcher in identifying multi-word terms. Therefore, the KeyWords and Concordance function are most relevant for the realisation of the aim of this study.

Because this is the first study in isiNdebele to be making use of the WordSmith Tools, the researcher had to manually verify the terms identified by the KeyWords function in order to establish to what extent the tool successfully identified term candidates. It is proven in this study that human beings will always remain the final judges in any terminological activity, whether that endeavour is manual or computational.

Chapter 4 presented the questionnaire as quantitative instrument used for collecting data in this study. Firstly, a first batch of bilingual pamphlets were obtained from the Mpumalanga Department of Health. These are the pamphlets from which terms that were listed in the questionnaires were manually sourced. They were the English-isiNdebele terms. The purpose was to verify if these terms were used on a day to day basis by the nurses in health facilities. Curriculum Implementers, isiNdebele translators as well as doctors were also sent the questionnaires.

More new terms were discovered, through questionnaires. This is because questionnaires were prepared in such a way that the respondents could either agree or not agree with the equivalent provided. In the case where the respondent did not agree with the equivalent provided, he/she had to provide an alternative term(s) deemed appropriate.

This chapter also discussed term-formation processes applicable in isiNdebele. In Chapter 2, the term-formation processes were explained under general but in Chapter 4, they were discussed with reference to isiNdebele. It was discovered in this study that amaNdebele prefer transliterated terms to coined ones. It was also discovered that factors that contributed to the usage of transliterated terms are age, qualifications in the language, geographical location, etcetera. For instance, it was discovered that younger nurses used transliterated terms more than the older nurses. Nurses whose health facilities are situated in rural areas used coined terms more than the ones whose health facilities are situated in semi-urban areas. Older nurses used mostly coined terms, unlike younger ones who had just graduated and were still under the influence of English.

Too much usage of transliteration as a term-formation technique contributes towards the under-development of the language. When one looks at the term *umhlengikazi* which is a coined term for the English 'nurse', it is very likely that most amaNdebele children do not even know that the coined term for a 'nurse' is *umhlengikazi*, instead they know the term *unese* or *usista* 'a sister'. Consequently these two terms are the ones that will be passed on from generation to generation. The term *umhlengikazi* may finally be missing in the vocabulary of the coming generation and examples like these will cause the language not to develop in the area of terminology.

The concept of taboo was discussed under general, in Chapter 2. How taboo is handled in isiNdebele was also discussed in this chapter. Taboo was seen to influence the preference of certain terms to others in the questionnaires. For instance, the term *ukuthulula* was preferred to *ukurhuda* for 'diarrhoea'. The term *ukurhuda* is regarded as vulgar and *ukuthulula*, as euphemistic. Factors such as age, qualification in isiNdebele and the geographical location of a health facility also played a role here. For instance, older nurses and those whose health facilities were in rural areas, preferred euphemism more than the younger ones and those that are in semi-urban areas.

Chapter 5 aimed at identifying term candidates that could be used in the compilation of an English-isiNdebele glossary of medical terms. The identification of term candidates was done semi-automatically, that is, manually and also by using the WordSmith Tools. The researcher had to use the hybrid approach to establish which approach would yield better results.

The alignment of the English-isiNdebele texts had to be done manually before the WordSmith Tools could be used, to ensure that the English sentences and paragraphs correlate with the isiNdebele ones. This proved that the human factor cannot be discarded in terminological practice.

Finally, an English-isiNdebele glossary of medical terms was compiled, and then compared with the Basic Health Terms NTS ENG-ZUL glossary of medical terms. The comparison revealed that there are terms in the NTS ENG-ZUL glossary, that are also available in the English-isiNdebele glossary (**Appendix G**).

The researcher aims to follow the submission procedures so that the English-isiNdebele glossary of medical terms compiled through this study is made available on the following websites: www.mpuhealth.gov.za>Resource as well as <https://sadilar.org/index.php/en/guidelines-standards/submit-a-resource>.

6.4 CONTRIBUTION OF THE STUDY

This study has proven that the use of corpora in compiling lexicographic products such as dictionaries and glossaries is possible in indigenous languages of South Africa. So far, no research has been conducted on the use of specialised parallel corpora as a resource for bilingual lexicography in isiNdebele. This study promotes synergy between different fields, corpus linguistics, translation studies, lexicography and computer science or technology. The study has also proven that the semi-automatic approach is possible. The study can be used by other researchers, translation and lexicography students, teachers and linguists to better understand the use of technology in resource development.

The study has illustrated that the Grounded theory can also be applied to quantitative data. This means that it can be applied to both the qualitative and quantitative data. This study will contribute to the development and standardisation of medical terminology. The type of terms and translation equivalents and their frequencies help in identifying those words that are commonly used in isiNdebele. The frequent words and highly frequent spellings facilitate the process of standardisation. It was found in the study that some spelling variations were not in line with the orthography and spelling rules of isiNdebele.

The researcher will make the isiNdebele corpora available on the SADiLaR web-page. They are valuable resources and will allow further research on the data by other researchers, too. The researcher would like the Department of Health and SADiLaR to benefit from this study. SADiLaR will therefore make the research data available for fellow students and academics. Through SADiLaR, the present study's data will be structured according to international standards. The data will be freely available in the SADiLaR repository <https://www.sadilar.org/index.php/en/resources/student-data-repository>.

Transcription applied in this study is denaturalised-literacised. It is denaturalised because fillers like telephone ringings, unnecessary repetitions, stuttering, etcetera have been removed. It is literacised because the spoken corpus has been transcribed to an extent that it became exactly like the written corpus.

As part of future work, since the data is freely available, others will also have the opportunity to expand on it in order to develop the language further.

This study will contribute to exemplifying corpus-driven lexicography in African languages. It has demonstrated that the exploitation of corpora is also possible in indigenous languages, like isiNdebele. The same approach followed in compiling the medical glossary of this study can be applied when compiling glossaries in other fields. This is because, so far, no research has been conducted on the use of specialised corpora as a resource for bilingual lexicography in isiNdebele.

The use of WordSmith Tools in extracting and analysing data from the corpora of this study, has demonstrated that it is possible to use the tool in other fields of the study. The study can be used by other researchers, isiNdebele translators, students of lexicography, teachers and linguists to better understand the use of technology in terminology resource development.

This study will contribute towards the development and standardisation of some medical terms that are inconsistently and randomly used. The study has addressed the confusion regarding the overlap in some terms, e.g. 'indigestion' and 'constipation'. For 'constipation', the appropriate isiNdebele equivalent is *ukubhinjidelwa* and the appropriate equivalents for 'indigestion' are both *ukuqurhelwa* and *ukuparelwa*.

Regarding 'germ', 'virus' and 'bacteria', the study concludes that the term *ingogwana* should be used to refer to a 'virus', *umulwana* should be used to refer to a 'germ'. The term *igciwana* should be used when referring to a 'bacterium'.

Regarding 'diarrhea' and 'cholera', the study has successfully made a distinction. The term *ukurhuda* or *ukuthulula* should be used when referring to 'diarrhoea' and the term *ubulwele bokurhuda* or *bokuthulula* should be used when referring to 'cholera'.

The glossary will contribute towards the terminology which is much needed in the Department of Health's institutions.

6.5 LIMITATIONS OF THE PRESENT STUDY

The study was limited to the field of medicine and also to the two stated languages, namely English and isiNdebele.

During the data collection phase, the researcher encountered several challenges: IsiNdebele is the official language in the Mpumalanga Province. However, most fliers that are available in Mpumalanga clinics and hospitals are in Siswati, the Sotho languages and English only. Some important notices in some of the clinics are written in Siswati instead of isiNdebele. The scarcity of isiNdebele medical fliers and pamphlets has contributed towards the number of terms put in the questionnaires for verification.

The researcher deliberately chose the isiNdebele speaking health workers, translators, Curriculum Implementers and doctors to participate in the study. This means those health workers that are non-isiNdebele speakers were deliberately excluded from participating in the study either through questionnaires or interviews.

The fact that isiNdebele is an agglutinating language has been a limitation also. There is no reliable tool available yet, which can be used to conduct quantitative analysis in the same way it is conducted in non-agglutinating languages.

It is also problematic to apply the frequency of occurrence WordList to a language like isiNdebele. During the preparation of the questionnaires, terms were manually sourced. The manual sourcing of terms does not, however, mean that the study was not corpus-driven. This was proven when the KeyWords function could not extract or identify most term candidates.

This study only focused on a few medical fields for instance, mental health, primary health care, cancer and gynecology. Even these fields could not be exhausted. Owing to the scarcity of isiNdebele medical terms, the researcher ended up lumping together terms that refer to medical and bio-medical technical conditions.

6.6 RECOMMENDATIONS

There is a need for resource development in isiNdebele, especially in technical and scientific fields. The resources to be developed include amongst others, the specialised, corpus-driven glossaries, dictionaries, terminology banks, etcetera. This will meet the communicative needs of African societies. All terminology activities in most African languages should be properly managed. The development of resources will further encourage the use of isiNdebele in scientific fields.

Terminology must match the development taking place in the world. Vocabulary must also grow and for that to happen, coining of terms and not random transliteration, should be a priority that must be aimed at. This will enable isiNdebele to be used in different fields such as medical institutions, financial institutions, legal institutions, etcetera and this will ensure that there are terms for technical language. Transliteration, which is the most dominant term-formation strategy in isiNdebele does not contribute towards the growing of this language's vocabulary.

The coining of terms must be done by various stakeholders through workshops and seminars. Stakeholders should include linguists, translators, government representatives, users of the language, etcetera. This will ensure that isiNdebele develops and is modernised in all spheres of life in a formal way, rather than being subject to individual efforts.

The Pan South African Language Board (PanSALB), the Provincial Language Committees (PLCs) and the National Lexicography Units (NLUs) work with terminology developments.

The Pan South African Language Board (PanSALB) plays a key role in the development and promotion of the indigenous official languages of South Africa, including the Khoi and San languages and South African Sign Language. This board promotes and creates the positive environment for multilingualism. It offers advice to government on language policy and planning of the official languages.

The Provincial Language Committees (PLCs) on the other hand, function on the provincial level. They work closely with provinces on language matters affecting their specific provinces. They also promote multilingualism. The PLCs promote literature for the previously marginalised languages. They also promote lexicography and terminology development; research; and projects. The role of the National Lexicography Units (NLUs) is to develop dictionaries in all the official languages. They work closely with the Terminology Coordination Section of the National Language Services (NLS). They are the authority in terms of approving lexicography and language standards.

The National Language Services promotes and facilitates communication across languages in South Africa. The NLS' core objective is to meet the language requirements of the constitution by facilitating, promoting and providing a translation and editing service in all the official languages. It manages language diversity through language planning and terminology projects. The NLS functions as the Government's professional language support system by translating official documents in all the official languages. Its terminology service assists with the development and modernisation of the technical vocabularies of the official languages. The language planning functions include advising the Government on the development of language policy and implementation strategies.

Financial resources should be set aside for purposes of translation and compilation of glossaries and dictionaries for special purposes. IsiNdebele terminology should be developed so as to enable the language to also be used in public services. Departments that deal with terminology development should work together. Working in silos has contributed towards the inconsistent use and duplication of terms.

IsiNdebele orthography and spelling rules need to be revised for the appropriate, quality and standardised writing of books, glossaries, dictionaries, etcetera. The isiNdebele orthography and spelling rules must be made available to school learners and educators.

IsiNdebele, like many other African languages, needs to be intellectualised. The availability of the resources, such as the glossary the researcher has compiled, would help reach this goal. The Mpumalanga government can contribute towards the intellectualisation of this language by committing huge amounts of financial and human resources for various programmes that will contribute towards the development of this language.

Workshops will have to be conducted where the attendees will be taught about the importance of their language (isiNdebele). The purpose will be to deal with the negative attitude people have towards African languages in general.

The researcher is aware that the government is faced with the task of developing a number of languages and that means that more resources will be required, but the government should focus on less developed languages of which isiNdebele is one. Universities must play a role also in terminology development by encouraging Honours students to do research in terminography and lexicography.

The Pan South African Language Board (PanSALB) should ensure that the Constitution of the Republic of South Africa is implemented by ensuring that isiNdebele, as one of the official languages of South Africa, is promoted. The promotion of isiNdebele should be encouraged at national, provincial and local government levels.

6.7 FUTURE RESEARCH

Since the data will be freely available, other researchers will have the opportunity to expand on it in order to develop the language further.

Another future direction that might be taken from the present study is the exploitation of corpora in designing an English-isiNdebele specialised dictionary. Researchers from other African languages can also utilise the approaches employed in this study to compile dictionaries and glossaries in different specialised fields such as health, law, technology, commerce, business, etcetera. This will be beneficial to many African languages.

With more medical challenges breaking out in South Africa, one will have to develop this study's glossary in the future and include the recent terms on diseases breaking out such as listeriosis and the Covid-19 pandemic, to mention but a few.

This study paved the way forward in the direction of practical applications and future research in isiNdebele terminological development. The methodology and insights gained in this study can be made available to other technical fields where terminological development is needed such as the legal and financial terminological fields.

Although an attempt was made to document term-formation processes used in isiNdebele, the processes still need further research. Aspects that need further research as far as term-formation processes are concerned, are processes such as compounding, loan-translation and semantic shift.

Although corpora of this study may not be as large in comparison to other corpora (from other languages), these corpora are exemplary of what can be achieved. More work still needs to be done in the area of spoken corpora. This study may prompt more involvement of researchers in obtaining spoken corpora, which indicates trends in the real language, instead of concentrating on written corpora only.

Parameters of corpus linguistics and its application are far from being exhausted. The application of corpus query tools, such as the WST KeyWords function should be exploited further, especially in the African languages, isiNdebele to be precise. This is because the automatic term extraction seems a very distant goal in isiNdebele.

Regarding transcription types, a study should be undertaken in the future where a naturalised type of translation is applied.

For future research, terms will not have to be lumped together. The researcher will have to compile either the medical or bio-medical technical terms and not both. The researcher will have to choose terms according to particular themes.

The researcher aims to work on the glossary she has compiled in this study to address some further theoretical issues.

BIBLIOGRAPHY

Aaron, P.G. and Joshi, R.M. 2006. Written language is as natural as spoken language: A biolinguistics perspective. *Reading Psychology*, 27(4):263-311.

Adom, K., Hussein, E.M. and Agyem, J.A. 2018. Theoretical and conceptual framework: Mandatory ingredients of a quality research. *International Journal of Scientific Research*, 7(1):438-441.

Alberts, M. 2014. Terminology development at tertiary institutions: A South African perspective. *Lexikos*, 24:1-26.

Alberts, M. 2017. *Terminology and terminography principles and practice: A South African perspective*. Milnerton: MLA Publications.

Alberts, M. and Mollema, N. 2013. Developing legal terminology in African languages as aid to the court interpreter: A South African perspective. *Lexikos*, 23:29-58.

Allan, K. and Burridge, K. 2006. *Forbidden words: Taboo and the censoring of language*. Cambridge, UK: Cambridge University Press. DOI: [10.1017/CBO9780511617881](https://doi.org/10.1017/CBO9780511617881).

Allwood, J. and Hendrikse, A.P. 2003. Spoken language corpora for the nine official African languages of South Africa. *Southern African Linguistics and Applied Language Studies*, 21(4): 189-201.

Andersen, G. 2010. How to use corpus linguistics in sociolinguistics. In: O'Keeffe, A. and McCarthy, M. (eds), *The Routledge handbook of corpus linguistics* (1st ed.). London: Routledge. Pp 547-562.

Bailey, J. 2008. First steps in qualitative data analysis: Transcribing. *Fam. Pract.* 25:127-131. Available at: <http://fampra.oxfordjournals.org/cgi/content/full/25/2/127#BIBL>. Accessed on 10/06/2019.

Baker, P. 2009. *Contemporary corpus linguistics*. London: Continuum.

Baker, P. 2010. *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.

Bennett, G.R. 2010. *Using corpora in the language learning classroom: Corpus linguistics for teachers*. Ann Arbor, MI: University of Michigan Press.

Bosch, S.E. and Pretorius, L. 2011. Towards Zulu corpus clean-up, lexicon development and corpus annotation by means of computational morphological analysis. *South African Journal of African Languages*, 31(1):138-158.

Bowker, L. and Pearson, J. 2002. *Working with specialized language: A practical guide to using corpora*. London: Routledge.

Bucholtz, M. 2000. The politics of transcription. *Journal of Pragmatics*, 32:1439-1465.

Buthelezi, T.M. 2008. Exploring the role of conceptual blending in developing the extension of terminology in isiZulu language. *Alternation*, 15(2):181-200.

Charmaz, K. 2000. *Grounded theory* (1st ed.). Thousand Oaks, Calif.: Sage.

Charmaz, K. 2006. *Constructing grounded theory: A practical guide through qualitative analysis*. Thousand Oaks: Sage.

Cluver, A.D. 1989. *A manual of terminography*. Pretoria: HRSC.

Corbin, J. and Strauss, A. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13(1):1-14.

Creswell, J.W. 1998. *Qualitative inquiry and research design: Choosing among five traditions*. London: Sage Publications.

Davidson, C. 2009. Transcription: Imperatives for qualitative research. *International Journal of Qualitative Methods*, 8(2):35-52.

Department of Arts and Culture, 2013a. Multilingual HIV/AIDS Terminology. Republic of South Africa.

Department of Arts and Culture, 2013b. Multilingual ICT dictionary. Republic of South Africa.

Department of Arts and Culture, 2013c. Multilingual mathematics R-6. Republic of South Africa.

Department of Arts and Culture, 2013d. Multilingual natural sciences and technology terminology list. Republic of South Africa.

Department of Arts and Culture, 2013e. Multilingual parliamentary/political term list. Republic of South Africa.

Department of Arts and Culture, 2013f. Multilingual soccer terminology list. Republic of South Africa.

Department of Arts and Culture, 2017a. Multilingual financial terminology list. Republic of South Africa.

Department of Arts and Culture, 2017b. Multilingual human, social, economic and management sciences. Republic of South Africa.

Department of Arts, Culture, Science and Technology. 1997. Basic Health Terms NTS-Eng-Zulu Language Committee. Republic of South Africa.

Duranti, A. 2007. Transcripts, like shadows on a wall. *Mind, Culture and Activity*, 13(4):301-310. Accessed on 25/10/2020.

Engelbrecht, C., Shangase, N.C., Majeke, S.J., Mthembu, S.Z. and Zondi, Z.M. 2010. IsiZulu terminology development in nursing and midwifery. *Alternation*, 17(1):249-272.

Finlayson, R. and Madiba, M. 2002. The intellectualisation of the indigenous languages of South Africa: Challenges and prospects. *Current Issues in Language Planning*, 3(1):40-61.

Foo, J. 2012. Computational terminology: Exploring bilingual and monolingual term extraction. PhD thesis, Linköping University. <http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-75243>. Accessed on 15/11/2020.

Goulding, C. 1999. Grounded theory: Some reflections on paradigm, procedures and misconceptions. *Working paper series*, WP006/99, Wolverhampton: University of Wolverhampton. [verified 24 May 2006] <http://www.wlv.ac.uk>. Accessed on 20/11/2020.

Greeff, M. 2005. Information collection: Interviewing. In: De Vos, A.S., Strydom, H., Fouché, C.B. and Delport, C.S.L. (eds), *Research at the grass roots for the social sciences and human service professions* (3rd ed.). Pretoria: JL Van Schaik Publishers.

Hadebe, S. 2000. Developing terminology in African languages with special reference to Ndebele. In: Chiwome, E.M., Mnguni, Z., Furusa, M. and Little, W.A. (eds), *Indigenous knowledge and technology in African and Diasporan communities: Multi-disciplinary approaches*. Harare: South African Association for Culture and Development Studies. Pp 228-232.

Hadebe, S. 2002. The standardisation of the Ndebele language through dictionary-making. Harare: ALLEX Project. Available from: <http://ir.uz.ac.zw/jspui/handle/10646/549>. Accessed on 20/01/2016.

Hanks, P. 2012. The corpus revolution in lexicography. *International Journal of Lexicography*, 25(4):398-436.

IsiNdebele Lexicography Unit/Iziko lesiHlathululi-mezwi sesiNdebele. 2006. IsiNdebele–English/English–IsiNdebele IsiHlathululimezwi/Bilingual and Explanatory Dictionary. Cape Town: Phumelela.

IsiNdebele Lexicography Unit/Iziko lesiHlathululi-mezwi sesiNdebele. 2014. isiNdebele monolingual dictionary. South African National Lexicography Units. South Africa.

IsiNdebele Lexicography Unit/Iziko lesiHlathululi-mezwi sesiNdebele. 2015. English-IsiNdebele-Afrikaans trilingual dictionary. South African Lexicography Units. South Africa.

Jiyane, D.M. 1994. Aspects of isiNdebele grammar. Unpublished MA dissertation. Pretoria: University of Pretoria.

Kalima, J. 2007. Word formation on Internet gaming forums: A Proseminar Paper in the Department of Languages. University of Jyväskylä, Finland.

Kenny, D. 2001. *Lexis and creativity in translation: A corpus-based study*. Manchester: St Jerome.

Kerlinger, F.N. 1986. *Foundations of behavioural research*, 3rd Edition, Holt Rhinehart & Winston: New York.

Khumalo, L. 2007. An analysis of the Ndebele passive construction. Unpublished DLitt et Phil thesis. Oslo: University of Oslo.

Khumalo, L. 2015. Semi-automatic term extraction for an isiZulu linguistic terms dictionary. *Lexikos*, 25:495-506.

Khumalo, L. 2017. Intellectualization through terminology development (AFRILEX- reeks/series) *Lexikos*, 27:252-264.

Khumalo, L. 2018. African Association for Lexicography 23rd International Conference: *Abstracts and programme*. Cape Town, South Africa: University of the Western Cape.

Krishnamurthy, R. 2008. Corpus-driven lexicography. *International Journal of Lexicography*. DOI: [10.1093/ijl/ecn028](https://doi.org/10.1093/ijl/ecn028), 1-12. Accessed on 1/11/2020.

Kruger, A. 2002. Corpus-based translation research: Its development and implications for general, literary and Bible translation. *Acta Theologica*, Supplementum, 2:70-106.

Krüger, C.J.H. 2006. *Introduction to the morphology of Setswana*. München: Lincom Europa.

Kübler, S. and Zinsmeister, H. 2015. *Corpus linguistics and linguistically annotated corpora*. London: Bloomsbury.

Laviosa, S. 2002. *Corpus-based translation studies: Theory, findings and applications*. Amsterdam: Rodopi.

Litosseliti, L. 2010. *Research methods in linguistics*. New York: Continuum.

Loubere, N. 2017. Questioning transcription: The case for the systematic and reflexive interviewing and reporting (SRIR) method. *FQS*, 18(2):1-22.

Mabasa, T.A. 2005. Translation equivalents for health/medical terminology in Xitsonga. Unpublished MA dissertation. Pretoria: University of Pretoria.

Mabule, D.R. 2009. The taboos attached to the translation of biological terms from English into Northern Sotho. *South African Journal of African Languages*, 29(1):43-53.

Madiba, M. 2004. Parallel corpora as tools for developing the indigenous languages of South Africa with special reference to Venda. *Language Matters*, 35(1):133-147.

Madiba, M.R. 2000. Strategies in modernization of Venda. Unpublished DLitt et Phil thesis. Pretoria: University of South Africa.

Madonsela, S. 2017. Word-formation strategies and processes in the creation of synsets for the African wordnet. *Southern African Linguistics and Applied Language Studies*, 35(2):201-210.

Madzimbamuto, F.D. 2012. Developing anatomical terms in an African language. *South African Medical Journal*, 102(3):132-135.

Mahlangu, K.S. 2007. Adoption of loan words in isiNdebele. Unpublished M.A. Mini Dissertation. Pretoria: University of Pretoria.

Mahlangu, K.S. 2013. Challenges to the usage of a hyphen in compound nouns in Southern Ndebele. *G.S.T.F. International Journal on Education*, 1(1):107-111.

Mahlangu, K.S. 2015. The growth and development of isiNdebele orthography and spelling (1921-2010). Unpublished DLitt et Phil thesis. Pretoria: University of Pretoria.

Mashiyane, Z.J. 2002. Some socio-linguistic aspects of Southern Ndebele as spoken in South Africa. Unpublished M.A. Mini Dissertation. KwaDlangezwa: University of Zululand.

Mawonga, S., Maseko, P. and Nkomo, D. 2014. The centrality of translation in the development of African languages for use in South African Higher Education Institutions: A case study of a political science English-isiXhosa glossary in a South African university. *Alternation Special Edition*, 13:55-79.

McEney, T. and Wilson, A. 2001. *Corpus linguistics* (2nd ed.). Edinburgh: Edinburgh University Press.

McEney, T., Xiao, R. and Tono, Y. 2006. *Corpus-based language studies: An advanced resource book*. London/New York: Routledge.

McMillan, J.H. and Schumacher, S. 2006. *Research in education: Evidence-based inquiry* (6th ed.). Boston, MA: Allyn and Bacon.

Mfazwe, L.R. 2003. Strategies for the translation of taboo words into IsiXhosa. Unpublished Masters Dissertation. Bloemfontein: University of the Free State.

Miller-Keane Encyclopaedia and Dictionary of Medicine, Nursing, and Allied Health (7th ed.) © 2003 by Saunders, an imprint of Elsevier, Inc. <https://www.elsevier.com/books/miller-keane-encyclopedia-and-dictionary-of-medicine-+nursing-and-allied-health/miller-keane/978-0-7216-9791-8>.

Mnguni, A. 2004. The use of the isiNdebele language in the South African public service. Unpublished DLitt et Phil thesis. Pretoria: Tshwane University of Technology.

Moore, F.O. 2016. Qualitative vs quantitative research. <https://www.researchgate.net/publication/310101530>. Accessed on 19/10/2020.

Moors, C., Wilken, I., Calteaux, K. and Gumede, T. 2018. Human Language Technology Audit 2018: Analysing the development trends in resource availability in all South African languages. *Proceedings of SAICSIT 2018*, 26-28 September 2018, Port Elizabeth, South Africa. Available: <https://dl.acm.org/doi/10.1145/3278681.3278716>. Accessed on 09/09/2020.

Moropa, C.K. 2007. Analysing the English-Xhosa parallel corpus of technical texts with Paraconc: A case study of term formation processes. *Southern African Linguistics and Applied Language Studies*, 25(2):183-205.

Mphahlele, M.C. 2004. The transliteration principle: Is this the best procedure in African language lexicography and terminology? (AFRILEX-reeks/series) *Lexikos*, 14:339-348.

Mpofu, N. and Mangoya, E. 2005. The compilation of the Shona–English biomedical dictionary: Problems and challenges. *Lexikos*, 15:117-131.

Mtintsilana, P.N. and Morris, R. 1988. Terminology in African languages in South Africa. *South African Journal of African Languages*, 8(4):109-113.

Ndhlovu, K. 2012. An investigation of strategies used by English-Ndebele translators in Zimbabwe in translating HIV/AIDS texts: A corpus-based approach. Unpublished PhD thesis. Alice, South Africa: University of Fort Hare.

Ndhlovu, K. 2014. Term-creation strategies used by Ndebele translators in Zimbabwe in the health sector: A corpus-based approach. *Stellenbosch Papers in Linguistics Plus*, 43:327-344.

Ndhlovu, K. and Botha, R. 2017. Euphemism vs explicitness: A corpus-based analysis of translated taboo words from English to Zimbabwean Ndebele. *South African Journal of African Languages*, 37(2):235-243.

Ndimande-Hlongwa, N. 2010. Corpus planning with specific reference to the use of standard isiZulu in media. *Alternation*, 17(1):207-224.

Ngcobo, M. and Nomdebevana, N. 2010. The role of spoken language corpora in the intellectualisation of indigenous languages in South Africa. *Alternation*, 17(1):186-206.

Neuman, W.L. 1997. *Social research methods*. Boston: Allyn and Bacon.

Newman, J. 2008. Spoken corpora: Rationale and application. *Taiwan Journal of Linguistics*, 6(2):27-58. Accessed on 20/11/2020.

Nkomo, D. and Madiba, M. 2011. The compilation of multilingual concept literacy glossaries at the University of Cape Town: A lexicographical function theoretical approach. *Lexikos*, 21:144-168.

Oliver, D.G., Serovich, J.M. and Mason, T.L. 2005. Constraints and opportunities with interview transcription: Towards reflection in qualitative research. *Social Forces*, 84(2):1273-1289. Accessed on 31/10/2020.

Olohan, M. 2004. *Introducing corpora in translation studies*. London and New York: Routledge.

Otlogetswe, T.J. 2007. Corpus design for Setswana lexicography. Unpublished PhD thesis. Pretoria: University of Pretoria.

PanSALB (Pan South African Language Board) isiGungu seliZwelope sesiNdebele (IKhwezi). 2008. *Imithetho yokuTlola nokuPeleda IsiNdebele*. Pretoria: PANSALB.

PanSALB (Pan South African Language Board). 2018. *Imithetho yokuTlola nokuPeleda IsiNdebele*. Pretoria: PANSALB.

Pretorius, L. and Bosch, S. 2012. 'Semi-automated extraction of morphological grammars for Nguni with special reference to Southern Ndebele'. A paper presented at The Workshop on Language Technology for Normalization of Less-Resourced Languages (SALTMIL8/AfLa2012). Available online: <https://pdfs.semanticscholar.org/333f/aad4a44c1d0497b1b7913956cb3885c5e145.pdf>. Accessed on 10/05/2018.

Pretorius, L., Viljoen, B., Pretorius, R. and Berg, A. 2008. Towards a computational morphological analysis of Setswana compounds. *Literator*, 29(1):1-20.

Prinsloo, D.J. 2015. Corpus-based lexicography for lesser-resourced languages maximizing the limited corpus. *Lexikos*, 25:285-300.

Prinsloo, D.J. and Prinsloo, D. 2011. WordSmith Tools. 5:95-106 <http://nflrc.hawaii.edu/lcd/>
<http://hdl.handle.net/10125/4493>.

Regmi, K., Naidoo, J. and Pilkington, P. 2010. Understanding the processes of translation and transliteration in qualitative research. *International Journal of Organizational Leadership*, 9(1):16-26. <https://doi.org/10.1177/160940691000900103>. Accessed on 07/02/2021.

Republic of South Africa. 1996. *The Constitution of the Republic of South Africa*. Pretoria: Government Printers.

Roets, S. 2001. Preliminary stage of a terminology project. Unpublished article. Pretoria: Department of Arts and Culture.

Sager, J.C. 1990. *A practical course in terminology processing*. Amsterdam: John Benjamins.

Scott, M. 2010. WordSmith Tools (Version 5.0). [Computer software]. Liverpool: Lexical Analysis Software.

Shabangu, T.M. and Swanepoel, J.J. 1989. *Isihlathululimezwi: An English-South Ndebele dictionary*. Cape Town: Maskew Miller Longman.

Siniscalco, M.T. and Auriat, N. 2005. Questionnaire design. In: Ross, K.N. (ed.), *Quantitative research methods in educational planning* (8). Paris: International Institute for Educational Planning/UNESCO. Retrieved from http://www.iiep.unesco.org/fileadmin/user_upload/Cap_Dev_Training/Training_Materials/Quality/Qu_Mod8.pdf

Skhosana, P.B. 1998. Foreign interferences in sound, grammatical and lexical system of Southern Ndebele. Unpublished M.A. dissertation, Pretoria: University of Pretoria.

Smith, K. and Davies, J. 2010. Qualitative data analysis. In: Dahlberg, L. and McCaig, C. (eds), *Practical researcher and evaluation: A start-to finish guide for practitioners*. London: Sage. Pp 145-158.

Strauss, A.L. and Corbin, J.M. 1996. *Grounded theory: Grundlagen qualitativer Sozialforschung* [Grounded theory: Foundations of qualitative social research]. Weinheim: Beltz, Psychologie Verlags Union.

Tagnin, S.E.O. and Teixeira, E.D. 2012. Translator-oriented, corpus-driven technical glossaries: The case of cooking terms. *Corpora*, 7(1):51-67.

Taljard, E. 2008. Issues in scientific terminology in African languages. In: Lafon, M. and Webb, V. (eds), *IFAS Working Paper Series*. Pp 88-92.

Taljard, E. and De Schryver, G. 2002. Semi-automatic term extraction for the African languages, with special reference to Northern Sotho. *Lexikos*, 12:44-74.

Thanh, N.C. 2015. The differences between spoken and written grammar in English, in comparison with Vietnamese. *GrST Education and Learning Research Journal*, 11:138-153. Accessed on 26/12/2020.

Theron, P.M. 2015. Coding and data analysis during qualitative empirical research in practical theology. *In die Skriflig*, 49(3):1-9.

Teurbet, W. 2001. Corpus linguistics and lexicography. *International Journal of Corpus Linguistics*, 6 (special issue): 125-153.

Tognini-Bonelli, E. 2001. *Corpus linguistics at work* (1st ed.). Amsterdam: Benjamins.

Van Huyssteen, L. 2003. A practical approach to the standardization and elaboration of Zulu as a technical language. Unpublished D Litt. et Phil thesis. Pretoria: University of South Africa.

Wallmach, K. and Kruger, A. 1999. 'Putting a sock on it': A contrastive analysis of problem solving translation strategies between African and European languages. *South African Journal of African Languages*, 19(4):275-289.

Wardhaugh, R. 1992. *An introduction to sociolinguistics*. Oxford: Blackwell.

World Health Organisation. 1978. 'Declaration of Alma Ata', Report on the international conference on primary health care, World Health Organisation.

Yamanti, K.P., Brata, I.F.M. and Sedeng, I.N. 2016. The English-Indonesian translation of taboo words in *Ted Movie*, 16(1):159-166. Accessed on 4/4/2019.

MEDICAL PAMPHLETS USED FOR QUESTIONNAIRES

Anxiety Disorders English-isiNdebele; Department of Health: Republic of South Africa
Healthy Life-style booklet: English-isiNdebele; Department of Health: Republic of South Africa.

High blood pressure: English-isiNdebele; Soul City Institute Health and Development
Communication: Republic of South Africa.

HiV/Aids and Treatment English-isiNdebele; Soul City Institute Health and Development
Communication: Republic of South Africa.

Living positively with HIV & AIDS: English-isiNdebele; Soul City Institute Health and Development
Communication: Republic of South Africa.

Patients' Rights Charter English-isiNdebele; Department of Health: Republic of South Africa.

Sexually Transmitted Infections (STIs) English-isiNdebele; Soul City Khomanani: Republic of
South Africa.

MEDICAL PAMPHLETS USED AS TEXT SOURCES

Chapter 2, Section 27 & 28 of the Constitution of the Republic of SA/ Isigaba 2 Somthetho-sisekelo we-Sewula Afrika.

English/Ndebele 16 days of activism: Republic of South Africa.

English Treatment adherence flyer/ Nde treatment adherence flyer: Association for Saving & Investment SA.

English mnt booklet / nde mnt booklet.

HIV testing information sheet/ Itjhidi lelwazi lokuhlololwa i-HIV: UNILIFE.

Palliative Care Training manual-Ibandulo labelaphi bendabuko: UNISA.

TB & Diabetes/ I-TB NeTjhukela. Department of Health, Republic of SA, sponsored by URC and SANOFI.

MEDICAL TEXTS FROM ISINDEBELE GRADE 12 QUESTION PAPERS

Abantu abaphila nokukhubazeka eendaweni zemisebenzi, 2015. Department of Basic Education, Grade 12 National Senior Certificate. IsiNdebele P1, HL.

Abonobangela nobungozi bokuba nomzimba omkhulu, 2014. Department of Basic Education, Grade 12 Senior Certificate isiNdebele Exemplar P1, HL.

Amalwele ayingozi ebantwini abatjha, 2010. Department of Basic Education, Grade 12 National Senior Certificate. IsiNdebele P1, HG.

Ikholera, 2010. Department of Basic Education, Grade 12 National Senior Certificate. IsiNdebele P1, HG.

Ukuhlongakala kwamasana, 2014. Department of Basic Education, Grade 12 Senior Certificate. IsiNdebele P1, HG.

Ukunikela ngeengazi, 2016. Department of Basic Education, Grade 12 National Senior Certificate. IsiNdebele P1, FAL.

Vuk'uzenzele June 2020. GCIS. English-isiNdebele. Government Communication & Information System (GCIS).

Websites

Dictionary.com. 2021. <https://www.dictionary.com/>. Accessed on 01/02/2021.

Merriam-Webster. 2021. <https://www.merriam-webster.com/>. Accessed on 01/02/2021.

SADiLaR. 2021. <https://repo.sadilar.org/handle/20.500.12185/272/>. Accessed on 01/02/2021.

APPENDICES

APPENDIX A: ETHICS CERTIFICATE FROM THE DEPARTMENT OF AFRICAN LANGUAGES, UNISA



COLLEGE OF HUMAN SCIENCES RESEARCH ETHICS REVIEW COMMITTEE

14 September 2015

Dear Mrs N J Malele

NHREC Registration # :
Rec-240816-052
CREC Reference # :
2015 -CHS - 90202392

Decision:
Ethics Approval from 14 September 2015 to 14 September 2017

Researcher(s): Mrs N J Malele

E-mail: 4558316@mylife.unisa.ac.za

Supervisor(s): Prof SE Bosch

(012 429 8253)

: Dr G Faass

E-mail: faassg@hildesheim.de

Title: *The Use of Bilingual Corpora in the Compilation of a Specialised isiNdebele Glossary of Medical Terms*

Degree Purpose: D Litt et Phil

Thank you for the application for research ethics clearance by the Unisa College of Human Science Ethics Committee. Ethics approval is granted for three years.

The **Low risk application** was reviewed by College of Human Sciences Research Ethics Committee, on **14 September 2015** in compliance with the Unisa Policy on Research Ethics and the Standard Operating Procedure on Research Ethics Risk Assessment.

The proposed research may now commence with the provisions that:

1. The researcher(s) will ensure that the research project adheres to the values and principles expressed in the UNISA Policy on Research Ethics.
2. Any adverse circumstance arising in the undertaking of the research project that is relevant to the ethicality of the study should be communicated in writing to the College Ethics Review Committee.
3. The researcher(s) will conduct the study according to the methods and procedures set out in the approved application.

4. Any changes that can affect the study-related risks for the research participants, particularly in terms of assurances made with regards to the protection of participants' privacy and the confidentiality of the data, should be reported to the Committee in writing, accompanied by a progress report.
5. The researcher will ensure that the research project adheres to any applicable national legislation, professional codes of conduct, institutional guidelines and scientific standards relevant to the specific field of study. Adherence to the following South African legislation is important, if applicable: Protection of Personal Information Act, no 4 of 2013; Children's act no 38 of 2005 and the National Health Act, no 61 of 2003.
6. Only de-identified research data may be used for secondary research purposes in future on condition that the research objectives are similar to those of the original research. Secondary use of identifiable human research data require additional ethics clearance.
7. No fieldwork activities may continue after the expiry date (**14 September 2017**). Submission of a completed research ethics progress report will constitute an application for renewal of Ethics Research Committee approval.

Note:

*The reference number **2020-CHS- 90202392** should be clearly indicated on all forms of communication with the intended research participants, as well as with the Committee.*

Yours sincerely,

Signature : PP 

Prof C.D. Ntuli
CHS Ethics Chairperson
Email: ntulicd@unisa.ac.za
Tel: (012) 429 4780

Signature : PP 

Prof R.M.H. Moeketsi
Executive Dean : CHS
E-mail: moekrmh@unisa.ac.za
Tel: (012) 429 2298



University of South Africa
Preller Street, Muckleneuk Ridge, City of Tshwane
PO Box 392 UNISA 0003 South Africa
Telephone: +27 12 429 3111 Facsimile: +27 12 429 4150
www.unisa.ac.za

APPENDIX B: LETTER TO MPUMALANGA REQUESTING ACCESS TO HEALTH FACILITIES

ACCESS LETTER REQUESTING PERMISSION TO CONDUCT RESEARCH IN NKANGALA DISTRICT'S HOSPITALS AND CLINICS



P.O. Box 392, UNISA, 0003, South Africa

Date: _____

District Manager

Dear Ms/Mr

REQUEST FOR PERMISSION TO CONDUCT RESEARCH

My name is Nomsebenzi Malele, a PhD candidate at the University of South Africa (UNISA). I am registered with the Department of African Languages where I am also an isiNdebele lecturer. My supervisor is Prof Bosch SE.

The proposed topic of my research is: **THE USE OF CORPORA IN THE COMPILATION OF A SPECIALISED ENGLISH-ISINDEBELE GLOSSARY OF MEDICAL TERMS**

The aim of this study is: to show how parallel corpora can be used to fill the gap caused by the lack of an isiNdebele dictionary for special purposes by developing an English-isiNdebele glossary of medical terms.

I am hereby seeking your consent to conduct interviews and also to send questionnaires to your clinics and hospitals. To assist you in reaching a decision, I have attached to this letter:

- (a) A copy of an ethical clearance certificate issued by the University.
- (b) A copy of the research instruments which I intend using in my research. Should you require any further information, please do not hesitate to contact me or my supervisor.

Our contact details are as follows:

Nomsebenzi Malele: 082 837 8734/012 429 2075/malelnj@unisa.ac.za

Prof Bosch SE (Supervisor): 012 429 8253/boschse@unisa.ac.za

Upon completion of the study, I undertake to provide you with a bound copy of the thesis.

Your permission to conduct this study will be greatly appreciated.

Yours sincerely,

Nomsebenzi Malele

APPENDIX C: APPLICATION FOR RESEARCH & ETHICS APPROVAL



health
MPUMALANGA PROVINCE
REPUBLIC OF SOUTH AFRICA

No.3, Government Boulevard, Riverside Park, Ext. 2, Mbombela, 1200, Mpumalanga Province
Private Bag X11285, Mbombela, 1200, Mpumalanga Province
Tel I: +27 (13) 766 3429, Fax: +27 (13) 766 3458

Litiko Letemphilo

Departement van Gesondheid

UmNyango WezeMaphilo

Enquiries: Themba Mulungo (013) 766 3511

Ms. Nomsebenzi Malele
P.O Box 392
UNISA
0003
South Africa

Dear Ms. Nomsebenzi Malele

APPLICATION FOR RESEARCH & ETHICS APPROVAL: THE USE OF BILINGUAL CORPORA IN THE COMPILATION OF A SPECIALISED ISINDEBELE GLOSSARY OF MEDICAL TERMS

The provincial health research committee has approved your research proposal in the latest format you sent.

- Approval Ref Number: MP_2017RP23_628
- Approval period: 03/07/2017 – 29/12/2017
- Facilities: Nkangala District
-

Kindly ensure that the study is conducted with minimal disruption and impact on our staff, and also ensure that you provide us with the soft or hard copy of the report once your research project has been completed.

Kind regards

MR J SIGUDLA
MPUMALANGA PHRC



03/07/2017
DATE



APPENDIX D: LETTER OF INFORMED CONSENT FOR PARTICIPANTS IN A SEMI-STRUCTURED INTERVIEW



P.O. Box 392, UNISA, 0003, South Africa

Date: _____

Dear Participant

RE: REQUEST TO PARTICIPATE IN A RESEARCH PROJECT

My name is Nomsebenzi Malele, a PhD candidate at the University of South Africa (UNISA). I am registered with the Department of African Languages where I am also an isiNdebele lecturer. Please read this consent document carefully before you decide to participate in this study.

You are requested to participate in a research study on **THE USE OF CORPORA IN THE COMPILATION OF A SPECIALISED ISINDEBELE GLOSSARY OF MEDICAL TERMS**

Purpose of the research study: It is to show how parallel corpora can be used to fill the gap caused by the lack of an isiNdebele dictionary for special purposes by developing an English-isiNdebele bilingual glossary. The glossary of medical terms will benefit health workers, translators, lexicographers, interpreters, educators, and learners, etc. The glossary will play a significant role in eliminating inconsistencies, particularly in respect of compounds. This will satisfy the needs of isiNdebele translators (primary user group), lexicographers (secondary user group) and interested laypersons (tertiary user group).

Duration: The interview session will take +/- 30 minutes and it will be recorded.

Risks: I will be speaking to you as a bearer of knowledge. My duty is to quote you as a primary source and not to pretend that your knowledge is mine.

Benefits: The benefit from this study is that your experiences and perceptions will be documented and captured in this research. As a result, you will be recognised as a source of knowledge. When this material becomes available, it may be read, quoted, or cited from and disseminated for education and scholarly purposes. You are welcome to request the completed study.

Confidentiality: I will protect the information and provide it only per request. Your names will not be revealed; I will use code names in the research.

Voluntary participation: Your participation in this study is completely voluntary.

Right to withdraw from the study: The participant is allowed to withdraw from this exercise if she/he feels uncomfortable at any time with no consequence. For example, you may terminate the interview at any point or later contact me to withdraw your participation. Your wish will be granted.

Agreement:

I _____ have read, understood and considered the above information which explains your intent, mission, and request for my participation in your research. I voluntarily agree to participate and I will retain a copy of this letter of consent. I show my willingness to participate by signing in the space provided below.

Participant: _____

Date: _____

Researcher: _____

Date: _____

APPENDIX E: QUESTIONNAIRES

Nomsebenzi Malele PhD Questionnaire 1

2017

The example of the questionnaire that was sent to Doctors who were not available for interviews

The given list of basic isiNdebele medical terms is an excerpt from isiNdebele Terminology and Orthography no.1 as well as Multilingual HIV/AIDS terminology.

ISIGABA A

IRHELO LEMIBUZO NGAMATHEMU APHATHELENE NEZAMAPHILO

Yeleda bona uhlanganyela kilelirhubhululo ngokuthanda kwakho begodu nangesikhathi sakho. Ibizo lakho alikazokufakwa kilelirhelo lemibuzo ukuqinisekisa bonyana imininingwana etholakele kilelirhubhululo ibulungeka iyifihlo. Isizo lakho njengomuntu okhuluma isiNdebele begodu nosebenza ngezamaphilo lizokuthatjelwa khulu ngombana imininingwana ozosinikela yona izokusiza ekuthuthukiseni amathemu wesiNdebele aphaathelene nezamaphilo. Nawuthanda ukuhlanganyela kilelirhubhululo, landela imilayo elandelako.

Please note that you participate in this research at your own will and at your own time. To protect your identity, you are requested to provide a pseudonym. Your participation as a mother-tongue speaker of isiNdebele and also as someone who works in Health will be highly appreciated. The information you are going to give us, will help in developing isiNdebele health terminology. If you want to participate in this research, follow the following instructions.

ISIGABA B

Kilesisigaba kunethemu le-English nethemu lesiNdebele. Faka itshwayo [√] nawuvumako, begodu ufake itshwayo [×] nawungavumiko. Lokha nawungavumelani nethemu lesiNdebele, elinikelweko, tlola elifaneleko.

In this section, there is an English term and its isiNdebele equivalent. Put a [√] where you agree and put an [×] where you disagree. Where you do not agree with the isiNdebele term provided, write the appropriate one.

English term	isiNdebele equivalent	Faka itshwayo [√] nawuvumako, begodu ufake itshwayo [×] nawungavumiko. Put a [√] where you agree and put an [×] where you disagree.	Lokha nawungavumelani nethemu lesiNdebele, elinikelweko, tlola elifaneleko. Where you do not agree with the isiNdebele term provided, write the appropriate one.
abdominal pain	ubuhlungu bamathumbu		
anaemia	i-anemiya		
accouchement	ukubeletha		
adverse reaction	ukwaliwa sihlahla		
antenatal care	ipheko lokuba sebantwini		
antibody	isivikelasifo		
antiseptic (n)	isivimbela kubhibhidlha		
aorta	i-ayotha		
arterial blood	iingazi ezithunyelwako		
asthma	isifuba somoya		
backstreet abortion	ukurhululwa kombungu okungasisemthethweni		

English term	isiNdebele equivalent	Faka itshwayo [√] nawuvumako, begodu ufake itshwayo [×] nawungavumiko. Put a [√] where you agree and put an [×] where you disagree.	Lokha nawungavumelani nethemu lesiNdebele, elinikelweko, tlola elifaneleko. Where you do not agree with the isiNdebele term provided, write the appropriate one.
bacterium	umulwana		
blood transfusion	ukufakelwa iingazi		
boil (tumour)	ithumba		
burn out	ukuphelelwa mamandla		
burning sensation	ubuhlungu obutjhisako		
cancer	Ikankere		
cardiac nerve	umuzwa wehliziyo		
clinical trial	ukulingelela ngomuntu		
community-based care	itjhejongulani lomphakathi		
contraception	ukukhandelimbeleko		
cracked nipple	ikono egazukileko		
diarrhoea	irhudo		
digestive system	irherho lokugaya		
discordant couple	iinthandani omunye anomulwana		
discordant results	imiphumela engafaniko		
dosage	umthamosilinganiso		
eczema	i-ekzema		
expressed breast milk	umkghatho okghanyiweko		
extrapulmonary tuberculosis	i-TB yezinye izitho		

English term	isiNdebele equivalent	Faka itshwayo [√] nawuvumako, begodu ufake itshwayo [×] nawungavumiko. Put a [√] where you agree and put an [×] where you disagree.	Lokha nawungavumelani nethemu lesiNdebele, elinikelweko, tlola elifaneleko. Where you do not agree with the isiNwebele term provided, Write the appropriate one.
epilepsy	isithunthwana		
foetus	umbungu		
genital sore	isilonda sezitho zobulili		
genital wart	isupa yezitho zobulili		
gonorrhoea	igonoriya		
grief	isililo		
groin	imbilapho		
health worker	isisebenzi sezamaphilo		
health-care provider	umtjheji wezamaphilo		
health facility	umtholapilo		
hepatitis virus	ingogwana yesifo sesibindi		
herpes	amatjhatjhazi		
herpes zoster	ibhande		
HIV infection	ukungenwa mbulalasilangu		
HIV negative	nganambulalasilangu ngana HIV		
HIV-negative result	imiphumela enganambulalasilangu		
home-based care	itjhejo lekhaya		
hypodermic needle	inalidi yokuhlaba		

English term	isiNdebele equivalent	Faka itshwayo [√] nawuvumako, begodu ufake itshwayo [×] nawungavumiko. Put a [√] where you agree and put an [×] where you disagree.	Lokha nawungavumelani nethemu lesiNdebele, elinikelweko, tlola elifaneleko. Where you do not agree with the isiNdebele term provided, write the appropriate one.
immune-boosting therapy	isizo elinikelelwa ukuvuselela umzimba		
immune system	irherho levikelomzimba		
immune deficiency	itlhayelo levikelomzimba		
immuno compromised patient	isiguli esitlhayela ivikelomzimba		
immunosuppression	igandeleleko levikelomzimba		
incubation period	ikhathi sokufukama kobulwele		
inflammation	umqubukofutho		
intravenous drug use	ukuzihlaba ngeendaki		
jaundice	ijondisi		
mastitis	ukutjhiswa mkghatho		
meningitis	imenenjayithisi		
mother-to-child transmission	ukudlulisela kukamma ubulwele emntwaneni		
multi-drug resistance	isifo esingalaphekiko		
multiple-drug-resistant tuberculosis	isifuba esingalaphekiko		
neonate	isana elibovana		

English term	isiNdebele equivalent	Faka itshwayo [√] nawuvumako, begodu ufake itshwayo [×] nawungavumiko. Put a [√] where you agree and put an [×] where you disagree.	Lokha nawungavumelani nethemu lesiNdebele, elinikelweko, tloa elifaneleko. Where you do not agree with the isiNdebele term provided, write the appropriate one.
neurological syndrome	amatshwayo wokungasebenzi kwemizwa		
nevirapine	inevaraphini		
occupational exposure	ukuba sebujameni obungakavikeleki emalelweni athathelwanako ngokomsebenzi		
opportunistic infection	ubulwele obubhudungelanako		
pain	ubuhlungu		
pandemic	ukuthuwelela kobulwele		
pap smear	ukuhlolwa komlomo wesibeletso		
post-exposure prophylaxis	ukukhandela ukutheleleka ngombulalasihlangu		
post-natal	ngemva kokubeletha		
prognosis	ihlolombono yobulwele		
pus	ubovu		
red blood-corpucle	ikhopheseli ebovu yeengazi		
respiratory disease	ubulwele bezitho zokuphefumula		
right of refusal	ilungelo lokwala		

English term	isiNdebele equivalent	Faka itshwayo [√] nawuvumako, begodu ufake itshwayo [×] nawungavumiko. Put a [√] where you agree and put an [×] where you disagree.	Lokha nawungavumelani nethemu lesiNdebele, elinikelweko, tlola elifaneleko. Where you do not agree with the isiNdebele term provided, write the appropriate one.
ringworm	ubuthi bembuzana		
seronegative	tlhogeke lamajoniwomzimba itlhogeke leemvikelimalwele		
seropositive	ubukhona bamajoniwomzimba ubukhona beemvikelimalwele		
terminally ill	ukugulela ukufa		
therapy	ukwelapha		
upper respiratory tract infection	ithelelwano lezitho zokuphefumula zangehla		
vaccine	umu-endo		
viral load	ubungako be-HIV		
virus	ivayirasi		
virulence	ubungako bengogwana		
window period	isikhathi sokufukama		
yellow fever	ifiva etjheli		

ISIGABA C

Kilesi sigaba kunamathemu wesiNdebele angaphezu kwelilodwa. Tlola itshwayo [√] lokha nawuvumelana nethemu elinikelweko, bese ufaka itshwayo [×] la ungavumelani khona nethemu elinikelweko. Ekholomini lokugcina, tlola ithemu olaziko.

In this section there is more than one isiNdebele equivalent. Put a [√] where you agree and put an [×] where you disagree. Write another isiNdebele term that you know, in the last column.

English term	isiNdebele equivalent 1	√ or ×	isiNdebele equivalent 2	√ or ×	Tlola elinye ithemu lesiNdebele olaziko. Write another isiNdebele term you know.
AIDS <Acronym>	intumbantonga		i-AIDS		
antenatal	sidisi		sebantwini		
antiretroviral drug	isigogobalisi sentumbantonga		isirhobhisi sentombantonga		
bacterial infection	ithelelwano yomulwana		ithelelwano yebhaktheriya		
CD4 count	isilinganiso se-CD4		imbalo ye-CD4		
cervix	umlomo wesibeletso		iseviksi		
cholera	lkhohla		ubulwele bokurhuda		
chronic disease	ubulwele oburhwahlaphazako		ubulwele obungalaphekiko		
communicable disease	ubulwele obuthelelanako		ubulwele obuthelelanako		
constipation	ukubhinjidelwa		ukuqurhelwa		
depression	ukudamba		ukuba phasi emoyeni		

English term	isiNdebele equivalent 1	√ or ×	isiNdebele equivalent 2	√ or ×	Tlola elinye ithubu lesiNdebele olaziko. Write another isiNdebele term you know.
dermatitis	ukuhlohlonya kwesikhumba		ubuzikazikani besikhumba		
diagnosis	miphumela yokupopola		imiphumela yokuphengula		
disinfectant	isibulalambewu sobulwele		isibulalambewu somulwana		
drug resistance	ukubhalelimithi		ukubhaleliinhlaha		
ear wax	iinkotokoto		iingonogono		
epidemic	umrhayilani		unorhatjhekile		
gall	inyongo		irhala		
genital herpes	inlonda zezitho zobulili		amatjhatjhazi wezitho zobulili		
germ	umulwana		imbewana yobulwele		
germ-cell	iseli lomulwana		iseli lembewana		
germicide	isibulaligciwani		isibulalimvunya		
gonorrhoea	igonoriya		idrophu		
haemorrhage	ukopha		ihomoreji		
HIV-positive result	umphumela otjengisa ukuba ne- HIV		umphumela otjengisa ukuba nombulalasihlangu		
HIV status	ubujamo bombulalasihlangu		ubujamo be-HIV		

English term	isiNdebele equivalent 1	√ or ×	isiNdebele equivalent 2	√ or ×	Tlola elinye ithemu lesiNdebele olaziko. Write another isiNdebele term you know.
HIV test	ihlahlubo yombulalasihlangu		ihlahlubo ye-HIV		
hepatitis	sifo sesibindi		ihephathayithisi		
hospice	elindelakufa		ihospisi		
hypodermic syringe	isirinji		isipeyidi		
Incubator	umtjhini wokuphandlusela		isifukamisi		
incurable disease	sindabelaphi		ubulwele obungala- phekiko		
infection	ukutshwayeleka		ukutheleleka		
kaposi's sarcoma	iKaposi-sakhoma		ikankeresikhumba kaKaposi		
lump	ngongoma		igugumba		
medicine	isihlahla		ipengu		
mental hygiene	ihlanzeko yengqondo		ihlanzeko yomkhumbulo		
mental dysfunction	ubulwele bengqondo		ubulwele bomkhumbulo		
menstruation	ukuba senya ngeni		umzombelezo wokuba sesikhathini		
miscarriage	ukubuya endleleni		ukonakalelwa		
malnutrition	ukungadli ngefanelo		ukungondleki		
mental health	ipilo yengqondo		ipilo yomkhumbulo		

English term	isiNdebele equivalent 1	√ or ×	isiNdebele equivalent 2	√ or ×	Tlola elinye ithemu lesiNdebele olaziko. Write another isiNdebele term you know.
oesophageal thrush	amalovula womrholo		amalovula womphimbo wokudla		
palliative care	ukugogobaja		ukuphungula		
PCR test	ukuhlolwa kombulalasihlangu		ukuhlolwa kwe-PCR		
pregnancy	ukuba sebantwini <euphemism>		ukuba sidisi <euphemism>		
pulse	ibetho		lphalsi		
syndrome	isindromu		ihlanganisela yamathshwayozifo		
ulcer	i-alsa		isilonda sangendeni		
uterus	isibeletho		isizalo		
virus	ivayirasi		umulwana othathelanako		
x-ray	i-eksireyi		ikhanyisizitho		

ISIGABA D: Lesi sigaba simayelana neminingwana yomuntu ozibandakanya kilelirhubhululo. (Ubulili, iminyaka njalonjalo). Indabuko yomuntu ohlanganyela kilelirhubhululo nokobana usebenze kimiphi imikhakha.

This section consists of Demographic issues (gender, age etc.). Questions are asked in order to establish the region of origin as well as the section of the hospital or clinic where they worked.

- c) Uhlala kuphi?
- d) Uthome nini ukusebenza kilesisibhedlela namkha itlinigi?
- e) Usebenza kimuphi umkhakha?
- f) Bala imikhakha okhe wasebenza kiyo ngaphambilini.
- g) Tlola ngamaboni wakho kileyo mikhakha.
- h) Uneminyaka emingaki?

SITHOKOZA KHULU ukobana uhlanganyele kilelirhubhululo. Nangu ongamthinta, ngaphandle kwami lokha nawunemibuzo ngerhubhululweli:

Prof. M.L. Mojapelo

Tel: +27 12 429 8038

Department of African Languages

University of South Africa

Preller Street, Muckleneuk Ridge, Pretoria

PO BOX 392, UNISA, 0003, South Africa

www.unisa.ac.za

APPENDIX F: IRHELO LAMAGAMA APHATHELENE NEZAMAPHILO: A LIST OF MEDICAL TERMS FROM THE WRITTEN AND THE SPOKEN CORPUS

A

abantu abaphila nokukhubazeka - people who live with disability
amalwele athathelanako - communicable diseases
amalwele angathathelaniko - non-communicable diseases
amalwele athathelana ngokomseme - sexually transmitted diseases
amanceba neembazi - wounds and scars
amathambo aphukileko - broken bones
amatshwayo - signs
amagciwana - germs
amakhondomu - condoms
amasotja womzima - antibodies
amahomowuni - hormones
amafayibhrothi - fibroids

I

ibhanditjhi - bandage
ipulasta - plaster
isilonda - wound
iHIV - HIV
iTB - TB
iTB neTjhukela - TB and Diabetes
i-AIDS - AIDS
iSANBS - SANBS
i-OCD - OCD
iHPV - HPV
iTOP - TOP
iFAST - FAST
iflu - flu
imalariya - malaria
iphenisilini - penicilin
imbilapho - lymph node
iinsumpa - warts
istresi - stress
imigreyini - migraine
idiphreshini - depression

ikhanselinghi - counselling
ihayiphehayidrosis - hyperhydrosis
iindakamizwa - drugs
igonoriya - gonorrhoea
intumbantonga - AIDS
ihlahlubo etjengisa ukutshayeleka - test that indicates infection
imitjhoga egogobalisa intumbantonga - antiretroviral drugs
isifundo esiphathelene nomseme - sex education
ilungelo lezamaphilo - right to health
i-aleji - allergy
iintjhubaba - skin pigmentation
imelanini - melanin
imelanon - imelanono/ikhenza yesikhumba
iinkhandelimbeleko - contraceptives
ikhenza yamathumbu - cancer of the stomach
ikhenza yomphimbo - cancer of the throat
ikhenza yomlomo wesibeletho - cancer of the cervix
ikhenza yesibeletho - cancer of the uterus
ikhenza yeboda lesibeletho - cancer of the endometrium
ihloko - headache
ikhemisi - chemist
iskizofreniya - schizophrenia
ivayithaligo - vitiligo
imbeleko - accouchement
ingogwana - virus
ikholikhi - colic
ipapsmiye - pap smear
ibhayopsi - biopsy
idawuni sindromu - down syndrome
ifobiya - phobia
ifiva - fever
igayini - gynecologist
inyumoniya – pneumonia

U

ukudambisa iinhlungu - pain

udorhodere wamazinyo - dentist

udorhodere womkhumbulo - psychologist

udorhodere wesikhumba - dermatologist

udorhodere wabafazi - gynecologist

udorhodere wesifo sehliziyo - cardiologist

ukubeletha ngokusikwa - c-section

ukwethuka - anxiety

ukuba semalangeneni - menstruation

ukumilisa - teething

ukuthulula - diarrhoea

umgomani - flu

umzombezezo wokuba sesikhathini - menstrual cycle

ubulwele betjhukela – diabetes

APPENDIX G: ENGLISH-ISINDEBELE GLOSSARY OF MEDICAL TERMS

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	accouchement	ukubeletha	ukubhebhula	ukuzala	accouchement: the action of giving birth to a baby	ukubeletha/ ukubhebhula: sisenzo sokubeletha isana	
	AIDS	intumbantonga	umbulala-sihlannomsegu	i-AIDS	AIDS: a disease in which there is a severe loss of the body's cellular immunity, greatly lowering the resistance to infection and malignancy	intumbantonga: bulwele obusahlela amasotja womzimba, ekugcineni lawo masotja womzimban abhalelwa kuvikela umzimba emalweleni	ingculazi
	AIDS virus	ingogwana ebanga intumbantonga	ingogwana ebanga umbulalasihlangu	ingogwana ebanga i-AIDS	human immune deficiency virus (HIV): the virus that causes AIDS, which is the most advanced stage of hiv infection	ingogwana ebanga intumbantonga: yingogwana ebanga intumbantonga sigaba esiphezulu sokutshayeleka ngentumbantonga	
	albinism	ubuswefe			albinism : genetic disorders that cause a person to produce little or no melanin	ubuswefe: kuphazamiseka kwefuzo okubangela ukobana umuntu akhiqize imelanini encani namkha angayikhiqizi nakancani imelanini	
	amnesia	i-amnesiya	ukuhlahlekelwa mkhumbulo		amnesia: a partial or total loss of memory	i-amnesiya: kulahlekelwa mkhumbulo ngokuyingcenyene namkha ngokupheleleko	i-amnesiya
	anemia	ukuthayelelwa ziingazi	i-anemiya		anaemia: a condition in which there is a deficiency of red cells or of haemoglobin in the blood, resulting in pallor and weariness	i-anemiya: bujamo bokuthayelelwa ziingazi, ubujamobu bubanga ukudinwa okukhulu	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	antibiotics	iimbulalamagciwana	ama-anthibhayothiksi		antibiotics: a medicine (such as penicillin) that inhibits the growth of or destroys micro-organisms	iimbulalamagciwana: mitjhoga (efana nephensilini) evimbela ukukhula kwamagciwana	
	antibody	isivikela sifo	isivikela malwele	amasotja womzimba	antibody: a blood protein produced in response to and counteracting a specific antigen	isivikela sifo: yiphrotheyini yeengazi ekhiqizelwa ukujamelana namkha ukulwisana ne-athijeni ethileko	
	antiretroviral drugs	iingogobalisi zentumbantonga			anti-retroviral: denoting or relating to a class of drugs which inhibit the activity of retroviruses such as hiv	iingogobalisi zentumbantonga: zijamele umkhakha wemitjhoga evimbela iingogwana ezifana nale yentumbantonga	
	antiseptic	isivimbelakubhibidlha			antiseptic: preventing the growth of disease-causing micro-organisms	isivimbelakubhibidlha: sivimbela ukukhula kwemulwana ebanga amalwele	
	anxiety	ukwethuka			anxiety: a feeling of worry, nervousness, or unease about something with an uncertain outcome	ukwethuka: umuzwa wokukhathazeka, ukungatjhaphuluki ngento ethileko onganasiqiniseko ngemiphumela yayo	
	asthma	i-asma	isifosesifuba		asthma: a condition in which your airways narrow and swell and may produce extra mucus, this can make breathing difficult and trigger coughing, a whistling sound (wheezing) when you breathe out and shortness of breath	i-asma: bujamo la iindledlana zommoya zincipha khona bezivuvuke. Zingakhiqiza amathimila. Lokhu kungenza ukuphefumula kube budisi bekukubangele ukukhohlela netjhada elitswninizako lokha nawuphefumulako. Lokhu kubanga nokuthayelelwa mummoya.	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	B						
	bacterium	umulwana	ibhaktheriya		bacteria: microscopic, single-celled organisms that thrive in diverse environments	umulwana: magciwana aneseli linye aphila ebujameni obuhlukeneko begodu abonakala ngemayikhroskopu kwaphela	Ibhaktheriya
	biopsy	ibhayopsi			biopsy: a medical test involving the extraction of sample cells or tissues for examination to determine the presence or extent of a disease	ibhayopsi: kuhlolwa kwezokwelapha okufaka hlangana ukudoswa kwesampula yamasele ukuze ayokuhlolwa ngehloso yokuqala ukobana ubulwele bukhona na nokuthi bukhambe kangangani	
	bipolar	ibhayiphola			bi-polar: a disorder associated with episodes of mood swings ranging from depressive lows to manic highs	ibhayiphola: bujamo bomkhumbulo obukhambelana nokutjhuguluka kwemizwa okungasuka la umuntu aphasi khulu emmoyeni ukuya la aphezulu khulu khona	
	bladder infection	ukutshayeleka kwesinye			bladder infection: type of urinary tract infection (uti). It is an infection anywhere in the urinary tract, such as the bladder, kidneys, ureters, or urethra. Most cases of bladder infections are acute, meaning they occur suddenly.	ukutshayeleka kwesinye: kutshayeleka kwendlela yomthundo. Ukutshayeleka kufaka hlangana isinye, iziso namkha isibeletho. Kanengi ukutshayeleka kwesinye kwenzeka msinyana, kungakalindeleki.	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	C						
	cancer	umdlavuza	ikhenza	ikankere	cancer: a disease in which abnormal cells divide uncontrollably and destroy body tissues	umdlavuza: bulwele la amaseli angakavami, azibuyelela khona ngokungawulekiko aze abulale amathitjhu womzimba	ikhensa
	cancer of the breast	umdlavuza webele	ikhenza yebele	ikankere yebele	breast cancer: cancer that forms in the cells of the breasts	umdlavuza webele: mdlavuza owakheka emaselini wamabele	umdlavuza wesifuba
	cancer of the cervix	umdlavuza womlomo wesibelesho	ikhenza yomlomo wesibelesho	ikankere yomlomo wesibelesho	cancer of the cervix: cancer that forms in the cells of the cervix	umdlavuza womlomo wesibelesho: mdlavuza owakheka emaselini womlomo wesibelesho	
	cancer of the endometrium	umdlavuza weboda lesibelesho	ikhenza/ikankere yeboda lesibelesho	ikhenza/ikankere ye-endomethriyam u	cancer of the endometrium: cancer that forms in the cells of the endometrium	umdlavuza weboda lesibelesho: mdlavuza owakhekha emaselini weboda lesibelesho	
	cancer of the prostate	umdlavuza wephrosteyiti	ikankere/ikhenza yephrosteyiti		cancer of the prostate: cancer that forms in the cells of the prostate	umdlavuza wephrosteyiti: mdlavuza owakheka emaselini wephrosteyiti	
	cancer of the skin	umdlavuza wesikhumba	ikankere/ikhenza yesikhumba	imelanono	cancer of the skin: cancer that forms in the cells of the skin	umdlavuza wesikhumba: mdlavuza owakhekha emaselini wesikhumba	
	cancer of the uterus	umdlavuza wesibelesho	ikhenza yesibelesho	ikankere yesibelesho	cancer of the uterus: cancer that forms in the cells of the uterus	umdlavuza wesibelesho: mdlavuza owakheka emaselini wesibelesho	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	cardiac arrest	ukujanyelwa yihliziyo	ukubotjhwa kwehliziyo		cardiac arrest: sudden, unexpected loss of heart function, breathing and consciousness	ukujanyelwa yihliziyo: kulokha ihliziyo ijama ingasasebenzi, ngokuphanya kwelihlo, lokhu kulandelwa kubhalelwa kuphefumula nokutsirimezeka	
	cardiology	ikhadiyoloji			cardiology is the study and treatment of disorders of the heart and the blood vessels	Ikhadiyoloji: sifundo nokwelatjhwa kwehliziyo nemithambo yegazi	
	cardiologist	udorhodere wehliziyo	Ikhadiyolojisti		a cardiologist: a doctor with special training and skill in finding, treating and preventing diseases of the heart and blood vessels	udorhodere wehliziyo: ngudorhodere obandulelwe ukuthola, ukwelapha namkha ukuvimbela amalwele wehliziyo nemithambo yeengazi	
	cardiovascular disease	ubulwele behliziyo			cardiovascular: heart and blood vessel disease	ubulwele behliziyo: bulwele obuphathelele nokusebenza kwehliziyo nemithambo yehliziyo	
	chest pain	ubuhlungu besifuba			chest pain: a discomfort in the chest including a dull ache, a crushing or burning feeling, a sharp stabbing pain and pain that radiates to the neck or shoulder	ubuhlungu besifuba: kungatjhaphuluki esifubeni, ukufaka hlangana ubuhlungu obubudisi, umuzwa otjhisako, ubuhlungu obuhlaba kabuhlungu obusuka entanyeni buye ehlombe	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	chicken pox	ipoksi			chicken pox: an infectious disease causing a mild fever and a rash of itchy inflamed pimples which turn to blisters and then loose scabs; it affects mainly children	ipoksi: bulwele obuthathelanako obubanga umtjhiso neratjhi kuvame ukobana kube neenlonjana ezizala amanzi; iphatha khulu abentwana	
	chemotherapy	ikhimotheraphi	ukwelatjhwa ngokutjhiswa		chemotherapy: the treatment of disease by the use of chemical substances, especially the treatment of cancer by cytotoxic and other drugs	ikhimotheraphi: kukwelatjhwa kobulwele ngokusebenzisa amakhemikhali, khulu khulu nakwelatjhwa umdlavuza	ikhimotheraphi
	cholera	ikholera	ubulwele bokurhuda		cholera: an infectious and often fatal bacterial disease of the small intestine, typically contracted from infected water supplies and causing severe vomiting and diarrhoea	ikholera: bulwele obuthelanelako bamathumbu amancani, butholakala emanzini atshayelekileko, bubanga nokuhlanza okukhulu ubulwelobu	ikholela/ihhudo
	chronic	ikhronikhi	ubulwele oburhwahlaphazako		chronic: (of an illness) persisting for a long time or constantly recurring	ikhronikhi: bulwele oburhwahlaphazako obuthatha isikhathi eside namkha obuhlala bubuyelela	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	Chronic Fatigue Syndrome (cfs)	i-CFS	ukudinwa okukhulu		Chronic Fatigue Syndrome (CFS): a complicated disorder characterised by extreme fatigue that lasts for at least six months and that can't be fully explained by an underlying medical condition; the fatigue worsens with physical or mental activity, but doesn't improve with rest	i-CFS: bujamo bokudinwa okukhulu okungahlala bekube ziinyanga ezisithandathu; ukudinokhu angeze wathi bubangelwa bujamo obuthileko bezamaphilo; lokhu kudinwa buba namandla lokha umuntu nakenza umsebenzi othileko wezandla namkha womkhumbulo; ukudinokhu akubi ngcono nanyana umuntu angaphumula kangangani	
	clinic	umtholapilo	Itlinigi		clinic: an establishment or hospital department where outpatients are given medical treatment or advice, especially of a specialist nature	umtholapilo/itlinigi: mumakhiwo namkha umnyango wesibhedlela la iinguli zangaphandle zinikelwa khona imitjhoga namkha iinyeleliso	
	colic	ikholikhi			colic: when a healthy baby cries a lot for a longer time than most babies	ikholikhi: kulokha isana eliphilileko lilila khulu isikhathi eside ukwedlula abanye abentwana	ikholikhi

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	communicable diseases	ubulwele obuthathelanako	ubulwele obuthelelanako		a communicable disease: disease that is spread from one person to another through a variety of ways that include: contact with blood and bodily fluids; breathing in an airborne virus; or by being bitten by an insect	ubulwele obuthathelanako: bulwele obusabalala ngokusuka komunye umuntu buye komunye ngebanga lokuthintana ngokweengazi namkha iketjezi, ukuphefumula, ukulunywa ziinunwana	
	constipation	ukubhinjidelwa			constipation: bowel movements that are infrequent or hard to pass	ukubhinjidelwa: kutjho ubuthuvi obungenzeki namkha obungezi ngendlela ejayelekileko namkha obungakhambi lula	ukuqumbelana
	contraceptives	iinkhandlimbeleko			contraceptives: a device or drug serving to prevent pregnancy	iinkhandlimbeleko: iqhinga namkha umtjhoga ovimbela ukuba sidisi	isivikela kukhulelwa
	counselling	ukululekwa ngokomkhumbulo	ikhanselinghi		counselling: the process that occurs when a client and counsellor set aside time to explore difficulties which may include the stressful or emotional feelings of the client; the act of helping the client to see things more clearly, possibly from a different viewpoint	ukululekwa ngokomkhumbulo: yikambiso eyenzekako lokha umluleki abekela ngeqadi isikhathi sokuhlola ubudisi obungafaka hlangana, imizwa ediselwako yomlulekwa; yikambiso yokusiza umlulekwa ukobana abone izinto ngendlela ecacileko nangakwelinye ihlangothi	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	c-section	ukubeletha ngokusikwa	i-c-section		cesarean delivery (C-section): a surgical procedure used to deliver a baby through incisions in the abdomen and uterus	ukubeletha ngokusikwa: kubeletha ngokobana umma asikwe isibeletho ukuze kukhutjwe isana	
	D						
	dehydration	ukuphelelwa mamanzi emzimbeni	ukoma		dehydration: when someone loses more fluids than he or she takes in	ukuthayelelwa mamanzi emzimbeni: kulokha umuntu aphelelwa mamanzi khulu okwedlula indlela awathola ngayo emzimbeni	
	dentist	udorhodere wamazinyo			a dentist, also known as a dental surgeon, is a surgeon who specialises in dentistry, the diagnosis, prevention, and treatment of diseases and conditions of the teeth	udorhodere wamazinyo: ngudorhodere obandulelwe amazinyo wabantu; uhlola, avikele bekalaphe amalwele wamazinyo	
	de-oxygenated blood	iingazi ezingena oksijini			deoxygenated blood :refers to the blood that has a low oxygen saturation relative to blood leaving the lungs	iingazi ezingena oksijini: ziingazi ezine-oksijini encani emaphatjhini	
	depression	ukugandeleleka ngokomkhumbulo	idiphreshini		depression (major depressive disorder): a common and serious medical illness that negatively affects how you feel, the way you think and how you act	ukugandeleleka ngokomkhumbulo: bulwele bomkhumbulo obunesithintela endleleni ozizwa ngayo, indlela ocabanga nowenza ngayo izinto	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	diabetes	ubulwele betjhukela	ubulwele beswigiri	itjhukela/ iswigiri	diabetes: a disease that occurs when your blood glucose, also called blood sugar, is too high. <i>type 1 diabetes</i> : a chronic condition in which the pancreas produces little or no insulin	itjhukela/ubulwele betjhukela: bulwele obenzeka lokha itjhukela eseengazini nayiphezulu. umhlobo-1 wetjhukela: lomhlobo wenzeka lokha amaphenkhrisi angakhiqizi i-insulini namkha akhiqiza i-insulini encani	isifo sikashukela
	diarrhoea	ukurhuda	ukuthulula		diarrhoea: loose, watery stools (bowel movements).	ukurhuda: buthuvi obumamanzi khulu	uhudo
	digestive system	ihlelo lokugaya ukudla	ihlelo lokusila ukudla		the digestive system is made up of the gastrointestinal tract — also called the gi tract or digestive tract — and the liver, pancreas, and gallbladder.	ihlelo lokugaya ukudla: lakhiwe ngezitho ezifaka hlangana isibindi, inyongo njalonjalo	uhlelo lokugaya ukudla
	diseases	amalwele			diseases: a disorder of structure or function in a human, animal, or plant, especially one that produces specific symptoms or that affects a specific location and is not simply a direct result of physical injury	amalwele: kungakhambi kuhle kwesakhiwo namkha kokusebenza emntwini, esibandaneni namkha esitjalweni lokhu kwenzeka esithweni esithileko, ngaphandle kokulimala	
	doctor	udorhodere	inyanga		doctor: a person who is qualified to treat people who are ill	udorhodere: mumuntu obandulelwe ukwelapha abantu abagulako	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	down syndrome	idawunisindromu			down syndrome: a congenital disorder arising from a chromosome defect, causing intellectual impairment and physical abnormalities including short stature and a broad facial profile	idawunisindromu: kulangahlangana kokuzalwa okuvela kumakhromozowumu angakalungi, enza kwenzeke okungakavami emzimbeni nemkhumbulweni. lokhu kwenza ukukhula kuriyade khulu	idawunisindromu
	E						
	ectopic pregnancy	isidisi samatjhubhu			ectopic pregnancy: a pregnancy in which the fertilised egg implants outside the uterus	isidisi samatjhubhu: iqanda elivundisweko lihlala ngaphandle kwesibeletso	
	eczema	i-ekzima			eczema: a condition where patches of skin become inflamed, itchy, red, cracked, and rough	i-ekzima: bujamo la amapetjhe wesikhumba aba neratjhi, alume, abe bovu, kube nokudabuka njll	
	epidemic	umbulalalizwe			an epidemic: the rapid spread of disease to a large number of people in a given population within a short period of time	umbulalalizwe: kurhatjheka msinyana kobulwele ebantwini abanengi ngesikhatjhana esincani	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	epilepsy	isithunhwani	ubulwele bokuwa	i-ephilepsi	epilepsy: a disorder in which nerve cell activity in the brain is disturbed, causing seizures	isithunhwani: kungasebenzi kuhle namkha ukuphazamiseka komthambo osebutjheni, lokhu kubanga ukutsirimezeka	Isithuthwane
	F						
	fallopian tubes	amatjhubhu wokudlulisa amaqanda			fallopian tubes: the female structures that transport the ova from the ovary to the uterus each month	amatjhubhu wokudlulisa amaqanda: matjhubhu akhambisa amaqanda, awase esibeletweni iinyanga zoke	
	FAST (Stroke)	I-FAST (Istrowuku)			FAST: an acronym for a device used to help detect and enhance responsiveness to the needs of a person having a stroke	I-FAST (Istrowuku): sirhunyezo esijamele isisetjenziswa esisetjenziselwa ukuthola ukobana umuntu usabela kangangani, sikhambe kangangani istrowuku	
	fetus	umbungu			fetus: an unborn or unhatched offspring of a mammal, in particular an unborn human	umbungu: mntwana womuntu ongakabelethwa	umbungu/ umntwana osakhula esiswini
	fever	umtjhisomkhulu womzimba			fever: a temporary increase in your body temperature, often due to an illness	umtjhisomkhulu womzimba: kukhuphuka kwamsinyana kwezinga lokutjhisomzimba, izingeli libangwa bulwele	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	fibroids	amafayibhrotsi	iimila zangesibelethweni		fibroids: non-cancerous growths in the uterus that can develop during a woman's childbearing years	amafayibhrotsi: ziimila ezingenamdlavuzwa ezikhula ngaphakathi kwesibelesho eminyakeni yomfazi yokubeletha	
	flu	umgomani	umkhuhlani	i-flu	flu: a common viral infection that can be deadly in high risk groups	umgomani: kutshayeleda ngengogwana okuvamileko okungabulala labo abanamasotja womzimba abuthakathaka	umkhuhlane/i-influenza
	flu vaccine	ukuhlabela umkhuhlani	ukuhlabela umgomani	ukuhlabela i-flu	flu-vaccine: vaccines that protect against infection by influenza viruses	ukuhlabela umgomani: mijovo evikela ukutshayeleda ngemulwana yomgomani	umgomo
	fontanelle	ihlokwana			fontanelle (or fontanel) (colloquially, soft spot): an anatomical feature of the infant human skull comprising any of the soft membranous gaps (sutures) between the cranial bones that make up the calvaria of a fetus or an infant	ihlokwana: litshwayo eliba khona ehlokweni yesana la ihloko itjhingela ngaphakathi kube ngasuthi kunomgodi ehlokweni yomntwana. kubangwa kungabi namanzi emzimbeni womntwana	ifontaneli

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	G						
	genital herpes	iinlonda zezitho zobulili	amatjhatjhazi weenlonda zobulili		genital herpes: a sexually transmitted disease; it causes herpes sores, which are painful blisters (fluid-filled bumps) that can break open and ooze fluid	iinlonda zezitho zobulili: bulwele obuthathwelana ngokomseme; bubanga iinlonda ezibuhlungu (ezizele amanzi) ezingaputjuka zivuze lawo manzana	
	germ	igciwana	igciwani		germ: micro-organisms known as pathogens or "germs" can lead to disease; these small organisms, too small to see without magnification, invade humans, animals, and other living hosts	igciwana: ziinunwani ezingabanga ukugula; iinunwanezi azibonakali ngamehlo wenyama; iinunwanezi zisahlela abantu, iimbandana neentjalo	igciwane
	gestational diabetes	itjhukela-sidisi			gestational diabetes: diabetes diagnosed for the first time during pregnancy (gestation)	itjhukela-sidisi: yitjhukela ephatha umfazi osidisi; kanengi ngemuva kokubeletha ayisabi khona	
	gonorrhoea	igonoriya			gonorrhea: a sexually transmitted bacterial infection, if untreated may cause infertility	igonoriya: bulwele obuthelana ngokomseme, nabungeze balatjhwa bungabanga ukungabelethi	igonoriya
	gynaecologist	udorhodere wabafazi	igayini	igayinakholojisti	gynecologist: a surgeon who specialises in the female reproductive system	udorhodere wabafazi: ngudorhodere obandulelwe izitho zabafazi zokubeletha	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	H						
	headache	ihloko	ihloko ebuhlungu		headache: a painful sensation in any part of the head, ranging from sharp to dull, that may occur with other symptoms	ihloko ebuhlungu: mumuzwa obuhlungu kinanyana ngiliphi iihlangothi lehloko; ubuhlungobu bungaba ngobuhlabako namkha obuzwakalela phasi; bungakhambisana namanye amatshwayo	
	HIV	iHIV			Human Immunodeficiency Virus (HIV): an infection that attacks the body's immune system, specifically the white blood cells called cd4 cells	iHIV: bulwele obulwa namasotja womzimba khulukhulu ama-white blood cells abizwa ngokobana ma cd4 seli	ingculazi
	HIV testing	ukuhlololwa iHIV			HIV testing: used to detect the presence of the HIV virus, the virus that causes AIDS syndrome	ukuhlololwa iHIV: kuhlolwa ngomnqopho wokuthola ukobana umuntu unayo na ingogwana ebanga intumbantonga	ukuhlolo-ngculazi
	hormones	amahomowuni			hormones: chemical messengers that are secreted directly into the blood, which carries them to organs and tissues of the body to exert their functions	amahomowuni: ziinthunywa ezibukhemikhali ezikhiqizwa ziingazi; zisa ezithweni zomzimba ukuze zenze umsebenzazo	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	hormonal pigmentation	iintjhubaba			hormonal pigmentation: appears as irregular patches of brown skin on the forehead, cheeks, upper lip, chin and nose and is more commonly seen in people with darker complexions	iintjhubaba: mapetjhi abhraweni atholakala epandla, eenhlathini, eendebeni, epumulweni; zibonakala khulu ebantwini abanebala elinzima	
	HPV	iHPV			Human Pilloma Virus (HPV): a viral infection that's passed between people through skin-to-skin contact	iHPV: kutshayeleva ngengogwana okwenzeka hlangana nabantu ngebanga lokuthintana ngesikhumba	
	hyperhidrosis	ihayiphehayidrosisi	ukungurumela khulu		hyperhidrosis: a condition characterised by excessive sweating	ihayiphehayidrosisi: bujamo obubangela ukobana umuntu angurumele khulu	
	hypertension	ihayiphathenshini			hypertension: another name for high blood pressure	i-hayiphathenshini: leli ngelinye ibizo lomfutho weengazi ophakemeko	
	hysteroscopy	ihistereskopi			hysteroscopy: a procedure that allows your doctor to look inside your uterus in order to diagnose and treat causes of abnormal bleeding	ihistereskopi: yikambiso eyenziwako, evumela udorhoderakho ukobana akwazi ukuqala ngaphakathi kwesibelevho sakho ukuze ahlole bekalaphe ukopha okungakajayeleki	
	hysterectomy	ukukhutjhwa kwesibelevho	ihisterektomi		hysterectomy: an operation to remove a woman's uterus	ihisterektomi: kukhutjhwa kwesibelevho ngokusikwa	ukususwa kwesibelevho

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	I						
	incubator	ibhodlelo	isifukameli		incubator: an insulated enclosure in which temperature, humidity, and other environmental conditions can be regulated at levels optimal for growth, hatching, or reproduction	ibhodlelo: simumathi la umtjhisio, umswakama namanye amajamo webhoduluko asetjenziselwa ukukhulisa, ukuphandlisela nokukhiqiza	isifukameli
	indigestion	ukuparelwa	ukuqurhelwa		indigestion: also called dyspepsia or an upset stomach is a general term that describes discomfort in your upper abdomen	ukuparelwa: kulokha kunganabunzindo kesingehla	ukuqumbelana
	infection	ukutshayeleka			infection: a disease caused by micro-organisms that invade tissue	ukutshayeleka: bulwele obubangwa magciwana angena kumathitjhu	
	injection	ukuhlaba			injection: the act of putting a liquid, especially a drug, into a person's body using a needle	ukuhlaba: sisenzo sokufaka imitjhoga emamanzi emzimbeni womuntu ngokusebenzisa inalidi	
	insulin	i-insulini			insulin: a hormone made by the pancreas that allows your body to use sugar (glucose) from carbohydrates in the food that you eat for energy or to store glucose for future use	i-insulini: yihomowuni eyenziwa yiphenkhri evumela umzimba ukusebenzisa itjhukela esuka kumakhabohayidretri ekudleni okudlela amandla namkha ukubeka itjhukela uyibekela isikhathi esizako	i-insulini

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	L						
	laboratory	ilaborathri			laboratory: a facility that provides controlled conditions in which scientific or technological research, experiments, and measurement may be performed	ilaborathri: umakhiwo wokusebenzela labo abaraga zesayensi ukufaka hlangana amarhubhululo wezesayensi newetheknoloji	
	laparoscopy	ilapharaskophi			laparoscopy: an operation performed in the abdomen or pelvis using small incisions with the aid of a camera	ilapharaskophi: kuhlinzwa okwenziwa ngemathunjini ngokusika okuncazana okunekhamera; ilapharaskopi isiza ekutholeni umraro ukuze kwelatjhwe	
	M						
	malaria	imalariya			malaria: a disease caused by a plasmodium parasite, transmitted by the bite of infected mosquitoes	imalariya: bulwele obusabalaliswa kulunywa bonompopoloza abatshayekileko, mgomani odengezelisako	umalaleveva
	malnutrition	ukungadli ukudla okunomsoco	ukungadli ukudla okunezakham zimba	imalnuthrishini	malnutrition: a condition that results from eating a diet in which one or more nutrients are either not enough or are too much such that the diet causes health problems	ukungadli ukudla okunomsoco: bujamo obuvela ngebanga lokobana ukudla okudlako kutjhoda izakhamzimba namkha kunezakha mzimba ezinengi, lokhu kubanga imiraro yezamaphilo	ukungondleki

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	medical care	itjhejo lezokwelapha			medical care: the provision of what is necessary for a person's health and well-being by a doctor, nurse, or other healthcare professional	itjhejo lezokwelapha: kunikelwa kwalokho isiguli esikutlhogako kwezokwelapha ngunese, udorhodere namkha isisebenzi nanyana ngisiphi sokwezokwelapha	
	medication	imitjhoga			medication (also referred to as medicine, pharmaceutical drug, or simply drug): a drug used to diagnose, cure, treat, or prevent disease	umtjhoga/ intatha: sigogobalisi esisetjenziselwa ukuhlola, ukwelapha nokuvikela ubulwele	
	melanin	imelanini			melanin: the pigment that gives human skin, hair, and eyes their colour	imelanini: yiphigmenti enikela isikhumba somuntu iinhluthu namehlo umbala	
	menopause	imenophosi	ukungasayi esikhathini		menopause: the time that marks the end of your menstrual cycles; it is diagnosed after you've gone 12 months without a menstrual period	imenophosi: sikhathi esiveza ukuphela komzombelezo wokuba sesikhathini; itholwa lokha nawuthethe iinyanga ezili-12 ungasi esikhathini	
	menstruation	ukuba sesikhathini	ukuba semalangeneni	ukuba senyangeneni	menstruation: the process in a woman of discharging blood and other material from the lining of the uterus at intervals of about one lunar month from puberty until the menopause, except during pregnancy	ukuba sesikhathini: kulokha umuntu wengubo nakavinyilika iingazi nokhunye okusuka ebodeni lesibeletho ngokuhlukana kwesikhathi okungaba yinyanga kusukela ebutlawaneni bekufike la umuntu angasayi khona esikhathini, ngaphandle kwalokha nakasidisi	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	menstrual cycle	umzombezezo wokuba sesikhathini	umzombezezo wesikhathi		menstrual cycle: a cycle is counted from the first day of 1 period to the first day of the next period. the average menstrual cycle is 28 days long	umzombezezo wokuba sesikhathini: ubalwa kusakela elangeni lokuthoma lokubona isikhathi ukuya elangeni lokuthoma lokubona amalanga ukuya esikhathini esilandelako; umzombezezo ovamileko uba malanga ama-28	
	migraine	imigreyini			migraine: headache of varying intensity, often accompanied by nausea and sensitivity to light and sound	imigreyini: yihloko ebuhlungu khulu, ekhambisana nokugonyuluka kanye nokungafuni umkhanyo netjhada	
	miscarriage	ukuphunyelwa sidisi	ukubuya endleleni		miscarriage: a spontaneous loss of a woman's pregnancy before the 20th week that can be both physically and emotionally painful	ukuphunyelwa sidisi: kulahlekelwa ngokungakahleli sidisi ngaphambi kobana sifike eemvekeni ezima-20; ukulahlekelokhu kuba buhlungu enyameni nemizweni	
	molar pregnancy	isidisi semola	isidisi seenyama		molar pregnancy: is a non-cancerous tumour that develops in the uterus as a result of a non-viable pregnancy	isidisi semola: sidisi seenyama ezinganamdlavuzwa ezenzeka ngaphakathi kwesibeetho ngebanga lesidisi esingakhange siphumelele	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
N							
	nausea	ukugonyuluka			nausea: stomach discomfort and the sensation of wanting to vomit	ukugonyuluka: kulokha kungasimnandi ngemathunjini ube ufune ukuhlanza	
	navarapine	inevaraphini	isidambisi sentumbantonga		nevirapine (viramune): an anti-hiv drug that reduces the amount of viruses in the body; anti-hiv drugs such as nevirapine slow down damage to the immune system and prevent the occurrence of AIDS-defining illnesses	inevaraphini: mtjhoga owehlisa inani lengogwana emzimbeni; umnqopho ehlisa izinga lobungozi emasotjeni womzimba ivimbele namalwele wentumbantonga	
	nervous system	ihlelo lemezwa			nervous system: the network of nerve cells and fibres which transmits nerve impulses between parts of the body	ihlelo lemezwa: lithungelelano lamaseli wemisipha aletha imilayezo hlangana nezitho zomzimba	uhlelo lemezwa
	non-communicable diseases	amalwele angathathelaniko			non-communicable disease (ncd): a disease that is not transmissible directly from one person to another	ubulwele obungathathelaniko: bulwele obungasukiko komunye umuntu buye komunye	
	nurse	umhlangikazi	unese		nurse: a person trained to care for the sick or infirm, especially in hospital	umhlangikazi: mumuntu obandulelwe ukuthogomela namkha ukutjheja abantu abagulako emitholapilo neembhedlela	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	O						
	obesity	ukunona			obesity: a medical condition in which excess body fat has accumulated to the extent that it may have an adverse effect on health	ukunona: bujamo bezokwelapha la amafutha womzimba abutheleleke ngendlela yokobana lokhu kunomthelela epilweni yomuntu	ukunona
	obsessive compulsive disorder(ocd)	i-ocd			obsessive compulsive disorder (ocd): a mental illness that causes repeated unwanted thoughts or sensations (obsessions) or the urge to do repetitive behaviours	i-ocd: bulwele bomkhumbulo obubanga imikhumbulo ebuyelelwako, engafunekiko; ubulwelobu benza umuntu enze izinto ngokubuyelelwe khulu.isb. ukuhlanza indawo ehlanzekileko	
	oral care	ukutlhogonyelwa komlomo	ukutjhejwa komlomo		oral care: the practice of keeping one's mouth clean and free of disease and other problems	ukutlhogonyelwa komlomo: mumukghwa wokuphatha kuhle umlomo, uhlale uhlanzekile unganamalwele neminye imiraro	
	oestrogens	ama-ustrejini			oestrogens: hormones that are important for sexual and reproductive development, mainly in women; they are also referred to as female sex hormones	ama-ustrejini: mahomowuni aqakathekileko kwezokulalana nokubeletha, atholakala ebantwini bengubo; abizwa godu nokuthi mahomowuni wabafazi wobulili	
	oxygenated blood	iingazi ezine-oksijini			oxygenated blood:the blood that has been exposed to oxygen in the lungs	iingazi ezine-oksijini: ziingazi ezithola ummoya ocwengileko/ iingazi ezithola i-oksijini emaphatjhini	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	P						
	pandemic	umabhubhisa			pandemic: a disease outbreak that spreads across countries or continents; it affects more people and takes more lives	umabhubhisa: bulwele oburhatjhekela kezinye iinarha namaphasi. bubamba abantu abanengi begodu buthatha nabantu abanengi	
	pap smear	ipapsmiye	ukuhlolwa komlomo wesibeletso		pap smear: a pap smear, also called a pap test, is a screening procedure for cervical cancer	ipapsmiye: kuhlolwa komlomo wesibeletso sihlolwa umdlavuza	ucwephano lomlomo wesibeletso
	patient	Isiguli	isigulana		patient: a person receiving or registered to receive medical treatment	isiguli: mumuntu owemukela namkha otloliselwe ukwemukela ukwelatjiswa kwezamaphilo	
	patient transfer	ukwedluliswa kwesiguli			patient transfer: moving a patient from one flat surface to another, such as from a bed to a stretcher	ukwedluliswa kwesiguli: kususwa kwesiguli kenye indawo sisiwe kenye indawo isib. sisuswa komunye umtholapilo ukuya komunye	
	penicillins	amaphenisilini			penicillins: a group of antibacterial drugs that attack a wide range of bacteria	amaphenisilini: libuthelelo lemitjhoga elisahlela amabhatktheriya amanengi	
	people living with disabilities	abantu abaphila nokukhubazeka			people living with disabilities: those who have long-term physical, mental, intellectual or sensory impairments which in interaction with various barriers may hinder their full and effective participation in society on an equal basis with others	abantu abaphila nokukhubazeka: babantu abanamalwele emzimbeni, emkhumbulweni, emizweni lawo malwele ayabavimbela ukuzibandakanya ngokupheleleko emphakathini njengabanye abantu	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	phantom pregnancy	isidisi esingasi ngesambala	isidisi sangemkhumbulweni		phantom pregnancy/ false pregnancy: the appearance of clinical or subclinical signs and symptoms associated with pregnancy when the woman is not actually pregnant	isidisi esingasingesamambala/ isidisi sangemkhumbulweni: isidisi okungasiso isidisi: kulokha umfazi anamatshwayo woke wokuba sidisi kodwana angasi sidisi	
	phobia	ifobiya			phobia: a type of anxiety disorder defined by a persistent and excessive fear of an object or situation	ifobiya: mhlobo wokwethuka namkha ukwesaba okudlulileko into ethileko namkha ubujamo isb. ukwesaba indawo evalelekileko	ifobiya
	pregnancy	ukuzithwala	ukuba nesidisi	ukuba sebantwini	pregnancy: the condition or period of being pregnant	ukuzithwala: bujamo namkha isikhathi sokuba sidisi	ukukhulelwa
	pre-diabetes	iphridayibhithisi			pre-diabetes: a condition in which blood sugar is high, but not high enough that it can be type 2 diabetes	iphridayibhithisi: bujamo la itjhukela iphakeme kodwana ingakaphami khulu bona kungaba mhlobo wesi-2 wobulwele betjhukela	
	primary health care	ukutlhogonyelwa kwepilo okusisekelo			primary health care, or phc: "essential health care" that is based on scientifically sound and socially acceptable methods and technology; this makes universal health care accessible to all individuals and families in a community	ukutlhogonyelwa kwepilo okusisekelo: litjhejo lezamaphilo elisekelwe kusayensi neendlela ezamukelekako ngekezokuhlalisana nekezetheknoloji; ukutlhogonyelwa kwepilo esisekelo kwenza zokwelapha kutjhideleleke kibo boke abantu kanye nemindenabo emphakathini	unakekelo lwempilo

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	progesterone	iphrojesteroni			progesterone: a hormone released by the corpus luteum in the ovary; it plays important roles in the menstrual cycle	iphrojesteroni: ihomowuni ekhutjha ma-ovari, edlala indima ekulu emzombelzweni wokuba sesikhathini	
	pneumonia	inyumoniya			pneumonia: lung inflammation caused by bacterial or viral infection, in which the air sacs fill with pus and may become solid	inyumoniya: bulwele bamaphaphu obubangelwa kutshayekeka ngengogwana; imigodla yomoya izala ubovu bese iyaqina	inyumoniya/ umkhuhlane wamaphaphu
	polio	ipholiyo			polio: a very contagious disease caused by a virus; it has no symptoms	ipholiyo: bulwele obuthelelanako obubangwa yingogwana; abunamatshwayo ubulwelobu	uvendle
	pregnancy test	ukuhlolelwa isidisi			pregnancy test: detects human pregnancy hormone to determine whether an individual is pregnant	ukuhlolelwa isidisi: kuhlola ihomowuni yesidisi ukuthola ukobana ingabe umuntu usidisi na	
	psychologist	udorhodere womkhumbulo	isayikholojisti		psychologist: someone who studies the mind and behaviour of people.	udorhodere womkhumbulo: mumuntu ofunde ngomkhumbulo nokuziphatha kwabantu	
	R						
	rash	iratjhi			rash: temporary outbreak of red, bumpy, itchy patches of skin	iratjhi: maqhuqhumba wesikhatjhana abovana, alumako avela esikhunjeni	ukuqubuka
	right to health	ilungelo lezamaphilo			right to health: a complete state of physical, mental and social well-being, and not merely the absence of disease or infirmity	ilungelo lezamaphilo: bujamo obupheleleko emzimbeni, emkhumbulweni nekezokuhlalisana; akutjho nokho ukungabi namalwele	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	ring-worm	ubuthi bembuzana	iringiwemu		ringworm: spread by skin-to-skin contact or by touching an infected animal or object	ubuthi bembuzana: busabalala ngokusuka kesinye isikhumba buye kesinye; ngokuthinta isibandana esitshayelekileko namkha into etshayelekileko	umbandamu
	S						
	schizophrenia	iskizofreniya			schizophrenia: long-term mental disorder of a type involving a breakdown in the relation between thought, emotion, and behaviour, leading to faulty perception, inappropriate actions and feelings, withdrawal from reality and personal relationships into fantasy and delusion, and a sense of mental fragmentation	iskizofreniya: bulwele bomkhumbulo obuthatha isikhathi eside obufaka hlangana, ukungasathelelan'amanzi phakathi kwemizwa, umkhumbulo nokuziphatha okwenza angasacabangi kuhle umuntu; umuntu ugcina enza izinto ezingakafaneli, uzithola angasanathintano nepilo yamambala, angasakwazi ukobona izinto ngendlela efaneleko	iskizo

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	sex education	isifundo sezobulili			sex education: high quality teaching and learning about a broad variety of topics related to sex and sexuality, exploring values and beliefs about those topics and gaining the skills that are needed to navigate relationships and manage one's own sexual health	isifundo sezobulili: kufunda nokufundisa ngemihlobo ehlukeneko ephathelene nobulili, umseme, ukuhlola amagugu neenkolelo ngalezo nhlobo nokuthola amakghono adingekako wokuphatha ubudlelwana nokulawula ipilo yobulili	
	sexually transmitted diseases	amalwele athathelana ngokomseme			sexually transmitted diseases: diseases transmitted by sexual contact, caused by micro-organisms that survive on the skin or mucus membranes of the genital	amalwele athathelana ngokomseme: malwele athathelwana ngokuthintana ngokomseme; lokhu kubangelwa ziinunwana eziphila esikhunjeni sezitho zobulili	isifo socansi esithathelanayo
	shirodkar stitch	istitjhi seshirokha			shirodkar stitch: the shirodkar stitch or mc donald circlage is a 'purse-string' stitch put into the cervix (mouth of the womb) to keep it tightly closed during pregnancy	istitjhi seshiroka: sitijhi namkha umthungo osasikhwanyana ofakwa ngaphakathi komlomo wesibelethe ukuze sivaleke ngci, ngesikhathi umma asazithwele	
	skin doctor	udorhodere wesikhumba			skin doctor: a doctor who specialises in the treatment of diseases of the skin	udorhodere wesikhumba: ngudorhodere obandululelwe ukwelapha amalwele wesikhumba	
	sore throat	umphimbo obuhlungu			sore throat: pain, scratchiness or irritation of the throat that often worsens when you swallow	umphimbo obuhlungu: buhlungu bomphimbo obenza kube budisi ukuginya	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	sterilisation	ukubotjhwa kwesibeletho	isterilizeyishini		sterilisation: a permanent method of birth control	isterilizeyishini: yindlela yaphakade yokuhlela umndeni	
	stress	istresi			stress: a state of mental or emotional strain or tension resulting from adverse or demanding circumstances	istresi: bujamo la umkhumbulo namkha imizwa ithwala kabudisi ngebanga lobujamo obubiza khulu	istresi
	stroke	istrowuku			a stroke: a sudden interruption in the blood supply of the brain; most strokes are caused by an abrupt blockage of arteries leading to the brain (ischemic stroke)	istrowuku: kuphazamiseka okumsinyana kokukhanjiswa kweengazi ebucotjheni; kanengi kwenzeka lokha nakuvaleka msinyana imithambo eya ebucotjheni	
	symptoms	amatshwayo			symptoms: physical or mental features which are regarded as indicating a condition of disease, particularly such features that are apparent to the patient	amatshwayo: matshwayo asenyameni namkha emkhumbulweni aveza ubujamo bobulwele, lamatshwayo abonakala eengulaneni	
	syphilis	isifilisi			syphilis: a highly contagious disease that's mostly spread through sexual activity	isifilisi: bulwele obuthelelana khulu, obusabalala ngokomseme	
	T						
	teething	ukumilisa			teething: when teeth first come through a baby's gums	ukumilisa: kulokha amazinyo nakavela kokuthoma eensinini zomntwana	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	tests	iinhlahlubo			tests: medical procedures performed to detect, diagnose, or monitor diseases, disease processes, susceptibility, or to determine a course of treatment	iinhlahlubo: iinkambiso zezokwelapha ezenzelwa ukuthola, zihlole namkha ukutjheja ubulwele nokuthola ukobana bungelatjhwa bunjani	
	TOP	iTOP	ukurhululwa kombungu		TOP (termination of pregnancy): a medical process of ending a pregnancy so it does not result in the birth of a baby	iTOP: yindlela yezokwelapha yokurhulula umbungu ukuze kungafiki la uba mntwana okumele abelethwe	
	trauma	ithroma			trauma: a deeply distressing or disturbing experience; a personal trauma like the death of a child	ithroma: amaboni azwisa ubuhlungu khulu begodu aphazamisako	
	thrush	ikwethu			thrush: a fungal (yeast) infection that can grow in your mouth, throat and other parts of your body	ikwethu: kutshayeleka okusamvubelo okwenzeka namkha okumila ngemlonyeni, emphinjeni nakezinye izitho zomzimba	
	to prick	ukuhlaba umuno	ukuphrikha		to prick: to make a small hole with a sharp point	ukuhlaba umuno: kukwenza imbotjana emnweni ngento ebukhali	
	U						
	ulcer	i-alsa	isilonda esivulekileko		ulcer: a sore that develops on the lining of the oesophagus, stomach or small intestine	i-alsa: silonda esivela ku-qhoqhoqho, emathunjini namkha emathunjini amancani	

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	V						
	vaccine	ivaksini	umu-endo		vaccine: a substance used to stimulate the production of antibodies and provide immunity against one or several diseases, prepared from the causative agent of a disease, its products, or a synthetic substitute, treated to act as an antigen without inducing the disease	ivaksini: ngilokho okusetjenziselwa ukugcugcuzela ukukhiqizwa kweemvikela izifo begodu inikela ivikeleko emalweleni. ivaksini yakhiwa ngaso isisabalalisi sobulwele. senziwa sibe yi-anthijeni ebulwelenobo	umgomo
	vasectomy	ivasekthomi	ukubotjhwa kwesidoda		vasectomy: minor surgery to block sperm from reaching the semen that is ejaculated from the penis	ivasekthomi: kuhlinzwa okuncani ukuze kubotjhwe amatjhubhu akhupha isidoda ukuze singahlangani neqanda	ivasekthomi
	viral load	ubungako be-hiv			viral load: the amount of virus in an infected person's blood	ubungako be-hiv: bungako be-hiv emuntwini oongazi zakhe zitshayelekile	
	virus	ingogwana	ivayirasi		virus: an infective agent that typically consists of a nucleic acid molecule in a protein coat, is too small to be seen by light microscopy, and is able to multiply only within the living cells of a host	ingogwana: sikhambisi esithwala ubulwele obuthelelanako; simumethe amamolekhyuli wesidi, yincani khulu ukobana ingabonwa ngamehlo begodu ikwazi ukuzibuyelela kwaphela, ngaphakathi kweseli ephilako	ivurisi

N	ENGLISH TERM	ISINDEBELE TERM	SYNONYM	SYNONYM	ENGLISH DEFINITION	ISINDEBELE DEFINITION	Basic Health Terms NTS-ENG-ZUL
	vitiligo	ivayithalayigo			vitiligo: a disease that causes the loss of skin colour in blotches	ivayithalayigo: bulwele obubanga ukulahlekelwa mbala ngeenkhekhe	
	vomit	ukuhlanza			vomit: matter from the stomach that has come up into and may be ejected through the mouth	ukuhlanza: izinto ezisuka emathunjini ziphuma ngomlomo, ngamandla	
	W						
	warts	iinsumpa			warts: small, fleshy bumps on the skin or mucous membrane caused by human papillomavirus	iinsumpa: magugumba amancani, anenyama atholakala enyameni abangelwa yingogwana i-human papilloma	ukuvuvuka
	wound	inceba	isilonda		wound: an injury to living tissue caused by a cut, blow or other impact	inceba: kulimala kwamathitjhu okubangelwa kusikeka, namkha ukuthinteka ngandlela thize	
	wound care	ukutjhejwa kwenceba	ukutjhejwa kwesilonda		wound care: to assess, treat and create care plans for the wound	ukutlhogonyelwa kwenceba: kuhlola, welaphe nokwenza iindlela zokutjheja inceba	

APPENDIX H: TURN-IT-IN DIGITAL RECEIPT



Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Nomsebenzi Johanna Malele
Assignment title:	Revision 3
Submission title:	The use of corpora in the compilatio...
File name:	Post_TiN_DLitt_etPhil_Nomsebenzi...
File size:	1.77M
Page count:	218
Word count:	55,568
Character count:	322,499
Submission date:	07-Feb-2021 10:13AM (UTC+0200)
Submission ID:	1503461107

APPENDIX I: TURN-IT-IN SIMILARITY INDEX

The use of corpora in the compilation of English-isiNdebele glossary of medical terms

ORIGINALITY REPORT

7 % SIMILARITY INDEX	7 % INTERNET SOURCES	1 % PUBLICATIONS	0 % STUDENT PAPERS
--------------------------------	--------------------------------	----------------------------	------------------------------

PRIMARY SOURCES

1	uir.unisa.ac.za Internet Source	4 %
2	repository.up.ac.za Internet Source	1 %
3	Moropa, Koliswa. "Analysing the English-Xhosa parallel corpus of technical texts with Paraconc: a case study of term formation processes", Southern African Linguistics and Applied Language Studies, 2007. Publication	1 %
4	lexikos.journals.ac.za Internet Source	1 %
5	hdl.handle.net Internet Source	1 %
6	spilplus.journals.ac.za Internet Source	1 %